Springer Texts in Business and Economics

Badi H. Baltagi

Econometric Analysis of Panel Data

Sixth Edition

Springer

Springer Texts in Business and Economics

Springer Texts in Business and Economics (STBE) delivers high-quality instructional content for undergraduates and graduates in all areas of Business/Management Science and Economics. The series is comprised of self-contained books with a broad and comprehensive coverage that are suitable for class as well as for individual self-study. All texts are authored by established experts in their fields and offer a solid methodological background, often accompanied by problems and exercises.

More information about this series at http://www.springer.com/series/10099

Badi H. Baltagi

Econometric Analysis of Panel Data

Sixth Edition

Springer

Badi H. Baltagi
Department of Economics
and The Center for Policy Research
Syracuse University
Syracuse, NY, USA

ISSN 2192-4333 ISSN 2192-4341 (electronic)
Springer Texts in Business and Economics
ISBN 978-3-030-53952-8 ISBN 978-3-030-53953-5 (eBook)
https://doi.org/10.1007/978-3-030-53953-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To My Wife, Phyllis

Preface

Panel data econometrics continues to be a hot topic in econometrics and has experienced a lot of growth over the last two decades. Micro- and Macro-panels are increasing in availability, and methods to deal with these data are in high demand from practitioners. Software programs helped this growth with freely available programs in R and a lot of available programs in Stata as well as EViews. Since the 5th edition, there have been six international conferences on panel data. These were held in London in 2013; Tokyo, 2014; Perth, Australia, 2016; Thessaloniki 2017; Seoul, Korea, 2018; and Vilnius, Lithuania, 2019. Panel data is widely applied in finance, development, trade, marketing, and micro- as well as macroeconomics. More specifically in health, public, labor, urban, and consumer economics. These attest to the usefulness of these panel methods in applied economics.

As of December 2020, I am proud to have more than 18,000 citations by Google Scholar to my book entitled *Econometric Analysis of Panel Data* published initially with Wiley and now updated in its 6th edition with Springer; see https://scholar.google.com/citations?user=XWrDL6IAAAAJ&hl=en.

This book is intended for an advanced undergraduate/graduate econometrics course focusing on panel data. The prerequisites include a good background in mathematical statistics and econometrics at the level of Greene (2003). Matrix algebra is necessary for the presentation of theoretical results. While this may be too technical for some readers, the book keeps an eye on real economic applications and replicates them using standard available software to help guide the applied researcher. I have taught panel data courses at Central Banks and Labor Institutes using this book focusing on the applied material and empirical applications.

Some of the major features of this book are that it provides an *up-to-date* coverage of basic panel data techniques, especially for serial correlation, spatial correlation, heteroskedasticity, seemingly unrelated regressions, simultaneous equations, dynamic panel models, incomplete panels, limited dependent variables, count and spatial panels, and nonstationary panels. I have tried to keep things *simple*, illustrating the basic ideas using the *same notation* for a diverse literature with heterogeneous notation. Many of the estimation and testing techniques are illustrated with *data sets* which are available for classroom use on the book website at https://www.springer.com/book/9783030539528. Other real economic applications are given in the text as well as in the problems sections at the end of each chapter with the data sets available from the journal's website where the article was published, like the data archive of *Journal of Applied Econometrics*. The book also summarizes several empirical studies using panel data techniques, so that the reader can relate the econometric methods with the economic applications. The book proceeds from single equation methods to simultaneous equation methods as in any standard econometrics text, so it should prove friendly to graduate students.

The book gives the basic coverage without being encyclopedic. There is an extensive amount of research in this area and not all topics are covered. Recent special issues on panel data include Baltagi and Maasoumi (2013) in *Econometric Reviews* and Bai, Baltagi and Pesaran (2016) in *Journal of Applied Econometrics*; a special issue of *Annals of Economics and Statistics* by Bonhomme and Davezies (2019), and a special issue of the *Journal of Econometrics* edited by Sarafidis and Wansbeek (2021).

I have used this book to teach panel data courses at Syracuse University; University of Leicester; Texas A&M University; University of California-San Diego; University of Cincinnati; University of Arizona; Aarhus University at the Center for Research in Econometric Analysis of Time Series, CREATES, the University of Macedonia, Thessaloniki, Greece, University of Padova, Universita Cattolica, Roma, University of Roma "La Sapienza" in Italy, University of Coimbra and the University of Minho in Portugal, Universidad del Rosario, Bogotá, Colombia; Cyprus University of Technology, University of Innsbruck and the University of Vienna in Austria; Carleton University, Canada; Universidad Autonoma de Madrid and Pompeau Fabra in Spain, Singapore Management University, Seoul National University in Korea, and yearly at the Barcelona Graduate School of Economics (2008–2019); and also at the International Monetary Fund (IMF), Washington D.C. (2004–2019); European Central Bank, Frankfurt (2001); Central Bank of Turkey (2010); Central Bank of Argentina (2009); Banco de Protugal (2008); Inter-American Development Bank, Washington D.C. (2005); Instituto de Estudios Fiscales, Madrid (2009–2011); The Marie Curie Training Programme in Applied Health Economics, Thessaloniki, Greece (2006); Center for Economic Studies (CES-Ifo) and the University of Munich (2002; 2005; 2009); Netherlands Network of Economics (NAKE), Utrecht University (2005); Institute for Economic Research (IWH)-Halle (1997), German Institute for Economic Research (DIW), Berlin (2004); Institute for Advanced Studies, Vienna (2001); Centro Interuniversitario de Econometria (CIDE)-Bertinoro (1998 and 2008); and The Study Center Gerzensee, a foundation of the Swiss National Bank (2014), to mention a few. See my webpage for a complete list https://pbaltagi.wix.com/badibaltagi.

The 6th edition continues to use empirical examples from the panel data literature to motivate the book. There are empirical illustrations and examples using Stata and EViews throughout the book. In teaching this material around the world, these useful applications are successful in the classroom and in teaching and applying the methods used in this book. The book has many problems at the end of each chapter, some of them published in *Econometric Theory* and I am grateful to

Peter C. B. Phillips for having that section and providing a valuable service to the profession.

I have written a *companion* to the 4th edition of this book; see Baltagi (2009). This is available from Wiley. The companion is written in a problem/solution format and solves a lot of the problems in this textbook. So it is a nice complement for graduate students and instructors interested in more proofs and supplemental material. The companion adds more empirical examples which are illustrated using Stata and EViews. Chapter 1 of the companion, for example, gives background material on partitioned regressions and the Frisch–Waugh–Lovell theorem that is useful for understanding fixed effects and the Within regression. It should also be helpful in explaining some of the technical material for those who need to see the proofs or check derivations, or do their own research extensions.

Virtually, every chapter was revised and updated in this new edition. References were updated and older ones were deleted when necessary due to space limitations. Older empirical applications were replaced by newer ones. New problems were added asking the reader to replicate recent panel data applications illustrating the methods in the book. I kept the old problems so that the reader can still use the companion Baltagi (2009) which accompanied the 4th edition. However, I have added new problems to many chapters including real empirical applications not presented in the text due to space limitations. The reader is asked to replicate these empirical applications. These new problems are of course not available in the Baltagi (2009) Wiley Companion.

Chapter 1 distinguishes between micro- and macro-panel data sources and discusses the benefits and limitations of using panel data in research. Both Chaps. 2 and 3 illustrate basic panel methods using a one-way (only individual effects) and two-way (both individual and time effects) error components models with fixed and random effects. Both chapters emphasize empirical examples and replicate their estimation using Stata and EViews. This is done with three empirical examples. The first is the Grunfeld (1958) investment equation. The second is the Baltagi and Griffin (1983) gasoline demand equation, and the third is Munnell's (1990) productivity of public capital in the private sector. Additional empirical examples are given in the problems section of Chap. 3. These include Ram's (2009) study on the relationship between openness, country size, and government size, and also Neumayer's (2003) study on the relationship between air pollution levels and left-wing party strength. Chapter 4 illustrates the testing procedures available with the standard software and highlight the need for more diagnostics and test statistics that have not yet made it to these packages. Tests for poolability, tests for random and fixed effects as well as Hausman's specification test are the focus of this chapter. These are illustrated with the three empirical examples used in Chap. 2. Additional empirical examples include the gravity trade equation by Glick and Rose (2002) which studies the effect of having the same currency on trade. Chapter 5 shows the reader how to deal with heteroskedasticity and serial correlation and how to test for it in a panel data context. Extensions of the Durbin–Watson test for serial correlation and the Breusch–Pagan test for heteroskedasticity to panel data are presented along with newly developed tests using panel data. These tests include (i) the *xttest1* command in Stata for jointly testing for serial correlation and random effects; (ii) the *xtserial* command which tests that the differenced fixed effects residuals have correlation of -0.5 with their lagged values. Also, (iii) the *xtqptest* and *xthrtest* commands which perform panel serial correlation tests suggested by Born and Breitung (2016). This chapter also discusses panel data methods that account for different autoregressive coefficient for each time series in the panel as well as heteroskedasticity across individuals. This is illustrated with *xtgls* in Stata. An empirical example estimating a dynamic unemployment rate equation using this method is replicated based on Nickell, Nunziata and Ochel (2005). Chapter 6 gives the estimation of seemingly unrelated regressions with panel data. This is done for the one-way and two-way error component models. Several applications are cited from the literature. Chapter 7 illustrates how one can deal with endogeneity of the simultaneous equation type using a crime example based on Cornwell and Trumbull (1994). This is illustrated with Stata using the *xtivreg* command with fixed effects as well as random effects. Another example on economic growth and foreign aid based on Bruckner (2013) is used to illustrate fixed effects 3SLS. The Hausman and Taylor (1981) method is also illustrated with Stata using PSID data applied to an earnings equation based on Cornwell and Rupert (1988), and also a gravity trade equation based on Serlenga and Shin (2007) studying the importance of common language on trade. Extensions of the Hausman and Taylor estimator to serial correlation, dynamics, and spatial correlation are discussed throughout this book. Chapter 8 treats the important dynamic panel data literature made popular with the Arellano and Bond (1991) paper. Two main empirical examples illustrate these methods for dynamic panel models. The first one estimates a dynamic demand for cigarettes across American states based on Baltagi and Levin (1986), and the second one looks at the relationship between democracy and education across countries based on Acemoglu et al. (2005). More empirical examples are added to the problems section for the reader to replicate. These include Tobin's q based on Schaller (1990); democracy and growth based on Acemoglu et al. (2019); dynamic Hausman and Taylor applied to the earnings equation of Cornwell and Rupert (1988); a gravity equation for foreign direct investment based on Egger and Pfaffermayr (2004); and the effect of a twin crisis (both currency and banking crisis) on GDP growth based on Hutchison and Noy (2005). Chapter 9 illustrates unbalanced panel data methods by estimating a hedonic housing equation and the public capital productivity puzzle using *xtmixed* in Stata. This chapter also discusses forecasting with unbalanced panels. Chapter 10 has selected topics including measurement error, rotating panels, pseudo-panels, heterogeneous panels, short-run versus long-run, and count panel data. Count panel is illustrated using the classic Hausman, Hall and Griliches (1984) study on the relationship between patents and R&D expenditures. Additional empirical examples on doctor's visits by Winkelmann (2004) and hospital visits by Geil et al. (1997) are in the problems section. Matched panels are highlighted in this chapter using the study of Abrevaya (2006) who estimates the effect of smoking on birth outcomes from panel data on mothers who give multiple births. Chapter 11 is dedicated to limited dependent variable panel data models. Featured estimation methods include the fixed effects probit and logit estimation suggested by Fernandez-Val and

Weidner (2016) and their accompanying Stata applications to female labor supply based on Fernandez-Val (2009) and a gravity trade example using *probitfe* and *logitfe* in Stata; see Cruz-Gonzalez, Fernandez-Val and Weidner (2017). This performs panel data Jackknife procedures and corrects for the bias of fixed effects in logit and probit panel models. Another empirical example features the grouped conditional logit estimator using Ruhm's (1996) study on the impact of beer taxes and a variety of alcohol-control policies on motor vehicle fatality rates. The chapter also highlights Wooldridge's (2005) simple approach for handling the initial conditions problem in dynamic nonlinear unobserved effects. This is illustrated with the Vella and Verbeek (1998) panel data set which estimated the union wage differential for working men using PSID. Finally, tests for sample selection are illustrated using nurses labor supply in Norway by Askildsen, Baltagi and Holmås (2003). Chapter 12 surveys the first-generation and second-generation panel unit roots along with the associated literature on panel cointegration. Three empirical examples illustrate these methods using EViews: the first on purchasing power parity using the study of Banerjee, Marcellino and Osbat (2005); the second on international R&D spillover using the data set by Coe and Helpman (1995); and the third on the relationship between real per capita healthcare expenditures and real per capita gross domestic product based on the study of Hansen and King (1996). Additional empirical examples of these panel unit-root methods using Stata are added to the problems section. These include Luintel's (2000) study on black markets for foreign exchange rates, and also test for unit roots using inflation rates by Culver and Papell (1997). Chapter 13 tackles spatial panel data methods using the spatial autoregressive models. Estimation methods using generalized moments and maximum likelihood are considered as well as forecasting with spatial panel data. Lagrange and likelihood ratio tests for spatial random effects panel models are discussed as well as panel data tests for cross-sectional dependence. These are illustrated with Stata using commands *xtcsd* and *xttest2* as well as in EViews. In addition, extensions to instrumental variables and dynamic spatial panel models are studied. An empirical example based on the residential demand for electricity by Belotti, Hughes and Piano Mortari (2017) is used to illustrate these methods with Stata using the command *xsmle*.

I would like to thank my students and co-authors for allowing me to draw freely on our joint work. In particular, I would like to thank Georges Bresson, Peter Egger, Qu Feng, Jim Griffin, Chihwa Kao, Walter Krämer, Dan Levin, Dong Li, Qi Li, Long Liu, Michael Pfaffermayr, Alain Pirotte, Seuck Heun Song, and Ping Wu. Many colleagues had direct and indirect influence on the contents of this book, most notably, Cheng Hsiao, Roberto Mariano, M. Hashem Pesaran, Peter C. B. Phillips, Peter Schmidt, and Tom Wansbeek, also, in memoriam, Gary Chamberlain, Arthur Goldberger, Clive Granger, Zvi Griliches, G. S. Maddala, Halbert White, and Arnold Zellner.

David Drukker provided help with Stata on the Hausman and Taylor procedure as well as EC2SLS in Chap. 7, and also the Baltagi and Wu LBI test in Chap. 5. Glenn Sueyoshi provided help with EViews on the panel unit-root tests in Chap. 12. Kevin Tappe provided valuable typos and suggestions after careful reading of an earlier edition of my book while programing with R. Thanks also go to Johannes

Glaeser and Judith Kripp at Springer for their editorial help. Last but not least, thanks go to my wife Phyllis whose encouragement and support gave me the required energy to complete this book. Responsibilities for errors and omissions are my own.

Syracuse, NY, USA Badi H. Baltagi

References

Abrevaya, J. 2006. Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics* 21: 489–519.

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2005. From education to democracy? *American Economic Review* 95: 44–49.

Acemoglu, D., S. Naidu, P. Restrepo, and J.A. Robinson. 2019. Democracy does cause growth. *Journal of Political Economy* 127 (1): 47–100.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Askildsen, J.E., B.H. Baltagi, and T.H. Holmås. 2003. Wage policy in the health care sector: A panel data analysis of nurses' labour supply. *Health Economics* 12: 705–719.

Bai, J., B.H. Baltagi, and M.H. Pesaran. (eds.). 2016. Cross-sectional dependence in panel data models: A special issue. *Journal of Applied Econometrics* 31: 1–297.

Baltagi, B.H. 2009. *A companion to econometric analysis of panel data*. Chichester: John Wiley.

Baltagi, B.H. 2013. *Panel Data Forecasting*, Chapter 18 in the *Handbook of Economic Forecasting*, Vol. 2B, Graham Elliott and Allan Timmermann, (eds.). North Holland, Amsterdam, 995–1024.

Baltagi, B.H., and J.M. Griffin. 1983. Gasoline demand in the OECD: An application of pooling and testing procedures. *European Economic Review* 22: 117–137.

Baltagi, B.H., and D. Levin. 1986. Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered. *Review of Economics and Statistics* 68: 148–155.

Baltagi, B.H., and E. Maasoumi. 2013. An overview of dependence in cross-section, time-series and panel data. *Econometric Reviews* 32: 543–546.

Banerjee, A., M. Marcellino, and C. Osbat. 2005. Testing for PPP: Should we use panel methods? *Empirical Economics* 30: 77–91.

Belotti, F., G. Hughes, and A. Piano Mortari. 2017. Spatial panel data models using Stata. *The Stata Journal* 17(1): 139–180.

Bonhomme, S., and L. Davezies. 2019. 2019: Panel data, old and new author(s). *Annals of Economics and Statistics* 134: 1–4.

Born, B., and J. Breitung. 2016. Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews* 35: 1290–1316.

Bruckner, M. 2013. On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics* 28: 126–150.

Coe, D., and E. Helpman. 1995. International R&D spillovers. *European Economic Review* 39: 859–887.

Cornwell, C., and P. Rupert. 1988. Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 3: 149–155.

Cornwell, C., and W.N. Trumbul. 1994. Estimating the economic model of crime with panel data. *Review of Economics and Statistics* 76: 360–366.

Cruz-Gonzalez, M., I. Fernandez-Val, and M. Weidner. 2017. Bias corrections for probit and logit models with two-way fixed effects. *The Stata Journal* 17(3): 517–545.

Culver, S.E., and D. Papell. 1997. Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics* 12: 435–444.
Egger, P., and M. Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19: 227–246.
Fernandez-Val, I. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150: 71–85.
Fernandez-Val, I., and M. Weidner. 2016. Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics* 192: 291–312.
Geil, P., A. Million, R. Rotte, and K.F. Zimmermann. 1997. Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics* 12: 295–311.
Glick, R., and A.K. Rose. 2002. Does a currency union affect trade? The time series evidence. *European Economic Review* 46: 1125–1151.
Greene, W.H. 2003. *Econometric Analysis*. New Jersey: Prentice Hall.
Grunfeld, Y. 1958. The determinants of corporate investment, unpublished Ph.D. dissertation (University of Chicago, Chicago).
Hansen, P., and A. King. 1996. The determinants of health care expenditure: A cointegration approach. *Journal of Health Economics* 15: 127–137.
Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.
Hausman, J.A., B.H. Hall, and Z. Griliches. 1984. Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* 52: 909–938.
Hutchison, M.M., and I. Noy. 2005. How bad are twins? Output costs of currency and banking crises. *Journal of Money, Credit, and Banking* 37: 725–752.
Luintel, K.B. 2000. Real exchange rate behaviour: Evidence from black markets. *Journal of Applied Econometrics* 15: 161–185.
Munnell, A. 1990. How does public infrastructure affect regional economic performance? *New England Economic Review* 3–22.
Neumayer, E. 2003. Are left-wing party strength and corporatism good for the environment? Evidence from panel analysis of air pollution in OECD countries. *Ecological Economics* 45: 203–220.
Nickell, S., Nunziata, L., and W. Ochel. 2005. Unemployment in the OECD since the 1960s. What do we know? *Economic Journal* 115: 1–27.
Ram, R. 2009. Openness, country size, and government size: Additional evidence from a large cross-country panel. *Journal of Public Economics* 93: 213–218.
Ruhm, C.J. 1996. Alcohol policies and highway vehicle fatalities. *Journal of Health Economics* 15: 435–454.
Sarafidis, V., and T. Wansbeek. 2021. Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics* 220: 215–226.
Schaller, H. 1990. A re-examination of the q theory of investment using U.S. firm data. *Journal of Applied Economics* 5: 209–325.
Serlenga, L., and Y. Shin. 2007. Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics* 22: 361–381.
Vella, F., and M. Verbeek. 1998. Whose wages do unions raise? A dynamic model of unionism and wage determination for young men. *Journal of Applied Econometrics* 13: 163–168.
Winkelmann, R. 2004. Health care reform and the number of doctor visits-an econometric analysis. *Journal of Applied Econometrics* 19: 455–472.
Wooldridge, J.M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20: 39–54.

Contents

1 Introduction

> "The increased availability of panel data from household surveys has been one of the most important developments in applied social research in the last thirty years."
>
> *Fitzgerald, Gottschalk and Moffitt (1998, p. 252)*

1.1 Panel Data: Some Examples

In this book, the term "panel data" refers to the pooling of observations on a cross-section of households, countries, firms, etc., over several time periods. This can be achieved by surveying a number of households or individuals and following them over time. The latter are known as *micro-panels* and are collected for a large number of N individuals (usually in the hundreds or thousands) over a short time period T (varying from a minimum of two years to a maximum rarely exceeding 10 or 20). In contrast, *macro-panels* usually involve a number of countries over time. These may have a moderate size N (varying from 7 countries say for the G7 countries to a larger set of say 20 OECD or European Union countries, or a mix of developed and developing countries, which could be as large as 100 or 200). These are usually observed annually over 20 to 60 years. Micro- and macro-panels require different econometric care. For example, the asymptotics for micro-panels has to be for large N and fixed T, whereas the asymptotics for macro-panels can be for large N and T. Also, with a long time series for macro-panels one has to deal with issues of nonstationarity in the time series, like unit-roots, structural breaks, and cointegration, see Chap. 12, whereas for micro-panels one does not deal with nonstationarity issues, especially since T is short for each individual or household. Also, in macro-panels, one has to deal with cross-country dependence. These are not usually an issue in micro-panels where the households are randomly sampled and

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_1

hence not likely correlated. However, Chap. 13 studies spatial dependence in panel data as a simple way to model externalities and spillovers across cross-sectional units.

1.1.1 Examples of Micro-panels

Two well-known examples of US micro-panel data are the Panel Study of Income Dynamics (PSID) collected by the Institute for Social Research at the University of Michigan (https://psidonline.isr.umich.edu) and the National Longitudinal Surveys (NLS) which is a set of surveys sponsored by the Bureau of Labor Statistics (https://www.bls.gov/nls/home.htm).

PSID began in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. It is the World's longest running household panel survey. The central focus of the data is economic and demographic. Information on these individuals and their descendants has been collected continuously, including data covering employment, income, poverty status, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, public assistance in the form of food or housing, other financial matters (e.g., taxes and inter-household transfers), family structure and demographic measures, housework time, housing, and numerous other topics. PSID is directed by faculty at the University of Michigan, and the data are available on the PSID website without cost to researchers and analysts.

NLS, on the other hand, are a set of surveys designed to gather information at multiple points in time on labor market activities and other significant life events of several groups of men and women. These include

(1) The NLSY 97 consisting of a nationally representative sample of young men and women who were 12–17 years old as of 1997.
(2) The NLSY 79 consisting of a nationally representative sample of young men and women who were 14–22 years old in 1979.
(3) The NLSY 79 children and young adults which includes the biological children born to women in the NLSY 79.

The list of variables includes information on schooling and career transitions, marriage and fertility, training investments, child care usage, and drug and alcohol use. A large number of studies have used the NLS and PSID data sets. The PSID applications cover a wide range of topics including inter-temporal models of labor supply; wages and employment over the business cycle; unemployment, job turnover, and labor mobility; consumption, income, and balance sheet dynamics; extended family behavior; poverty, welfare, and income dynamics; intergenerational transmission of economic status; and antecedents of economic and demographic events.

Panels can also be constructed from the Current Population Survey (CPS), a monthly national household survey of about 50,000 households conducted by the

Bureau of Census for the Bureau of Labor Statistics (www.census.gov/cps). CPS is the primary source of information on the labor force characteristics of the U.S. population. Compared with the NLS and PSID data, CPS contains fewer variables, spans a shorter period, and does not follow movers. However, it covers a much larger sample and is representative of all demographic groups. CPS provides estimates of employment, unemployment, earnings, hours of work, and other indicators. These are available by a variety of demographic characteristics including age, sex, race, marital status, and educational attainment. They are also available by occupation, industry, and class of worker.

Another important source of household survey data for developing countries is the World Bank's Living Standards Measurement Study (LSMS) which was established in the early 1980s (www.worldbank.org/LSMS). Since 1985, LSMS has conducted surveys in about 20 developing countries from Albania to Vietnam. These tend to be small samples of the order of 2000 to 5000 households. In some countries this could be one survey or multiple surveys. In other countries it could be a two to a four-year panel. Three types of questionnaires were conducted: a household, a community, and a price questionnaire. In some cases, a school or health facility questionnaire was added. The LSMS data has focused mostly on documenting regularities concerning the nature of poverty. Repeated surveys, like the LSMS, even though may not constitute a genuine panel, can be used to construct a *pseudo-panel* as we will see in Chap. 10.

Although the US panels started in the 1960s, it was only in the 1980s that the European panels began setting up. In 1989, a special section of *European Economic Review* published papers using the German Social Economic Panel, the Swedish study of household market and nonmarket activities, and the Intomart Dutch panel of households. The first wave of the German Socio-Economic Panel (GSOEP) was collected by the DIW (German Institute for Economic Research, Berlin) in 1984 and included 5921 West German households (www.diw.de/soep). This included 12290 respondents. Standard demographic variables as well as wages, income, benefit payments, level of satisfaction with various aspects of life, hopes and fears, political involvement, etc., are collected. In 1990, 4453 adult respondents in 2179 households from East Germany were included in GSOEP due to German unification. The attrition rate has been relatively low in GSOEP. Wagner, Burkhauser and Behringer (1993) report that through eight waves of the GSOEP, 54.9% of the original panel respondents have records without missing years. The British Household Panel Survey (BHPS) is an annual survey of private households in Britain first collected in 1991 by the Institute for Social and Economic Research at the University of Essex (www.iser.essex.ac.uk/ulsc/bhps/). This is a national representative sample of some 5500 households and 10,300 individuals drawn from 250 areas of Great Britain. In 1999, additional samples of 1,500 households in each of Scotland and Wales were added to the main sample, as well as a sample of 2,000 households in 2001 from Northern Ireland. Data collected includes demographic and household characteristics, household organization, labor market, health, education, housing, consumption, and income, and social and political values. The Swedish Panel Study Market and Non-market Activities (HUS) were collected in

1984, 1986, 1988, 1991, 1993, 1996, and 1998 (https://snd.gu.se/en/catalogue/study/SND0277). Data were collected on child care, housing, market work, income and wealth, tax reform (1993), willingness to pay for a good environment (1996), local taxes, public services, and activities in the black economy (1998).

The European Community Household Panel (ECHP) is centrally designed and coordinated by the Statistical Office of the European Communities (EuroStat), (https://ec.europa.eu/eurostat/web/microdata/european-community-household-panel). ECHP spans 8 years, running from 1994 to 2001. This involved the member states including Belgium, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Luxembourg, the Netherlands, Austria, Portugal, Sweden, and the United Kingdom. The project was launched to obtain comparable information across member countries on income, work and employment, poverty and social exclusion, housing, health, and many other diverse social indicators indicating living conditions of private households and persons. Other panel studies include the following: The Russian Longitudinal Monitoring Survey (RLMS) collected in 1992 by the Carolina Population Center at the University of North Carolina (https://www.cpc.unc.edu/projects/rlms-hse/index.html). RLMS is a nationally representative household survey designed to measure the effects of Russian reforms on economic well-being. Data includes individual health and dietary intake, measurement of expenditures and service utilization, and community-level data including region-specific prices and community infrastructure. The Korea Labor and Income Panel Study (KLIPS) is available since 1998 (https://www.kli.re.kr/klips). The Household, Income and Labour Dynamics in Australia (HILDA) is a household panel survey whose first wave was conducted by Melbourne Institute of Applied Economic and Social Research in 2001 (https://melbourneinstitute.unimelb.edu.au/hilda). The Indonesia Family Life Survey (https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html), whose sample is representative of about 83% of the Indonesian population and contains over 30,000 individuals living in 13 of the 26 provinces in the country. This list of panel data sets is by no means exhaustive but provides a good selection of panel data sets readily accessible for economic research.

1.1.2 Examples of Macro-panels

In contrast to micro-panel surveys, there are several *macro-panels* for countries over time, and hence they have to be expressed in the same currency and in real terms. These include (i) The Penn World Table (PWT) available at (https://www.rug.nl/ggdc/productivity/pwt/). PWT version 9.1 provides purchasing power parity and national income accounts converted to international prices for 182 countries for some or all of the years 1950–2017. In addition, the European Union or the OECD provide detailed purchasing power and real product estimates for their countries and the World Bank makes current price estimates for most PWT countries at the GDP level. (ii) The World Bank is a great source of macro-panels including the World Development Indicators (WDI) available at (https://databank.worldbank.org/source/world-development-indicator). (iii) The International Monetary Fund (http://www.

imf.org) provides several sources of macro-panel data. These include World Economic Outlook Databases and International Financial Statistics which provide time-series data for GDP growth, inflation, unemployment, payments balances, exports, imports, external debt, capital flows, commodity prices, etc., IMF Statistics Data, Principal Global Indicators, and Global Housing Watch. The latter is a website that tracks developments in housing markets around the world: Balance of Payments Statistics, Direction of Trade Statistics, Government Financial Statistics, among others. This is a rich source that includes exchange rates, fund accounts, and the main global and country economic indicators. (iv) United Nations provides a wealth of macro-country panel data at (https://unstats.un.org/databases.htm). (v) The Organization for Economic Co-operation and Development (OECD) data is available at (http://www.oecd.org). (vi) The European Central Bank (ECB) provides data on the European Union member countries at (http://www.ecb.int). (vii) The Central Intelligence Agency's World Factbook is available on the Web at https://www.cia.gov/library/publications/resources/the-world-factbook/index.html.

These are but few of the agencies providing macro-data on individual countries over time, which can be pooled and used in panel studies.

We will study several types of panel data encountered in practice including *unbalanced panels* in Chap. 9, *nested panels* in Sect. 9.7, *unequally spaced panels* in Sect. 5.2.5, *rotating panels* in Sect. 10.2, *pseudo-panels* in Sect. 10.3, *spatial panels* in Chap. 13, *count panels* in Sect. 10.6, and *heterogeneous panels* in Sect. 10.5.

1.1.3 Some Basic References

Virtually, every graduate text in econometrics contains a chapter or a major section on the econometrics of panel data. Recommended readings on this subject include Hsiao's (2003) Econometric Society monograph along with two chapters in the Handbook of Econometrics: Chapter 22 by Chamberlain (1984) and chapter 53 by Arellano and Honoré (2001). Maddala (1993) edited two volumes collecting some of the classic articles on the subject. This collection of readings was updated with two more volumes covering the period 1992–2002 and edited by Baltagi (2002). Other books on the subject include Arellano (2003), Wooldridge (2010), and a handbook on the econometrics of panel data edited by Mátyás and Sevestre (2008) and more recently by Baltagi (2015). Special issues of journals dedicated to panel data include two special issues of *Journal of Econometrics*. The first one edited by Baltagi (1995) and a more recent one by Sarafidis and Wansbeek (2021). Two volumes of *Annales D'Economie et de Statistique* edited by Sevestre (1999), and a more recent one in the *Annals of Economics and Statistics* by Bonhomme and Davezies (2019). A special issue of *Oxford Bulletin of Economics and Statistics* edited by Banerjee (1999). Three special issues of *Econometric Reviews*. Two were edited by Maasoumi and Heshmati (2000) and the third by Baltagi and Maasoumi (2013). A special issue of *Advances in Econometrics* edited by Baltagi, Fomby and Hill (2000). Two special issues of *Empirical Economics*. One edited by Baltagi

(2004) and the second by Baltagi and Breitung (2011). Two special issues of the *Journal of Applied Econometrics*. The first one edited by Baltagi and Pesaran (2007) and the second by Bai, Baltagi and Pesaran (2016).

The objective of this book is to provide a simple introduction to some of the basic issues of panel data analysis. It is intended for economists and social scientists with the usual background in statistics and econometrics. Panel data methods have been used in political science, see Beck and Katz (1995), in sociology, finance, and marketing; see Keane (2015). While restricting the focus of the book to basic topics may not do justice to this rapidly growing literature, it is nevertheless unavoidable in view of the space limitations of the book. Topics not covered in this book include duration models and hazard functions (see Heckman and Singer 1985), and also the frontier production function literature using panel data (see Kumbhakar and Lovell 2000; Koop and Steel 2001), the literature on time-varying parameters, random coefficients, and Bayesian models, see Swamy and Tavlas (2001) and Hsiao (2003), and nonparametric and semi-parametric panels; see Li and Racine (2007).

1.2 Why Should We Use Panel Data? Their Benefits and Limitations

Hsiao (2003) lists several benefits from using panel data. These include the following:

(1) Controlling for *individual heterogeneity*. Panel data suggests that individuals, firms, states, or countries are heterogeneous. Time-series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results, e.g., see Moulton (1986, 1987). Let us demonstrate this with an empirical example. Baltagi and Levin (1986) consider panel data estimation of cigarette demand across 46 American states. Consumption is modeled as a function of lagged consumption, price, and income. These variables vary with states and time. However, there are a lot of other variables that may be state-invariant or time-invariant that may affect consumption. Let us call these Z_i and W_t, respectively. Examples of Z_i are religion and education. For the religion variable, one may not be able to get the percentage of the population that is, say, Mormon in each state for every year, nor does one expect that to change much across time. The same holds true for the percentage of the population completing high school or a college degree. Examples of W_t include advertising on TV and radio. This advertising is nationwide and does not vary across states. In addition, some of these variables are difficult to measure or hard to obtain so that not all the Z_i or W_t variables are available for inclusion in the consumption equation. Omission of these variables leads to bias in the resulting estimates. Panel data are able to control for these state- and time-invariant variables whereas a time-series study or a cross-section study cannot. In fact, from the data one observes that Utah has less than half the average per capita consumption of cigarettes in the USA. This is because it is mostly a Mormon state, a religion that

prohibits smoking. Controlling for Utah in a cross-section regression may be done with a dummy variable which has the effect of removing that state's observation from the regression. This would not be the case for panel data as we will shortly discover. In fact, with panel data, one might first difference the data to get rid of all Z_i type variables and hence effectively control for all state-specific characteristics. This holds whether the Z_i are observable or not. Alternatively, the dummy variable for Utah controls for every state-specific effect that is distinctive of Utah without omitting the observations for Utah.

Another example is given by Hajivassiliou (1987) who studies the external debt repayments problem using a panel of 79 developing countries observed over the period 1970–82. These countries differ in terms of their colonial history, financial institutions, religious affiliations, and political regimes. All of these country-specific variables affect the attitudes that these countries have with regards to borrowing and defaulting and the way they are treated by the lenders. Not accounting for this country heterogeneity causes serious misspecification.

Deaton (1995) gives another example from agricultural economics. This pertains to the question of whether small farms are more productive than large farms. OLS regressions of yield per hectare on inputs such as land, labor, fertilizer, and farmer's education usually find that the sign of the estimate of the land coefficient is negative. These results imply that smaller farms are more productive. Some explanations from economic theory argue that higher output per head is an optimal response to uncertainty by small farmers, or that hired labor requires more monitoring than family labor. Deaton (1995) offers an alternative explanation. This regression suffers from the omission of unobserved heterogeneity; in this case "land quality", and this omitted variable is systematically correlated with the explanatory variable (farm size). In fact, farms in low-quality marginal areas (semi-desert) are typically large, while farms in high-quality land areas are often small. Deaton argues that while gardens add more value per hectare than a sheep station; this does not imply that sheep stations should be organized as gardens. In this case, differencing may not resolve the "small farms are productive" question since farm size will usually change little or not at all over short periods.

(2) Panel data give *more informative data, more variability, less collinearity among the variables, more degrees of freedom, and more efficiency*. Time-series studies are plagued with multicollinearity; for example, in the case of demand for cigarettes above, there is high collinearity between price and income in the aggregate time series for the USA. This is less likely with a panel across American states since the cross-section dimension adds a lot of variability, adding more informative data on price and income. In fact, the variation in the data can be decomposed into variation between states of different sizes and characteristics, and variation within states. The former variation is usually bigger. With additional, more informative data one can produce more reliable parameter estimates. Of course, the same relationship has to hold for each state, i.e., the data have to be poolable. This is a testable assumption and one that we will tackle in due course.

(3) Panel data are better able to study the *dynamics of adjustment*. Cross-sectional distributions that look relatively stable hide a multitude of changes. Spells of unemployment, job turnover, residential and income mobility are better studied with panels. Panel data are also well suited to study the duration of economic states like unemployment and poverty, and if these panels are long enough, they can shed light on the speed of adjustments to economic policy changes. For example, in measuring unemployment, cross-sectional data can estimate what proportion of the population is unemployed at a point in time. Repeated cross-sections can show how this proportion changes over time. Only panel data can estimate what proportion of those who are unemployed in one period can remain unemployed in another period. Important policy questions like determining whether families' experiences of poverty, unemployment, and welfare dependence are transitory or chronic necessitate the use of panels. Deaton (1995) argues that, unlike cross-sections, panel surveys yield data on *changes* for individuals or households. It allows us to observe *how* the individual living standards change during the development process. It enables us to determine *who* is benefiting from development. It also allows us to observe whether poverty and deprivation are transitory or long-lived, the income-dynamics question. Panels are also necessary for the estimation of intertemporal relations, life-cycle and intergenerational models. In fact, panels can relate the individual's experiences and behavior at one point in time to other experiences and behavior at another point in time. For example, in evaluating training programs, a group of participants and non-participants are observed before and after the implementation of the training program. This is a panel of at least two time periods and the basis for the "difference-in-differences" estimator; see Chap. 2.

(4) Panel data are better able to *identify and measure effects that are simply not detectable in pure cross-section or pure time-series data*. Suppose that we have a cross-section of women with a 50% average yearly labor force participation rate. This might be due to (a) each woman having a 50% chance of being in the labor force, in any given year, or (b) 50% of the women working all the time and 50% not at all. Case (a) has high turnover, while case (b) has no turnover. Only panel data could discriminate between these cases. Another example is the determination of whether union membership increases or decreases wages. This can be better answered as we observe a worker moving from union to nonunion jobs or vice versa. Holding the individual's characteristics constant, we will be better equipped to determine whether union membership affects wage and by how much. This analysis extends to the estimation of other types of wage differentials holding individuals' characteristics constant, for example, the estimation of wage premiums paid in dangerous or unpleasant jobs.

Economist studying workers level of satisfaction run into the problem of anchoring in a cross-section study; see Winkelmann and Winkelmann (1998) in Chap. 11. The survey usually asks the question: "how satisfied are you with your life"? with zero meaning completely dissatisfied and 10 meaning completely satisfied. The problem is that each individual anchors their scale at different levels, rendering interpersonal comparisons of responses meaningless. However, in a panel study, where the metric used by individuals is time-invariant over the period of observation, one can avoid this problem since a difference (or fixed effects) estimator will make inference based only on intra rather than interpersonal comparison of satisfaction.

(5) Panel data models allow us to *construct and test more complicated behavioral models than purely cross-section or time-series data*. For example, technical efficiency is better studied and modeled with panels (see Kumbhakar and Lovell 2000, and Koop and Steel 2001).

(6) Micro-panel data gathered on individuals, firms, and households may be more accurately measured than similar variables measured at the macro level. *Biases resulting from aggregation over firms or individuals may be reduced or eliminated*.

(7) Macro-panel data on the other hand have a longer time series and unlike the problem of nonstandard distributions typical of unit roots tests in time-series analysis; Chap. 12 shows that panel unit root tests have standard asymptotic distributions.

Limitations of panel data include

(1) *Design and data collection problems.* For an extensive discussion of problems that arise in designing panel surveys as well as data collection and data management issues, see Kasprzyk et al. (1989). These include problems of coverage (incomplete account of the population of interest), nonresponse (due to lack of cooperation of the respondent or because of interviewer error), recall (respondent not remembering correctly), frequency of interviewing, interview spacing, reference period, the use of bounding and time-in-sample bias.[1]
(2) *Distortions of measurement errors*. Measurement errors may arise because of faulty responses due to unclear questions, memory errors, deliberate distortion of responses (e.g., prestige bias), inappropriate informants, and misrecording of responses and interviewer effects (see Kalton, Kasprzyk and McMillen 1989). The validation study by Duncan and Hill (1985) on PSID illustrates the significance of the measurement error problem. They compare the responses of the employees of a large firm with the records of the employer. They find small response biases except for work hours which are overestimated. The ratio of measurement error variance to the true variance is found to be 15% for annual earnings, 37% for annual work hours, and 184% for average hourly earnings. These figures are for a one-year recall, i.e., 1983 for 1982, and are more than doubled with two years' recall. Brown and Light (1992) investigate the inconsistency in job tenure responses in PSID and NLS. Cross-section data users have little choice but to believe the reported values of tenure (unless they have external information) while users of panel data can check for inconsistencies of tenure responses with elapsed time between interviews. For example, a respondent may claim to have three years of tenure in one interview and a year later claim six years. This should alert the user of this panel to the presence of measurement error. Brown and Light (1992) show that failure to

use internally consistent tenure sequences can lead to misleading conclusions about the slope of wage-tenure profiles. Section 10.1 deals with measurement error in panel data.

(3) *Selectivity problems*. These include

  (a) *Self-selectivity*. People choose not to work because the reservation wage is higher than the offered wage. In this case, we observe the characteristics of these individuals but not their wage. Since only their wage is missing, the sample is censored. However, if we do not observe all data on these people, this would be a truncated sample. An example of truncation is the New Jersey negative income tax experiment. We are only interested in poverty, and people with income larger than 1.5 times the poverty level are dropped from the sample. Inference from this truncated sample introduces bias that is not helped by more data, because of the truncation (see Hausman and Wise, 1979). Chapter 11 deals with selectivity problems in panel data.
  (b) *Nonresponse*. This can occur at the initial wave of the panel due to refusal to participate, nobody at home, untraced sample unit, and other reasons. Item (or partial) nonresponse occurs when one or more questions are left unanswered or are found not to provide a useful response. Complete nonresponse occurs when no information is available from the sampled household. Besides the efficiency loss due to missing data, this nonresponse can cause serious identification problems for the population parameters. The seriousness of the problem is directly proportional to the amount of nonresponse. Nonresponse rates in the first wave of the European panels vary across countries from 10% in Greece and Italy where participation is compulsory to 52% in Germany and 60% in Luxembourg. The overall nonresponse rate is 28%; see Peracchi (2002). The comparable nonresponse rate for the first wave of the PSID is 24%, for the BHPS (26%) and for the GSOEP (38%).
  (c) *Attrition*. While nonresponse occurs also in cross-section studies, it is a more serious problem in panels because subsequent waves of the panel are still subject to nonresponse. Respondents may die, or move, or find that the cost of responding is high. See Chap. 11 on the consequences of attrition in panels. The degree of attrition varies depending on the panel studied; see Kalton, Kasprzyk and McMillen (1989) for several examples. In general, the overall rates of attrition increase from one wave to the next, but the rate of increase declines over time. Becketti et al. (1988) study the representativeness of the PSID 14 years after it had started. The authors find that only 40% of those originally in the sample in 1968 remained in the sample in 1981. However, they do find that as far as the dynamics of entry and exit are concerned, PSID is still representative. The most potentially damaging threat to the value of panel data is the presence of biasing attrition. Fitzgerald, Gottschalk and Moffitt (1998) report that by 1989, 51% of the original sample had attrited. The major reasons were

family unit nonresponse, death, or because of a residential move. Attritors were found to have lower earnings, lower education levels, and lower marriage propensities. Despite the large amount of attrition, Fitzgerald, Gottschalk and Moffitt (1998) report that there is no strong evidence that this attrition had seriously distorted the representativeness of PSID through 1989. In the same vain of research, Lillard and Panis (1998) find evidence of significant selectivity in attrition for PSID. For example, they find that less-educated individuals and older people are more likely to drop out. Married people are more likely to continue. This propensity to participate in the survey diminishes the longer the duration of the respondent in the sample. Despite this, the effects of ignoring this selective attrition on household income dynamics, marriage formation and dissolution, and adult mortality risk are mild. In Europe, the comparable attrition rates (between the first and second wave) vary from 6% in Italy to 24% in the UK. The average attrition rate is about 10%. For BHPS, attrition from the first to the second wave is 12%. For GSOEP, attrition is 12.4% for the West German sample and 8.9% for the East German sample; see Peracchi (2002). In order to counter the effects of attrition, rotating panels are sometimes used, where a fixed percentage of the respondents are replaced in every wave to replenish the sample. More on rotating and pseudo-panels in Chap. 10. A special issue of the *Journal of Human Resources*, Spring 1998, is dedicated to attrition in longitudinal surveys.

(4) *Short time-series dimension*. Typical micro-panels involve annual data covering a short time span for each individual. This means that asymptotic arguments rely crucially on the number of individuals tending to infinity. Increasing the time span of the panel is not without cost either. In fact, this increases the chances of attrition and increases the computational difficulty for limited dependent variable panel data models (see Chap. 11).

(5) *Cross-section dependence.* Macro-panels on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference. Chapter 12 shows that several panel unit-root tests suggested in the literature assumed cross-section independence. Accounting for cross-section dependence turns out to be important and affects inference. Alternative panel unit-root tests are suggested that account for this dependence. Chapter 13 surveys tests for cross-sectional dependence in panels.

Panel data is not a panacea and will not solve all the problems that a time-series or a cross-section study could not handle. Examples are given in Chap. 12, where we cite econometric studies arguing that panel data will yield more powerful unit root tests than individual time series. This in turn should help shed more light on the purchasing power parity (PPP) and the growth convergence questions. In fact, this lead to a flurry of empirical applications along with some skeptics who argued that panel data did not save the PPP or the growth convergence problem; see Maddala, Wu and Liu (2000) and Banerjee, Marcellino and Osbat (2004, 2005). Collecting panel data is quite costly, and there is always the question of how often should one interview respondents. Deaton (1995) argues that economic development is far from instantaneous, so that changes from one year to the next are probably too noisy and too short term to be really useful. He concludes that the payoff for panel data is over long time periods, five years, ten years, or even longer. In contrast, for health and nutrition issues, especially those of children, one could argue the opposite case, i.e., those panels with a shorter time span are needed in order to monitor the health and development of these children.

This book will make the case that panel data provides several advantages worth its cost. However, as Zvi Griliches argued about economic data in general, the more we have of it, the more we demand of it. The economist using panel data or any data for that matter has to know its limitations.

1.3 Note

1. Bounding is used to prevent the shifting of events from outside the recall period into the recall period. Time-in-sample bias is observed when a significantly different level for a characteristic occurs in the first interview than in later interviews, when one would expect the same level.

References

Arellano, M. 2003. *Panel data econometrics*. Oxford: Oxford University Press.

Arellano, M., and B. Honoré. 2001. Panel data models: Some recent developments. Chapter 53. In *Handbook of econometrics*, (eds.) J. Heckman and E. Leamer, 3229–3296. North-Holland, Amsterdam.

Bai, J., B.H. Baltagi, and M.H. Pesaran. (eds.). 2016. Cross-sectional dependence in panel data models: A special issue. *Journal of Applied Econometrics* 31: 1-297.

Baltagi, B.H. 1995. Introduction: Panel data. *Journal of Econometrics* 68: 1–4.

Baltagi, B.H. (ed.). 2002. *Recent developments in the econometrics of panel data, volumes I and II*. Cheltenham: Edward Elgar Publishing.

Baltagi, B.H. (ed.). 2004. Editor's introduction: Panel data, theory and application. *Empirical Economics* 29: 1–4.

Baltagi, B.H. (ed.). 2015. *Handbook of panel data*. Oxford: Oxford University Press.

Baltagi, B.H., and J. Breitung. 2011. Editors' introduction: Panel data econometrics. *Empirical Economics* 40: 1–4.

Baltagi, B.H., T.B. Fomby, and R.C. Hill (eds.). 2000. *Nonstationary panels, panel cointegration, and dynamic panels, Advances in Econometrics*, vol. 15. Amsterdam: Elsevier Science.

Baltagi, B.H., and D. Levin. 1986. Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered. *Review of Economics and Statistics* 68: 148–155.

Baltagi, B.H., and E. Maasoumi. 2013. An overview of dependence in cross-section, time-series and panel data. *Econometric Reviews* 32: 543–546.

Baltagi, B.H., and M.H. Pesaran. 2007. Heterogeneity and cross section dependence in panel data models: Theory and applications introduction. *Journal of Applied Econometrics* 22: 229–232.

Banerjee, A. 1999. Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics* 61: 607–629.

Banerjee, A., M. Marcellino, and C. Osbat. 2004. Some cautions on the use of panel methods for integrated series of macro-economic data. *Econometrics Journal* 7: 322–340.

Banerjee, A., M. Marcellino, and C. Osbat. 2005. Testing for PPP: Should we use panel methods? *Empirical Economics* 30: 77-91.

Beck, N., and J. Katz. 1995. What to do (and not to do) with time-series-cross-section data in comparative politics. *American Political Science Review* 89: 634–647.

Becketti, S., W. Gould, L. Lillard, and F. Welch. 1988. The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics* 6: 472–492.

Bonhomme, S., and L. Davezies. 2019. 2019: Panel data, old and new author(s). *Annals of Economics and Statistics* 134: 1–4.

Brown, J.N., and A. Light. 1992. Interpreting panel data on job tenure. *Journal of Labor Economics* 10: 219–257.

Chamberlain, G. 1984. Panel data, Chapter 22. In *Handbook of Econometrics*, (ed.) Z. Griliches, and M. Intrilligator, 1247–1318. Amsterdam: North-Holland.

Deaton, A. 1995. Data and econometric tools for development analysis, Chapter 33 in Behrman, J., and Srinivasan T.N. (eds.). *Handbook of development economics*. Amsterdam: Elsevier Science, 1785–1882.

Duncan, G.J., and D.H. Hill. 1985. An investigation of the extent and consequences of measurement error in labor economic survey data. *Journal of Labor Economics* 3: 508–532.

Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. An analysis of sample attrition in panel data: The Michigan panel study of income dynamics. *The Journal of Human Resources* 33: 251–299.

Hajivassiliou, V.A. 1987. The external debt repayments problems of LDC's: An econometric model based on panel data. *Journal of Econometrics* 36: 205–230.

Hausman, J.A., and D. Wise. 1979. Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica* 47: 455–473.

Heckman, J.J., and B. Singer. 1985. *Longitudinal analysis of labor market data*. Cambridge: Cambridge University Press.

Hsiao, C. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.

Kalton, G., D. Kasprzyk, and D. McMillen. 1989. Nonsampling errors in panel surveys. In Kasprzyk, D., Duncan, G.J., Kalton, G., and Singh M.P.(eds.). *Panel surveys*. New York: John Wiley, 249–270.

Kasprzyk, D., G.J. Duncan, G. Kalton and M.P. Singh. 1989. *Panel Surveys*. New York: John Wiley.

Keane, M. 2015. Discrete choice models of consumer demand, Chapter 18 in Baltagi, B.H. (ed.) *Oxford handbook of panel data*. Oxford: Oxford University Press, 548–582.

Koop, G., and M.F.J. Steel. 2001. Bayesian analysis of stochastic frontier models, Chapter 24 in Baltagi, B.H. (ed.) *A companion to theoretical econometrics*. Massachusetts: Blackwell Publishers, 520–537.

Kumbhakar, S.C., and C.A.K. Lovell. 2000. *Stochastic frontier analysis*. Cambridge: Cambridge University Press.

Li, Q., and J.C. Racine. 2007. *Nonparametric econometrics*. New Jersey: Princeton University Press.

Lillard, L.A., and C.W.A. Panis. 1998. Panel attrition from the panel study of income dynamics: Household income, marital status, and mortality. *The Journal of Human Resources* 33: 437–457.

Maasoumi, E., and A. Heshmati. 2000. Introduction. *Econometric Reviews* 19(4): 5-5.

Maddala, G.S. (ed.). 1993. *The econometrics of panel data, vols I and II*. Cheltenham: Edward Elgar Publishing.

Maddala, G.S., S. Wu and P.C. Liu. 2000. Do panel data rescue purchasing power parity (PPP) theory? Chapter 2 In *Panel data econometrics: Future directions*, (eds.) J. Krishnakumar and E. Ronchetti, 35–51. North-Holland, Amsterdam.

Mátyás, L., and P. Sevestre (eds.). 2008. *The econometrics of panel data: Fundamentals and recent developments in theory and practice*. Berlin: Springer.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Moulton, B.R. 1987. Diagnostics for group effects in regression analysis. *Journal of Business and Economic Statistics* 5: 275–282.

Peracchi, F. 2002. The European community household panel: A review. *Empirical Economics* 27: 63–90.

Sarafidis, V., and T. Wansbeek. 2021. Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics* 220: 215–226.

Sevestre, P. 1999. 1977–1997: Changes and continuities in panel data econometrics. *Annales D'Économie et de Statistique* 55–56: 15–25.

Swamy, P.A.V.B., and G.S. Tavlas. 2001. Random coefficient models, Chapter 19 in Baltagi, B.H. (ed.) *A companion to theoretical econometrics*. Massachusetts: Blackwell Publishers, 410–428.

Wagner, G.G., R.V. Burkhauser, and F. Behringer. 1993. The English public use file of the German socio-economic panel. *The Journal of Human Resources* 28: 429–433.

Winkelmann, L., and R. Winkelmann. 1998. Why are the unemployed so unhappy? Evidence from panel data. *Economica* 65: 1–15.

Wooldridge, J.M. 2010. *Econometric analysis of cross-section and panel data*. Massachusetts: MIT Press.

2 The One-Way Error Component Regression Model

2.1 Introduction

A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables, i.e.,

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{2.1}$$

with i denoting households, individuals, firms, countries, etc., and t denoting time. The i subscript, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. α is a scalar, β is $K \times 1$, and X_{it} is the itth observation on K explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, with

$$u_{it} = \mu_i + \nu_{it} \tag{2.2}$$

where μ_i denotes the *unobservable* individual specific effect and ν_{it} denotes the remainder disturbance. For example, in an earnings equation in labor economics, y_{it} will measure earnings of the head of the household, whereas X_{it} may contain a set of variables like experience, education, union membership, sex, race, etc. Note that μ_i is time-invariant and it accounts for any individual specific effect that is not included in the regression. In this case, we could think of it as the individual's unobserved ability. The remainder disturbance ν_{it} varies with individuals and time and can be thought of as the usual disturbance in the regression. Alternatively, for a production function utilizing data on firms across time, y_{it} will measure output and X_{it} will measure inputs. The unobservable firm-specific effects will be captured by the μ_i, and we can think of these as the unobservable entrepreneurial or managerial skills of the firm's executives. In vector form (2.1) can be written as

$$y = \alpha\iota_{NT} + X\beta + u = Z\delta + u \tag{2.3}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [\iota_{NT}, X]$, $\delta' = (\alpha', \beta')$, and ι_{NT} is a vector of ones of dimension NT. Also, (2.2) can be written as

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_2

$$u = Z_\mu \mu + \nu \tag{2.4}$$

where $u' = (u_{11}, \ldots, u_{1T}, u_{21}, \ldots, u_{2T}, \ldots, u_{N1}, \ldots, u_{NT})$ with the observations stacked such that the slower index is over individuals and the faster index is over time. $Z_\mu = I_N \otimes \iota_T$ where I_N is an identity matrix of dimension N, ι_T is a vector of ones of dimension T, and $\otimes$ denotes Kronecker product. Z_μ is a selector matrix of ones and zeros, or simply the matrix of individual dummies that one may include in the regression to estimate the μ_i if they are assumed to be fixed parameters. $\mu' = (\mu_1, \ldots, \mu_N)$ and $\nu' = (\nu_{11}, \ldots, \nu_{1T}, \ldots, \nu_{N1}, \ldots, \nu_{NT})$. Note that, $Z_\mu Z'_\mu = I_N \otimes J_T$ where J_T is a matrix of ones of dimension T, and $P = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu$; the projection matrix on Z_μ reduces to $I_N \otimes \bar{J}_T$ where $\bar{J}_T = J_T/T$. P is a matrix which averages the observation across time for each individual, and $Q = I_{NT} - P$ is a matrix which obtains the deviations from individual means. For example, regressing y on the matrix of dummy variables Z_μ gets the predicted values Py which have a typical element $\bar{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$ repeated T times for each individual. The residuals of this regression are given by Qy which have a typical element $\left(y_{it} - \bar{y}_{i.}\right)$. P and Q are (i) symmetric idempotent matrices, i.e., $P' = P$ and $P^2 = P$. This means that the rank$(P) = tr(P) = N$ and rank$(Q) = tr(Q) = N(T-1)$. This uses the result that rank of an idempotent matrix is equal to its trace (see Graybill (1961), Theorem 1.63). Also, (ii) P and Q are orthogonal, i.e., $PQ = 0$ and (iii) they sum to the identity matrix $P + Q = I_{NT}$. In fact, any two of these properties imply the third (see Graybill (1961), Theorem 1.68).

2.2 The One-Way Fixed Effects Model

In this case, the μ_i are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with v_{it} independent and identically distributed IID$\left(0, \sigma_\nu^2\right)$. The X_{it} are assumed independent of the v_{it} for all i and t. The fixed effects model is an appropriate specification if we are focusing on a specific set of N firms, say, IBM, GE, Westinghouse, etc., and our inference is restricted to the behavior of these sets of firms. Alternatively, it could be a set of N OECD countries, or N American states. Inference in this case is conditional on the particular N firms, countries or states that are observed. One can substitute the disturbances given by (2.4) into (2.3) to get

$$y = \alpha\iota_{NT} + X\beta + Z_\mu\mu + \nu = Z\delta + Z_\mu\mu + \nu \tag{2.5}$$

and then perform ordinary least squares (OLS) on (2.5) to get estimates of α, β, and μ. Note that Z is $NT \times (K+1)$ and Z_μ , the matrix of individual dummies, is $NT \times N$. If N is large, (2.5) will include too many individual dummies, and the matrix to be inverted by OLS is large and of dimension $(N+K)$. In fact, since α and β are the parameters of interest, one can obtain the least squares dummy variables (LSDV) estimator from (2.5), by premultiplying the model by Q and performing OLS on the resulting transformed model:

$$Qy = QX\beta + Qv \tag{2.6}$$

This uses the fact that $QZ_\mu = Q\iota_{NT} = 0$, since $PZ_\mu = Z_\mu$. In other words, the Q matrix wipes out the individual effects. This is a regression of $\widetilde{y} = Qy$ with typical element $(y_{it} - \bar{y}_{i.})$ on $\widetilde{X} = QX$ with typical element $(X_{it,k} - \bar{X}_{i.,k})$ for the kth regressor, $k = 1, 2, \ldots, K$. This involves the inversion of a $(K \times K)$ matrix rather than $(N + K) \times (N + K)$ as in (2.5). The resulting OLS estimator is

$$\widetilde{\beta} = \left(X'QX\right)^{-1} X'Qy \tag{2.7}$$

with $\text{var}(\widetilde{\beta}) = \sigma_\nu^2 (X'QX)^{-1} = \sigma_\nu^2 (\widetilde{X}'\widetilde{X})^{-1}$. $\widetilde{\beta}$ could have been obtained from (2.5) using results on partitioned inverse or the Frisch-Waugh-Lovell theorem discussed in Davidson and MacKinnon (1993, p. 19). This uses the fact that P is the projection matrix on Z_μ and $Q = I_{NT} - P$ (see problem 2.1). In addition, generalized least squares (GLS) on (2.6), using generalized inverse, will also yield $\widetilde{\beta}$ (see problem 2.2).

Note that for the simple regression

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \nu_{it} \tag{2.8}$$

and averaging over time gives

$$\bar{y}_{i.} = \alpha + \beta \bar{x}_{i.} + \mu_i + \bar{\nu}_{i.} \tag{2.9}$$

Therefore, subtracting (2.9) from (2.8) gives

$$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + (\nu_{it} - \bar{\nu}_{i.}) \tag{2.10}$$

Also, averaging across all observations in (2.8) gives

$$\bar{y}_{..} = \alpha + \beta \bar{x}_{..} + \bar{\nu}_{..} \tag{2.11}$$

where we utilized the restriction that $\sum_{i=1}^N \mu_i = 0$. This is an arbitrary restriction on the dummy variable coefficients to avoid the dummy variable trap, or perfect multicollinearity. In fact only β and $(\alpha + \mu_i)$ are estimable from (2.8), and not α and μ_i separately, unless a restriction like $\sum_{i=1}^N \mu_i = 0$ is imposed. In this case, $\widetilde{\beta}$ is obtained from regression (2.10), $\widetilde{\alpha} = \bar{y}_{..} - \widetilde{\beta}\bar{x}_{..}$ can be recovered from (2.11) and $\widetilde{\mu}_i = \bar{y}_{i.} - \widetilde{\alpha} - \widetilde{\beta}\bar{x}_{i.}$ from (2.9). For large labor or consumer panels, where N is very large, regressions like (2.5) may not be feasible, since one is including $(N - 1)$ dummies in the regression. This fixed effects (FE) least squares, also known as least squares dummy variables (LSDV), suffers from a large loss of degrees of freedom. We are estimating $(N - 1)$ extra parameters, and too many dummies may aggravate the problem of multicollinearity among the regressors. In addition, this FE estimator cannot estimate the effect of any time-invariant variable like sex, race, religion, and schooling or union participation. These time-invariant variables are wiped out by the Q transformation, the deviations from means transformation (see (2.10)). Alternatively, one can see that these time-invariant variables are spanned by the individual dummies in (2.5) and therefore any regression package attempting (2.5) will fail, signaling perfect multicollinearity. If (2.5) is the true model, LSDV is the best linear unbiased estimator (BLUE) as long as ν_{it} is the standard classical disturbance with mean 0 and variance–covariance matrix $\sigma_\nu^2 I_{NT}$. Note that as $T \to \infty$, the FE estimator is consistent. However, if T is fixed and $N \to \infty$ as typical in short labor panels, then only the FE estimator of β is consistent; the FE estimators of the individual effects $(\alpha + \mu_i)$ are not consistent since the number of these parameters increases as N increases. Note that when the true model is fixed effects as in (2.5), OLS on (2.1) yields biased and inconsistent estimates of the regression parameters. This is an omission variables bias due to the fact that OLS deletes the individual dummies when in fact they are relevant.

(1) *Testing for fixed effects*. One could test the joint significance of these dummies, i.e., H_0; $\mu_1 = \mu_2 = \cdots = \mu_{N-1} = 0$, by performing an F-test. (Testing for individual effects will be extensively treated in Chap. 4.) This is a simple Chow test with the restricted residual sums of squares (RRSS) being that of OLS on the pooled model and the unrestricted residual sums of squares (URSS) being that of the LSDV regression. If N is large, one can perform the Within transformation and use that residual sum of squares as the URSS. In this case

$$F_0 = \frac{(RRSS - URSS)/N - 1)}{URSS/(NT - N - K)} \overset{H_0}{\sim} F_{N-1,N(T-1)-K} \tag{2.12}$$

(2) *Computational warning*. One computational caution for those using the Within regression is given by (2.10). The s^2 of this regression as obtained from a typical regression package divides the residual sums of squares by $NT - K$ since the intercept and the dummies are not included. The proper s^2, say s^{*2} from the LSDV regression in (2.5), would divide the same residual sums of squares by $N(T - 1) - K$. Therefore, one has to adjust the variances obtained from the Within regression (2.10) by multiplying the variance–covariance matrix by (s^{*2}/s^2) or simply by multiplying by $[NT - K]/[N(T - 1) - K]$.

(3) *Robust estimates of the standard errors*. For the Within estimator, Arellano (1987) suggests a simple method for obtaining robust estimates of the standard errors that allow for a general variance–covariance matrix on the ν_{it} as in White (1980). One would stack the panel as an equation for each individual:

$$y_i = Z_i\delta + \mu_i\iota_T + \nu_i \tag{2.13}$$

where y_i is $(T \times 1)$, $Z_i = [\iota_T, X_i]$, X_i is $(T \times K)$, μ_i is a scalar, $\delta' = (\alpha, \beta')$, ι_T is a vector of ones of dimension T, and ν_i is $(T \times 1)$. In general, $E(\nu_i\nu_i') = \Omega_i$ for $i = 1, 2, \ldots, N$, where Ω_i is a positive definite matrix of dimension T. We still assume $E(\nu_i\nu_j') = 0$, for $i \neq j$. T is assumed small and N large as in household or company panels, and the asymptotic results are performed for $N \to \infty$ and T fixed. Performing the Within transformation on this set of equations (2.13), one gets

$$\widetilde{y}_i = \widetilde{X}_i\beta + \widetilde{\nu}_i \tag{2.14}$$

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$, and $\widetilde{\nu} = Q\nu$, with $\widetilde{y} = (\widetilde{y}_1', \ldots, \widetilde{y}_N')'$ and $\widetilde{y}_i = (I_T - \bar{J}_T)y_i$. Computing robust least squares on this system, as described by White (1980), under the restriction that each equation has the same β one gets the Within estimator of β which has the following asymptotic distribution:

$$N^{1/2}(\widetilde{\beta} - \beta) \sim N(0, M^{-1}VM^{-1}) \tag{2.15}$$

where $M = \text{plim}(\widetilde{X}'\widetilde{X})/N$ and $V = \text{plim}\sum_{i=1}^{N}(\widetilde{X}_i'\Omega_i\widetilde{X}_i)/N$. Note that $\widetilde{X}_i = (I_T - \bar{J}_T)X_i$ and $\widetilde{X}'Q\ \text{diag}[\Omega_i]Q\widetilde{X} = \widetilde{X}'\text{diag}[\Omega_i]\widetilde{X}$ (see problem 2.3). In this case, V is estimated by $\widetilde{V} = \sum_{i=1}^{N} \widetilde{X}_i' \widetilde{u}_i \widetilde{u}_i' \widetilde{X}_i / N$, where $\widetilde{u}_i = \widetilde{y}_i - \widetilde{X}_i \widetilde{\beta}$. Therefore, the robust asymptotic variance–covariance matrix of β is estimated by

$$\operatorname{var}(\widetilde{\beta}) = (\widetilde{X}'\widetilde{X})^{-1} \left[\sum_{i=1}^{N} \widetilde{X}_i' \widetilde{u}_i \widetilde{u}_i' \widetilde{X}_i \right] (\widetilde{X}'\widetilde{X})^{-1}. \tag{2.16}$$

Stock and Watson (2008) consider the panel data model in (2.5) with serially uncorrelated errors, and apply the conventional White (1980) *cross-section* heteroskedasticity-robust variance–covariance matrix estimator to the fixed effects regression. This amounts to replacing the term in brackets in (2.16) by

$$V_{CS} = \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{X}_{it} \widetilde{X}_{it}' \widetilde{u}_{it}^2 / (NT - N - K).$$

They show that V_{CS} is *inconsistent* if T is fixed (and greater than 2) and $N \to \infty$. They suggest a bias adjusted estimator given by

$$V_{FE} = \frac{(T-1)}{(T-2)} \left[V_{CS} - \frac{1}{N(T-1)} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \widetilde{X}_{it} \widetilde{X}_{it}' \right) \left(\frac{1}{(T-1)} \sum_{s=1}^{T} \widetilde{u}_{is}^2 \right) \right]$$

Using Monte Carlo experiments, they show that tests based on V_{FE} are recommended especially if T is moderate or large.

Hansen (2007b) studies the properties of the FE estimator and its robust variance–covariance matrix not only when N is large but also when T may be large. He shows that tests based on these robust standard errors are consistent as long as $N \to \infty$, regardless of the relative size of N and T even in cases where the data is equicorrelated.

Bramati and Croux (2007) focus on robust alternatives to the Within estimator. The resulting estimator is robust in the sense that it is not altered too much by removing or modifying a small percentage of the observations. The basic idea is to center the variables by the median instead of the mean, since the median is known to be min-max robust. After centering, one runs a robust regression estimator such as the Least Trimmed Squares estimator. This estimator minimizes the sum of the smallest h squared residuals, where $1 \leq h \leq NT$ is a truncation value. A default choice is $h = [3NT/4]$, making it possible to cope with up to 25% of outliers.

Throughout this book, the reader will encounter several empirical examples where fixed effects give different results from pooled OLS. Some prominent examples are (1) the effect of real beer taxes on the US states motor vehicle fatality rates from drunken driving, see Ruhm (1996) in Chap. 11, where OLS yields a positive (0.012) and significant effect of real beer taxes on motor vehicle fatality rates, whereas FE obtains a negative (–0.324) and significant effect of real beer taxes on motor vehicle fatality rates. (2) The productivity puzzle example 3 in this chapter, where Munnell (1990) gets a positive (0.155) and significant effect of public Capital on productivity in the private sector using OLS, whereas FE obtains a negative (–0.026) and insignificant effect of public capital on productivity in the private sector, and hence the productivity puzzle. (3) Mothers who smoke during pregnancy are more likely to adopt other behaviors like drinking, poor nutritional intake, etc., that could have a negative impact on birthweight. Abrevaya (2006) argues that if these omitted variables are positively correlated with smoking, then OLS will result in an overestimate of the effect of smoking on birthweight. Typical OLS estimates yield a lowering birthweight of 230–250 grams. Using a matched panel on mothers with multiple birth, Abrevaya (2006) shows that fixed effects yields a much lower birthweight effect of smoking mothers of about 144 grams; see problem 10.7.

Difference-in-Differences

Note that the fixed effects (FE) transformation ($\widetilde{y}_{it} = y_{it} - \bar{y}_{i.}$) is not the only transformation that will wipe out the individual effects. In fact, first differencing (FD) will also do the trick ($\triangle y_{it} = y_{it} - y_{i,t-1}$). This is a crucial tool used in the *Difference-in-Differences* estimator. Before the approval of any drug, it is necessary to assign patients *randomly* to receive the drug or a placebo, and the drug is approved or disapproved depending on the difference in the health outcome between these two groups. In this case, the FDA is concerned with the drug's safety and its effectiveness. However, we run into problems in setting this experiment. How can we hold other factors constant? Even twins which have been used in economic studies are not identical and may have different life experiences. With panel data, observations on the same subjects before and after a health policy change allow us to estimate the effectiveness of this policy on the treated and control groups without the contamination of individual effects. In simple regression form, assuming the assignment to the control and treatment groups is *random*; one regresses the change in the health outcome before and after the health policy is enacted on a dummy variable which takes the value 1 if the individual is in the affected (treatment) group and zero if the individual is in the unaffected (control) group. This regression computes the average change in the health outcome for the treatment group before and after the policy change and subtracts that from the average change in the health outcome for the control group. One can include additional regressors which measure the individual characteristics prior to the policy change. Examples are gender, race, education, and age of the individual. This is known as the *difference-in-differences* (DID) estimator in econometrics. Alternatively, one can regress the health outcome y on d_g, d_t, and their interaction $d_t\text{x}d_g$. d_g is a dummy variable that takes the value 1 if the subject is in the *treatment group*, and zero otherwise; d_t is a dummy variable which takes the value 1 for the *post-treatment period*, and zero otherwise. In this case, $d_t\text{x}d_g$ takes the value one only for observations in the treatment group *and* in the post-treatment period. The OLS estimate of the coefficient of $d_t\text{x}d_g$ yields the DID estimator. Another advantage of running this regression is that one can robustify the standard errors with standard software.

In economics, one cannot conduct medical experiments. Card (1990) used a *natural experiment* to see whether immigration reduces wages. Taking advantage of the "Mariel boatlift" where a large number of Cuban immigrants entered Miami, Card (1990) compared the change in wages of low-skilled workers in Miami to the change in wages of similar workers in other comparable U.S. cities over the same period. Card concluded that the influx of Cuban immigrants had a negligible effect on wages of less-skilled workers. Gruber and Poterba (1994) use the DID estimator to show that a change in the tax law did increase the purchase of health insurance among the self-employed. They compared the fraction of the self-employed who had health insurance before the tax change 1985–1986 with the period after the tax change 1988–1989. The control group was the fraction of employed (not self-employed) workers with health insurance in those years.

Donald and Lang (2007) warn that the standard asymptotics for the DID estimator cannot be applied when the number of groups is *small*, as in the case where one compares two states in two years, or self-employed workers and employees over a small number of years. They reconsider the Gruber and Poterba (1994) paper on health insurance and self-employment and Card (1990) study of the Mariel boatlift. They show that analyzing the t-statistic, taking into account a possible group-error component, dramatically reduces the precision of their results. In fact for Card (1990) Mariel boatlift study, their findings suggest that the data cannot exclude large effects of the migration on blacks in Miami.

Bertrand, Duflo and Mullainathan (2004) argue that several DID studies in economics rely on a *long time series*. They warn that in this case, *serial correlation* will understate the standard error of the estimated treatment effects, leading to overestimation of t-statistics and significance levels. They show that the block bootstrap (taking into account the autocorrelation of the data) works well when the number of states is large enough.[1] Hausman and Kuersteiner (2008) warn that both the DID and the fixed effects estimators are not efficient if the stochastic disturbances are serially correlated. The optimal estimator in this case is generalized least squares (GLS), but this is rarely used in applications of DID studies. Hausman and Kuersteiner (2008) use higher order Edgeworth expansion to construct a size corrected t-statistic (based on feasible GLS) for the significance of treatment variables in DID regressions. They find that size corrected t-statistic based on feasible GLS yields accurate size and is significantly more powerful than robust OLS when serial correlation in the level data is high.

Conley and Taber (2011) consider the case where there are only a small number N_1 of treatment groups, say states, that change a law or policy within a fixed time span T. Let N_0 denote the number of control groups (states) that do not change their policy. Conley and Taber argue that the standard large-sample approximations used for inference can be misleading especially in the case of non-Gaussian or serially correlated errors. They suggest an alternative approach to inference under the assumption that N_1 is finite, using asymptotic approximations that let N_0 grow large, with T fixed. Point estimators of the treatment effect parameter(s) are not consistent since N_1 and T are fixed. However, they use information from the N_0 control groups to consistently estimate the distribution of these point estimators up to the true values of the parameter.

DID estimation has its benefits and limitations. It is simple to compute and controls for heterogeneity of the individuals or the groups considered before and after the policy change. However, it does not account for the possible *endogeneity* of the interventions themselves. Abadie (2005) discusses how well the comparison groups used in nonexperimental studies approximate appropriate control groups. Athey and

Imbens (2006) critique the linearity assumptions used in DID estimation and provide a general changes-in-changes (CIC) estimator that does not require such assumptions.

The DID estimator requires that, in the absence of the treatment, the average outcomes for the treated and control groups would have followed parallel paths over time. This assumption may be too restrictive. Abadie (2005) considers the case in which differences in observed characteristics create non-parallel outcome dynamics between treated and controls. He proposes a family of semiparametric DID estimators that can be used to estimate the average effect of the treatment for the treated. Abadie, Diamond and Hainmueller (2010) advocate the use of data-driven procedures to construct suitable comparison groups. Data-driven procedures reduce discretion in the choice of the comparison control units, forcing researchers to demonstrate the affinities between the affected and unaffected units using observed quantifiable characteristics. The idea behind the *synthetic control* approach is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone. They apply the synthetic control method to study the effects of California's Proposition 99, a large-scale tobacco control program implemented in California in 1988. They demonstrate that following the passage of Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region. They estimate that by the year 2000, annual per capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99.

Athey and Imbens (2006) generalize the DID methodology to what they call the changes-in-changes (CIC) methodology. Their approach allows the effects of both time and the treatment to differ systematically across individuals, as when new medical technology differentially benefits sicker patients. They propose an estimator for the entire counterfactual distribution of effects of the treatment on the treatment group as well as the distribution of effects of the treatment on the control group, where the two distributions may differ from each other in arbitrary ways. They provide conditions under which the proposed model is identified nonparametrically and extend the model to allow for discrete outcomes. They also provide extensions to settings with multiple groups and multiple time periods. They revisit the effects of disability insurance on injury durations. They show that the CIC approach leads to results that differ from the standard DID results in terms of magnitude and significance. They attribute this to the restrictive assumptions required for the standard DID methods.

Laporte and Windmeijer (2005) show that the FE and FD estimators lead to very different estimates of treatment effects when these are not constant over time and treatment is a state that only changes occasionally. They suggest allowing for flexible time-varying treatment effects when estimating panel data models with binary indicator variables. They illustrate this by looking at the effect of divorce on mental well-being using the British Household Panel Survey. They show that divorce has an adverse effect on mental well-being that starts before the actual divorce, peaks in the year of the divorce, and diminishes rapidly thereafter. A model that implies a constant instantaneous effect of divorce leads to very different FD and FE estimates, while a model that allows for flexibility in these effects leads to similar results. In general, the FE estimator is more efficient than the FD estimator when the remainder disturbance $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$. The FD estimator is more efficient than the FE estimator when the remainder disturbance ν_{it} is a random walk; see Wooldridge (2002). These estimators are affected differently by measurement error, see Chap. 10, and by nonstationarity; see Chap. 12.

Of course, this analysis can be refined to account for perhaps better control and treatment groups. If a policy is enacted by state s to reduce teenage smoking or motor vehicle fatality due to alcohol consumption or healthcare service for the elderly, then, for the two periods case, d_t takes the value 1 for the post-policy *period*, and zero otherwise; d_s takes the value 1 if the *state* has implemented this policy and zero otherwise; and d_g takes the value 1 for the treatment *group* affected by this policy like the elderly, and zero otherwise. In this case, one regresses healthcare outcome on d_t, d_s, d_g, d_txd_g, d_txd_s, d_sxd_g, and d_txd_sxd_g. The OLS estimate of the coefficient of d_txd_sxd_g yields the *difference-in-difference-in-differences* estimator of this policy. This estimator computes the average change in the health outcome for the elderly in the treatment state before and after the policy is implemented, and then subtracts from that the average change in the health outcome for the elderly in the control state, as well as the average change in the health outcome for the non-elderly in the treatment state.

Carpenter (2004) studied the effect of zero-tolerance (ZT) driving laws on alcohol-related behaviors of 18–20-year-olds, controlling for macroeconomic conditions, other alcohol policies, state fixed effects, survey year and month effects, and linear state-specific time trends. ZT Laws make it illegal for drivers under age 21 to have measurable amounts of alcohol in their blood, resulting in immediate license suspension and fines. Carpenter uses the Behavioral Risk Factor Surveillance System, which includes information on alcohol consumption and drunk driving behavior for young adults over age 18 for the years 1984–2001. He estimates the effects of ZT Laws using the difference-in-differences approach. The control group is composed of 22–24-year-olds who are otherwise similar to treated individuals (18–20-year-olds) but who should have been unaffected by the ZT policies. Let d_{ZT} be a dummy variable that takes the value 1 if the state has ZT in that year and zero otherwise; and d_g is a dummy variable that takes the value 1 if the subject is in the *treatment group*, and zero otherwise. Alcohol consumption is regressed on d_{ZT}, d_{1820} and d_{ZT}xd_{1820}, and other control variables mentioned above. The OLS estimate of the coefficient of d_{ZT}xd_{1820} yields the *difference-in-differences* estimator of the zero tolerance laws. Carpenter's results indicate that the laws reduced heavy episodic drinking (five or more drinks at one sitting) among underage males by 13%.

There is now a huge literature on treatment effects and we cannot do justice here; we refer the reader to Chap. 9 of the Oxford Handbook of Panel Data by Lechner (2015) entitled treatment effects and panel data.

2.3 The One-Way Random Effects Model

There are too many parameters in the fixed effects model and the loss of degrees of freedom can be avoided if the μ_i can be assumed random. In this case $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, and the μ_i are independent of the ν_{it}. In addition, the X_{it} are independent of the μ_i and ν_{it}, for all i and t. The random effects model is an appropriate specification if we are drawing N individuals randomly from a large population. This is usually the case for household panel studies. Care is taken in the design of the panel to make it "representative" of the population we are trying to make inferences about. In this case, N is usually large and a fixed effects model would lead to an enormous loss of degrees of freedom. The individual effect is characterized as random, and inference pertains to the population from which this sample was randomly drawn. From (2.4), one can compute the variance–covariance matrix

$$\begin{aligned}\Omega &= E(uu') = Z_\mu E(\mu\mu')Z_\mu' + E(\nu\nu') \\ &= \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes I_T)\end{aligned} \tag{2.17}$$

This implies a homoskedastic variance $\text{var}(u_{it}) = \sigma_\mu^2 + \sigma_\nu^2$ for all i and t, and an equicorrelated block-diagonal covariance matrix which exhibits serial correlation over time only between the disturbances of the same individual. In fact,

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\nu^2 \quad \text{for} \quad i = j, t = s \\ &= \sigma_\mu^2 \qquad\quad \text{for} \quad i = j, t \neq s\end{aligned}$$

and zero otherwise. This also means that the correlation coefficient between u_{it} and u_{js} is

$$\begin{aligned}\rho = \text{correl}(u_{it}, u_{js}) &= 1 \qquad\qquad\qquad \text{for } i = j, t = s \\ &= \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2) \text{ for } i = j, t \neq s\end{aligned}$$

and zero otherwise. In order to obtain the GLS estimator of the regression coefficients, we need Ω^{-1}. This is a huge matrix for typical panels and is of dimension $(NT \times NT)$, so no brute force inversion should be attempted. We will follow a simple trick devised by Wansbeek and Kapteyn (1982) that allows the derivation of Ω^{-1} and $\Omega^{-1/2}$.[2] Essentially, one replaces J_T by $T\bar{J}_T$, and I_T by $(E_T + \bar{J}_T)$ where E_T is by definition $(I_T - \bar{J}_T)$. In this case

$$\Omega = T\sigma_\mu^2(I_N \otimes \bar{J}_T) + \sigma_\nu^2(I_N \otimes E_T) + \sigma_\nu^2(I_N \otimes \bar{J}_T)$$

Collecting terms with the same matrices, we get

$$\Omega = (T\sigma_\mu^2 + \sigma_\nu^2)(I_N \otimes \bar{J}_T) + \sigma_\nu^2(I_N \otimes E_T) = \sigma_1^2 P + \sigma_\nu^2 Q \tag{2.18}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$. (2.18) is the spectral decomposition representation of Ω, with σ_1^2 being the first unique characteristic root of Ω of multiplicity N, and σ_ν^2 is the second unique characteristic root of Ω of multiplicity $N(T-1)$. It is easy to verify, using the properties of P and Q, that

$$\Omega^{-1} = \frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\nu^2}Q \tag{2.19}$$

and

$$\Omega^{-1/2} = \frac{1}{\sigma_1}P + \frac{1}{\sigma_\nu}Q \tag{2.20}$$

In fact, $\Omega^r = (\sigma_1^2)^r P + (\sigma_\nu^2)^r Q$ where r is an arbitrary scalar. Now we can obtain GLS as a weighted least squares. Fuller and Battese (1973, 1974) suggested pre-multiplying the regression equation given in (2.3) by $\sigma_\nu \Omega^{-1/2} = Q + (\sigma_\nu/\sigma_1)P$ and performing OLS on the resulting transformed regression. In this case, $y^* = \sigma_\nu \Omega^{-1/2} y$ has a typical element $y_{it} - \theta \bar{y}_{i.}$ where $\theta = 1 - (\sigma_\nu/\sigma_1)$ (see problem 2.4). This transformed regression inverts a matrix of dimension $(K+1)$ and can be easily implemented using any regression package.

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the spectral decomposition of Ω. In fact, $Pu \sim (0, \sigma_1^2 P)$ and $Qu \sim (0, \sigma_\nu^2 Q)$ and

$$\widehat{\sigma}_1^2 = \frac{u'Pu}{tr(P)} = T \sum_{i=1}^{N} \bar{u}_{i.}^2 / N \tag{2.21}$$

and

$$\widehat{\sigma}_\nu^2 = \frac{u'Qu}{tr(Q)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (u_{it} - \bar{u}_{i.})^2}{N(T-1)} \tag{2.22}$$

provide the BQU estimators of σ_1^2 and σ_ν^2, respectively (see problem 2.5).

These are analyses of variance-type estimators of the variance components and are minimum variance unbiased under normality of the disturbances (see Graybill, 1961). The true disturbances are not known and therefore (2.21) and (2.22) are not feasible. Wallace and Hussain (1969) suggest substituting OLS residual $\widehat{u}_{OLS}$ instead of the true u. After all, under the random effects model, the OLS estimates are still unbiased and consistent, but no longer efficient. Amemiya (1971) shows that these estimators of the variance components have a different asymptotic distribution from that knowing the true disturbances. He suggests using the LSDV residuals instead of the OLS residuals. In this case, $\widetilde{u} = y - \widetilde{\alpha}\iota_{NT} - X\widetilde{\beta}$ where $\widetilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\widetilde{\beta}$ and $\bar{X}'_{..}$ is a $1 \times K$ vector of averages of all regressors. Substituting these $\widetilde{u}$ for u in (2.21) and (2.22), we get the Amemiya-type estimators of the variance components. The resulting estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}(\widehat{\sigma}_\nu^2 - \sigma_\nu^2) \\ \sqrt{N}(\widehat{\sigma}_\mu^2 - \sigma_\mu^2) \end{pmatrix} \sim N\left(0, \begin{pmatrix} 2\sigma_\nu^4 & 0 \\ 0 & 2\sigma_\mu^4 \end{pmatrix}\right) \tag{2.23}$$

where $\widehat{\sigma}_\mu^2 = (\widehat{\sigma}_1^2 - \widehat{\sigma}_\nu^2)/T$.[3]

Swamy and Arora (1972) suggest running two regressions to get estimates of the variance components from the corresponding mean square errors of these regressions. The first regression is the Within regression, given in (2.10), which yields the following s^2:

$$\widehat{\widehat{\sigma}}_\nu^2 = [y'Qy - y'QX(X'QX)^{-1}X'Qy]/[N(T-1) - K] \tag{2.24}$$

The second regression is the Between regression which runs the regression of averages across time, i.e.,

$$\bar{y}_{i.} = \alpha + \bar{X}_{i.}'\beta + \bar{u}_{i.} \qquad i = 1, \ldots, N \tag{2.25}$$

This is equivalent to premultiplying the model in (2.5) by P and running OLS. The only caution is that the latter regression has NT observations because it repeats the averages T times for each individual, while the cross-section regression in (2.25) is based on N observations. To remedy this, one can run the cross-section regression

$$\sqrt{T}\bar{y}_{i.} = \alpha\sqrt{T} + \sqrt{T}\bar{X}_{i.}'\beta + \sqrt{T}\bar{u}_{i.} \tag{2.26}$$

where one can easily verify that $\text{var}(\sqrt{T}\bar{u}_{i.}) = \sigma_1^2$. This regression will yield an s^2 given by

$$\widehat{\widehat{\sigma}}_1^2 = (y'Py - y'PZ(Z'PZ)^{-1}Z'Py)/(N - K - 1) \tag{2.27}$$

Note that stacking the following two transformed regressions we just performed yields

$$\begin{pmatrix} Qy \\ Py \end{pmatrix} = \begin{pmatrix} QZ \\ PZ \end{pmatrix}\delta + \begin{pmatrix} Qu \\ Pu \end{pmatrix} \tag{2.28}$$

and the transformed error has mean 0 and variance–covariance matrix given by

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2 P \end{pmatrix}$$

Problem 2.7 asks the reader to verify that OLS on this system of $2NT$ observations yields OLS on the pooled model (2.3). Also, GLS on this system yields GLS on (2.3). Alternatively, one could get rid of the constant α by running the following stacked regressions:

$$\begin{pmatrix} Qy \\ (P - \bar{J}_{NT})y \end{pmatrix} = \begin{pmatrix} QX \\ (P - \bar{J}_{NT})X \end{pmatrix}\beta + \begin{pmatrix} Qu \\ (P - \bar{J}_{NT})u \end{pmatrix} \tag{2.29}$$

This follows from the fact that $Q\iota_{NT} = 0$ and $(P - \bar{J}_{NT})\iota_{NT} = 0$. The transformed error has zero mean and variance–covariance matrix

$$\begin{pmatrix} \sigma_\nu^2 Q & 0 \\ 0 & \sigma_1^2(P - \bar{J}_{NT}) \end{pmatrix}$$

OLS on this system yields OLS on (2.3) and GLS on (2.29) yields GLS on (2.3). In fact,

$$\begin{aligned}\widehat{\beta}_{GLS} &= [(X'QX/\sigma_\nu^2) + X'(P - \bar{J}_{NT})X/\sigma_1^2]^{-1}[(X'Qy/\sigma_\nu^2) + X'(P - \bar{J}_{NT})y/\sigma_1^2] \\ &= [W_{XX} + \phi^2 B_{XX}]^{-1}[W_{Xy} + \phi^2 B_{Xy}]\end{aligned} \tag{2.30}$$

with $\text{var}(\widehat{\beta}_{GLS}) = \sigma_\nu^2[W_{XX} + \phi^2 B_{XX}]^{-1}$. Note that $W_{XX} = X'QX$, $B_{XX} = X'(P - \bar{J}_{NT})X$, and $\phi^2 = \sigma_\nu^2/\sigma_1^2$. Also, the Within estimator of β is $\widetilde{\beta}_{Within} = W_{XX}^{-1}W_{Xy}$ and the Between estimator of β is $\widehat{\beta}_{Between} = B_{XX}^{-1}B_{Xy}$. This shows that $\widehat{\beta}_{GLS}$ is a matrix-weighted average of $\widetilde{\beta}_{Within}$ and $\widehat{\beta}_{Between}$ weighing each estimate by the inverse of its corresponding variance. In fact

$$\widehat{\beta}_{GLS} = W_1\widetilde{\beta}_{Within} + W_2\widehat{\beta}_{Between} \tag{2.31}$$

where

$$W_1 = [W_{XX} + \phi^2 B_{XX}]^{-1} W_{XX}$$

and

$$W_2 = [W_{XX} + \phi^2 B_{XX}]^{-1}(\phi^2 B_{XX}) = I - W_1$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_\mu^2 = 0$ then $\phi^2 = 1$ and $\widehat{\beta}_{GLS}$ reduces to $\widehat{\beta}_{OLS}$. (ii) If $T \to \infty$, then $\phi^2 \to 0$ and $\widehat{\beta}_{GLS}$ tends to $\widetilde{\beta}_{Within}$. Also, if W_{XX} is huge compared to B_{XX} then $\widehat{\beta}_{GLS}$ will be close to $\widetilde{\beta}_{Within}$. However, if B_{XX} dominates W_{XX} then $\widehat{\beta}_{GLS}$ tends to $\widehat{\beta}_{Between}$. In other words, the Within estimator ignores the Between variation, and the Between estimator ignores the Within variation. The OLS estimator gives equal weight to the Between and Within variations. From (2.30), it is clear that $\text{var}(\widetilde{\beta}_{Within}) - \text{var}(\widehat{\beta}_{GLS})$ is a positive semi-definite matrix, since ϕ^2 is positive. However, as $T \to \infty$ for any fixed N, $\phi^2 \to 0$ and both $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$ have the same asymptotic variance.

Another estimator of the variance components was suggested by Nerlove (1971). His suggestion is to estimate σ_μ^2 as $\sum_{i=1}^{N}(\widehat{\mu}_i - \overline{\widehat{\mu}})^2/(N-1)$ where $\widehat{\mu}_i$ are the dummy coefficients estimates from the LSDV regression. σ_ν^2 is estimated from the Within residual sums of squares divided by NT without correction for degrees of freedom.[4]

Note that, except for Nerlove (1971) method, one has to retrieve $\widehat{\sigma}_\mu^2$ as $(\widehat{\sigma}_1^2 - \widehat{\sigma}_\nu^2)/T$. In this case, there is no guarantee that the estimate of $\widehat{\sigma}_\mu^2$ would be nonnegative. Searle (1971) has an extensive discussion of the problem of negative estimates of the variance components in the biometrics literature. One solution is to replace these negative estimates by zero. This in fact is the suggestion of the Monte Carlo study by Maddala and Mount (1973). This study finds that negative estimates occurred only when the true σ_μ^2 was small and close to zero. In these cases OLS is still a viable estimator. Therefore, replacing negative $\widehat{\sigma}_\mu^2$ by zero is not a sin after all, and the problem is dismissed as not being serious.

How about the properties of the various feasible GLS estimators of β? Under the random effects model, GLS based on the true variance components is BLUE, and all the feasible GLS estimators considered are asymptotically efficient as either N or $T \to \infty$. Maddala and Mount (1973) compared OLS, Within, Between, feasible GLS methods, MINQUE, Henderson's method III, true GLS, and maximum likelihood estimation using their Monte Carlo study. They found little to choose among the various feasible GLS estimators in small samples and argued in favor of methods that were easier to compute. MINQUE was dismissed due to being more difficult to compute, and the applied researcher given one shot at the data was warned to compute at least two methods of estimation, like an ANOVA feasible GLS and maximum likelihood to ensure that they do not yield drastically different results. When they do give different results, the authors diagnose misspecification.

Taylor (1980) derived exact finite sample results for the one-way error component model. He compared the Within estimator with the Swamy–Arora feasible GLS estimator. He found the following important results:

(1) Feasible GLS is more efficient than LSDV for all but the fewest degrees of freedom.

(2) The variance of feasible GLS is never more than 17% above the Cramer–Rao lower bound.
(3) More efficient estimators of the variance components do not necessarily yield more efficient feasible GLS estimators.

These finite sample results are confirmed by the Monte Carlo experiments carried out by Maddala and Mount (1973) and Baltagi (1981).[5]

Fixed versus Random
Having discussed the fixed effects and the random effects models and the assumptions underlying them, the reader is left with the daunting question, which one to choose? This is not as easy a choice as it might seem. In fact, the fixed versus random effects issue has generated a hot debate in the biometrics and statistics literature which has spilled over into the panel data econometrics literature. Wallace and Hussain (1969) and Mundlak (1978) were early proponents of the fixed effects model. In Chap. 4, we will study a specification test proposed by Hausman (1978) which is based on the difference between the fixed and random effects estimators. Unfortunately, applied researchers have interpreted a rejection as an adoption of the fixed effects model and non-rejection as an adoption of the random effects model.[6] Chamberlain (1984) showed that the fixed effects model imposes testable restrictions on the parameters of the reduced form model and one should check the validity of these restrictions before adopting the fixed effects model (see Chap. 4). Mundlak (1978) argued that the random effects model assumes exogeneity of *all* the regressors with the random individual effects. In contrast, the fixed effects model allows for endogeneity of *all* the regressors with these individual effects. So, it is an "all" or "nothing" choice of exogeneity of the regressors and the individual effects; see Chap. 7 for a more formal discussion of this subject. Hausman and Taylor (1981) allowed for *some* of the regressors to be correlated with the individual effects, as opposed to the all or nothing choice. These over-identification restrictions are testable using a Hausman-type test (see Chap. 7). For the applied researcher, performing fixed effects and random effects and the associated Hausman test reported in standard packages like Stata, LIMDEP, TSP, etc., the message is clear: Do not stop here. Test the restrictions implied by the fixed effects model derived by Chamberlain (1984) (see Chap. 4) and check whether a Hausman and Taylor (1981) specification might be a viable alternative (see Chap. 7).

2.4 Maximum Likelihood Estimation

Under normality of the disturbances, one can write the likelihood function as

$$L(\alpha, \beta, \phi^2, \sigma_\nu^2) = \text{ constant } -\frac{NT}{2}\log\sigma_\nu^2 + \frac{N}{2}\log\phi^2 - \frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \qquad (2.32)$$

where $\Omega = \sigma_\nu^2\Sigma$, $\phi^2 = \sigma_\nu^2/\sigma_1^2$, and $\Sigma = Q + \phi^{-2}P$ from (2.18). This uses the fact that $|\ \Omega\ |$= product of its characteristic roots $= (\sigma_\nu^2)^{N(T-1)}\ (\sigma_1^2)^N = (\sigma_\nu^2)^{NT}\ (\phi^2)^{-N}$.

Note that there is a one-to-one correspondence between ϕ^2 and σ_μ^2. In fact, $0 \leqslant \sigma_\mu^2 < \infty$ translates into $0 < \phi^2 \leqslant 1$. Brute force maximization of (2.32) leads to nonlinear first-order conditions (see Amemiya (1971)). Instead, Breusch (1987) concentrates the likelihood with respect to α and σ_ν^2. In this case, $\widehat{\alpha}_{mle} = \bar{y}_{..} - \bar{X}'_{..}\widehat{\beta}_{mle}$ and $\widehat{\sigma}^2_{\nu,mle} = (1/NT)\widehat{u}'\widehat{\Sigma}^{-1}\widehat{u}$ where $\widehat{u}$ and $\widehat{\Sigma}$ are based on maximum likelihood estimates of β, ϕ^2, and α. Let $d = y - X\widehat{\beta}_{mle}$ then $\widehat{\alpha}_{mle} = (1/NT)\iota'_{NT}d$ and $\widehat{u} = d - \iota_{NT}\widehat{\alpha}_{mle} = d - \bar{J}_{NT}d$. This implies that $\widehat{\sigma}^2_{\nu,mle}$ can be rewritten as

$$\widehat{\sigma}^2_{\nu,mle} = d'[Q + \phi^2(P - \bar{J}_{NT})]d/NT \tag{2.33}$$

and the concentrated likelihood becomes

$$L_C(\beta, \phi^2) = \text{ constant } - \frac{NT}{2}\log\{d'[Q + \phi^2(P - \bar{J}_{NT})]d\} + \frac{N}{2}\log\phi^2 \tag{2.34}$$

Maximizing (2.34) over ϕ^2, given β (see problem 2.9), yields

$$\widehat{\phi}^2 = \frac{d'Qd}{(T-1)d'(P - \bar{J}_{NT})d} = \frac{\sum\sum(d_{it} - \bar{d}_{i.})^2}{T(T-1)\sum(\bar{d}_{i.} - \bar{d}_{..})^2} \tag{2.35}$$

Maximizing (2.34) over β, given ϕ^2, yields

$$\widehat{\beta}_{mle} = [X'(Q + \phi^2(P - \bar{J}_{NT}))X]^{-1}X'[Q + \phi^2(P - \bar{J}_{NT})]y \tag{2.36}$$

One can iterate between β and ϕ^2 until convergence. Breusch (1987) shows that provided $T > 1$, any ith iteration β, call it β_i, gives $0 < \phi^2_{i+1} < \infty$ in the $(i+1)$th iteration. More importantly, Breusch (1987) shows that these ϕ_i^2 have a "remarkable property" of forming a monotonic sequence. In fact, starting from the Within estimator of β, for $\phi^2 = 0$, the next ϕ^2 is finite and positive and starts a monotonically increasing sequence of ϕ^2. Similarly, starting from the Between estimator of β, for $(\phi^2 \to \infty)$ the next ϕ^2 is finite and positive and starts a monotonically decreasing sequence of ϕ^2. Hence, to guard against the possibility of a local maximum, Breusch (1987) suggests starting with $\widetilde{\beta}_{Within}$ and $\widehat{\beta}_{Between}$ and iterating. If these two sequences converge to the same maximum, then this is the global maximum. If one starts with $\widehat{\beta}_{OLS}$ for $\phi^2 = 1$, and the next iteration obtains a larger ϕ^2, then we have a local maximum at the boundary $\phi^2 = 1$. Maddala (1971) finds that there are at most two maxima for the likelihood $L(\phi^2)$ for $0 < \phi^2 \leqslant 1$. Hence, we have to guard against one local maximum.

2.5 Prediction

Suppose we want to predict S periods ahead for the ith individual. For the GLS model, knowing the variance–covariance structure of the disturbances, Goldberger (1962) showed that the best linear unbiased predictor (BLUP) of $y_{i,T+S}$ is

$$\widehat{y}_{i,T+S} = Z'_{i,T+S}\widehat{\delta}_{GLS} + w'\Omega^{-1}\widehat{u}_{GLS} \quad \text{for } s \geqslant 1 \tag{2.37}$$

where $\widehat{u}_{GLS} = y - Z\widehat{\delta}_{GLS}$ and $w = E(u_{i,T+S}u)$. Note that for period $T + S$

$$u_{i,T+S} = \mu_i + \nu_{i,T+S} \tag{2.38}$$

and $w = \sigma_\mu^2(l_i \otimes \iota_T)$ where l_i is the ith column of I_N, i.e., l_i is a vector that has 1 in the ith position and zero elsewhere. In this case

$$w'\Omega^{-1} = \sigma_\mu^2(l_i' \otimes \iota_T')\left[\frac{1}{\sigma_1^2}P + \frac{1}{\sigma_\nu^2}Q\right] = \frac{\sigma_\mu^2}{\sigma_1^2}(l_i' \otimes \iota_T') \tag{2.39}$$

since $(l_i' \otimes \iota_T')P = (l_i' \otimes \iota_T')$ and $(l_i' \otimes \iota_T')Q = 0$. Using (2.39), the typical element of $w'\Omega^{-1}\widehat{u}_{GLS}$ becomes $((T\sigma_\mu^2/\sigma_1^2)\overline{\widehat{u}}_{i.,GLS})$ where $\overline{\widehat{u}}_{i.,GLS} = \sum_{t=1}^{T}\widehat{u}_{it,GLS}/T$. Therefore, in (2.37), the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual. This predictor was considered by Taub (1979). The BLUP are optimal assuming true values of the variance components. In practice, these are replaced with estimated values that yield empirical BLUP. Kackar and Harville (1984) propose inflation factors that account for the additional uncertainty introduced by estimating these variance components.

Baillie and Baltagi (1999) consider the practical situation of prediction from the error component regression model when the variance components are not known. They derive both theoretical and simulation evidence as to the relative efficiency of four alternative predictors: (i) an ordinary predictor, based on the optimal predictor given in (2.37), but with MLEs replacing population parameters, (ii) a truncated predictor that ignores the error component correction, given by the last term in (2.37), but uses MLEs for its regression parameters, (iii) a misspecified predictor which uses OLS estimates of the regression parameters, and (iv) a fixed effects predictor which assumes that the individual effects are fixed parameters that can be estimated. The asymptotic formula for MSE prediction is derived for all four predictors. Using numerical and simulation results, these are shown to perform adequately in realistic sample sizes ($N = 50$ and 500 and $T = 10$ and 20). Both the analytical and sampling results show that there are substantial gains in mean square error prediction by using the ordinary predictor instead of the misspecified or the truncated predictors, especially with increasing $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ values. The reduction in MSE is about tenfold for $\rho = 0.9$ and a little more than twofold for $\rho = 0.6$ for various values of N and T. The fixed effects predictor performs remarkably well being a close second to the ordinary predictor for all experiments. Simulation evidence confirm the importance of taking into account the individual effects when making predictions. The ordinary predictor and the fixed effects predictor outperform the truncated and misspecified predictors and are recommended in practice.[7]

2.6 Examples

2.6.1 Example 1: Investment Equation

Grunfeld (1958) considered the following investment equation:

$$I_{it} = \alpha + \beta_1 F_{it} + \beta_2 C_{it} + u_{it} \tag{2.40}$$

where I_{it} denotes real gross investment for firm i in year t, F_{it} is the real value of the firm (shares outstanding), and C_{it} is the real value of the capital stock. These panel data consist of 10 large US manufacturing firms over 20 years, 1935-54, and are available on the Springer website as Grunfeld.fil. This data set, even though dated, is of manageable size for classroom use and has been used by Zellner (1962) and Taylor (1980). Table 2.1 gives the OLS, Between and Within estimators for the slope coefficients along with their standard errors. The Between estimates are different from the Within estimates and a Hausman (1978) test based on their difference is given in Chap. 4. OLS and feasible GLS are matrix-weighted combinations of these two estimators. Table 2.1 reports three feasible GLS estimates of the regression coefficients along with the corresponding estimates of ρ, σ_μ, and σ_ν. These are WALHUS, AMEMIYA, and SWAR. EViews computes the Wallace and Hussain (1969) estimator as an option under the random effects panel data procedure. This EViews output is reproduced in Table 2.2. Similarly, Table 2.3 gives the EViews output for the Amemiya (1971) procedure which is named Wansbeek and Kapteyn (1989) in EViews, since the latter paper generalizes the Amemiya method to deal with unbalanced or incomplete panels; see Chap. 9. Table 2.4 gives the EViews output for the Swamy and Arora (1972) procedure. Note that in Table 2.4, $\widehat{\sigma}_\mu = 84.2$, $\widehat{\sigma}_\nu = 52.77$, and $\widehat{\rho} = \widehat{\sigma}^2_\mu/(\widehat{\sigma}^2_\mu + \widehat{\sigma}^2_\nu) = 0.72$. This is not $\widehat{\theta}$, but the latter can be obtained as $\widehat{\theta} = 1 - (\widehat{\sigma}_\nu/\widehat{\sigma}_1) = 0.86$. Next, Breusch (1987) iterative maximum likelihood

Table 2.1 Grunfeld's data. One-way error component results

	β_1	β_2	ρ	σ_μ	σ_ν
OLS	0.116 (0.006)*	0.231 (0.025)*			
Between	0.135 (0.029)	0.032 (0.191)			
Within	0.110 (0.012)	0.310 (0.017)			
WALHUS	0.110 (0.011)	0.308 (0.017)	0.73	87.36	53.75
AMEMIYA	0.110 (0.010)	0.308 (0.017)	0.71	83.52	52.77
SWAR	0.110 (0.010)	0.308 (0.017)	0.72	84.20	52.77
IMLE	0.110 (0.010)	0.308 (0.017)	0.70	80.30	52.49

*These are biased standard errors when the true model has error component disturbances (see Moulton 1986)

Table 2.2 Grunfeld's data: Wallace and Hussain RE estimator

Dependent Variable: I
Method: Panel EGLS (Cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wallace and Hussain estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-57.86253	29.90492	-1.934883	0.0544
F	0.109789	0.010725	10.23698	0.0000
K	0.308183	0.017498	17.61207	0.0000
Effects Specification				
Cross-section random S.D. / Rho			87.35803	0.7254
Idiosyncratic random S.D. / Rho			53.74518	0.2746
Weighted Statistics				
R-squared	0.769410	Mean dependent var		19.89203
Adjusted R-squared	0.767069	S.D. dependent var		109.2808
S.E. of regression	52.74214	Sum squared resid		548001.4
F-statistic	328.6646	Durbin-Watson stat		0.683829
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.803285	Mean dependent var		145.9582
Sum squared resid	1841243.	Durbin-Watson stat		0.203525

estimation is performed (IMLE). This procedure converged to a global maximum in three to four iterations depending on whether one started from the Between or Within estimators. There is not much difference among the feasible GLS estimates or the iterative MLE and they are all close to the Within estimates. This is understandable given that $\widehat{\theta}$ for these estimators is close to 1.

Table 2.3 Grunfeld's data: Amemiya/Wansbeek and Kapteyn RE estimator

Dependent Variable: I
Method: Panel EGLS (Cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wansbeek and Kapteyn estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-57.82187	28.68562	-2.015710	0.0452
F	0.109778	0.010471	10.48387	0.0000
K	0.308081	0.017172	17.94062	0.0000
Effects Specification				
Cross-section random S.D. / Rho			83.52354	0.7147
Idiosyncratic random S.D. / Rho			52.76797	0.2853
Weighted Statistics				
R-squared	0.769544	Mean dependent var		20.41664
Adjusted R-squared	0.767205	S.D. dependent var		109.4431
S.E. of regression	52.80503	Sum squared resid		549309.2
F-statistic	328.9141	Durbin-Watson stat		0.682171
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.803313	Mean dependent var		145.9582
Sum squared resid	1840981.	Durbin-Watson stat		0.203545

2.6.2 Example 2: Gasoline Demand Equation

Baltagi and Griffin (1983) considered the following gasoline demand equation:

$$\ln \frac{Gas}{Car} = \alpha + \beta_1 \ln \frac{Y}{N} + \beta_2 \ln \frac{P_{MG}}{P_{GDP}} + \beta_3 \ln \frac{Car}{N} + u \qquad (2.41)$$

where Gas/Car is motor gasoline consumption per auto, Y/N is real per capita income, P_{MG}/P_{GDP} is real motor gasoline price, and Car/N denotes the stock of cars per capita. This panel consists of annual observations across 18 OECD countries, covering the period 1960–78. The data for this example are given as Gasoline.dat on the Springer website. Table 2.5 gives the parameter estimates for OLS, Between,

Table 2.4 Grunfeld's data: Swamy and Arora RE estimator

Dependent Variable: I
Method: Panel EGLS (Cross-section random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-57.83441	28.88930	-2.001932	0.0467
F	0.109781	0.010489	10.46615	0.0000
K	0.308113	0.017175	17.93989	0.0000
Effects Specification				
Cross-section random S.D. / Rho			84.20095	0.7180
Idiosyncratic random S.D. / Rho			52.76797	0.2820
Weighted Statistics				
R-squared	0.769503	Mean dependent var		20.25556
Adjusted R-squared	0.767163	S.D. dependent var		109.3928
S.E. of regression	52.78556	Sum squared resid		548904.1
F-statistic	328.8369	Durbin-Watson stat		0.682684
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.803304	Mean dependent var		145.9582
Sum squared resid	1841062.	Durbin-Watson stat		0.203539

Within, and three feasible GLS estimates of the slope coefficients along with their standard errors, and the corresponding estimates of ρ, σ_μ, and σ_ν. Breusch's (1987) iterative maximum likelihood converged to a global maximum in four to six iterations depending on whether one starts from the Between or Within estimators. For the SWAR procedure, $\widehat{\sigma}_\mu = 0.196$, $\widehat{\sigma}_\nu = 0.092$, $\widehat{\rho} = 0.82$, and $\widehat{\theta} = 0.89$. Once again the estimates of θ are closer to 1 than 0, which explains why feasible GLS is closer to the Within estimator than the OLS estimator. The Between and OLS price elasticity estimates of gasoline demand are more than double than those for the Within and feasible GLS estimators.

2.6.3 Example 3: Public Capital Productivity

Following Munnell (1990), Baltagi and Pinnoi (1995) considered the following Cobb–Douglas production function relationship investigating the productivity of public capital in private production:

$$\ln Y = \alpha + \beta_1 \ln K_1 + \beta_3 \ln K_2 + \beta_3 \ln L + \beta_4 \text{ Unemp } + u \tag{2.42}$$

where Y is gross state product, K_1 is public capital which includes highways and streets, water and sewer facilities, and other public buildings and structures. K_2 is the private capital stock based on the Bureau of Economic Analysis national stock estimates, L is labor input measured as employment in nonagricultural payrolls. Unemp is the state unemployment rate included to capture business cycle effects. This panel consists of annual observations for 48 contiguous states over the period 1970-86. This data set was provided by Munnell (1990) and is given as Produc.prn on the Springer website. Table 2.6 gives the estimates for a one-way error component model. Note that both OLS and the Between estimators report that public capital is productive and significant in the states private production. In contrast, the Within and feasible GLS estimators find that public capital is insignificant. This result was also reported by Holtz-Eakin (1994) who found that after controlling for state-specific effects, the public-sector capital has no role in affecting private production.

Tables 2.7 and 2.8 give the Stata output reproducing the Between and within estimates in Table 2.6. This is done using the *xtreg* command with options (,be) for Between and (,fe) for fixed effects. Note that the fixed effects regression prints out the F-test for the significance of the state effects at the bottom of the output. This is the F-test described in (2.12). It tests whether all state dummy coefficients are

Table 2.5 Gasoline demand data. One-way error component results

	β_1	β_2	β_3	ρ	σ_μ	σ_ν
OLS	0.890 (0.036)*	–0.892 (0.030)*	–0.763 (0.019)*			
Between	0.968 (0.156)	–0.964 (0.133)	–0.795 (0.082)			
Within	0.662 (0.073)	–0.322 (0.044)	–0.640 (0.030)			
WALHUS	0.545 (0.066)	–0.447 (0.046)	–0.605 (0.029)	0.75	0.197	0.113
AMEMIYA	0.602 (0.066)	–0.366 (0.042)	–0.621 (0.027)	0.93	0.344	0.092
SWAR	0.555 (0.059)	–0.420 (0.042)	–0.607 (0.026)	0.82	0.196	0.092
IMLE	0.588 (0.066)	–0.378 (0.044)	–0.616 (0.027)	0.91	0.292	0.092

*These are biased standard errors when the true model has error component disturbances (see Moulton, 1986)

Source Baltagi and Griffin (1983). Reproduced by permission of Elsevier Science Publishers B.V.(North-Holland)

Table 2.6 Public capital productivity data. One-way error component results

	β_1	β_2	β_3	β_4	ρ	σ_μ	σ_ν
OLS	0.155 (0.017)*	0.309 (0.010)*	0.594 (0.014)*	-0.007 (0.001)*			
Between	0.179 (0.072)	0.302 (0.042)	0.576 (0.056)	–0.004 (0.010)			
Within	–0.026 (0.029)	0.292 (0.025)	0.768 (0.030)	–0.005 (0.001)			
WALHUS	0.006 (0.024)	0.311 (0.020)	0.728 (0.025)	–0.006 (0.001)	0.82	0.082	0.039
AMEMIYA	0.002 (0.024)	0.309 (0.020)	0.733 (0.025)	–0.006 (0.001)	0.84	0.088	0.038
SWAR	0.004 (0.023)	0.311 (0.020)	0.730 (0.025)	–0.006 (0.001)	0.82	0.083	0.038
IMLE	0.003 (0.024)	0.310 (0.020)	0.731 (0.026)	–0.006 (0.001)	0.83	0.085	0.038

* These are biased standard errors when the true model has error component disturbances (see Moulton (1986))

Table 2.7 Public capital productivity data: the Between estimator

```
. xtreg lny lnk1 lnk2 lnl u, be

Between regression (regression on group means)  Number of obs      =       816
Group variable (i) : stid                       Number of groups   =        48

R-sq:  within  = 0.9330                         Obs per group: min =        17
       between = 0.9939                                        avg =      17.0
       overall = 0.9925                                        max =        17

                                                F(4,43)            =   1754.11
sd(u_i + avg(e_i.))=  .0832062                  Prob > F           =    0.0000

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnk1 |   .1793651   .0719719     2.49    0.017      .0342199    .3245104
        lnk2 |   .3019542   .0418215     7.22    0.000      .2176132    .3862953
         lnl |   .5761274   .0563746    10.22    0.000      .4624372    .6898176
           u |  -.0038903   .0099084    -0.39    0.697     -.0238724    .0160918
       _cons |   1.589444   .2329796     6.82    0.000      1.119596    2.059292
------------------------------------------------------------------------------
```

equal and in this case it yields an F(47,764) = 75.82 which is statistically significant. This indicates that the state dummies are jointly significant. It also means that the OLS estimates which omit these state dummies suffer from an omission variables problem rendering them biased and inconsistent. Table 2.9 gives the Swamy and Arora (1972) estimate of the random effects model. This is the default option in

Table 2.8 Public capital productivity data: the Within estimator

```
. xtreg lny lnk1 lnk2 lnl u, fe

Fixed-effects (within) regression               Number of obs      =       816
Group variable (i) : stid                       Number of groups   =        48

R-sq:  within  = 0.9413                         Obs per group: min =        17
       between = 0.9921                                        avg =      17.0
       overall = 0.9910                                        max =        17

                                                F(4,764)           =   3064.81
corr(u_i, Xb)  = 0.0608                         Prob > F           =    0.0000

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnk1 |  -.0261493   .0290016    -0.90   0.368    -.0830815    .0307829
        lnk2 |   .2920067   .0251197    11.62   0.000     .2426949    .3413185
         lnl |   .7681595   .0300917    25.53   0.000     .7090872    .8272318
           u |  -.0052977   .0009887    -5.36   0.000    -.0072387   -.0033568
       _cons |   2.352898   .1748131    13.46   0.000     2.009727    2.696069
-------------+----------------------------------------------------------------
     sigma_u |  .09057293
     sigma_e |  .03813705
         rho |   .8494045   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(47, 764) =    75.82             Prob > F = 0.0000
```

Table 2.9 Public capital productivity data: the Swamy and Arora estimator

```
. xtreg lny lnk1 lnk2 lnl u, re theta

Random-effects GLS regression                   Number of obs      =       816
Group variable (i) : stid                       Number of groups   =        48

R-sq:  within  = 0.9412                         Obs per group: min =        17
       between = 0.9928                                        avg =      17.0
       overall = 0.9917                                        max =        17

Random effects u_i ~ Gaussian                   Wald chi2(4)       =  19131.09
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
theta              = .8888353

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnk1 |   .0044388   .0234173     0.19   0.850    -.0414583    .0503359
        lnk2 |   .3105483   .0198047    15.68   0.000     .2717317    .3493649
         lnl |   .7296705   .0249202    29.28   0.000     .6808278    .7785132
           u |  -.0061725   .0009073    -6.80   0.000    -.0079507   -.0043942
       _cons |   2.135411   .1334615    16.00   0.000     1.873831     2.39699
-------------+----------------------------------------------------------------
     sigma_u |   .0826905
     sigma_e |  .03813705
         rho |  .82460109   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 2.10 Public capital productivity data: the maximum likelihood estimator

```
. xtreg lny lnk1 lnk2 lnl u, mle

Random-effects ML regression                    Number of obs      =       816
Group variable (i) : stid                       Number of groups   =        48

Random effects u_i ~ Gaussian                   Obs per group: min =        17
                                                               avg =      17.0
                                                               max =        17

                                                LR chi2(4)         =   2412.91
Log likelihood  =  1401.9041                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnk1 |   .0031446   .0239185     0.13   0.895    -.0437348     .050024
        lnk2 |    .309811    .020081    15.43   0.000      .270453     .349169
         lnl |   .7313372   .0256936    28.46   0.000     .6809787    .7816957
           u |  -.0061382   .0009143    -6.71   0.000    -.0079302   -.0043462
       _cons |   2.143865   .1376582    15.57   0.000      1.87406    2.413671
-------------+----------------------------------------------------------------
    /sigma_u |    .085162   .0090452     9.42   0.000     .0674337    .1028903
    /sigma_e |   .0380836   .0009735    39.12   0.000     .0361756    .0399916
-------------+----------------------------------------------------------------
         rho |   .8333481   .0304597                      .7668537    .8861754
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0: chibar2(01)= 1149.84 Prob>=chibar2 = 0.000
```

Stata and is obtained from the *xtreg* command with option (,re). Finally, Table 2.10 gives the Stata output for the maximum likelihood estimator. These are obtained from the *xtreg* command with option (,mle).

2.7 Selected Applications

There are far too many applications of the error component model in economics to be exhaustive and here we only want to refer the reader to a few applications. These include the following:

(1) Cornwell and Rupert (1997) used panel data from the NLSY to show that much of the wage premium normally attributed to marriage is associated with unobservable individual effects that are correlated with marital status and wages. Their fixed effects estimates of the marriage premium is no more than 5% to 7%. This cast doubt on the interpretation that marriage enhances productivity through specialization.

(2) Lundberg and Rose (2002) used panel data from the PSID to estimate the effects of children and the differential effects of sons and daughters on men's labor supply and hourly wage rate. Their fixed effects estimates indicate that, on average, a child increases a man's wage rate by 4.2% and his annual hours of work by 38 hours per year.

(3) Glick and Rose (2002) studied the question of whether leaving a currency union reduces international trade. They used panel data on bilateral trade among 217 countries over the period 1948–1997. Their fixed effects estimates show that currency union more than doubles trade.

2.8 Computational Note

There is no magical software written explicitly for all panel data estimation and testing procedures. Simple panel data estimators can be done with LIMDEP, RATS, SAS, TSP, R, EViews, or Stata. In fact, the results reported in examples 2.1, 2.2, and 2.3 have been verified using EViews and Stata. Unfortunately, not all panel data methods discussed in this book have made it into these standard software packages, and researchers had to program them with the help of GAUSS, OX, MATLAB, and R, to mention a few. A recent book using the splm package in R which illustrates them using several examples from this book is Croissant and Millo (2019).

2.9 Notes

1. See also Hansen (2007a) for inference in panel models with serial correlation and fixed effects, and Stock and Watson (2008) for a heteroskedasticity-robust variance matrix estimator for the fixed effects estimator.
2. See also Searle and Henderson (1979) for a systematic method for deriving the characteristic roots and vectors of Ω for any balanced error component model.
3. It is important to note that once one substitutes OLS or LSDV residuals in (2.21) and (2.22), the resulting estimators of the variance components are no longer unbiased. The degrees of freedom corrections required to make these estimators unbiased involve traces of matrices that depend on the data. These correction terms are given in Wallace and Hussain (1969) and Amemiya (1971), respectively. Alternatively, one can infer these correction terms from the more general unbalanced error component model considered in Chap. 9.
4. One can also apply Rao (1971a, b) minimum norm quadratic unbiased estimation (MINQUE) procedure or Henderson's method III as described by Fuller and Battese (1973). These methods are studied in detail in Baltagi (1995, Appendix 3) for the two-way error component model and in Chap. 9 for the unbalanced error component model. Unfortunately, these methods have not been widely used in the empirical economics literature.
5. For the estimation of fixed effects nonparametric and semi-parametric partially linear panel data models, see Li and Racine (2007) and the references cited there.

6. Hsiao and Sun (2000) argue that fixed versus random effects specification is better treated as an issue of model selection rather than hypothesis testing. They suggest a recursive predictive density ratio as well as the Akaike and Schwartz information criteria for model selection. Monte Carlo results indicate that all three criteria perform well in finite samples. However, the Schwartz criterion was found to be the more reliable of the three.
7. For a survey on forecasting using panel data, see Baltagi (2013).

2.10 Problems

2.1 *LSDV is identical to the Within estimator*. Prove that $\widetilde{\beta}$ given in (2.7) can be obtained from OLS on (2.5) using results on partitioned inverse. This can be easily obtained using the Frisch–Waugh–Lovell theorem of Davidson and MacKinnon (1993, p. 19). Hint: this theorem states that the OLS estimate of β from (2.5) will be identical to the OLS estimate of β from (2.6). Also, the least squares residuals will be the same. See Chap. 1 of Baltagi (2009).

2.2 *OLS and GLS are equivalent for the Within transformed model*.

(a) Using generalized inverse, show that OLS or GLS on (2.6) yields $\widetilde{\beta}$, the Within estimator given in (2.7).
(b) Show that (2.6) satisfies the necessary and sufficient condition for OLS to be equivalent to GLS (see Baltagi (1989)). Hint: show that $\text{var}(Q\nu) = \sigma_\nu^2 Q$ which is positive semi-definite and then use the fact that Q is idempotent and is its own generalized inverse.

2.3 *Robust FE variance–covariance estimates*. Verify that by stacking the panel as an equation for each individual in (2.13) and performing the Within transformation as in (2.14), one gets the Within estimator as OLS on this system. Verify that the robust asymptotic $\text{var}(\widetilde{\beta})$ is the one given by (2.16).

2.4 Fuller and Battese (1973) *transformation for the one-way random effects model*.

(a) Verify (2.17) and check that $\Omega^{-1}\Omega = I$ using (2.18).
(b) Verify that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$ using (2.20) and (2.19).
(c) Premultiply y by $\sigma_\nu \Omega^{-1/2}$ from (2.20), and show that the typical element is $y_{it} - \theta \bar{y}_{i.}$ where $\theta = 1 - (\sigma_\nu/\sigma_1)$.

2.5 *Unbiased estimates of the variance components: The one-way model*. Using (2.21) and (2.22), show that $E(\widehat{\sigma}_1^2) = \sigma_1^2$ and $E(\widehat{\sigma}_\nu^2) = \sigma_\nu^2$, Hint: $E(u'Qu) = E\{\text{tr}(u'Qu)\} = E\{\text{tr}(uu'Q)\} = \text{tr}\{E(uu')Q\} = \text{tr}(\Omega Q)$.

2.6 Swamy and Arora (1972) *estimates of the variance components: The one-way model*.

(a) Show that $\widehat{\widehat{\sigma}}_\nu^2$ given in (2.24) is unbiased for σ_ν^2.
(b) Show that $\widehat{\widehat{\sigma}}_1^2$ given in (2.27) is unbiased for σ_1^2.

2.7 *System estimation for the one-way model: OLS versus GLS.*

(a) Perform OLS on the system of equations given in (2.28) and show that the resulting estimator is pooled OLS $\widehat{\delta}_{OLS} = (Z'Z)^{-1}Z'y$.
(b) Perform GLS on the system of equations given in (2.28) and show that the resulting estimator is the random effects estimator $\widehat{\delta}_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$ where Ω^{-1} is given in (2.19).

2.8 *GLS is more efficient than Within.* Using the var($\widehat{\beta}_{GLS}$) expression below (2.30) and $\text{var}(\widetilde{\beta}_{Within}) = \sigma_\nu^2 W_{XX}^{-1}$, show that

$$(\text{var}(\widehat{\beta}_{GLS}))^{-1} - (\text{var}(\widetilde{\beta}_{Within}))^{-1} = \phi^2 B_{XX}/\sigma_\nu^2$$

which is positive semi-definite. Conclude that $\text{var}(\widetilde{\beta}_{Within}) - \text{var}(\widehat{\beta}_{GLS})$ is positive semi-definite.

2.9 *Maximum likelihood estimation of the random effects model.*

(a) Using the concentrated likelihood function in (2.34), solve $\partial L_C/\partial\phi^2 = 0$ and verify (2.35).
(b) Solve $\partial L_C/\partial\beta = 0$ and verify (2.36).

2.10 *Prediction in the one-way random effects model.*

(a) For the predictor $y_{i,T+S}$ given in (2.37), compute $E(u_{i,T+S}u_{it})$ for $t = 1, 2, \ldots, T$ and verify that $w = E(u_{i,T+S}u) = \sigma_\mu^2(l_i \otimes \iota_T)$ where l_i is the ith column of I_N.
(b) Verify (2.39) by showing that $(l_i' \otimes \iota_T')P = (l_i' \otimes \iota_T')$.

2.11 Using *Grunfeld's data* given as Grunfeld.fil on the Springer website, reproduce Table 2.1.
2.12 Using the *gasoline data* of Baltagi and Griffin (1983), given as Gasoline.dat on the Springer website, reproduce Table 2.5.
2.13 Using the Monte Carlo setup for the one-way error component model, given in Maddala and Mount (1973), compare the various estimators of the variance components and regression coefficients studied in this chapter.

2.14 *Bounds for s^2 in a one-way random effects model.* For the random one-way error component model given in (2.1) and (2.2), consider the OLS estimator of $\text{var}(u_{it}) = \sigma^2$, which is given by $s^2 = \widehat{u}'_{OLS}\widehat{u}_{OLS}/(n-K')$, where $n = NT$ and $K' = K+1$.

(a) Show that $E(s^2) = \sigma^2 + \sigma_\mu^2[K' - tr(I_N \otimes J_T)P_x]/(n-K')$.

(b) Consider the inequalities given by Kiviet and Krämer (1992) which state that

$$0 \leqslant \text{mean of } (n-K') \text{ smallest roots of } \Omega \leqslant E(s^2)$$
$$\leqslant \text{mean of } (n-K') \text{ largest roots of } \Omega \leqslant tr(\Omega)/(n-K')$$

where $\Omega = E(uu')$. Show that for the one-way error component model, these bounds are

$$0 \leqslant \sigma_\nu^2 + (n-TK')\sigma_\mu^2/(n-K') \leqslant E(s^2) \leqslant \sigma_\nu^2 + n\sigma_\mu^2/(n-K')$$
$$\leqslant n\sigma^2/(n-K')$$

As $n \to \infty$, both bounds tend to σ^2, and s^2 is asymptotically unbiased, irrespective of the particular evolution of X. See Baltagi and Krämer (1994) for a proof of this result.

2.15 Using the *public capital data* of Munnell (1990) given as Produc.prn on the Springer website, reproduce Table 2.6.

2.16 Using the Monte Carlo design of Baillie and Baltagi (1999), compare the four predictors described in Sect. 2.5.

2.17 *Heteroskedastic fixed effects models*. This is based on problem 96.5.1 in *Econometric Theory* by Baltagi (1996). Consider the fixed effects model

$$y_{it} = \alpha_i + u_{it} \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T_i$$

where y_{it} denotes output in industry i at time t and α_i denotes the industry fixed effect. The disturbances u_{it} are assumed to be independent with heteroskedastic variances σ_i^2. Note that the data are unbalanced with different number of observations for each industry.

(a) Show that OLS and GLS estimates of α_i are identical.

(b) Let $\sigma^2 = \sum_{i=1}^N T_i\sigma_i^2/n$ where $n = \sum_{i=1}^N T_i$, be the average disturbance variance. Show that the GLS estimator of σ^2 is unbiased, whereas the OLS estimator of σ^2 is biased. Also show that this bias disappears if the data are balanced or the variances are homoskedastic.

(c) Define $\lambda_i^2 = \sigma_i^2/\sigma^2$ for $i = 1, 2\ldots, N$. Show that for $\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_N)$

$$E[\text{estimated var}(\widehat{\alpha}_{OLS}) - \text{ true var}(\widehat{\alpha}_{OLS})]$$
$$= \sigma^2[(n - \sum_{i=1}^N \lambda_i^2)/(n-N)] \text{ diag } (1/T_i) - \sigma^2 \text{ diag } (\lambda_i^2/T_i)$$

This problem shows that in case there are no regressors in the unbalanced panel data model, fixed effects with heteroskedastic disturbances can be estimated by OLS, but one has to correct the standard errors. See solution 96.5.1 in *Econometric Theory* by Kleiber (1997).

References

Abadie, A. 2005. Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72: 1–19.

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105: 493–505.

Abrevaya, J. 2006. Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics* 21: 489–519.

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13.

Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49: 431–434.

Athey, S., and G.W. Imbens. 2006. Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74: 431–497.

Baillie, R.T., and Baltagi, B.H. 1999. Prediction from the regression model with one-way error components, Chapter 10 in Hsiao, C., Lahiri, K., Lee, L.F., and Pesaran, H. (eds.). *Analysis of panels and limited dependent variable models*. Cambridge: Cambridge University Press, 255–267.

Baltagi, B.H. 1981. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error components model. *Journal of Econometrics* 17: 21–49.

Baltagi, B.H. 1989. Applications of a necessary and sufficient condition for OLS to be BLUE. *Statistics and Probability Letters* 8: 457–461.

Baltagi, B.H. 1995. *Econometric analysis of panel data*. Chichester: Wiley.

Baltagi, B.H. 1996. Heteroskedastic fixed effects models, Problem 96.5.1. *Econometric Theory* 12: 867.

Baltagi, B.H. 2009. *A companion to econometric analysis of panel data*. Chichester: Wiley.

Baltagi, B.H. 2013. *Panel data forecasting*, Chapter 18 in the *Handbook of economic forecasting*, Vol. 2B, Graham Elliott and Allan Timmermann, (eds.). North Holland, Amsterdam, 995–1024.

Baltagi, B.H., and J.M. Griffin. 1983. Gasoline demand in the OECD: An application of pooling and testing procedures. *European Economic Review* 22: 117–137.

Baltagi, B.H., and W. Krämer. 1994. Consistency, asymptotic unbiasedness and bounds on the bias of s^2 in the linear regression model with error component disturbances. *Statistical Papers* 35: 323–328.

Baltagi, B.H., and N. Pinnoi. 1995. Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics* 20: 351–359.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119: 249–275.

Bramati, M.C., and C. Croux. 2007. Robust estimators for the fixed effects panel data model. *Econometrics Journal* 10: 521–540.

Breusch, T.S. 1987. Maximum likelihood estimation of random effects models. *Journal of Econometrics* 36: 383–389.

Card, D. 1990. The impact of the Mariel boat lift on the Miami labor market. *Industrial and Labor Relations Review* 43: 245–253.

Carpenter, C. 2004. How do zero tolerance drunk driving laws work? *Journal of Health Economics* 23: 61–83.
Chamberlain, G. 1984. Panel data, Chapter 22 in Griliches, Z., and Intrilligator, M. (eds.). *Handbook of econometrics*. North-Holland, Amsterdam, 1247–1318.
Conley, T.G., and C.R. Taber. 2011. Inference with "Difference in Differences" with a small number of policy changes. *Review of Economics and Statistics* 93: 113–125.
Cornwell, C., and P. Rupert. 1997. Unobservable individual effects, marriage, and then the earnings of young men. *Economic Inquiry* 35: 285–294.
Croissant, Y., and G. Millo. 2019. *Panel data econometrics with R*. Chichester: John Wiley.
Davidson, R., and J.G. MacKinnon. 1993. *Estimation and inference in econometrics*. New York: Oxford University Press.
Donald, S.G., and K. Lang. 2007. Inference with difference in differences and other panel data. *Review of Economics and Statistics* 89: 221–233.
Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.
Fuller, W.A., and G.E. Battese. 1974. Estimation of linear models with cross-error structure. *Journal of Econometrics* 2: 67–78.
Glick, R., and A.K. Rose. 2002. Does a currency union affect trade? *The time series evidence, European Economic Review* 46: 1125–1151.
Goldberger, A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57: 369–375.
Graybill, F.A. 1961. *An introduction to linear statistical models*. New York: McGraw-Hill.
Gruber, J., and J. Poterba. 1994. Tax incentives and the decision to purchase health insurance: Evidence from the self-employed. *Quarterly Journal of Economics* 109: 701–734.
Grunfeld, Y. 1958. The determinants of corporate investment, unpublished Ph.D. dissertation, Chicago: University of Chicago.
Hansen, C.B. 2007a. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics* 140: 670–694.
Hansen, C.B. 2007b. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141: 597–620.
Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.
Hausman, J.A., and G. Kuersteiner. 2008. Difference in difference meets generalized least squares: Higher order properties of hypotheses tests. *Journal of Econometrics* 144: 371–391.
Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.
Holtz-Eakin, D. 1994. Public-sector capital and the productivity puzzle. *Review of Economics and Statistics* 76: 12–21.
Hsiao, C., and Sun, B. 2000. To pool or not to pool panel data, Chapter 9 in Krishnakumar, J., and Ronchetti, E. (eds.). *Panel data econometrics: Future directions*, North-Holland, Amsterdam, 181–198.
Kackar, R.N., and D. Harville. 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* 79: 853–862.
Kiviet, J.F., and W. Krämer. 1992. Bias of s^2 in the linear regression model with correlated errors. *Empirical Economics* 16: 375–377.
Kleiber, C. 1997. Heteroskedastic fixed effects models, Solution 96.5.1. *Econometric Theory* 13: 891–893.
Laporte, A., and F. Windmeijer. 2005. Estimation of panel data models with binary indicators when treatment effects are not constant over time. *Economics Letters* 88: 389–396.
Lechner, M. 2015. Treatment effects and panel data, Chapter 9 in Baltagi, B.H. (ed.). *Oxford handbook of panel data*. Oxford: Oxford University Press, 257–284.
Li, Q., and J.C. Racine. 2007. *Nonparametric econometrics*. New Jersey: Princeton University Press.

Lundberg, S., and E. Rose. 2002. The effects of sons and daughters on men's labor supply and wages. *The Review of Economics and Statistics* 84: 251–268.

Maddala, G.S. 1971. The use of variance components models in pooling cross section and time series data. *Econometrica* 39: 341–358.

Maddala, G.S., and T.D. Mount. 1973. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association* 68: 324–328.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Mundlak, Y. 1978. On the pooling of time series and cross-section data. *Econometrica* 46: 69–85.

Munnell, A. 1990. How does public infrastructure affect regional economic performance? *New England Economic Review* 3–22.

Nerlove, M. 1971. Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. *Econometrica* 39: 359–382.

Rao, C.R. 1971a. Estimation of variance and covariance components -MINQUE theory. *Journal of Multivariate Analysis* 1: 257–275.

Rao, C.R. 1971b. Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1: 445–456.

Ruhm, C.J. 1996. Alcohol policies and highway vehicle fatalities. *Journal of Health Economics* 15: 435–454.

Searle, S.R. 1971. *Linear models*. New York: Wiley.

Searle, S.R., and H.V. Henderson. 1979. Dispersion matrices for variance components models. *Journal of the American Statistical Association* 74: 465–470.

Stock, J.H., and M.W. Watson. 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica* 76: 155–174.

Swamy, P.A.V.B., and S.S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40: 261–275.

Taub, A.J. 1979. Prediction in the context of the variance-components model. *Journal of Econometrics* 10: 103–108.

Taylor, W.E. 1980. Small sample considerations in estimation from panel data. *Journal of Econometrics* 13: 203–223.

Wallace, T.D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37: 55–72.

Wansbeek, T.J., and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.

Wansbeek, T.J., and A. Kapteyn. 1989. Estimation of the error components model with incomplete panels. *Journal of Econometrics* 41: 341–361.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–368.

3 The Two-Way Error Component Regression Model

3.1 Introduction

Consider the regression model given by (2.1), but with *two-way* error component disturbances:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \qquad i = 1, \ldots, N \qquad t = 1, \ldots, T \tag{3.1}$$

where μ_i denotes the unobservable *individual effect* discussed in Chap. 2, λ_t denotes the unobservable *time effect*, and ν_{it} is the remainder stochastic disturbance term. Note that λ_t is individual-invariant and it accounts for any time-specific effect that is not included in the regression. For example, it could account for strike year effects that disrupt production, oil embargo effects that disrupt the supply of oil and affect its price, Surgeon General reports on the ill-effects of smoking, or government laws restricting smoking in public places, all of which could affect consumption behavior. In vector form, (3.1) can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu \tag{3.2}$$

where Z_μ, μ, and ν were defined earlier. $Z_\lambda = \iota_N \otimes I_T$ is the matrix of time dummies that one may include in the regression to estimate the λ_t if they are fixed parameters, and $\lambda' = (\lambda_1, \ldots, \lambda_T)$. Note that $Z_\lambda Z_\lambda' = J_N \otimes I_T$ and the projection on Z_λ is $Z_\lambda(Z_\lambda' Z_\lambda)^{-1} Z_\lambda' = \bar{J}_N \otimes I_T$. This last matrix averages the data over individuals, i.e., if we regress y on Z_λ, the predicted values are given by $(\bar{J}_N \otimes I_T)y$ which have a typical element $\bar{y}_{.t} = \sum_{i=1}^{N} y_{it}/N$.

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_3

3.2 The Two-Way Fixed Effects Model

If the μ_i *and* λ_t are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, then (3.1) represents a *two-way* fixed effects error component model. The X_{it} are assumed independent of the ν_{it} for all i and t. Inference in this case is conditional on the particular N individuals and over the specific time periods observed. Recall that Z_λ, the matrix of time dummies, is $NT \times T$. If N or T is large, there will be too many dummy variables in the regression $\{(N-1)+(T-1)\}$ of them, and this causes an enormous loss in degrees of freedom. In addition, this attenuates the problem of multicollinearity among the regressors. Rather than invert a large $(N+T+K-1)$ matrix, one can obtain the fixed effects estimates of β by performing the following Within transformation given by Wallace and Hussain (1969):

$$Q = E_N \otimes E_T = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \tag{3.3}$$

where $E_N = I_N - \bar{J}_N$ and $E_T = I_T - \bar{J}_T$. This transformation "sweeps" the μ_i and λ_t effects. In fact, $\widetilde{y} = Qy$ has a typical element $\widetilde{y}_{it} = (y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..})$ where $\bar{y}_{..} = \sum_i \sum_t y_{it}/NT$, and one would perform the regression of $\widetilde{y} = Qy$ on $\widetilde{X} = QX$ to get the Within estimator $\widetilde{\beta} = (X'QX)^{-1}X'Qy$.

Note that by averaging the simple regression given in (2.8) over individuals, we get

$$\bar{y}_{.t} = \alpha + \beta\bar{x}_{.t} + \lambda_t + \bar{\nu}_{.t} \tag{3.4}$$

where we have utilized the restriction that $\sum_i \mu_i = 0$ to avoid the dummy variable trap. Similarly, the averages defined in (2.9) and (2.11) still hold using $\sum_t \lambda_t = 0$, and one can deduce that

$$(y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{y}_{..}) = (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})\beta + (\nu_{it} - \bar{\nu}_{i.} - \bar{\nu}_{.t} + \bar{\nu}_{..}) \tag{3.5}$$

OLS on this model gives $\widetilde{\beta}$, the Within estimator for the two-way model. Once again, the Within estimate of the intercept can be deduced from $\widetilde{\alpha} = \bar{y}_{..} - \widetilde{\beta}\bar{x}_{..}$ and those of μ_i and λ_t are given by

$$\widetilde{\mu}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \widetilde{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \tag{3.6}$$

$$\widetilde{\lambda}_t = (\bar{y}_{.t} - \bar{y}_{..}) - \widetilde{\beta}(\bar{x}_{.t} - \bar{x}_{..}) \tag{3.7}$$

Note that the Within estimator cannot estimate the effect of time-invariant and individual-invariant variables because the Q transformation wipes out these variables. If the true model is a two-way fixed effects model as in (3.2), then OLS on (2.1) yields biased and inconsistent estimates of the regression coefficients. OLS ignores both sets of dummy variables, whereas the one-way fixed effects estimator considered in Chap. 2 ignores only the time dummies. If these time dummies are statistically significant, the one- way fixed effects estimator will also suffer from omission bias.

3.2.1 Testing for Fixed Effects

As in the one-way error component model case, one can test for joint significance of the dummy variables:

$$H_0;\ \mu_1 = \ldots = \mu_{N-1} = 0 \qquad \text{and} \qquad \lambda_1 = \ldots = \lambda_{T-1} = 0$$

The restricted residual sums of squares (RRSS) is that of pooled OLS and the unrestricted residual sums of squares (URSS) is that from the Within regression in (3.5). In this case,

$$F_1 = \frac{(RRSS - URSS)/(N + T - 2)}{URSS/(N - 1)(T - 1) - K} \overset{H_0}{\sim} F_{(N+T-2),(N-1)(T-1)-K} \tag{3.8}$$

Next, one can test for the existence of individual effects allowing for time effects, i.e.,

$$H_2;\ \mu_1 = \ldots = \mu_{N-1} = 0 \qquad \text{allowing} \qquad \lambda_t \neq 0 \quad \text{for} \quad t = 1, \ldots, T - 1$$

The URSS is still the Within residual sum of squares. However, the RRSS is the regression with time-series dummies only, or the regression based upon

$$(y_{it} - \bar{y}_{.t}) = (x_{it} - \bar{x}_{.t})\beta + (u_{it} - \bar{u}_{.t}) \tag{3.9}$$

In this case, the resulting F-statistic is $F_2 \overset{H_0}{\sim} F_{(N-1),(N-1)(T-1)-K}$. Note that F_2 differs from F_0 in (2.12) in testing for $\mu_i = 0$. The latter tests $H_0;\ \mu_i = 0$ assuming that $\lambda_t = 0$, whereas the former tests $H_2;\ \mu_i = 0$ allowing $\lambda_t \neq 0$ for $t = 1, \ldots, T - 1$. Similarly, one can test for the existence of time effects allowing for individual effects, i.e.,

$$H_3;\ \lambda_1 = \ldots = \lambda_{T-1} = 0 \qquad \text{allowing} \qquad \mu_i \neq 0,\ i = 1, \ldots, (N - 1)$$

The RRSS is given by the regression in (2.10), while the URSS is obtained from the regression (3.5). In this case, the resulting F-statistic is $F_3 \overset{H_0}{\sim} F_{(T-1),(N-1)(T-1)-K}$. These conditional F-tests are applied to the Grunfeld data in Chap. 4.

Computational warning

As in the one-way model, s^2 from the regression in (3.5) as obtained from any standard regression package has to be adjusted for loss of degrees of freedom. In this case, one divides by $(N - 1)(T - 1) - K$ and multiplies by $(NT - K)$ to get the proper variance–covariance matrix of the Within estimator.

Empirical Applications

(i) Ram (2009) questions the body of influential research that suggests that there is a negative association between country size (as measured by logarithm of population) and government size (as proxied by government consumption as percent of GDP), and also between country size and trade openness (as measured by the sum of imports and exports as a percent of GDP). Ram uses a 41-year panel data (1960–2000) for over 150 countries from the Penn World Tables 6.1. The pooled OLS results support the foregoing scenario, whereas the two-way fixed effects results find little evidence of a negative association of country size with either government size or trade openness. Both country and time dummies are significant and indicate that pooled OLS is biased and inconsistent. Problem 3.17 asks the reader to replicate these results.

(ii) Neumayer (2003a) investigates the effect of left-wing party strength on air pollution levels using a panel of 21 OECD countries observed over the period 1980–1999. Neumayer reports the two-way fixed effects estimates of several measures of air pollution levels (like carbon dioxide emissions) regressed on measures of scale: (GDP and vehicle use), measures of composition (share of manufacturing and fossil fuels), and a measure of efficiency, as well as three measures of left-wing party strength, and one indicator of corporatism. The results find that parliamentary green/left-libertarian party strength is associated with lower pollution levels. Problem 3.18 asks the reader to replicate these results.

(iii) Neumayer (2003b) provides empirical evidence that good political governance and good economic policies can lower homicide rates. This is based on two-way fixed effects estimates using a panel of homicide data from up to 117 countries over the period 1980–97. The results suggest that economic growth, higher income levels, respect for human rights, and the abolition of the death penalty are all associated with lower homicide rates. The same is true for democracy but only at high levels of democracy. Problem 3.19 asks the reader to replicate these results.

3.3 The Two-Way Random Effects Model

If $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda^2)$ and $\nu_{it} \sim \text{IID}\ (0, \sigma_\nu^2)$ independent of each other, then this is the two-way *random* effects model. In addition, X_{it} is independent of μ_i, λ_t, and ν_{it} for all i and t. Inference in this case pertains to the large population from which this sample was randomly drawn. From (3.2), one can compute the variance–covariance matrix

$$\begin{aligned}\Omega = E(uu') &= Z_\mu E(\mu\mu')Z_\mu' + Z_\lambda E(\lambda\lambda')Z_\lambda' + \sigma_\nu^2 I_{NT} \\ &= \sigma_\mu^2(I_N \otimes J_T) + \sigma_\lambda^2(J_N \otimes I_T) + \sigma_\nu^2(I_N \otimes I_T) \qquad (3.10)\end{aligned}$$

The disturbances are homoskedastic with $\text{var}(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2$ for all i and t,

$$\begin{aligned} \text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 \; i = j, t \neq s \\ &= \sigma_\lambda^2 \; i \neq j, t = s \end{aligned} \tag{3.11}$$

and zero otherwise. This means that the correlation coefficient

$$\begin{aligned} \text{correl}(u_{it}, u_{js}) &= \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2) \; i = j, t \neq s \\ &= \sigma_\lambda^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\nu^2) \; i \neq j, = s \\ &= 1 \qquad\qquad\qquad\qquad\; i = j, t = s \\ &= 0 \qquad\qquad\qquad\qquad\; i \neq j, t \neq s \end{aligned} \tag{3.12}$$

In order to get Ω^{-1}, we replace J_N by $N\bar{J}_N$, I_N by $E_N + \bar{J}_N$, J_T by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$, and collect terms with the same matrices. This gives

$$\Omega = \sum_{i=1}^{4} \lambda_i Q_i \tag{3.13}$$

where $\lambda_1 = \sigma_\nu^2$, $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$, $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$, and $\lambda_4 = T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\nu^2$. Correspondingly, $Q_1 = E_N \otimes E_T$, $Q_2 = E_N \otimes \bar{J}_T$, $Q_3 = \bar{J}_N \otimes E_T$ and $Q_4 = \bar{J}_N \otimes \bar{J}_T$, respectively. The λ_i are the distinct characteristic roots of Ω and the Q_i are the corresponding matrices of eigenprojectors. λ_1 is of multiplicity $(N-1)(T-1)$, λ_2 is of multiplicity $(N-1)$, λ_3 is of multiplicity $(T-1)$, and λ_4 is of multiplicity 1.[1]. Each Q_i is symmetric and idempotent with its rank equal to its trace. Moreover, the Q_i are pairwise orthogonal and sum to the identity matrix. The advantages of this spectral decomposition are that

$$\Omega^r = \sum_{i=1}^{4} \lambda_i^r Q_i \tag{3.14}$$

where r is an arbitrary scalar so that

$$\sigma_\nu \Omega^{-1/2} = \sum_{i=1}^{4} (\sigma_\nu/\lambda_i^{1/2}) Q_i \tag{3.15}$$

and the typical element of $y^* = \sigma_\nu \Omega^{-1/2} y$ is given by

$$y_{it}^* = y_{it} - \theta_1 \bar{y}_{i.} - \theta_2 \bar{y}_{.t} + \theta_3 \bar{y}_{..} \tag{3.16}$$

where $\theta_1 = 1 - (\sigma_\nu/\lambda_2^{1/2})$, $\theta_2 = 1 - (\sigma_\nu/\lambda_3^{1/2})$, and $\theta_3 = \theta_1 + \theta_2 + (\sigma_\nu/\lambda_4^{1/2}) - 1$. As a result, GLS can be obtained as OLS of y^* on Z^*, where $Z^* = \sigma_\nu \Omega^{-1/2} Z$. This transformation was first derived by Fuller and Battese (1974).

The best quadratic unbiased (BQU) estimators of the variance components arise naturally from the fact that $Q_i u \sim (0, \lambda_i Q_i)$. Hence,

$$\widehat{\lambda}_i = u' Q_i u / \text{tr}(Q_i) \tag{3.17}$$

is the BQU estimator of λ_i for $i = 1, 2, 3$. These ANOVA estimators are minimum variance unbiased (MVU) under normality of the disturbances (see Graybill 1961). As in the one-way error component model, one can obtain feasible estimates of the variance components by replacing the true disturbances by OLS residuals (see Wallace and Hussain 1969). OLS is still an unbiased and consistent estimator under the random effects model but it is inefficient and results in biased standard errors and t-statistics. Alternatively, one could substitute the Within residuals with $\widetilde{u} = y - \widetilde{\alpha}\iota_{NT} - X\widetilde{\beta}$, where $\widetilde{\alpha} = \bar{y}_{..} - \bar{X}'_{..}\widetilde{\beta}$ and $\widetilde{\beta}$ is obtained by the regression in (3.5). This is the method proposed by Amemiya (1971). In fact, Amemiya (1971) shows that the Wallace and Hussain (1969) estimates of the variance components have a different asymptotic distribution from that knowing the true disturbances, while the Amemiya (1971) estimates of the variance components have the same asymptotic distribution as that knowing the true disturbances:

$$\begin{pmatrix} \sqrt{NT}(\widehat{\sigma}_\nu^2 - \sigma_\nu^2) \\ \sqrt{N}(\widehat{\sigma}_\mu^2 - \sigma_\mu^2) \\ \sqrt{T}(\widehat{\sigma}_\lambda^2 - \sigma_\lambda^2) \end{pmatrix} \sim N\left(0, \begin{pmatrix} 2\sigma_\nu^4 & 0 & 0 \\ 0 & 2\sigma_\mu^4 & 0 \\ 0 & 0 & 2\sigma_\lambda^4 \end{pmatrix}\right) \tag{3.18}$$

Substituting OLS or Within residuals instead of the true disturbances in (3.17) introduces bias in the corresponding estimates of the variance components. The degrees of freedom corrections that make these estimates unbiased depend upon traces of matrices that involve the matrix of regressors X. These corrections are given in Wallace and Hussain (1969) and Amemiya (1971), respectively. Alternatively, one can infer these correction terms from the more general unbalanced error component model considered in Chap. 9.

Swamy and Arora (1972) suggest running three least squares regressions and estimating the variance components from the corresponding mean square errors of these regressions. The first regression corresponds to the Within regression which transforms the original model by $Q_1 = E_N \otimes E_T$. This is equivalent to the regression in (3.5) and yields the following estimate of σ_ν^2:

$$\widehat{\widehat{\lambda}}_1 = \widehat{\widehat{\sigma}}_\nu^2 = [y'Q_1y - y'Q_1X(X'Q_1X)^{-1}X'Q_1y]/[(N-1)(T-1) - K] \tag{3.19}$$

The second regression is the Between individuals regression which transforms the original model by $Q_2 = E_N \otimes \bar{J}_T$. This is equivalent to the regression of $(\bar{y}_{i.} - \bar{y}_{..})$ on $(\bar{X}_{i.} - \bar{X}_{..})$ and yields the following estimate of $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$:

$$\widehat{\widehat{\lambda}}_2 = [y'Q_2y - y'Q_2X(X'Q_2X)^{-1}X'Q_2y]/[(N-1) - K] \tag{3.20}$$

from which one obtains $\widehat{\widehat{\sigma}}_\mu^2 = (\widehat{\widehat{\lambda}}_2 - \widehat{\widehat{\sigma}}_\nu^2)/T$. The third regression is the Between time periods regression which transforms the original model by $Q_3 = \bar{J}_N \otimes E_T$. This is equivalent to the regression of $(\bar{y}_{.t} - \bar{y}_{..})$ on $(\bar{X}_{.t} - \bar{X}_{..})$ and yields the following estimate of $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$:

$$\widehat{\widehat{\lambda}}_3 = [y'Q_3y - y'Q_3X(X'Q_3X)^{-1}X'Q_3y]/[(T-1) - K] \tag{3.21}$$

from which one obtains $\widehat{\widehat{\sigma}}_{\lambda}^2 = (\widehat{\widehat{\lambda}}_3 - \widehat{\widehat{\sigma}}_{\nu}^2)/N$. Stacking the three transformed regressions just performed yields

$$\begin{pmatrix} Q_1 y \\ Q_2 y \\ Q_3 y \end{pmatrix} = \begin{pmatrix} Q_1 X \\ Q_2 X \\ Q_3 X \end{pmatrix} \beta + \begin{pmatrix} Q_1 u \\ Q_2 u \\ Q_3 u \end{pmatrix} \tag{3.22}$$

since $Q_i \iota_{NT} = 0$ for $i = 1, 2, 3$, and the transformed error has mean 0 and variance–covariance matrix given by $\text{Diag}[\lambda_i Q_i]$ with $i = 1, 2, 3$. Problem 3.4 asks the reader to show that OLS on this system of $3NT$ observations yields the same estimator of β as OLS on the pooled model (2.3). Also, GLS on this system of equations (3.22) yields the same estimator of β as GLS on (2.3). In fact,

$$\begin{aligned} \widehat{\beta}_{GLS} &= [(X'Q_1X)/\sigma_{\nu}^2 + (X'Q_2X)/\lambda_2 + (X'Q_3X)/\lambda_3]^{-1} \\ &\quad \times [(X'Q_1y)/\sigma_{\nu}^2 + (X'Q_2y)/\lambda_2 + (X'Q_3y)/\lambda_3] \\ &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}[W_{Xy} + \phi_2^2 B_{Xy} + \phi_3^2 C_{Xy}] \end{aligned} \tag{3.23}$$

with $\text{var}(\widehat{\beta}_{GLS}) = \sigma_{\nu}^2[W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}$. Note that $W_{XX} = X'Q_1X$, $B_{XX} = X'Q_2X$, and $C_{XX} = X'Q_3X$ with $\phi_2^2 = \sigma_{\nu}^2/\lambda_2$, $\phi_3^2 = \sigma_{\nu}^2/\lambda_3$. Also, the Within estimator of β is $\widetilde{\beta}_W = W_{XX}^{-1} W_{Xy}$, the Between individuals estimator of β is $\widehat{\beta}_B = B_{XX}^{-1} B_{Xy}$, and the Between time periods estimator of β is $\widehat{\beta}_C = C_{XX}^{-1} C_{Xy}$. This shows that $\widehat{\beta}_{GLS}$ is a matrix-weighted average of $\widetilde{\beta}_W$, $\widehat{\beta}_B$, and $\widehat{\beta}_C$. In fact,

$$\widehat{\beta}_{GLS} = W_1 \widetilde{\beta}_W + W_2 \widehat{\beta}_B + W_3 \widehat{\beta}_C \tag{3.24}$$

where

$$\begin{aligned} W_1 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} W_{XX} \\ W_2 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} (\phi_2^2 B_{XX}) \\ W_3 &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1} (\phi_3^2 C_{XX}) \end{aligned}$$

This was demonstrated by Maddala (1971). Note that (i) if $\sigma_{\mu}^2 = \sigma_{\lambda}^2 = 0$, $\phi_2^2 = \phi_3^2 = 1$ and $\widehat{\beta}_{GLS}$ reduces to $\widehat{\beta}_{OLS}$, (ii) as T and $N \to \infty$, ϕ_2^2 and $\phi_3^2 \to 0$ and $\widehat{\beta}_{GLS}$ tends to $\widetilde{\beta}_W$, (iii) if $\phi_2^2 \to \infty$ with ϕ_3^2 finite, then $\widehat{\beta}_{GLS}$ tends $\widehat{\beta}_B$, (iv) if $\phi_3^2 \to \infty$ with ϕ_2^2 finite, then $\widehat{\beta}_{GLS}$ tends to $\widehat{\beta}_C$.

Wallace and Hussain (1969) compare $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$ in the case of nonstochastic (repetitive) X and find that both are (i) asymptotically normal, (ii) consistent and unbiased, and that (iii) $\widehat{\beta}_{GLS}$ has a smaller generalized variance (i.e,. more efficient) in finite samples. In the case of nonstochastic (nonrepetitive) X, they find that both $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$ are consistent, asymptotically unbiased, and have equivalent asymptotic variance–covariance matrices, as both N and $T \to \infty$. The last statement can be proved as follows: the limiting variance of the GLS estimator is

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} (X'\Omega^{-1}X/NT)^{-1} = \frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \left[\sum_{i=1}^{3} \frac{1}{\lambda_i} (X'Q_iX/NT) \right]^{-1} \tag{3.25}$$

But the limit of the inverse is the inverse of the limit, and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{X'Q_iX}{NT} \qquad \text{for} \quad i = 1, 2, 3 \tag{3.26}$$

all exist and are positive semi-definite, since $\lim_{\substack{N\to\infty \\ T\to\infty}}(X'X/NT)$ is assumed finite and positive definite. Hence

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{(N\sigma_\lambda^2 + \sigma_\nu^2)} \left(\frac{X'Q_3X}{NT}\right) = 0$$

and

$$\lim_{\substack{N\to\infty \\ T\to\infty}} \frac{1}{(T\sigma_\mu^2 + \sigma_\nu^2)} \left(\frac{X'Q_2X}{NT}\right) = 0$$

Therefore, the limiting variance of the GLS estimator becomes

$$\frac{1}{NT} \lim_{\substack{N\to\infty \\ T\to\infty}} \sigma_\nu^2 \left(\frac{X'Q_1X}{NT}\right)^{-1}$$

which is the limiting variance of the Within estimator.

One can extend Nerlove (1971a) method for the one-way model, by estimating σ_μ^2 as $\sum_{i=1}^N(\widehat{\mu}_i - \overline{\widehat{\mu}})^2/(N-1)$ and σ_λ^2 as $\sum_{t=1}^T(\widehat{\lambda}_t - \overline{\widehat{\lambda}})^2/(T-1)$ where the $\widehat{\mu}_i$ and $\widehat{\lambda}_t$ are obtained as coefficients from the least squares dummy variables regression (LSDV). σ_ν^2 is estimated from the Within residual sums of squares divided by NT. Baltagi (1995, Appendix 3) develops two other methods of estimating the variance components. The first is Rao's (1971) minimum norm quadratic unbiased estimation (MINQUE) and the second is Henderson's method III as described by Fuller and Battese (1973). These methods require more notation and development and may be skipped in a brief course on this subject. Chapter 9 studies these estimation methods in the context of an unbalanced error component model.

Baltagi (1981) performed a Monte Carlo study on a simple regression equation with two-way error component disturbances and studied the properties of the following estimators: ordinary least squares (OLS), the Within estimator, and six feasible GLS estimators denoted by WALHUS, AMEMIYA, SWAR, MINQUE, FUBA, and NERLOVE corresponding to the methods developed by Wallace and Hussain (1969), Amemiya (1971), Swamy and Arora (1972), Rao (1971), Fuller and Battese (1974), and Nerlove (1971a), respectively. The mean square error (MSE) of these estimators was computed relative to that of true GLS, i.e., GLS knowing the true variance components.

To review some of the properties of these estimators, OLS is unbiased, but asymptotically inefficient, and its standard errors are biased; see Moulton (1986) for the extent of this bias in empirical applications. In contrast, the Within estimator is unbiased whether or not prior information about the variance components is available.

It is also asymptotically equivalent to the GLS estimator in case of weakly nonstochastic exogenous variables. Early in the literature, Wallace and Hussain (1969) recommended the Within estimator for the practical researcher, based on theoretical considerations but more importantly for its ease of computation. In Wallace and Hussain's (1969, p. 66) words the "covariance estimators come off with a surprisingly clear bill of health." True GLS is BLUE, but the variance components are usually not known and have to be estimated. All of the feasible GLS estimators considered are asymptotically efficient. In fact, Swamy and Arora (1972) proved the existence of a family of asymptotically efficient two-stage feasible GLS estimators of the regression coefficients. Therefore, based on asymptotics only, one cannot differentiate among these two-stage GLS estimators. This leaves undecided the question of which estimator is the best to use. Swamy and Arora (1972) derived the relative efficiencies of (i) SWAR with respect to OLS, (ii) SWAR with respect to Within, and (iii) Within with respect to OLS. Then, for various values of N, T, the variance components, the Between groups, Between time periods, and Within groups sums of squares of the independent variable, they tabulated these relative efficiency values (see Swamy and Arora, 1972, p. 272). Among their basic findings is the fact that, for small samples, SWAR is less efficient than OLS if σ_μ^2 and σ_λ^2 are small. Also, SWAR is less efficient than Within if σ_μ^2 and σ_λ^2 are large. The latter result is disconcerting, since Within which uses only a part of the available data is more efficient than SWAR, a feasible GLS estimator, which uses all of the available data.

3.3.1 Monte Carlo Results

Baltagi (1981) considered the following simple regression equation:

$$y_{it} = \alpha + \beta x_{it} + u_{it} \tag{3.27}$$

with

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{3.28}$$

The exogenous variable x was generated by a similar method to that of Nerlove (1971a). Throughout the experiment, $\alpha = 5$, $\beta = 0.5$, $N = 25$, $T = 10$, and $\sigma^2 = 20$. However, $\rho = \sigma_\mu^2/\sigma^2$ and $\omega = \sigma_\lambda^2/\sigma^2$ were varied over the set (0, 0.01, 0.2, 0.4, 0.6, 0.8) such that $(1 - \rho - \omega)$ is always positive. In each experiment, 100 replications were performed. For every replication, $(NT + N + T)$ independent and identically distributed Normal IIN(0, 1) random numbers were generated. The first N numbers were used to generate the μ_i as IIN(0, σ_μ^2). The second T numbers were used to generate the λ_t as IIN(0, σ_λ^2), and the last NT numbers were used to generate the ν_{it} as IIN(0, σ_ν^2). For the estimation methods considered, the Monte Carlo results show the following:

(1) For the two-way model, the researcher should not label the problem of negative variance estimates "not serious" as in the one-way model. This is because we cannot distinguish between the case where the model is misspecified (i.e., with at least one of the variance components actually equal to zero), and the case where the model is properly specified (i.e., with at least one of the variance components relatively small but different from zero). Another important reason is that we may not be able to distinguish between a case where OLS is equivalent to GLS according to the MSE criterion and a case where it is not. For these cases, the practical solution seems to be the replacement of a negative estimate by zero. Of course, this will affect the properties of the variance components estimates especially if the actual variances are different from zero. The Monte Carlo results of Baltagi (1981) report that the performance of the two-stage GLS methods is not seriously affected by this substitution.
(2) As long as the variance components are not relatively small and close to zero, there is always gain according to the MSE criterion in performing feasible GLS rather than least squares or least squares with dummy variables.
(3) All the two-stage GLS methods considered performed reasonably well according to the relative MSE criteria. However, none of these methods could claim to be the best for all the experiments performed. Most of these methods had relatively close MSEs which therefore made it difficult to choose among them. This same result was obtained in the one-way model by Maddala and Mount (1973).
(4) Better estimates of the variance components do not necessarily give better second-round estimates of the regression coefficients. This confirms the finite sample results obtained by Taylor (1980) and extends them from the one-way to the two-way model.

Finally, the recommendation given in Maddala and Mount (1973) is still valid, i.e., always perform more than one of the two-stage GLS procedures to see whether the estimates obtained differ widely.

3.4 Maximum Likelihood Estimation

In this case, the normality assumption is needed on our error structure. The log-likelihood function is given by

$$\log L = \text{ constant } - \frac{1}{2}\log \mid \Omega \mid - \frac{1}{2}(y - Z\gamma)'\Omega^{-1}(y - Z\gamma) \tag{3.29}$$

where Ω and Ω^{-1} were given in (3.13) and (3.14). The maximum likelihood estimators of $\gamma, \sigma_\nu^2, \sigma_\mu^2$ and σ_λ^2 are obtained by simultaneously solving the following normal equations:

$$\frac{\partial \log L}{\partial \gamma} = Z'\Omega^{-1}y - (Z'\Omega^{-1}Z)\gamma = 0$$
$$\frac{\partial \log L}{\partial \sigma_\nu^2} = -\frac{1}{2}\,\mathrm{tr}\,\Omega^{-1} + \frac{1}{2}u'\Omega^{-2}u = 0$$
$$\frac{\partial \log L}{\partial \sigma_\mu^2} = -\frac{1}{2}\,\mathrm{tr}\,\Omega^{-1}(I_N \otimes J_T) + \frac{1}{2}u'\Omega^{-2}(I_N \otimes J_T)u = 0$$
$$\frac{\partial \log L}{\partial \sigma_\lambda^2} = -\frac{1}{2}\,\mathrm{tr}\,\Omega^{-1}(J_N \otimes I_T) + \frac{1}{2}u'\Omega^{-2}(J_N \otimes I_T)u = 0 \qquad (3.30)$$

Even if the u were observable, these would still be highly nonlinear and difficult to solve explicitly. However, Amemiya (1971) suggests an iterative scheme to solve (3.30). The resulting maximum likelihood estimates of the variance components are shown to be consistent and asymptotic normal with an asymptotic distribution given by (3.18).

Following Breusch (1987), one can write the likelihood for the two-way model as

$$\begin{aligned} L(\alpha, \beta, \sigma_\nu^2, \phi_2^2, \phi_3^2) = {} & \text{constant} \; - (NT/2)\log\sigma_\nu^2 + (N/2)\log\phi_2^2 + (T/2)\log\phi_3^2 \\ & -(1/2\log[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2] - (1/2\sigma_\nu^2)u'\Sigma^{-1}u \end{aligned} \qquad (3.31)$$

where $\Omega = \sigma_\nu^2\Sigma = \sigma_\nu^2(\sum_{i=1}^4 Q_i/\phi_i^2)$ from (3.13) with $\phi_i^2 = \sigma_\nu^2/\lambda_i$ for $i = 1, \ldots, 4$. The likelihood (3.31) uses the fact that $|\,\Omega\,|^{-1} = (\sigma_\nu^2)^{-NT}(\phi_2^2)^{N-1}(\phi_3^2)^{T-1}\phi_4^2$. The feasibility conditions $\infty > \lambda_i \geqslant \sigma_\nu^2$ are equivalent to $0 < \phi_i^2 \leqslant 1$ for $i = 1, 2, 3, 4$. Following Breusch (1987), we define $d = y - X\beta$, therefore $u = d - \iota_{NT}\alpha$. Given arbitrary values of β,ϕ_2^2,ϕ_3^2, one can concentrate this likelihood function with respect to α and σ_ν^2. Estimates of α and σ_ν^2 are obtained later as $\widehat{\alpha} = \iota'_{NT}d/NT$ and $\widehat{\sigma}_\nu^2 = (u'\Sigma^{-1}u/NT)$. Substituting the maximum value of α in u one gets $u = d - \iota_{NT}\widehat{\alpha} = (I_{NT} - \bar{J}_{NT})d$. Also, using the fact that

$$(I_{NT} - \bar{J}_{NT})\Sigma^{-1}(I_{NT} - \bar{J}_{NT}) = Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3$$

one gets $\widehat{\sigma}_\nu^2 = d'[Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3]d/NT$, given β, ϕ_2^2, and ϕ_3^2. The concentrated likelihood function becomes

$$\begin{aligned} L_C(\beta, \phi_2^2, \phi_3^2) = {} & \text{constant} \; - (NT/2)\log[d'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)d] \\ & +(N/2)\log\phi_2^2 + (T/2)\log\phi_3^2 \\ & -(1/2)\log[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2] \end{aligned} \qquad (3.32)$$

Maximizing L_C over β, given ϕ_2^2 and ϕ_3^2, Baltagi and Li (1992a) get

$$\widehat{\beta} = [X'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)X]^{-1}X'(Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3)y \qquad (3.33)$$

which is the GLS estimator knowing ϕ_2^2 and ϕ_3^2. Similarly, maximizing L_C over ϕ_2^2, given β and ϕ_3^2, one gets[2]

$$\frac{\delta L_C}{\delta \phi_2^2} = -\frac{NT}{2}\frac{d'Q_2d}{d'[Q_1 + \phi_2^2 Q_2 + \phi_3^2 Q_3]d} + \frac{N}{2}\frac{1}{\phi_2^2} - \frac{1}{2}\frac{(1-\phi_3^2)}{[\phi_2^2 + \phi_3^2 - \phi_2^2\phi_3^2]} = 0 \tag{3.34}$$

which can be written as

$$a\phi_2^4 + b\phi_2^2 + c = 0 \tag{3.35}$$

where $a = -[N(T-1)+1](1-\phi_3^2)(d'Q_2d)$, $b = (1-\phi_3^2)(N-1)d'[Q_1 + \phi_3^2 Q_3]d - \phi_3^2(T-1)N(d'Q_2d)$, and $c = N\phi_3^2 d'[Q_1 + \phi_3^2 Q_3]d$. We will *fix* ϕ_3^2, where $(0 < \phi_3^2 < 1)$ and focus on iterating between β and ϕ_2^2.[3] For a fixed ϕ_3^2, if $\phi_2^2 = 0$, then (3.33) becomes $\widehat{\beta}_{BW} = [X'(Q_1 + \phi_3^2 Q_3)X]^{-1}X'(Q_1 + \phi_3^2 Q_3)y$, which is a matrix-weighted average of the Within estimator $\widehat{\beta}_W = (X'Q_1X)^{-1}X'Q_1y$ and the Between time periods estimator $\widehat{\beta}_C = (X'Q_3X)^{-1}X'Q_3y$. If $\phi_2^2 \to \infty$, with ϕ_3^2 fixed, then (3.33) reduces to the Between individuals estimator $\widehat{\beta}_B = (X'Q_2X)^{-1}X'Q_2y$. Using standard assumptions, Baltagi and Li (1992a) show that $a < 0$ and $c > 0$ in (3.35). Hence $b^2 - 4ac > b^2 > 0$, and the unique positive root of (3.35) is

$$\widehat{\phi_2^2} = \left[-b - \sqrt{b^2 - 4ac}\right]/2a = \left[b + \sqrt{b^2 + 4 \mid a \mid c}\right]/2 \mid a \mid \tag{3.36}$$

Since ϕ_3^2 is fixed, we let $\bar{Q}_1 = Q_1 + \phi_3^2 Q_3$, then (3.33) becomes

$$\widehat{\beta} = [X'(\bar{Q}_1 + \phi_2^2 Q_2)X]^{-1}X'(\bar{Q}_1 + \phi_2^2 Q_2)y \tag{3.37}$$

Iterated GLS can be obtained through the successive application of (3.36) and (3.37). Baltagi and Li (1992a) show that the update of $\phi_2^2(i+1)$ in the $(i+1)$th iteration will be positive and finite even if the initial $\beta(i)$ value is $\widehat{\beta}_{BW}$ (from $\phi_2^2(i) = 0$) or $\widehat{\beta}_B$ (from the limit as $\phi_2^2(i) \to \infty$). More importantly, Breusch (1987) "remarkable property" extends to the two-way error component model in the sense that the ϕ_2^2 form a monotonic sequence. Therefore, if one starts with $\widehat{\beta}_{BW}$, which corresponds to $\phi_2^2 = 0$, the sequence of ϕ_2^2 is strictly increasing. On the other hand, starting with $\widehat{\beta}_B$, which corresponds to $\phi_2^2 \to \infty$, the sequence of ϕ_2^2 is strictly decreasing. This remarkable property allows the applied researcher to check for the possibility of multiple local maxima. For a fixed ϕ_3^2, starting with *both* $\widehat{\beta}_{BW}$ and $\widehat{\beta}_B$ as initial values, there is a single maximum if and only if both sequences of iterations converge to the same ϕ_2^2 estimate.[4] Since this result holds for any arbitrary ϕ_3^2 between zero and one, a search over ϕ_3^2 in this range will guard against multiple local maxima. Of course, there are other computationally more efficient maximum likelihood algorithms. In fact, two-way MLE can be implemented using TSP. The iterative algorithm described here is of value for pedagogical reasons as well as for guarding against a local maximum.

3.5 Prediction

How does the best linear unbiased predictor (BLUP) look like for the ith individual, S periods ahead for the two-way model? From (3.1), for period $T+S$

$$u_{i,T+S} = \mu_i + \lambda_{T+S} + \nu_{i,T+S} \tag{3.38}$$

and

$$\begin{aligned} E(u_{i,T+S}u_{jt}) &= \sigma_\mu^2 \text{ for } i = j \\ &= 0 \quad \text{for } i \neq j \end{aligned} \tag{3.39}$$

and $t = 1, 2, \ldots, T$. Hence, for the BLUP given in (2.37), $w = E(u_{i,T+S}u) = \sigma_\mu^2(l_i \otimes \iota_T)$ remains the same where l_i is the ith column of I_N. However, Ω^{-1} is given by (3.14), and

$$w'\Omega^{-1} = \sigma_\mu^2(l_i' \otimes \iota_T')\left[\sum_{i=1}^{4} \frac{1}{\lambda_i} Q_i\right] \tag{3.40}$$

Using the fact that

$$\begin{aligned} (l_i' \otimes \iota_T')Q_1 = 0 \quad (l_i' \otimes \iota_T')Q_2 = (l_i' \otimes \iota_T') - \iota_{NT}'/N \\ (l_i' \otimes \iota_T')Q_3 = 0 \quad (l_i' \otimes \iota_T')Q_4 = \iota_{NT}'/N \end{aligned} \tag{3.41}$$

one gets

$$w'\Omega^{-1} = \frac{\sigma_\mu^2}{\lambda_2}[(l_i' \otimes \iota_T') - \iota_{NT}'/N] + \frac{\sigma_\mu^2}{\lambda_4}(\iota_{NT}'/N) \tag{3.42}$$

Therefore, the typical element of $w'\Omega^{-1}\widehat{u}_{GLS}$ where $\widehat{u}_{GLS} = y - Z\widehat{\delta}_{GLS}$ is

$$\frac{T\sigma_\mu^2}{(T\sigma_\mu^2 + \sigma_\nu^2)}(\overline{\widehat{u}}_{i.,GLS} - \overline{\widehat{u}}_{..,GLS}) + \frac{T\sigma_\mu^2}{(T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\nu^2)}\overline{\widehat{u}}_{..,GLS} \tag{3.43}$$

or

$$\frac{T\sigma_\mu^2}{(T\sigma_\mu^2 + \sigma_\nu^2)}\overline{\widehat{u}}_{i.,GLS} + T\sigma_\mu^2\left[\frac{1}{\lambda_4} - \frac{1}{\lambda_2}\right]\overline{\widehat{u}}_{..,GLS}$$

where $\overline{\widehat{u}}_{i.,GLS} = \sum_{t=1}^{T} \widehat{u}_{it,GLS}/T$ and $\overline{\widehat{u}}_{..,GLS} = \sum_i \sum_t \widehat{u}_{it,GLS}/NT$. See problem 88.1.1 in *Econometric Theory* by Baltagi (1988) and its solution 88.1.1 by Koning (1989). In general, $\overline{\widehat{u}}_{..,GLS}$ is not necessarily zero. The GLS normal equations are $Z'\Omega^{-1}\widehat{u}_{GLS} = 0$. However, if Z contains a constant, then $\iota_{NT}'\Omega^{-1}\widehat{u}_{GLS} = 0$, and using the fact that $\iota_{NT}'\Omega^{-1} = \iota_{NT}'/\lambda_4$ from (3.14), one gets $\overline{\widehat{u}}_{..,GLS} = 0$. Hence, for the two-way model, if there is a constant in the model, the BLUP for $y_{i,T+S}$ corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual

$$\widehat{y}_{i,T+S} = Z_{i,T+S}'\widehat{\delta}_{GLS} + \left(\frac{T\sigma_\mu^2}{T\sigma_\mu^2 + \sigma_\nu^2}\right)\overline{\widehat{u}}_{i.,GLS} \tag{3.44}$$

This looks exactly like the BLUP for the one-way model but with a different Ω. If there is no constant in the model, the last term in (3.44) should be replaced by (3.43).

How would one forecast with a two-way fixed effects model with both country and time effects? After all, future coefficients of time dummies cannot be estimated unless more structure can be placed on the model. One example is the study by Schmalensee, Stoker, and Judson (1998) which forecasted the world carbon dioxide emissions through 2050 using national-level panel data over the period 1950–1990. This consisted of 4018 observations. In 1990, this data covered 141 countries which accounted for 98.6% of the world's population. This paper estimated a reduced form model relating per capita CO_2 emissions from energy consumption to a flexible functional form of real GDP per capita using time and period fixed effects. Schmalensee, Stoker, and Judson (1998) forecasted the time effects using a linear spline model with different growth rates prior to 1970 and after 1970, i.e., $\lambda_t = \gamma_1 + \gamma_2 t + \gamma_3 (t - 1970).1[t \geqq 1970]$, with the last term being an indicator function which is 1 when $t \geqq 1970$, and also using a nonlinear trend model including a logarithmic term, i.e., $\lambda_t = \delta_1 + \delta_2 t + \delta_3 \ln(t - 1940)$. Although these two time effects specifications had essentially the same goodness-of-fit performance, they resulted in different out of sample projections. The linear spline projected the time effects by continuing the estimated 1970–1990 trend to 2050, while the nonlinear trend projected a flattening trend consistent with the trend deceleration from 1950 to 1990. An earlier study by Holtz-Eakin and Selden (1995) employed 3754 observations over the period 1951–1986. For their main case, they simply set the time effect at its value in the last year in their sample.

3.6 Examples

3.6.1 Example 1: Investment Equation

For Grunfeld's (1958) example considered in Chap. 2, the investment equation is estimated using a two-way error component model. Table 3.1 gives OLS, Within, three feasible GLS estimates, and the iterative MLE for the slope coefficients. The Within estimator yields a $\widetilde{\beta}_1$ estimate at 0.118 (0.014) and $\widetilde{\beta}_2$ estimate at 0.358 (0.023). In fact, Table 3.2 gives the EViews output for the two-way fixed effects estimator. This is performed under the panel option with fixed individual and fixed time effects. For the random effects estimators, both the SWAR and WALHUS report negative estimates of σ^2_λ and this is replaced by zero. Table 3.3 gives the EViews output for the random effects estimator of the two-way error component model for the Wallace and Hussain (1969) option. Table 3.4 gives the EViews output for the Amemiya (1971) estimator. In this case, the estimate of σ_λ is 15.8, the estimate of σ_μ is 89.3, and the estimate of σ_ν is 51.7. This means that the variance of the time effects is only 2.3% of the total variance, while the variance of the firm effects is 73.1% of the total variance, and the variance of the remainder effects is 24.6% of the total variance. Table 3.5 gives the EViews output for the Swamy and Arora (1972) estimator. The iterative maximum likelihood method yields $\widehat{\beta}_1$ at 0.110 (0.010) and $\widehat{\beta}_2$ estimate at 0.309 (0.020). This was performed using TSP.

Table 3.1 Grunfeld's data: Two-way error component results

	β_1	β_2	σ_μ	σ_λ	σ_ν
OLS	0.116 (0.006)[a]	0.231 (0.025)[a]			
Within	0.118 (0.014)	0.358 (0.023)			
WALHUS	0.110 (0.010)	0.308 (0.017)	87.31	0	55.33
AMEMIYA	0.111 (0.011)	0.324 (0.019)	89.26	15.78	51.72
SWAR	0.110 (0.011)	0.308 (0.017)	84.23	0	51.72
IMLE	0.110 (0.010)	0.309 (0.020)	80.41	3.87	52.35

[a]These are biased standard errors when the true model has error component disturbances (see Moulton 1986)

Table 3.2 Grunfeld's data. Two-way Within estimator

Dependent Variable: I
Method: Panel Least Squares

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-80.16380	14.84402	-5.400409	0.0000
F	0.117716	0.013751	8.560354	0.0000
K	0.357916	0.022719	15.75404	0.0000

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.951693	Mean dependent var	145.9582
Adjusted R-squared	0.943118	S.D. dependent var	216.8753
S.E. of regression	51.72452	Akaike info criterion	10.87132
Sum squared resid	452147.1	Schwarz criterion	11.38256
Log likelihood	-1056.132	F-statistic	110.9829
Durbin-Watson stat	0.719087	Prob(F-statistic)	0.000000

Table 3.3 Grunfeld's data. Two-way Wallace and Hussain estimator

Dependent Variable: I
Method: Panel EGLS (Two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wallace and Hussain estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-57.81705	28.63258	-2.019275	0.0448
F	0.109776	0.010473	10.48183	0.0000
K	0.308069	0.017186	17.92575	0.0000

Effects Specification

Cross-section random S.D. / Rho	87.31428	0.7135
Period random S.D. / Rho	0.000000	0.0000
Idiosyncratic random S.D. / Rho	55.33298	0.2865

Weighted Statistics

R-squared	0.769560	Mean dependent var	20.47837
Adjusted R-squared	0.767221	S.D. dependent var	109.4624
S.E. of regression	52.81254	Sum squared resid	549465.3
F-statistic	328.9438	Durbin-Watson stat	0.681973
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.803316	Mean dependent var	145.9582
Sum squared resid	1840949.	Durbin-Watson stat	0.203548

3.6.2 Example 2: Gasoline Demand Equation

For the motor gasoline data in Baltagi and Griffin (1983) considered in Chap. 2, the gasoline demand equation is estimated using a two-way error component model. Table 3.6 gives OLS, Within, three feasible GLS estimates, and iterative MLE for

Table 3.4 Grunfeld's data. Two-way Amemiya/Wansbeek and Kapteyn estimator

Dependent Variable: I
Method: Panel EGLS (Two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Wansbeek and Kapteyn estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-63.89217	30.53284	-2.092573	0.0377
F	0.111447	0.010963	10.16577	0.0000
K	0.323533	0.018767	17.23947	0.0000
Effects Specification				
Cross-section random S.D. / Rho			89.26257	0.7315
Period random S.D. / Rho			15.77783	0.0229
Idiosyncratic random S.D. / Rho			51.72452	0.2456
Weighted Statistics				
R-squared	0.748982	Mean dependent var		18.61292
Adjusted R-squared	0.746433	S.D. dependent var		101.7143
S.E. of regression	51.21864	Sum squared resid		516799.9
F-statistic	293.9017	Durbin-Watson stat		0.675336
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.798309	Mean dependent var		145.9582
Sum squared resid	1887813.	Durbin-Watson stat		0.199923

the slope coefficients. The Within estimator is drastically different from OLS. The WALHUS and SWAR methods yield negative estimates of σ_λ^2 and this is replaced by zero. IMLE is obtained using TSP.

Table 3.5 Grunfeld's data. Two-way Swamy and Arora estimator

Dependent Variable: I
Method: Panel EGLS (Two-way random effects)

Sample: 1935 1954
Cross-sections included: 10
Total panel (balanced) observations: 200
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-57.86538	29.39336	-1.968655	0.0504
F	0.109790	0.010528	10.42853	0.0000
K	0.308190	0.017171	17.94833	0.0000
Effects Specification				
Cross-section random S.D. / Rho			84.23332	0.7262
Period random S.D. / Rho			0.000000	0.0000
Idiosyncratic random S.D. / Rho			51.72452	0.2738
Weighted Statistics				
R-squared	0.769400	Mean dependent var		19.85502
Adjusted R-squared	0.767059	S.D. dependent var		109.2695
S.E. of regression	52.73776	Sum squared resid		547910.4
F-statistic	328.6473	Durbin-Watson stat		0.683945
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.803283	Mean dependent var		145.9582
Sum squared resid	1841262.	Durbin-Watson stat		0.203524

3.6.3 Example 3: Public Capital Productivity

For the Munnell (1990) public capital data considered by Baltagi and Pinnoi (1995) in Chap. 2, the Cobb–Douglas production function is estimated using a two-way error component model. Table 3.7 gives OLS, Within, three feasible GLS estimates,

Table 3.6 Gasoline demand data. Two-way error component results

	β_1	β_2	β_3	σ_μ	σ_λ	σ_ν
OLS	0.889 (0.036)[a]	−0.892 (0.030)[a]	−0.763 (0.019)[a]			
Within	0.051 (0.091)	−0.193 (0.043)	−0.593 (0.028)			
WALHUS	0.545 (0.056)	−0.450 (0.039)	−0.605 (0.025)	0.197	0	0.115
AMEMIYA	0.170 (0.080)	−0.233 (0.041)	-0.602 (0.026)	0.423	0.131	0.081
SWAR	0.565 (0.061)	−0.405 (0.040)	−0.609 (0.026)	0.196	0	0.081
IMLE	0.231 (0.091)	−0.254 (0.045)	−0.606 (0.026)	0.361	0.095	0.082

[a]These are biased standard errors when the true model has error component disturbances (see Moulton 1986)

Table 3.7 Public capital data. Two-way error component results

	β_1	β_2	β_3	β_4	σ_μ	σ_λ	σ_ν
OLS	0.155 (0.017)[a]	0.309 (0.010)[a]	0.594 (0.014)[a]	-0.007 (0.001)[a]			
Within	-0.030 (0.027)	0.169 (0.028)	0.769 (0.028)	−0.004 (0.001)			
WALHUS	0.026 (0.023)	0.258 (0.021)	0.742 (0.024)	−0.005 (0.001)	0.082	0.016	0.036
AMEMIYA	0.002 (0.025)	0.217 (0.024)	0.770 (0.026)	−0.004 (0.001)	0.154	0.026	0.034
SWAR	0.018 (0.023)	0.266 (0.021)	0.745 (0.024)	−0.005 (0.001)	0.083	0.010	0.034
IMLE	0.020 (0.024)	0.250 (0.023)	0.750 (0.025)	−0.004 (0.001)	0.091	0.017	0.035

[a]These are biased standard errors when the true model has error component disturbances (see Moulton 1986)

and iterative MLE for the slope coefficients. With the exception of OLS, estimates of the public capital coefficient are insignificant in this production function. Also, none of the feasible GLS estimators yield negative estimates of the variance components.

3.7 Computational Note

EViews allows easy estimation of two-way random effects error component models using two drop down windows for *period* and *cross-section* effects which can be chosen as fixed or random. The two-way random effects specification can be done with Wallace and Hussain (1969), Amemiya (1971) or Swamy and Arora (1972) estimation as illustrated in this chapter. When one effect is *random* and the other is *fixed*, this is denoted as a *mixed* model; see problem 3.16. For an extension to the *three-way error component model,* see problem 3.15. For the *nested* error components models, see problem 3.14. These extensions are taken up again in Chap. 9 when we deal with *unbalanced* panel data. The reader is referred to Mátyás (2017) for insight into the econometrics of multi-dimensional panels.

3.8 Notes

1. These characteristic roots and eigenprojectors were first derived by Nerlove (1971b) for the two-way error component model. More details are given in Appendix 1 of Baltagi (1995).
2. Alternatively, one can maximize L_C over ϕ_3^2, given β and ϕ_2^2. The results are symmetric and are left as an exercise. In fact, one can show (see problem 3.6) that ϕ_3^2 will satisfy a quadratic equation like (3.35) with N exchanging places with T, ϕ_2^2 replacing ϕ_3^2, and Q_2 exchanging places with Q_3 in a, b, and c, respectively.
3. The case where $\phi_3^2 = 1$ corresponds to $\sigma_\lambda^2 = 0$, i.e., the one-way error component model where Breusch's (1987) results apply.
4. There will be no local maximum interior to $0 < \phi_2^2 \leqslant 1$, if starting from $\widehat{\beta}_{BW}$ we violate the nonnegative variance component requirement, $\phi_2^2 \leqslant 1$. In this case, one should set $\phi_2^2 = 1$.

3.9 Problems

3.1 *Two-way fixed effects regression.*

(a) Prove that the Within estimator $\widetilde{\beta} = (X'QX)^{-1}X'Qy$ with Q defined in (3.3) can be obtained from OLS on the panel regression model (2.3) with disturbances defined in (3.2). Hint: Use the Frisch–Waugh–Lovell theorem of Davidson and MacKinnon (1993, p. 19), and also the generalized inverse matrix result given in problem 9.6. See the complete solution in Chap. 3 of the companion, Baltagi (2009).

(b) *Within two-way is equivalent to two Withins one-way*. This is based on problem 98.5.2 in *Econometric Theory* by Baltagi (1998). Show that the Within two-way estimator of β can be obtained by applying two Within (one-way) transformations. The first is the Within transformation ignoring the time effects followed by the Within transformation ignoring the individual effects. Show that the order of these two Within (one-way) transformations is unimportant. Give an intuitive explanation for this result. See solution 98.5.2 in *Econometric Theory* by Li (1999).

3.2 *OLS and GLS are equivalent for the two-way Within transformed model.*

(a) Using generalized inverse, show that OLS or GLS on (2.6) with Q defined in (3.3) yields $\widetilde{\beta}$, the Within estimator.
(b) Show that (2.6) with Q defined in (3.3) satisfies the necessary and sufficient condition for OLS to be equivalent to GLS (see Baltagi 1989).

3.3 *Fuller and Battese* (1973) *transformation for the two-way random effects model.*

(a) Verify (3.10) and (3.13) and check that $\Omega^{-1}\Omega = I$ using (3.14).
(b) Verify that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$ using (3.14).
(c) Premultiply y by $\sigma_\nu\Omega^{-1/2}$ from (3.15) and show that the typical element is given by (3.16).

3.4 *System estimation of the two-way model: OLS versus GLS.*

(a) Perform OLS on the system of equations given in (3.22) and show that the resulting estimate is $\widehat{\beta}_{OLS} = (X(I_{NT} - \bar{J}_{NT})X)^{-1}X'(I_{NT} - \bar{J}_{NT})y$.
(b) Perform GLS on this system of equations and show that $\widehat{\beta}_{GLS}$ reduces to the expression given by (3.23).

3.5 *Unbiased estimates of the variance components: The two-way model.* Show that the Swamy and Arora (1972) estimators of λ_1, λ_2 and λ_3 given by (3.19), (3.20), and (3.21) are unbiased for σ_ν^2, λ_2, and λ_3, respectively.

3.6 *Maximum likelihood estimation of the two-way random effects model.*

(a) Using the concentrated likelihood function in (3.32), solve $\partial L_C/\partial\beta = 0$, given ϕ_2^2 and ϕ_3^2, and verify (3.33).
(b) Solve $\partial L_C/\partial\phi_2^2 = 0$, given ϕ_3^2 and β, and verify (3.34).
(c) Solve $\partial L_C/\partial\phi_3^2 = 0$, given ϕ_2^2 and β, and show that the solution ϕ_3^2 satisfies

$$\overline{a}\phi_3^4 + \overline{b}\phi_3^2 + \overline{c} = 0$$

where

$$\overline{a} = -[T(N-1)+1](1-\phi_2^2)(d'Q_3 d)$$
$$\overline{b} = (1-\phi_2^2)(T-1)d'[Q_1 + \phi_2^2 Q_2]d - \phi_2^2 T(N-1)d'Q_3 d$$

and

$$\overline{c} = T\phi_2^2 d'(Q_1 + \phi_2^2 Q_2)d$$

Note that this is analogous to (3.35), with ϕ_2^2 replacing ϕ_3^2, N replacing T, and Q_2 replacing Q_3 and vice versa, wherever they occur.

3.7 *Prediction in the two-way random effects model.*

(a) For the two-way error component model in (3.1), verify (3.39) and (3.42).
(b) Also, show that if there is a constant in the regression $\iota'_{NT}\Omega^{-1}\widehat{u}_{GLS} = 0$, and $\bar{\bar{\widehat{u}}}_{..,GLS} = 0$.

3.8 Using *Grunfeld's data* given on the Springer website as Grunfeld.fil, reproduce Table 3.1.

3.9 Using the *gasoline data* of Baltagi and Griffin (1983), given as Gasoline.dat on the Springer website, reproduce Table 3.6.

3.10 Using the *public capital data* of Munnell (1990) given as Produc.prn on the Springer website, reproduce Table 3.7.

3.11 Using the Monte Carlo setup for the two-way error component model given in (3.27) and (3.28) (see Baltagi 1981), compare the various estimators of the variance components and regression coefficients studied in this chapter.

3.12 *Variance component estimation under misspecification*. This is based on problem 91.3.3 in *Econometric Theory* by Baltagi and Li (1991). This problem investigates the consequences of under- or overspecifying the error component model on the variance components estimates. Since the one-way and two-way error component models are popular in economics, we focus on the following two cases:

(1) *Underspecification*: In this case, the true model is two-way; see (3.1):

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \qquad i = 1, \ldots, N \qquad t = 1, \ldots, T$$

while the estimated model is one-way; see (2.2):

$$u_{it} = \mu_i + \nu_{it}$$

$\mu_i \sim \text{IID}(0, \sigma^2_\mu)$, $\lambda_t \sim \text{IID}(0, \sigma^2_\lambda)$, and $\nu_{it} \sim \text{IID}(0, \sigma^2_\nu)$ independent of each other and among themselves.

(a) Knowing the true disturbances (u_{it}), show that the BQUE of σ^2_ν for the misspecified one-way model is biased upwards, while the BQUE of σ^2_μ remains unbiased.
(b) Show that if the u_{it} are replaced by the one-way least squares dummy variables (LSDV) residuals, the variance component estimate of σ^2_ν given in part (a) is inconsistent, while that of σ^2_μ is consistent.

(2) *Overspecification*: In this case, the true model is one-way, given by (2.2), while the estimated model is two-way, given by (3.1).

(c) Knowing the true disturbances (u_{it}), show that the BQUE of σ^2_μ, σ^2_λ and σ^2_ν for the misspecified two-way model remain unbiased.

(d) Show that if the u_{it} are replaced by the two-way (LSDV) residuals, the variance components estimates given in part (c) remain consistent. Hint: see solution 91.3.3 in *Econometric Theory* by Baltagi and Li (1992b).

3.13 *Bounds for s^2 in a two-way random effects model*. For the random two-way error component model described by (2.1) and (3.1), consider the OLS estimator of $\text{var}(u_{it}) = \sigma^2$, which is given by $s^2 = \widehat{u}'_{OLS}\widehat{u}_{OLS}/(n - K')$ where $n = NT$ and $K' = K + 1$.

(a) Show that

$$E(s^2) = \sigma^2 - \sigma_\mu^2[\text{tr}(I_N \otimes J_T)P_x - K']/(n - K')$$
$$-\sigma_\lambda^2[\text{tr}(J_N \otimes I_T)P_x - K']/(n - K')$$

(b) Consider the inequalities given by Kiviet and Krämer (1992) which are reproduced in problem 2.14, part (b). Show that for the two-way error component model, these bounds are given by the following two cases:

(1) For $T\sigma_\mu^2 < N\sigma_\lambda^2$:

$$0 \leqslant \sigma_\nu^2 + \sigma_\mu^2(n - T)/(n - K') + \sigma_\lambda^2(n - NK')/(n - K') \leqslant E(s^2)$$
$$\leqslant \sigma_\nu^2 + \sigma_\mu^2[n/(n - K')] + \sigma_\lambda^2[n/(n - K')] \leqslant \sigma^2(n/n - K')$$

(2) For $T\sigma_\mu^2 > N\sigma_\lambda^2$:

$$0 \leqslant \sigma_\nu^2 + \sigma_\mu^2(n - TK')/(n - K') + \sigma_\lambda^2(n - N)/(n - K') \leqslant E(s^2)$$
$$\leqslant \sigma_\nu^2 + \sigma_\mu^2[n/(n - K')] + \sigma_\lambda^2[n/(n - K')] \leqslant \sigma^2(n/n - K')$$

In either case, as $n \to \infty$, both bounds tend to σ^2 and s^2 is asymptotically unbiased, irrespective of the particular evolution of X. See Baltagi and Krämer (1994) for a proof of this result.

3.14 *Nested effects*. This is based on problem 93.4.2 in *Econometric Theory* by Baltagi (1993). In many economic applications, the data may contain nested groupings. For example, data on firms may be grouped by industry and data on states by region and data on individuals by profession. In this case, one can control for unobserved industry and firm effects using a nested error component model. Consider the regression equation

$$y_{ijt} = x'_{ijt}\beta + u_{ijt} \quad \text{for } i = 1, \ldots, M; j = 1, \ldots, N \quad \text{and} \quad t = 1, 2, \ldots, T$$

where y_{ijt} could denote the output of the jth firm in the ith industry for the tth time period. x_{ijt} denotes a vector of k inputs, and the disturbance is given by

$$u_{ijt} = \mu_i + \nu_{ij} + \epsilon_{ijt}$$

where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{ij} \sim \text{IID}(0, \sigma_\nu^2)$, and $\epsilon_{ijt} \sim \text{IID}(0, \sigma_\epsilon^2)$, independent of each other and among themselves. This assumes that there are M industries with N firms in each industry observed over T periods.

(1) Derive $\Omega = E(uu')$ and obtain $\Omega^{-1}$and $\Omega^{-1/2}$.
(2) Show that $y^* = \sigma_\epsilon \Omega^{-1/2} y$ has a typical element

$$y^*_{ijt} = (y_{ijt} - \theta_1 \bar{y}_{ij.} + \theta_2 \bar{y}_{i..})$$

where $\theta_1 = 1 - (\sigma_\epsilon/\sigma_1)$ with $\sigma_1^2 = (T\sigma_\nu^2 + \sigma_\epsilon^2)$; $\theta_2 = -(\sigma_\epsilon/\sigma_1) + (\sigma_\epsilon/\sigma_2)$ with $\sigma_2^2 = (NT\sigma_\mu^2 + T\sigma_\nu^2 + \sigma_\epsilon^2)$; $\bar{y}_{ij.} = \sum_{t=1}^T y_{ijt}/T$ and $\bar{y}_{i..} = \sum_{j=1}^N \sum_{t=1}^T y_{ijt}/NT$. See solution 93.4.2 in *Econometric Theory* by Xiong (1995).

3.15 *Three-way error component model.* Ghosh (1976) considered the following error component model:

$$u_{itq} = \mu_i + \lambda_t + \eta_q + \nu_{itq}$$

where $i = 1, \ldots, N$; $T = 1, \ldots, T$; and $q = 1, \ldots, M$. Ghosh (1976) argued that in international or interregional studies, there might be two rather than one cross-sectional component; for example, i might denote countries and q might be regions within that country. These four *independent* error components are assumed to be random with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda^2)$, $\eta_q \sim \text{IID}(0, \sigma_\eta^2)$ and $\nu_{itq} \sim \text{IID}(0, \sigma_\nu^2)$. Order the observations such that the faster index is q, while the slowest index is t, so that

$$u' = (u_{111}, \ldots, u_{11M}, u_{121}, \ldots, u_{12M}, \ldots, u_{1N1}, \ldots, \\ u_{1NM}, \ldots, u_{T11}, \ldots, u_{T1M}, \ldots, u_{TN1}, \ldots, u_{TNM})$$

(a) Show that the error has mean zero and variance–covariance matrix

$$\Omega = E(uu') = \sigma_\nu^2 (I_T \otimes I_N \otimes I_M) + \sigma_\lambda^2 (I_T \otimes J_N \otimes J_M) \\ + \sigma_\mu^2 (J_T \otimes I_N \otimes J_M) + \sigma_\eta^2 (J_T \otimes J_N \otimes I_M)$$

(b) Using the Wansbeek and Kapteyn (1982b) trick, show that $\Omega = \sum_{j=1}^5 \xi_j V_j$ where $\xi_1 = \sigma_\nu^2$, $\xi_2 = NM\sigma_\lambda^2 + \sigma_\nu^2$, $\xi_3 = TM\sigma_\mu^2 + \sigma_\nu^2$, $\xi_4 = NT\sigma_\eta^2 + \sigma_\nu^2$, and $\xi_5 = NM\sigma_\lambda^2 + TM\sigma_\mu^2 + NT\sigma_\eta^2 + \sigma_\nu^2$. Also

$$V_1 = I_T \otimes I_N \otimes I_M - I_T \otimes \bar{J}_N \otimes \bar{J}_M - \bar{J}_T \otimes I_N \otimes \bar{J}_M \\ -\bar{J}_T \otimes \bar{J}_N \otimes I_M + 2\bar{J}_T \otimes \bar{J}_N \otimes \bar{J}_M$$
$$V_2 = E_T \otimes \bar{J}_N \otimes \bar{J}_M \quad \text{where} \quad E_T = I_T - \bar{J}_T$$
$$V_3 = \bar{J}_T \otimes E_N \otimes \bar{J}_M$$
$$V_4 = \bar{J}_T \otimes \bar{J}_N \otimes E_M \quad \text{and} \, V_5 = \bar{J}_T \otimes \bar{J}_N \otimes \bar{J}_M$$

all symmetric and idempotent and sum to the identity matrix.
(c) Conclude that $\Omega^{-1} = \sum_{j=1}^5 (1/\xi_j) V_j$ and $\sigma_\nu \Omega^{-1/2} = \sum_{j=1}^5 (\sigma_\nu/\sqrt{\xi_j}) V_j$ with the typical element of $\sigma_\nu \Omega^{-1/2} y$ being

$$y_{tiq} - \theta_1 \bar{y}_{t..} - \theta_2 \bar{y}_{.i.} - \theta_3 \bar{y}_{..q} - \theta_4 \bar{y}_{...}$$

where the dot indicates a sum over that index and a bar means an average. Here, $\theta_j = 1 - \sigma_\nu/\sqrt{\xi_{j+1}}$ for $j = 1, 2, 3$ while $\theta_4 = \theta_1 + \theta_2 + \theta_3 - 1 + (\sigma_\nu/\sqrt{\xi_5})$.

(d) Show that the BQU estimator of ξ_j is given by $u'V_ju/\text{tr}(V_j)$ for $j = 1, 2, 3, 4$. Show that BQU estimators of σ_ν^2, σ_μ^2, σ_η^2, and σ_λ^2 can be obtained using the one-to-one correspondence between the ξ_j and σ^2.
This problem is based on Baltagi (1987). For a generalization of this four-component model as well as an alternative class of decompositions of the variance–covariance matrix, see Wansbeek and Kapteyn (1982a), and also Davis (2002) who gives an elegant generalization to the multi-way unbalanced error component model; see Chap. 9.

3.16 *A mixed-error component model.* This is based on problem 95.1.4 in *Econometric Theory* by Baltagi and Krämer (1995). Consider the panel data regression equation with a two-way mixed error component model described by (3.1) where the individual specific effects are assumed to be random, with $\mu_i \sim (0, \sigma_\mu^2)$ and $\nu_{it} \sim (0, \sigma_\nu^2)$ independent of each other and among themselves. The time-specific effects, i.e., the λ_t's are assumed to be fixed parameters to be estimated. In vector form, this can be written as

$$y = X\beta + Z_\lambda\lambda + w \tag{1}$$

where $Z_\lambda = \iota_N \otimes I_T$, and

$$w = Z_\mu\mu + \nu \tag{2}$$

with $Z_\mu = I_N \otimes \iota_T$. By applying the Frisch–Waugh–Lovell (FWL) theorem, one gets

$$Q_\lambda y = Q_\lambda X\beta + Q_\lambda w \tag{3}$$

where $Q_\lambda = E_N \otimes I_T$ with $E_N = I_N - \bar{J}_N$ and $\bar{J}_N = \iota_N\iota_N'/N$. This is the familiar Within time-effects transformation, with the typical element of $Q_\lambda y$ being $y_{it} - \bar{y}_{.t}$ and $\bar{y}_{.t} = \sum_{i=1}^N y_{it}/N$. Let $\Omega = E(ww')$, this is the familiar one-way error component variance–covariance matrix given in (2.17).

(a) Show that GLS estimator of β obtained from (1) by premultiplying by $\Omega^{-1/2}$ first and then applying the FWL theorem yields the same estimator as GLS on (3) using the generalized inverse of $Q_\lambda\Omega Q_\lambda$. This is a special case of a more general result proved by Fiebig, Bartels and Krämer (1996).
(b) Show that pseudo-GLS on (3) using Ω rather than $Q_\lambda\Omega Q_\lambda$ for the variance of the disturbances yields the same estimator of β as found in part (a). In general, pseudo-GLS may not be the same as GLS, but Fiebig, Bartels and Krämer (1996) provided a necessary and sufficient condition for this equivalence that is easy to check in this case. In fact, this amounts to checking whether $X'Q_\lambda\Omega^{-1}Z_\lambda = 0$. See solution 95.1.4 in *Econometric Theory* by Xiong (1996).

For computational purposes, these results imply that one can perform the Within time-effects transformation to wipe out the matrix of time dummies and then do the usual Fuller and Battese (1974) transformation without worrying about the loss in efficiency of not using the proper variance–covariance matrix of the transformed disturbances.

3.17 *Openness, country size, and government size*. Ram (2009) questions the body of influential research that suggests that there is a negative association between *country size* (as measured by logarithm of population) and *government size* (as proxied by government consumption as percent of GDP). Also, between *country size* and *trade openness* (as measured by the sum of imports and exports as percent of GDP). Ram uses a 41-year panel data (1960–2000) for over 150 countries from the Penn World Tables 6.1. The pooled OLS results support the foregoing scenario, whereas the *two-way* fixed-effects results find little evidence of a negative association of *country size* with either *government size* or *trade openness*. You are asked to replicate Tables 1, 2, and 3 of Ram (2009; pp. 215–216). Also, test for the significance of time and country dummies.

3.18 *Air pollution levels and left-wing party strength*. Neumayer (2003a) investigates the effect of *left-wing* party strength on *air pollution* levels using a panel of 21 OECD countries observed over the period 1980–1999. Neumayer reports the *two-way* fixed-effects estimates of several measures of air pollution levels (like carbon dioxide emissions) regressed on measures of scale: (GDP and vehicle use); measures of composition (share of manufacturing and fossil fuels); and a measure of efficiency, as well as three measures of left-wing party strength, and one indicator of corporatism. The results find that parliamentary green/left-libertarian party strength is associated with lower pollution levels. The data set and Stata code are provided on the author's university webpage. (http://www2.lse.ac.uk/geographyAndEnvironment/whosWho/profiles/neumayer/replicationdatasets2.aspx).

(a) Replicate Table 2 of Neumayer (2003a; p. 213) which reports the fixed and random effects estimates for carbon dioxide emissions for the period 1980–1999 as well as 1990–1999. Report the tests for country and time effects.
(b) Replicate Tables 3, 4, 5, and 6 of Neumayer (2003a; pp. 214–215) which report the fixed and random effects estimates for sulfur dioxide, nitrogen dioxide, carbon monoxide, and volatile organic compound emissions, respectively.

3.19 *Political governance, economic policies, and homicide rates*. Neumayer (2003b) provides empirical evidence that good political governance and good economic policies can lower homicide rates. This is based on two-way fixed-effects estimates using a panel of homicide data from up to 117 countries over the period 1980–97. The results suggest that economic growth, higher income levels, respect for human rights, and the abolition of the death penalty are all associated with lower homicide rates. The same is true for democracy but only at high levels of democracy. The data set and Stata code are provided on the author's university webpage. (http://www2.lse.ac.uk/geographyAndEnvironment/whosWho/profiles/neumayer/replicationdatasets2.aspx).

(a) Replicate Table I of Neumayer (2003b; p. 629) which reports the descriptive statistics.
(b) Replicate Table II of Neumayer (2003b; p. 630) which reports the correlation coefficients matrix of the variables after the Within transformation.
(c) Replicate columns 1, 2, 4, and 5 of Table III of Neumayer (2003b; p. 632) which report the two-way fixed effects estimates using different homicide measures and regressors.

References

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13.

Baltagi, B.H. 1981. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error components model. *Journal of Econometrics* 17: 21–49.

Baltagi, B.H. 1987. On estimating from a more general time-series cum cross-section data structure. *The American Economist* 31: 69–71.

Baltagi, B.H. 1988. Prediction with a two-way error component regression model, Problem 88.1.1. *Econometric Theory 4*: 171.

Baltagi, B.H. 1989. Applications of a necessary and sufficient condition for OLS to be BLUE. *Statistics and Probability Letters* 8: 457–461.

Baltagi, B.H. 1993. Nested effects, Problem 93.4.2. *Econometric Theory* 9: 687–688.

Baltagi, B.H. 1995. *Econometric analysis of panel data*. Chichester: Wiley.

Baltagi, B.H. 1998. Within two way is equivalent to two withins one way, Problem 98.5.2. *Econometric Theory 14*: 687.

Baltagi, B.H. 2009. *A companion to econometric analysis of panel data*. Chichester: Wiley.

Baltagi, B.H., and J.M. Griffin. 1983. Gasoline demand in the OECD: An application of pooling and testing procedures. *European Economic Review* 22: 117–137.

Baltagi, B.H., and W. Krämer. 1994. Consistency, asymptotic unbiasedness and bounds on the bias of s^2 in the linear regression model with error component disturbances. *Statistical Papers* 35: 323–328.

Baltagi, B.H., and W. Krämer. 1995. A mixed error component model, Problem 95.1.4. *Econometric Theory* 11: 192–193.

Baltagi, B.H., and Q. Li. 1991. Variance component estimation under misspecification, Problem 91.3.3. *Econometric Theory* 7: 418–419.

Baltagi, B.H., and Q. Li. 1992a. A monotonic property for iterative GLS in the two-way random effects model. *Journal of Econometrics* 53: 45–51.

Baltagi, B.H., and Q. Li. 1992b. Variance component estimation under misspecification, Solution 91.3.3. *Econometric Theory* 8: 430–433.

Baltagi, B.H., and N. Pinnoi. 1995. Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics* 20: 351–359.

Breusch, T.S. 1987. Maximum likelihood estimation of random effects models. *Journal of Econometrics* 36: 383–389.

Davidson, R., and J.G. MacKinnon. 1993. *Estimation and inference in econometrics*. New York: Oxford University Press.

Davis, P. 2002. Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics* 106: 67–95.

Fiebig, D., R. Bartels, and W. Krämer. 1996. The Frisch-Waugh theorem and generalized least squares. *Econometric Reviews* 15: 431–443.

Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.

Fuller, W.A., and G.E. Battese. 1974. Estimation of linear models with cross-error structure. *Journal of Econometrics* 2: 67–78.

Ghosh, S.K. 1976. Estimating from a more general time-series cum cross-section data structure. *The American Economist* 20: 15–21.

Graybill, F.A. 1961. *An introduction to linear statistical models*. New York: McGraw-Hill.

Grunfeld, Y. 1958. The determinants of corporate investment. Unpublished Ph.D. dissertation, University of Chicago, Chicago.

Holtz-Eakin, D., and T.M. Selden. 1995. Stocking the fires? CO_2 emissions and economic growth. *Journal of Public Economics* 57: 85–101.

Kiviet, J.F., and W. Krämer. 1992. Bias of s^2 in the linear regression model with correlated errors. *Empirical Economics* 16: 375–377.

Koning, R.H. 1989. Prediction with a two-way error component regression model, Solution 88.1.1. *Econometric Theory 5*: 175.

Li, D. 1999. Within two way is equivalent to two withins one-way, Solution 98.5.2. *Econometric Theory* 15: 781–783.

Maddala, G.S. 1971. The use of variance components models in pooling cross section and time series data. *Econometrica* 39: 341–358.

Maddala, G.S., and T.D. Mount. 1973. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association* 68: 324–328.

Mátyás, L. (ed.). 2017. *The econometrics of multi-dimensional panels, Advanced studies in theoretical and applied econometrics*, vol. 50. Berlin: Springer.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Munnell, A. 1990. How does public infrastructure affect regional economic performance? *New England Economic Review* 3–22.

Nerlove, M. 1971a. Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. *Econometrica* 39: 359–382.

Nerlove, M. 1971b. A note on error components models. *Econometrica* 39: 383–396.

Neumayer, E. 2003a. Are left-wing party strength and corporatism good for the environment? Evidence from panel analysis of air pollution in OECD countries. *Ecological Economics* 45: 203–220.

Neumayer, E. 2003b. Good policy can lower violent crime: Evidence from a cross-national panel of homicide rates, 1980–97. *Journal of Peace Research* 40: 619–640.

Ram, R. 2009. Openness, country size, and government size: Additional evidence from a large cross-country panel. *Journal of Public Economics* 93: 213–218.

Rao, C.R. 1971. Estimation of variance and covariance components—MINQUE theory. *Journal of Multivariate Analysis* 1: 257–275.

Schmalensee, R., T.M. Stoker, and R.A. Judson. 1998. World carbon dioxide emissions: 1950–2050. *Review of Economics and Statistics* 80: 15–27.

Swamy, P.A.V.B., and S.S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40: 261–275.

Taylor, W.E. 1980. Small sample considerations in estimation from panel data. *Journal of Econometrics* 13: 203–223.

Wallace, T.D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37: 55–72.

Wansbeek, T.J., and A. Kapteyn. 1982a. A class of decompositions of the variance-covariance matrix of a generalized error components model. *Econometrica* 50: 713–724.

Wansbeek, T.J., and A. Kapteyn. 1982b. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.

Xiong, W. 1995. Nested effects, Solution 93.4.2. *Econometric Theory* 11: 658–659.

Xiong, W. 1996. A mixed-error component model, Solution 95.1.4. *Econometric Theory* 12: 401–402.

4 Test of Hypotheses with Panel Data

4.1 Tests for Poolability

The question of whether to pool the data or not naturally arises with panel data. The restricted model is the pooled model given by (2.3) representing a behavioral equation with the same parameters over time and across regions. The unrestricted model, however, is the same behavioral equation but with different parameters across time or across regions. For example, Cornwell and Rupert (1988) estimate a mincer wage equation based on a panel of 595 individuals drawn from the Panel Study of Income Dynamics (PSID) and observed over the period 1976–82. In this case, the question of whether to pool or not boils down to the question of whether the parameters of this mincer wage equation vary from one year to the other over the seven years of available data. One can have a behavioral equation whose parameters may vary across countries. For example, Baltagi and Griffin (1983) considered panel data on motor gasoline demand for 18 OECD countries. In this case, one is interested in testing whether the behavioral relationship predicting demand is the same across the 18 OECD countries, i.e., the parameters of the prediction equation do not vary from one country to the other.

These are but two examples of many economic applications where time-series and cross-section data may be pooled. Generally, most economic applications tend to be of the first type, i.e., with a large number of observations on individuals, firms, economic sectors, regions, industries, and countries but only over a few time periods. In what follows, we study the tests for the poolability of the data for the case of pooling across regions keeping in mind that the other case of pooling over time can be obtained in a similar fashion.

For the unrestricted model, we have a regression equation for each region given by

$$y_i = Z_i \delta_i + u_i \quad \text{for} \quad i = 1, 2, \ldots, N \tag{4.1}$$

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_4

where $y_i' = (y_{i1}, \ldots, y_{iT})$, $Z_i = [\iota_T, X_i]$, and X_i is $(T \times K)$. δ_i' is $1 \times (K+1)$ and u_i is $T \times 1$. The important thing to notice is that δ_i is different for every regional equation. We want to test the hypothesis $H_0 : \delta_i = \delta$ for all i, so that under H_0 we can write the restricted model given in (4.1) as

$$y = Z\delta + u \tag{4.2}$$

where $Z' = (Z_1', Z_2', \ldots, Z_N')$ and $u' = (u_1', u_2', \ldots, u_N')$. The unrestricted model can also be written as

$$y = \begin{pmatrix} Z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & Z_N \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{pmatrix} + u = Z^*\delta^* + u \tag{4.3}$$

where $\delta^{*\prime} = (\delta_1', \delta_2', \ldots, \delta_N')$ and $Z = Z^* I^*$ with $I^* = (\iota_N \otimes I_{K'})$, an $NK' \times K'$ matrix, with $K' = K + 1$. Hence the variables in Z are all linear combinations of the variables in Z^*.

4.1.1 Test for Poolability Under $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I_{NT}})$

Assumption 4.1 $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I_{NT}})$.

Under Assumption 4.1, the minimum variance unbiased (MVU) estimator for δ in Eq. (4.2) is

$$\widehat{\delta}_{OLS} = \widehat{\delta}_{mle} = (Z'Z)^{-1}Z'y \tag{4.4}$$

and therefore

$$y = Z\widehat{\delta}_{OLS} + e \tag{4.5}$$

implying that $e = (I_{NT} - Z(Z'Z)^{-1}Z')y = My = M(Z\delta + u) = Mu$ since $MZ = 0$. Similarly, under Assumption 4.1, the MVU for δ_i is given by

$$\widehat{\delta}_{i,OLS} = \widehat{\delta}_{i,mle} = (Z_i'Z_i)^{-1}Z_i'y_i \tag{4.6}$$

and therefore

$$y_i = Z_i\widehat{\delta}_{i,OLS} + e_i \tag{4.7}$$

implying that $e_i = (I_T - Z_i(Z_i'Z_i)^{-1}Z_i')y_i = M_i y_i = M_i(Z_i\delta_i + u_i) = M_i u_i$ since $M_i Z_i = 0$, and this is true for $i = 1, 2, \ldots, N$. Also, let

$$M^* = I_{NT} - Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = \begin{pmatrix} M_1 & 0 & \ldots & 0 \\ 0 & M_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & M_N \end{pmatrix}$$

One can easily deduce that $y = Z^*\widehat{\delta}^* + e^*$ with $e^* = M^*y = M^*u$ and $\widehat{\delta}^* = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y$. Note that both M and M^* are symmetric and idempotent with $MM^* = M^*$. This easily follows since

$$\begin{aligned} Z(Z'Z)^{-1}Z'Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} &= Z(Z'Z)^{-1}I^{*\prime}Z^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} \\ &= Z(Z'Z)^{-1}Z' \end{aligned}$$

This uses the fact that $Z = Z^* I^*$. Under Assumption 4.1, $e'e - e^{*\prime}e^* = u'(M - M^*)u$ and $e^{*\prime}e^* = u'M^*u$ are independent since $(M - M^*)M^* = 0$. Also, both quadratic forms when divided by σ^2 are distributed as χ^2 since $(M - M^*)$ and M^* are idempotent. Dividing these quadratic forms by their respective degrees of freedom, and taking their ratio leads to the following test statistic:[1]

$$F_{obs} = \frac{(e'e - e^{*\prime}e^*)/(\mathrm{tr}(M) - \mathrm{tr}(M^*))}{e^{*\prime}e^*/\mathrm{tr}(M^*)}$$

$$F_{obs} = \frac{(e'e - e_1'e_1 - e_2'e_2 - \cdots - e_N'e_N)/(N-1)K'}{(e_1'e_1 + e_2'e_2 + \cdots + e_N'e_N)/N(T-K')} \tag{4.8}$$

Under H_0, F_{obs} is distributed as an $F((N-1)K', N(T-K'))$. Hence the critical region for this test is defined as

$$\{F_{obs} > F((N-1)K', NT - NK'; \alpha_0)\}$$

where α_0 denotes the level of significance of the test. This is exactly the Chow test presented by Chow (1960) extended to the case of N linear regressions. Therefore, if an economist has a reason to believe that Assumption 4.1 is true, and wants to pool his data across regions, then it is recommended that he or she test for the poolability of the data using the Chow test given in (4.8). However, for the variance component model, $u \sim (0, \Omega)$ and not $(0, \sigma^2 I_{NT})$. Therefore, even if we assume normality on the disturbances, two questions remain: (1) Is the Chow test still the right test to perform when $u \sim N(0, \Omega)$? and (2) does the Chow statistic still have an F-distribution when $u \sim N(0, \Omega)$? The answer to the first question is no, the Chow test given in (4.8) is not the right test to perform. However, as will be shown later, a generalized Chow test will be the right test to perform. As for the second question, it is still relevant to ask because it highlights the problem of economists using the Chow test assuming erroneously that u is $N(0, \sigma^2 I_{NT})$ when in fact it is not.

Having posed the two questions above, we can proceed along two lines: the first is to find the approximate distribution of the Chow statistic (4.8) in case $u \sim N(0, \Omega)$ and therefore show how erroneous it is to use the Chow test in this case (this is not pursued in this book). The second route, and the more fruitful, is to derive the right test to perform for pooling the data in case $u \sim N(0, \Omega)$. This is done in the next subsection.

4.1.2 Test for Poolability Under the General Assumption $u \sim N(0, \Omega)$

Assumption 4.2 $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega})$.

In case Ω is known up to a scalar factor, the test statistic employed for the poolability of the data would be simple to derive. All we need to do is transform our model (under both the null and alternative hypotheses) such that the transformed disturbances have a variance of $\sigma^2 I_{NT}$, then apply the Chow test on the transformed model. The later step is legitimate because the transformed disturbances have homoskedastic variances and the analysis of the previous section applies in full. Given $\Omega = \sigma^2\Sigma$, we premultiply the restricted model given in (4.2) by $\Sigma^{-1/2}$ and we call $\Sigma^{-1/2}y = \dot{y}$, $\Sigma^{-1/2}Z = \dot{Z}$, and $\Sigma^{-1/2}u = \dot{u}$. Hence

$$\dot{y} = \dot{Z}\delta + \dot{u} \tag{4.9}$$

with $E(\dot{u}\dot{u}') = \Sigma^{-1/2}E(uu')\Sigma^{-1/2'} = \sigma^2 I_{NT}$. Similarly, we premultiply the unrestricted model given in (4.3) by $\Sigma^{-1/2}$, and we call $\Sigma^{-1/2}Z^* = \dot{Z}^*$. Therefore

$$\dot{y} = \dot{Z}^*\delta^* + \dot{u} \tag{4.10}$$

with $E(\dot{u}\dot{u}') = \sigma^2 I_{NT}$.

At this stage, we can test H_0; $\delta_i = \delta$ for every $i = 1, 2, \ldots, N$, simply by using the Chow statistic, only now on the transformed models (4.9) and (4.10) since they satisfy Assumption 4.1 of homoskedasticity of the normal disturbances. Note that $\dot{Z} = \dot{Z}^* I^*$ which is simply obtained from $Z = Z^* I^*$ by premultiplying by $\Sigma^{-1/2}$. Defining $\dot{M} = I_{NT} - \dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'$ and $\dot{M}^* = I_{NT} - \dot{Z}^*(\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}$, it is easy to show that $\dot{M}$ and $\dot{M}^*$ are both symmetric and idempotent such that $\dot{M}\dot{M}^* = \dot{M}^*$. Once again the conditions for lemma 2.2 of Fisher (1970) are satisfied, and the test statistic

$$\dot{F}_{obs} = \frac{(\dot{e}'\dot{e} - \dot{e}^{*\prime}\dot{e}^*)/(\text{tr}(\dot{M}) - \text{tr}(\dot{M}^*))}{\dot{e}^{*\prime}\dot{e}^*/\text{tr}(\dot{M}^*)} \sim F((N-1)K', N(T-K')) \tag{4.11}$$

where $\dot{e} = \dot{y} - \dot{Z}\dot{\delta}_{OLS}$ and $\widehat{\dot{\delta}}_{OLS} = (\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}$ implying that $\dot{e} = \dot{M}\dot{y} = \dot{M}\dot{u}$. Similarly, $\dot{e}^* = \dot{y} - \dot{Z}^*\widehat{\dot{\delta}}^*_{OLS}$ and $\widehat{\dot{\delta}}^*_{OLS} = (\dot{Z}^{*\prime}\dot{Z}^*)^{-1}\dot{Z}^{*\prime}\dot{y}$ implying that $\dot{e}^* = \dot{M}^*\dot{y} = \dot{M}^*\dot{u}$. Using the fact that $\dot{M}$ and $\dot{M}^*$ are symmetric and idempotent, we can rewrite (4.11) as

$$\begin{aligned}\dot{F}_{obs} &= \frac{(\dot{y}'\dot{M}\dot{y} - \dot{y}'\dot{M}^*\dot{y})/(N-1)K'}{\dot{y}'\dot{M}^*\dot{y}/N(T-K')} \\ &= \frac{(y'\Sigma^{-1/2}\dot{M}\Sigma^{-1/2}y - y'\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2}y)/(N-1)K'}{y'\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2}y/N(T-K')}\end{aligned} \tag{4.12}$$

But

$$\dot{M} = I_{NT} - \Sigma^{-1/2}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1/2'} \tag{4.13}$$

and

$$\dot{M}^* = I_{NT} - \Sigma^{-1/2}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1/2'}$$

so that

$$\Sigma^{-1/2}\dot{M}\Sigma^{-1/2} = \Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}$$

and

$$\Sigma^{-1/2}\dot{M}^*\Sigma^{-1/2} = \Sigma^{-1} - \Sigma^{-1}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1}$$

Hence we can write (4.12) in the form

$$\dot{F}_{obs} = \frac{y'[\Sigma^{-1}(Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime} - Z(Z'\Sigma^{-1}Z)^{-1}Z')\Sigma^{-1}]y/(N-1)K'}{(y'\Sigma^{-1}y - y'\Sigma^{-1}Z^*(Z^{*\prime}\Sigma^{-1}Z^*)^{-1}Z^{*\prime}\Sigma^{-1}y)/N(T-K')} \tag{4.14}$$

and $\dot{F}_{obs}$ has an F-distribution with $((N-1)K', N(T-K'))$ degrees of freedom. It is important to emphasize that (4.14) is operational only when Σ is known. This test is a special application of a general test for linear restrictions described in Roy (1957) and used by Zellner (1962) to test for aggregation bias in a set of seemingly unrelated regressions. In case Σ is unknown, we replace Σ in (4.14) by a consistent estimator (say $\widehat{\Sigma}$) and call the resulting test statistics $\widehat{F}_{obs}$.

One of the main motivations behind pooling a time-series of cross-sections is to widen our database in order to get better and more reliable estimates of the parameters of our model. Using the Chow test, the question of whether "to pool or not" reduced to a test of the validity of the null hypothesis H_0; $\delta_i = \delta$ for all i. Imposing these restrictions (true or false) will reduce the variance of the pooled estimator, but may introduce bias if these restrictions are false. This motivated Toro-Vizcarrondo and Wallace (1968, p. 560) to write, "if one is willing to accept some bias in trade for a reduction in variance, then even if the restriction is not true one might still prefer the restricted estimator." Baltagi (1995b, pp. 54–58) discusses three mean square error (MSE) criteria used in the literature to test whether the pooled estimator restricted by H_0 is better than the unrestricted estimator of δ^*. It is important to emphasize that these MSE criteria do not test whether H_0 is true or false, but help us to choose on "pragmatic grounds" between two sets of estimators of δ^* and hence achieve, in a sense, one of the main motivations behind pooling. A summary table of these MSE criteria is given by Wallace (1972, p. 697). McElroy (1977) extends these MSE criteria to the case where $u \sim N(0, \sigma^2\Sigma)$.

Monte Carlo Evidence

In the Monte Carlo study by Baltagi (1981), the Chow test is performed given that the data are poolable and the model is generated as a two-way error component model. This test gave a high frequency of rejecting the null hypothesis when true. The reason for the poor performance of the Chow test is that it is applicable only under Assumption 4.1 on the disturbances. This is violated under a random effects model with large variance components. For example, in testing the stability of cross-section regressions over time, the high frequency of type I error occurred whenever the variance components due to the time effects are not relatively small. Similarly, in testing the stability of time-series regressions across regions, the high frequency of type I error occurred whenever the variance components due to the cross-section effects are not relatively small.

Under this case of nonspherical disturbances, the proper test to perform is the Roy–Zellner test given by (4.14). Applying this test knowing the true variance components or using the Amemiya (1971) and the Wallace and Hussain (1969) type estimates of the variance components leads to low frequencies of committing a type I error. Therefore, if pooling is contemplated using an error component model, then the Roy–Zellner test should be used rather than the Chow test.

The alternative MSE criteria, developed by Toro-Vizcarrondo and Wallace (1968) and Wallace (1972), were applied to the error component model in order to choose between the pooled and the unpooled estimators. These weaker criteria gave a lower frequency of committing a type I error than the Chow test, but their performance was still poor when compared to the Roy–Zellner test. McElroy's (1977) extension of these weaker MSE criteria to the case of nonspherical disturbances performed well when compared with the Roy–Zellner test, and is recommended.

4.1.3 Examples

Example 1 For the Grunfeld data, Chow's test for poolability across firms as in (4.1) gives an observed F-statistic of 27.75 and is distributed as $F(27, 170)$ under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. The RRSS $= 1755850.48$ is obtained from pooled OLS, and the URSS $= 324728.47$ is obtained from summing the RSS from 10 individual firm OLS regressions, each with 17 degrees of freedom. There are 27 restrictions and the test rejects poolability across firms for all coefficients. One can test for poolability of slopes only, allowing for varying intercepts. The restricted model is the Within regression with firm dummies. The RRSS $= 523478$, while the unrestricted regression is the same as above. The observed F-statistic is 5.78 which is distributed as $F(18, 170)$ under H_0; $\beta_i = \beta$ for $i = 1, \ldots, N$. This again is significant at the 5% level and rejects poolability of the slopes across firms. Note that one could have tested poolability across time. The Chow test gives an observed value of 1.12 which is distributed as $F(57, 140)$. This does not reject poolability across time, but the unrestricted model is based on 20 regressions each with only 7 degrees of freedom. As clear from the numerator degrees of freedom, this F-statistic tests 57 restrictions. The Roy–Zellner test for poolability across firms, allowing for one-way error component disturbances, yields an observed F-value of 4.35 which is distributed as $F(27, 170)$ under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. This still rejects poolability across firms even after allowing for one-way error component disturbances. The Roy–Zellner test for poolability over time, allowing for a one-way error component model, yields an F-value of 2.72 which is distributed as $F(57, 140)$ under H_0; $\delta_t = \delta$ for $t = 1, \ldots, T$.

Example 2 For the gasoline demand data in Baltagi and Griffin (1983), Chow's test for poolability across countries yields an observed F-statistic of 129.38 and is distributed as $F(68, 270)$ under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. This tests the stability of four time-series regression coefficients across 18 countries. The unrestricted SSE is based upon 18 OLS time-series regressions, one for each country. For the stability of the slope coefficients only, H_0; $\beta_i = \beta$, an observed F-value of 27.33 is obtained which is distributed as $F(51, 270)$ under the null. Chow's test for poolability across time yields an F-value of 0.276 which is distributed as $F(72, 266)$ under H_0; $\delta_t = \delta$ for $t = 1, \ldots, T$. This tests the stability of four cross-section regression coefficients across 19 time periods. The unrestricted SSE is based upon 19 OLS cross-section regressions, one for each year. This does not reject poolability across time periods. The Roy–Zellner test for poolability across countries, allowing for a one-way error component model, yields an F-value of 21.64 which is distributed as $F(68, 270)$

under H_0; $\delta_i = \delta$ for $i = 1, \ldots, N$. The Roy–Zellner test for poolability across time yields an F-value of 1.66 which is distributed as $F(72, 266)$ under H_0; $\delta_t = \delta$ for $t = 1, \ldots, T$. This rejects H_0 at the 5% level.

4.2 Tests for Individual and Time Effects

4.2.1 The Breusch–Pagan Test

For the random two-way error component model, Breusch and Pagan (1980) derived a Lagrange multiplier (LM) test to test H_0; $\sigma_\mu^2 = \sigma_\lambda^2 = 0$. The log-likelihood function under normality of the disturbances is given by (3.29) as

$$L(\delta, \theta) = \text{constant} - \frac{1}{2} log \mid \Omega \mid - \frac{1}{2} u'\Omega^{-1}u \tag{4.15}$$

where $\theta' = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$ and Ω is given by (3.10) as

$$\Omega = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\lambda^2 (J_N \otimes I_T) + \sigma_\nu^2 I_{NT} \tag{4.16}$$

The information matrix is block-diagonal between θ and δ. Since H_0 involves only θ, the part of the information matrix due to δ is ignored. In order to reconstruct the Breusch and Pagan (1980) LM statistic, we need the score $D(\widetilde{\theta}) = (\partial L/\partial\theta) \mid_{\widetilde{\theta}_{mle}}$, the first derivative of the likelihood with respect to θ, evaluated at the restricted MLE of θ under H_0, which is denoted by $\widetilde{\theta}_{mle}$. Hartley and Rao (1967) or Hemmerle and Hartley (1973) give a useful general formula to obtain $D(\theta)$:

$$\partial L/\partial\theta_r = \frac{1}{2}\text{tr}[\Omega^{-1}(\partial\Omega/\partial\theta_r)] + \frac{1}{2}[u'\Omega^{-1}(\partial\Omega/\partial\theta_r)\Omega^{-1}u] \tag{4.17}$$

for $r = 1, 2, 3$. Also, from (4.16), $(\partial\Omega/\partial\theta_r) = (I_N \otimes J_T)$ for $r = 1$, $(J_N \otimes I_T)$ for $r = 2$, and I_{NT} for $r = 3$. The restricted MLE of Ω under H_0 is $\widetilde{\Omega} = \widetilde{\sigma}_\nu^2 I_{NT}$ where $\widetilde{\sigma}_\nu^2 = \widetilde{u}'\widetilde{u}/NT$ and $\widetilde{u}$ are the OLS residuals. Using $\text{tr}(I_N \otimes J_T) = \text{tr}(J_N \otimes I_T) = \text{tr}(I_{NT}) = NT$, one gets

$$D(\widetilde{\theta}) = \begin{bmatrix} -\frac{1}{2}\text{tr}[(I_N \otimes J_T)/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'(I_N \otimes J_T)\widetilde{u}/\widetilde{\sigma}_\nu^4] \\ -\frac{1}{2}\text{tr}[(J_N \otimes I_T)/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'(J_N \otimes I_T)\widetilde{u}/\widetilde{\sigma}_\nu^4] \\ -\frac{1}{2}\text{tr}[I_{NT}/\widetilde{\sigma}_\nu^2] + \frac{1}{2}[\widetilde{u}'\widetilde{u}/\widetilde{\sigma}_\nu^4] \end{bmatrix}$$

$$= \frac{-NT}{2\widetilde{\sigma}_\nu^2} \begin{bmatrix} 1 - \dfrac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} \\ 1 - \dfrac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} \\ 0 \end{bmatrix} \tag{4.18}$$

The information matrix for this model is

$$J(\theta) = E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right] = [J_{rs}] \quad \text{for } r, s = 1, 2, 3$$

where

$$J_{rs} = E[-\partial^2 L/\partial\theta_r\partial\theta_s] = \frac{1}{2}tr[\Omega^{-1}(\partial\Omega/\partial\theta_r)\Omega^{-1}(\partial\Omega/\partial\theta_s)] \tag{4.19}$$

(see Harville (1977). Using $\widetilde{\Omega}^{-1} = (1/\widetilde{\sigma}_\nu^2)I_{NT}$ and $\text{tr}[(I_N \otimes J_T)(J_N \otimes I_T)] = \text{tr}(J_{NT}) = NT$, $\text{tr}(I_N \otimes J_T)^2 = NT^2$, and $\text{tr}(J_N \otimes I_T)^2 = N^2T$, one gets

$$\begin{aligned}\widetilde{J} &= \frac{1}{2\widetilde{\sigma}_\nu^4}\begin{bmatrix} \text{tr}(I_N \otimes J_T)^2 & \text{tr}(J_{NT}) & \text{tr}(I_N \otimes J_T) \\ \text{tr}(J_{NT}) & \text{tr}(J_N \otimes I_T)^2 & \text{tr}(J_N \otimes I_T) \\ \text{tr}(I_N \otimes J_T) & \text{tr}(J_N \otimes I_T) & \text{tr}(I_{NT}) \end{bmatrix} \\ &= \frac{NT}{2\widetilde{\sigma}_\nu^4}\begin{bmatrix} T & 1 & 1 \\ 1 & N & 1 \\ 1 & 1 & 1 \end{bmatrix}\end{aligned} \tag{4.20}$$

with

$$\widetilde{J}^{-1} = \frac{2\widetilde{\sigma}_\nu^4}{NT(N-1)(T-1)}\begin{bmatrix} (N-1) & 0 & (1-N) \\ 0 & (T-1) & (1-T) \\ (1-N) & (1-T) & (NT-1) \end{bmatrix} \tag{4.21}$$

Therefore

$$\begin{aligned}LM &= \widetilde{D}'\widetilde{J}^{-1}\widetilde{D} \\ &= \frac{NT}{2(N-1)(T-1)}\left[(N-1)\left[1 - \frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2\right. \\ &\quad \left. +(T-1)\left[1 - \frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2\right] \\ LM &= LM_1 + LM_2\end{aligned} \tag{4.22}$$

where

$$LM_1 = \frac{NT}{2(T-1)}\left[1 - \frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2 \tag{4.23}$$

and

$$LM_2 = \frac{NT}{2(N-1)}\left[1 - \frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}}\right]^2 \tag{4.24}$$

Under H_0, LM is asymptotically distributed as a χ_2^2. This LM test requires only OLS residuals and is easy to compute. This may explain its popularity. In addition, if one wants to test $H_0^a; \sigma_\mu^2 = 0$, following the derivation given above, one gets LM_1 which is asymptotically distributed under H_0^a as χ_1^2. Similarly, if one wants to test $H_0^b; \sigma_\lambda^2 = 0$, by symmetry, one gets LM_2 which is asymptotically distributed as χ_1^2 under H_0^b. This LM test performed well in Monte Carlo studies (see Baltagi (1981)), except for small values of σ_μ^2 and σ_λ^2 close to zero. These are precisely the cases where negative estimates of the variance components are most likely to occur.

4.2.2 Honda, King and Wu, and the Standardized Lagrange Multiplier Tests

One problem with the Breusch–Pagan test is that it assumes that the alternative hypothesis is two-sided when we know that the variance components are nonnegative. This means that the alternative hypotheses should be one-sided. Honda (1985) suggests a *uniformly most powerful* test for H_0^a; $\sigma_\mu^2 = 0$ which is based upon

$$HO \equiv A = \sqrt{\frac{NT}{2(T-1)}}\left[\frac{\widetilde{u}'(I_N \otimes J_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} - 1\right] \overset{H_0^a}{\rightarrow} N(0,1) \qquad (4.25)$$

Note that the square of this $N(0, 1)$ statistic is the Breusch and Pagan (1980) LM_1 test statistic given in (4.23). Honda (1985) finds that this test statistic is robust to nonnormality. Moulton and Randolph (1989) showed that the asymptotic $N(0, 1)$ approximation for this one-sided LM statistic can be poor even in large samples. This occurs when the number of regressors is large or the intra-class correlation of some of the regressors is high. They suggest an alternative standardized Lagrange multiplier (SLM) test whose asymptotic critical values are generally closer to the exact critical values than those of the LM test. This SLM test statistic centers and scales the one-sided LM statistic so that its mean is zero and its variance is one:

$$SLM = \frac{\text{HO} - E(HO)}{\sqrt{\text{var}(HO)}} = \frac{d - E(d)}{\sqrt{\text{var}(d)}} \qquad (4.26)$$

where $d = \widetilde{u}'D\widetilde{u}/\widetilde{u}'\widetilde{u}$ and $D = (I_N \otimes J_T)$. Using the results on moments of quadratic forms in regression residuals (see, e.g., Evans and King (1985)), we get

$$E(d) = tr(D\bar{P}_Z)/p \qquad (4.27)$$

and

$$\text{var}(d) = 2\{p\,\text{tr}(D\bar{P}_Z)^2 - [tr(D\bar{P}_Z)]^2\}/p^2(p+2) \qquad (4.28)$$

where $p = n - (K + 1)$ and $\bar{P}_Z = I_n - Z(Z'Z)^{-1}Z'$. Under the null hypothesis H_0^a, SLM has an asymptotic $N(0, 1)$ distribution. King and Wu (1997) suggest a locally mean most powerful (LMMP) one-sided test, which for H_0^a coincides with Honda's (1985) uniformly most powerful test (see Baltagi, Chang and Li (1992)).

Similarly, for H_0^b; $\sigma_\lambda^2 = 0$, the one-sided Honda-type LM test statistic is

$$B = \sqrt{\frac{NT}{2(N-1)}}\left[\frac{\widetilde{u}'(J_N \otimes I_T)\widetilde{u}}{\widetilde{u}'\widetilde{u}} - 1\right] \qquad (4.29)$$

which is asymptotically distributed as $N(0, 1)$. Note that the square of this statistic is the corresponding two-sided LM test given by LM_2 in (4.24). This can be standardized as in (4.26) with $D = (J_N \otimes I_T)$. Also, the King and Wu (1997) LMMP test for H_0^b coincides with Honda's uniformly most powerful test given in (4.29).

For H_0^c; $\sigma_\mu^2 = \sigma_\lambda^2 = 0$, the two-sided LM test, given by Breusch and Pagan (1980), is $A^2 + B^2 \sim \chi^2(2)$. Honda (1985) does not derive a uniformly most powerful one-sided test for H_0^c, but he suggests a "handy" one-sided test given by $(A + B)/\sqrt{2}$

which is distributed as $N(0, 1)$ under H_0^c. Following King and Wu (1997), Baltagi, Chang and Li (1992) derived the LMMP one-sided test for H_0^c . This is given by

$$KW = \frac{\sqrt{T-1}}{\sqrt{N+T-2}}A + \frac{\sqrt{N-1}}{\sqrt{N+T-2}}B \tag{4.30}$$

which is distributed as $N(0, 1)$ under H_0^c. See problem 4.6.

Following the Moulton and Randolph (1989) standardization of the LM test for the one-way error component model, Honda (1991) suggested a standardization of his "handy" one-sided test for the two-way error component model. In fact, for $HO = (A + B)/\sqrt{2}$, the SLM test is given by (4.26) with $d = \widetilde{u}'D\widetilde{u}/\widetilde{u}'\widetilde{u}$, and

$$D = \frac{1}{2}\sqrt{\frac{NT}{(T-1)}}(I_N \otimes J_T) + \frac{1}{2}\sqrt{\frac{NT}{(N-1)}}(J_N \otimes I_T) \tag{4.31}$$

Also, one can similarly standardize the KW test given in (4.30) by subtracting its mean and dividing by its standard deviation, as in (4.26), with $d = \widetilde{u}'D\widetilde{u}/\widetilde{u}'\widetilde{u}$ and

$$D = \frac{\sqrt{NT}}{\sqrt{2}\sqrt{N+T-2}}[(I_N \otimes J_T) + (J_N \otimes I_T)] \tag{4.32}$$

With this new D matrix, $E(d)$ and $\mathrm{var}(d)$ can be computed using (4.27) and (4.28). Under $H_0^c; = \sigma_\mu^2 = \sigma_\lambda^2 = 0$, these SLM statistics are asymptotically $N(0, 1)$ and their asymptotic critical values should be closer to the exact critical values than those of the corresponding unstandardized tests.

4.2.3 Gourieroux, Holly and Monfort Test

Note that A or B can be negative for a specific application, especially when one or both variance components are small and close to zero. Following Gourieroux, Holly and Monfort (1982), hereafter GHM, Baltagi, Chang and Li (1992) proposed the following test for H_0^c:

$$\chi_m^2 = \begin{cases} A^2 + B^2 & \text{if } A > 0, B > 0 \\ A^2 & \text{if } A > 0, B \leqslant 0 \\ B^2 & \text{if } A \leqslant 0, B > 0 \\ 0 & \text{if } A \leqslant 0, B \leqslant 0 \end{cases} \tag{4.33}$$

χ_m^2 denotes the mixed χ^2 distribution. Under the null hypothesis,

$$\chi_m^2 \sim \left(\frac{1}{4}\right)\chi^2(0) + \left(\frac{1}{2}\right)\chi^2(1) + \left(\frac{1}{4}\right)\chi^2(2)$$

where $\chi^2(0)$ equals zero with probability one.[2] The weights $\left(\frac{1}{4}\right)$, $\left(\frac{1}{2}\right)$, and $\left(\frac{1}{4}\right)$ follow from the fact that A and B are asymptotically independent of each other and the results in Gourieroux, Holly and Monfort (1982). This proposed test has the advantage over the Honda and KW tests in that it is immune to the possible negative values of A and B.

4.2.4 Conditional LM Tests

When one uses HO given in (4.25) to test $H_0^a; \sigma_\mu^2 = 0$, one implicitly assumes that the time-specific effects do not exist. This may lead to incorrect decisions especially when the variance of the time effects (assumed to be zero) is large. To overcome this problem, Baltagi, Chang and Li (1992) suggest testing the individual effects conditional on the time-specific effects (i.e., allowing $\sigma_\lambda^2 > 0$). The corresponding LM test for testing $H_0^d; \sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$) is derived in Appendix 2 of Baltagi, Chang and Li (1992) and is given by

$$LM_\mu = \frac{\sqrt{2}\widetilde{\sigma}_2^2\widetilde{\sigma}_\nu^2}{\sqrt{T(T-1)[\widetilde{\sigma}_\nu^4 + (N-1)\widetilde{\sigma}_2^4]}}\widetilde{D}_\mu \tag{4.34}$$

where

$$\widetilde{D}_\mu = \frac{T}{2}\left\{\frac{1}{\widetilde{\sigma}_2^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes \bar{J}_T)\widetilde{u}}{\widetilde{\sigma}_2^2} - 1\right] + \frac{(N-1)}{\widetilde{\sigma}_\nu^2}\left[\frac{\widetilde{u}'(E_N \otimes \bar{J}_T)\widetilde{u}}{(N-1)\widetilde{\sigma}_\nu^2} - 1\right]\right\} \tag{4.35}$$

with $\widetilde{\sigma}_2^2 = \widetilde{u}'(\bar{J}_N \otimes I_T)\widetilde{u}/T$ and $\widetilde{\sigma}_\nu^2 = \widetilde{u}'(E_N \otimes I_T)\widetilde{u}/T(N-1)$. LM_μ is asymptotically distributed as $N(0,1)$ under H_0^d. The estimated disturbances $\widetilde{u}$ denote the one-way GLS residuals using the maximum likelihood estimates $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_2^2$. One can easily check that if $\widetilde{\sigma}_\lambda^2 \to 0$, then $\widetilde{\sigma}_2^2 \to \widetilde{\sigma}_\nu^2$ and LM_μ given in (4.34) tends to the one-sided Honda test given in (4.25).

Similarly, the alternative LM test statistic for testing $H_0^e; \sigma_\lambda^2 = 0$ (allowing $\sigma_\mu^2 > 0$) can be obtained as follows:

$$LM_\lambda = \frac{\sqrt{2}\widetilde{\sigma}_1^2\widetilde{\sigma}_\nu^2}{\sqrt{N(N-1)[\widetilde{\sigma}_\nu^4 + (T-1)\widetilde{\sigma}_1^4]}}\widetilde{D}_\lambda \tag{4.36}$$

where

$$\widetilde{D}_\lambda = \frac{N}{2}\left\{\frac{1}{\widetilde{\sigma}_1^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes \bar{J}_T)\widetilde{u}}{\widetilde{\sigma}_1^2} - 1\right] + \frac{(T-1)}{\widetilde{\sigma}_\nu^2}\left[\frac{\widetilde{u}'(\bar{J}_N \otimes E_T)\widetilde{u}}{(T-1)\widetilde{\sigma}_\nu^2} - 1\right]\right\} \tag{4.37}$$

with $\widetilde{\sigma}_1^2 = \widetilde{u}'(I_N \otimes \bar{J}_T)\widetilde{u}/N$ and $\widetilde{\sigma}_\nu^2 = \widetilde{u}'(I_N \otimes E_T)\widetilde{u}/N(T-1)$. The test statistic LM_λ is asymptotically distributed as $N(0,1)$ under H_0^e.

4.2.5 ANOVA F and the Likelihood Ratio Tests

Moulton and Randolph (1989) found that the ANOVA F-test which tests the significance of the fixed effects performs well for the one-way error component model. The ANOVA F-test statistics have the following familiar general form:

$$F = \frac{y'MD(D'MD)^{-}D'My/(p-r)}{y'Gy/[NT - (\widetilde{k} + p - r)]} \tag{4.38}$$

Under the null hypothesis, this statistic has a central F-distribution with $p - r$ and $NT - (\widetilde{k} + p - r)$ degrees of freedom. For $H_0^a; \sigma_\mu^2 = 0, D = I_N \otimes \iota_T, M =$

$\bar{P}_Z, \widetilde{k} = K', p = N, r = K' + N -$ rank(Z, D), and $G = \bar{P}_{(Z,D)}$ where $\bar{P}_Z = I - P_Z$ and $P_Z = Z(Z'Z)^{-1}Z'$. For details regarding other hypotheses, see Baltagi, Chang and Li (1992).

The one-sided likelihood ratio (LR) tests all have the following form:

$$LR = -2 \log \frac{l(\text{res})}{l(\text{unres})} \tag{4.39}$$

where l(res) denotes the restricted maximum likelihood value (under the null hypothesis), while l(unres) denotes the unrestricted maximum likelihood value. The LR tests require MLE estimators of the one-way and the two-way models and are comparatively more expensive than their LM counterparts. Under the null hypotheses considered, the LR test statistics have the same asymptotic distributions as their LM counterparts (see Gourieroux, Holly and Monfort (1982)). More specifically, for H_0^a, H_0^b, H_0^d, and H_0^e, $LR \sim (\frac{1}{2})\chi^2(0) + (\frac{1}{2})\chi^2(1)$ and for H_0^c, $LR \sim (\frac{1}{4})\chi^2(0) + (\frac{1}{2})\chi^2(1) + (\frac{1}{4})\chi^2(2)$.

4.2.6 Monte Carlo Results

Baltagi, Chang and Li (1992) compared the performance of the above tests using Monte Carlo experiments on the two-way error component model described in Baltagi (1981). Each experiment involved 1000 replications. For each replication, the following test statistics were computed: BP, Honda, KW, SLM, LR, GHM, and the F-test statistics. The results can be summarized as follows: when $H_0^a; \sigma_\mu^2 = 0$ is true but σ_λ^2 is large, all the usual tests for H_0^a perform badly since they ignore the fact that $\sigma_\lambda^2 > 0$. In fact, the two-sided BP test performs the worst, over-rejecting the null, while HO, SLM, LR, and F underestimate the nominal size. As σ_μ^2 gets large, all the tests perform well in rejecting the null hypothesis H_0^a. But, for small $\sigma_\mu^2 > 0$, the power of all the tests considered deteriorates as σ_λ^2 increases.

For testing $H_0^d; \sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$), LM$_\mu$, LR, and F perform well with their estimated size not significantly different from their nominal size. Also, for large σ_μ^2 all these tests have high power rejecting the null hypothesis in 98–100% of the cases. The results also suggest that overspecifying the model, i.e., assuming the model is two-way ($\sigma_\lambda^2 > 0$) when in fact it is one-way ($\sigma_\lambda^2 = 0$) does not seem to hurt the power of these tests. Finally, the power of all tests improves as σ_λ^2 increases. This is in sharp contrast to the performance of the tests that ignore the fact that $\sigma_\lambda^2 > 0$. The Monte Carlo results strongly support the fact that one should not ignore the possibility that $\sigma_\lambda^2 > 0$ when testing $\sigma_\mu^2 = 0$. In fact, it may be better to overspecify the model rather than underspecify it in testing the variance components.

For the joint test $H_0^c; \sigma_\mu^2 = \sigma_\lambda^2 = 0$, the BP, HO, KW, and LR significantly underestimate the nominal size, while GHM and the F-test have estimated sizes that are not significantly different from the nominal size. Negative values of A and B make the estimated size for HO and KW underestimate the nominal size. For these cases, the GHM test is immune to negative values of A and B, and performs well in the

Monte Carlo experiments. Finally, the ANOVA F-tests perform reasonably well when compared to the LR and LM tests, for both the one-way and two-way models and are recommended. This confirms similar results on the F-statistic by Moulton and Randolph (1989).

4.2.7 An Illustrative Example

The Monte Carlo results show that the test statistics A and/or B take on large negative values quite often under some designs. A natural question is whether a large negative A and/or B is possible for real data. In this subsection, we apply the tests considered above to the Grunfeld (1958) investment equation. Table 4.1 gives the observed test statistics. The null hypotheses H_0^c; $\sigma_\mu^2 = \sigma_\lambda^2 = 0$, as well as H_0^a; $\sigma_\mu^2 = 0$ and H_0^d; $\sigma_\mu^2 = 0$ (allowing $\sigma_\lambda^2 > 0$), are rejected by all tests considered. Clearly, all the tests strongly suggest that there are individual specific effects. However, for testing time-specific effects, except for the two-sided LM (BP) test which rejects H_0^b; $\sigma_\lambda^2 = 0$, all the tests suggest that there are no time-specific effects for this data. The conflict occurs because B takes on a large negative value (-2.540) for this data set. This means that the two-sided LM test is $B^2 = 6.454$ which is larger than the χ_1^2 critical value (3.841) whereas, the one-sided LM, SLM, LR, and F-tests for this hypothesis do not reject H_0^b. In fact, the LM_λ test proposed by Baltagi, Chang and Li (1992) for testing H_0^e; $\sigma_\lambda^2 = 0$ (allowing $\sigma_\mu^2 > 0$) as well as the LR and F-tests for this hypothesis do not reject H_0^e. These data clearly support the use of the one-sided test in empirical applications. Stata reports the LM (BP) test for H_0^a; $\sigma_\mu^2 = 0$ using the command (*xttest0*) after running the random effects specification. This computes the A^2 term in (4.23) which is 798.16 for the Grunfeld data as reported in Table 4.1. Using EViews, one can replicate the first 5 rows of Table 4.1 after running OLS. The results are given in Table 4.2.

Table 4.1 Test results for the grunfeld example

Null Hypothesis	H_0^a	H_0^b	H_0^c	H_0^d	H_0^e
BP	798.162 (3.841)	6.454 (3.841)	804.615 (5.991)	–	–
HO	28.252 (1.645)	– 2.540 (1.645)	18.181 (1.645)	–	–
KW	28.252 (1.645)	– 2.540 (1.645)	21.832 (1.645)	–	–
SLM	32.661 (1.645)	– 2.433 (1.645)	–	–	–
GHM	–	–	798.162 (4.231)	–	–
F	49.177 (1.930)	0.235 (1.645)	17.403 (1.543)	52.362 (1.648)	1.403 (1.935)
LR	193.091 (2.706)	0 (2.706)	193.108 (4.231)	193.108 (2.706)	0.017 (2.706)
LM_μ	–	–	–	28.252 (2.706)	–
LM_λ	–	–	–	–	0.110 (2.706)

Numbers in parentheses are asymptotic critical values at the 5% level
Source Baltagi, Chang and Li (1992). Reproduced by permission of Elsevier Science Publishers B.V. (North Holland)

Table 4.2 Grunfeld's data: Lagrangian multiplier tests

Lagrange Multiplier Tests for Random Effects
Null hypotheses: No effects
Alternative hypotheses: Two-sided (Breusch-Pagan) and one-sided
(all others) alternatives

	Test Hypothesis		
	Cross-section	Time	Both
Breusch-Pagan	798.1615 (0.0000)	6.453882 (0.0111)	804.6154 (0.0000)
Honda	28.25175 (0.0000)	-2.540449 (0.9945)	18.18064 (0.0000)
King-Wu	28.25175 (0.0000)	-2.540449 (0.9945)	21.83221 (0.0000)
Standardized Honda	32.66605 (0.0000)	-2.432565 (0.9925)	16.29814 (0.0000)
Standardized King-Wu	32.66605 (0.0000)	-2.432565 (0.9925)	20.96591 (0.0000)
Gourieroux, et al.*	--	--	798.1615 (0.0000)

Stata reports the LR test for H_0^a at the bottom of the MLE results using (*xtreg, mle*). This replicates the observed LR test statistic of 193.04 in Table 4.1. The Stata output is not reproduced here but one can refer to the Stata results in Table 2.10 where we reported the MLE for the public capital productivity data. The bottom of Table 2.10 reports the observed LR test statistic of 1149.84. This shows that the random *state* effects are significant and their variance is not 0. Also note that the fixed effects Stata output (*xtreg, fe*) reports the F-test for the significance of the fixed individual effects. For the Grunfeld data, this replicates the F (9,188) value of 49.18 which is reported in Table 4.1. The Stata output is not reproduced here, but one can refer to the Stata results in Table 2.8 where we reported the fixed effects estimates for the public capital productivity data. The bottom of Table 2.8 reports the observed F (47,764) value of 75.82. This shows that the fixed state effects are significant.

EViews computes the F-tests for redundant fixed effects after performing one-way or two-way fixed effects. Table 4.3 reports these results for the Grunfeld data where the observed F-statistics of 52.362, 1.403, and 17.403 replicate those in Table 4.1.

Table 4.3 Grunfeld's data: F-tests for redundant fixed effects

Redundant Fixed Effects Tests
Equation: EQFE
Test cross-section and period fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	52.362355	(9,169)	0.0000
Cross-section Chi-square	266.395500	9	0.0000
Period F	1.403241	(19,169)	0.1309
Period Chi-square	29.297556	19	0.0614
Cross-Section/Period F	17.403146	(28,169)	0.0000
Cross-Section/Period Chi-square	271.340224	28	0.0000

In fact, the F-statistic given in (3.8) which tests the null hypothesis $H_0^{c'}$; $\mu_1 = \ldots = \mu_{N-1} = 0$ and $\lambda_1 = \ldots = \lambda_{T-1} = 0$ yields 17.403. Since EViews already ran the unrestricted model with both time period and individual fixed effects, it reports the restricted regression for this hypothesis which is OLS. Also, the F-statistic for the null hypothesis $H_0^{d'}$; $\mu_1 = \ldots = \mu_{N-1} = 0$ allowing $\lambda_t \neq 0$ for $t = 1, \ldots, T-1$, described below (3.8), yields 52.362. EViews reports the restricted regression for this hypothesis which is a one-way time period fixed effects model. Finally, the F-statistic for the null hypothesis $H_0^{e'}$; $\lambda_1 = \ldots = \lambda_{T-1} = 0$ allowing $\mu_i \neq 0$ for $i = 1, \ldots, (N-1)$, described below (3.9), yields 1.403. EViews reports the restricted regression for this hypothesis which is a one-way individual fixed effects model.

4.3 Hausman's Specification Test

A critical assumption in the error component regression model is that $E(u_{it}/X_{it}) = 0$. This is important given that the disturbances contain individual effects (the μ_i) which are unobserved and may be correlated with the X_{it}. For example, in an earnings equation these μ_i may denote unobservable ability of the individual, and this may be correlated with the schooling variable included on the right-hand side of this equation. In this case, $E(u_{it}/X_{it}) \neq 0$ and the GLS estimator $\widehat{\beta}_{GLS}$ becomes biased and inconsistent for β. However, the Within transformation wipes out these μ_i and leaves the Within estimator $\widetilde{\beta}_{Within}$ unbiased and consistent for β. Hausman (1978) suggests comparing $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$, both of which are consistent under the null hypothesis H_0; $E(u_{it}/X_{it}) = 0$, but will have different probability limits if H_0 is not true. In fact, $\widetilde{\beta}_{Within}$ is consistent whether H_0 is true or not, while $\widehat{\beta}_{GLS}$ is BLUE, consistent, and asymptotically efficient under H_0 , but is inconsistent when H_0 is false. A natural test statistic would be based on $\widehat{q}_1 = \widehat{\beta}_{GLS} - \widetilde{\beta}_{Within}$. Under H_0, plim $\widehat{q}_1 = 0$, and $\text{cov}(\widehat{q}_1, \widehat{\beta}_{GLS}) = 0$.

Using the fact that $\widehat{\beta}_{GLS} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$ and $\widetilde{\beta}_{Within} - \beta = (X'QX)^{-1}X'Qu$, one gets $E(\widehat{q}_1) = 0$ and

$$\begin{aligned}\text{cov}(\widehat{\beta}_{GLS}, \widehat{q}_1) &= \text{var}(\widehat{\beta}_{GLS}) - \text{cov}(\widehat{\beta}_{GLS}, \widetilde{\beta}_{Within})\\ &= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(uu')QX(X'QX)^{-1}\\ &= (X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1} = 0\end{aligned}$$

Using the fact that $\widetilde{\beta}_{Within} = \widehat{\beta}_{GLS} - \widehat{q}_1$, one gets

$$\text{var}(\widetilde{\beta}_{Within}) = \text{var}(\widehat{\beta}_{GLS}) + \text{var}(\widehat{q}_1)$$

since $\text{cov}(\widehat{\beta}_{GLS}, \widehat{q}_1) = 0$. Therefore

$$\text{var}(\widehat{q}_1) = \text{var}(\widetilde{\beta}_{Within}) - \text{var}(\widehat{\beta}_{GLS}) = \sigma_\nu^2(X'QX)^{-1} - (X'\Omega^{-1}X)^{-1} \tag{4.40}$$

Hence, the Hausman test statistic is given by

$$m_1 = \hat{q}_1'[\text{var}(\widehat{q}_1)]^{-1}\widehat{q}_1 \tag{4.41}$$

and under H_0, it is asymptotically distributed as χ^2_K where K denotes the dimension of slope vector β. In order to make this test operational, Ω is replaced by a consistent estimator $\widehat{\Omega}$, and GLS by its corresponding feasible GLS.

An alternative asymptotically equivalent test can be obtained from the augmented regression

$$y^* = X^*\beta + \widetilde{X}\gamma + w \tag{4.42}$$

where $y^* = \sigma_\nu\Omega^{-1/2}y$, $X^* = \sigma_\nu\Omega^{-1/2}X$, and $\widetilde{X} = QX$. Hausman's test is now equivalent to testing whether $\gamma = 0$. This is a standard Wald test for the omission of the variables $\widetilde{X}$ from (4.42).[3] It is worthwhile to rederive this test. In fact, performing OLS on (4.42) yields

$$\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\end{pmatrix} = \begin{bmatrix} X'(Q+\phi^2P)X & X'QX\\ X'QX & X'QX\end{bmatrix}^{-1}\begin{pmatrix} X'(Q+\phi^2P)y\\ X'Qy\end{pmatrix} \tag{4.43}$$

where $\sigma_\nu\Omega^{-1/2} = Q + \phi P$ and $\phi = \sigma_\nu/\sigma_1$ (see (2.20)). Using partitioned inverse formulas, one can show that

$$\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\end{pmatrix} = \begin{bmatrix} E & -E\\ -E & (X'QX)^{-1} + E\end{bmatrix}\begin{pmatrix} X'(Q+\phi^2P)y\\ X'Qy\end{pmatrix} \tag{4.44}$$

where $E = (X'PX)^{-1}/\phi^2$. This reduces to

$$\widehat{\beta} = \widehat{\beta}_{Between} = (X'PX)^{-1}X'Py \tag{4.45}$$

and

$$\widehat{\gamma} = \widetilde{\beta}_{Within} - \widehat{\beta}_{Between} \tag{4.46}$$

Substituting the Within and Between estimators of β into (4.46), one gets

$$\widehat{\gamma} = (X'QX)^{-1}X'Q\nu - (X'PX)^{-1}X'Pu \tag{4.47}$$

It is easy to show that $E(\widehat{\gamma}) = 0$ and

$$\begin{aligned} \text{var}(\widehat{\gamma}) &= E(\widehat{\gamma}\widehat{\gamma}') = \sigma_\nu^2(X'QX)^{-1} + \sigma_1^2(X'PX)^{-1} \\ &= \text{var}(\widetilde{\beta}_{Within}) + \text{var}(\widehat{\beta}_{Between}) \end{aligned} \tag{4.48}$$

since the cross-product terms are zero. The test for $\gamma = 0$ is based on $\widehat{\gamma} = \widetilde{\beta}_{Within} - \widehat{\beta}_{Between} = 0$ and the corresponding test statistic would therefore be $\widehat{\gamma}'(\text{var}(\widehat{\gamma}))^{-1}\widehat{\gamma}$, which looks different from the Hausman m_1 statistic given in (4.41). These tests are numerically exactly identical (see Hausman and Taylor (1981)). In fact, Hausman and Taylor (1981) showed that H_0 can be tested using any of the following three paired differences: $\widehat{q}_1 = \widehat{\beta}_{GLS} - \widetilde{\beta}_{Within}$, $\widehat{q}_2 = \widehat{\beta}_{GLS} - \widehat{\beta}_{Between}$, or $\widehat{q}_3 = \widetilde{\beta}_{Within} - \widehat{\beta}_{Between}$. The corresponding test statistics can be computed as $m_i = \widehat{q}_i'V_i^{-1}\widehat{q}_i$, where $V_i = \text{var}(\widehat{q}_i)$. These are asymptotically distributed as χ^2_K for $i = 1, 2, 3$ under H_0.[4] Hausman and Taylor (1981) proved that these three tests differ from each other by nonsingular matrices. This easily follows from the fact that

$$\widehat{\beta}_{GLS} = W_1\widetilde{\beta}_{Within} + (I - W_1)\widehat{\beta}_{Between}$$

derived in (2.31). So $\widehat{q}_1 = \widehat{\beta}_{GLS} - \widetilde{\beta}_{Within} = (I - W_1)(\widehat{\beta}_{Between} - \widetilde{\beta}_{Within}) = \Gamma\widehat{q}_3$, where $\Gamma = W_1 - I$. Also, $\text{var}(\widehat{q}_1) = \Gamma\text{var}(\widehat{q}_3)\Gamma'$ and

$$\begin{aligned} m_1 &= \widehat{q}_1'[\text{var}(\widehat{q}_1)]^{-1}\widehat{q}_1 = \widehat{q}_3'\Gamma'[\Gamma\text{var}(\widehat{q}_3)\Gamma']^{-1}\Gamma\widehat{q}_3 \\ &= \widehat{q}_3'[\text{var}(\widehat{q}_3)]^{-1}\widehat{q}_3 = m_3 \end{aligned}$$

This proves that m_1 and m_3 are numerically exactly identical. Similarly, one can show that m_2 is numerically exactly identical to m_1 and m_3. In fact, problem 4.13 shows that these m_i are also exactly numerically identical to $m_4 = \widehat{q}_4'V_4^{-1}\widehat{q}_4$ where $\widehat{q}_4 = \widehat{\beta}_{GLS} - \widehat{\beta}_{OLS}$ and $V_4 = \text{var}(\widehat{q}_4)$. In the Monte Carlo study by Baltagi (1981), the Hausman test is performed given that the exogeneity assumption is true. This test performed well with a low frequency of type I errors.

Arellano (1993) provided an alternative variable addition test to the Hausman test which is robust to autocorrelation and heteroskedasticity of arbitrary form. In particular, Arellano (1993) suggests constructing the following regression:

$$\begin{pmatrix} y_i^+ \\ \bar{y}_i \end{pmatrix} = \begin{bmatrix} X_i^+ & 0 \\ \bar{X}_i' & \bar{X}_i' \end{bmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} u_i^+ \\ \bar{u}_i \end{pmatrix} \tag{4.49}$$

where $y_i^+ = (y_{i1}^+, \ldots, y_{iT}^+)'$ and $X_i^+ = (X_{i1}^+, \ldots, X_{iT}^+)'$ is a $T \times K$ matrix and $u_i^+ = (u_{i1}^+, \ldots, u_{iT}^+)'$. Also

$$y_{it}^+ = \left[\frac{T-t}{T-t+1}\right]^{1/2}\left[y_{it} - \frac{1}{(T-t)}(y_{i,t+1} + .. + y_{iT})\right] \quad t = 1, 2, \ldots, T-1$$

defines the forward orthogonal deviations operator, $\bar{y}_i = \Sigma_{t=1}^T y_{it}/T$, X_{it}^+, $\bar{X}_i$, u_{it}^+, and $\bar{u}_i$ are similarly defined. OLS on this model yields $\widehat{\beta} = \widetilde{\beta}_{Within}$ and $\widehat{\gamma} = \widehat{\beta}_{Between} - \widetilde{\beta}_{Within}$. Therefore, Hausman's test can be obtained from the artificial regression

(4.49) by testing for $\gamma = 0$. If the disturbances are heteroskedastic and/or serially correlated, then neither $\widetilde{\beta}_{Within}$ nor $\widehat{\beta}_{GLS}$ are optimal under the null or alternative. Also, the standard formulae for the asymptotic variances of these estimators are no longer valid. Moreover, these estimators cannot be ranked in terms of efficiency so the var(q) is not the difference of the two variances var($\widetilde{\beta}_W$)− var($\widehat{\beta}_{GLS}$). Arellano (1993) suggests using White (1984) robust variance–covariance matrix from OLS on (4.49) and applying a standard Wald test for $\gamma = 0$ using these robust standard errors. This can be easily calculated using any standard regression package that computes White robust standard errors. This test is asymptotically distributed as χ^2_K under the null. In fact, a simpler robust Hausman test can be obtained by testing for $\gamma = 0$ from the artificial regression (4.42) using the robust variance–covariance matrix option; see examples 2 and 3 below for illustrations.

Chamberlain (1982) showed that the fixed effects specification imposes testable restrictions on the coefficients from regressions of all leads and lags of dependent variables on all leads and lags of independent variables. Chamberlain specified the relationship between the unobserved individual effects and X_{it} as follows:

$$\mu_i = X'_{i1}\lambda_1 + .. + X'_{iT}\lambda_T + \varepsilon_i \tag{4.50}$$

where each λ_t is of dimension $K \times 1$ for $t = 1, 2, \ldots, T$. Let $y'_i = (y_{i1}, \ldots, y_{iT})$ and $X'_i = (X'_{i1}, \ldots, X'_{iT})$ and denote the "reduced form" regression of y'_i on X'_i by

$$y'_i = X'_i\pi + \eta_i \tag{4.51}$$

The restrictions between the reduced form and structural parameters are given by

$$\pi = (I_T \otimes \beta) + \lambda\iota'_T \tag{4.52}$$

with $\lambda' = (\lambda'_1, \ldots, \lambda'_T)$. Chamberlain (1982) suggested estimation and testing be carried out using the minimum chi-square method where the minimand is a χ^2 goodness-of-fit statistic for the restrictions on the reduced form. However, Angrist and Newey (1991) showed that this minimand can be obtained as the sum of T terms. Each term of this sum is simply the degrees of freedom times the R^2 from a regression of the Within residuals for a particular period on all leads and lags of the independent variables. Angrist and Newey (1991) illustrate this test using two examples. The first example estimates and tests a number of models for the union wage effect using five years of data from the National Longitudinal Survey of Youth (NLSY). They find that the assumption of fixed effects in an equation for union wage effects is not at odds with the data. The second example considers a conventional human capital earnings function. They find that the fixed effects estimates of the return to schooling in the NLSY are roughly twice those of ordinary least squares. However, the over-identification test suggests that the fixed effects assumption may be inappropriate for this model. Carey (1997) applies the Chamberlain minimum chi-square method to the estimation of a multiple output hospital cost function using a panel of 1733 facilities over the period 1987–1991. OLS (year by year), fixed effects, seemingly unrelated regressions, and Chamberlain's minimum chi-square method are reported. In this application, the minimum chi-squared test rejects the restrictions imposed by the null hypothesis.

Unfortunately, this careful testing of the FE restrictions has not been the usual practice in empirical work. In fact, the standard practice has been to run a Hausman (1978) test. Not rejecting this null, the applied researcher reports the RE estimator. Otherwise, the researcher reports the FE estimator; see Owusu-Gyapong (1986) and Cardellichio (1990) for two such applications. Rejecting the null of the Hausman test implies that the RE estimator is not consistent. This does not necessarily mean that the FE restrictions are satisfied. Therefore, a natural next step would be to test the FE restrictions before settling on this estimator as the preferred one. In fact, Baltagi, Bresson and Pirotte (2009) argue that one should run the Chamberlain (1982) test or its Angrist–Newey (1991) alternative to check that the restrictions imposed by an FE model are valid. Their Monte Carlo results show that these tests yield the same decision and are in conflict at most 2.2 % of the time. One caveat is that like the Sargan over-identification test for dynamic panels, (see Chap. 8), the MCS test tends to understate the true variance of the test statistic as T gets large. This is because as T gets large, the number of testable restrictions increase and the variance of the test statistic is understated. They suggest careful examination of which regressors may or may not be correlated with the individual effects. In this case, one should be willing to entertain a more restricted model where only a subset of the regressors are correlated with the individual effects as proposed by Hausman and Taylor (1981); see Chap. 7. This would impose less restrictions than the general Chamberlain model and is also testable with a Hausman test. Alternatively, one could question the endogeneity of the regressors with the disturbances, not only with the individual effects. This endogeneity leads to inconsistency of the FE estimator and invalidates the Hausman test performed based on the fixed effects versus the random effects estimator, (see Chap. 7).

Modifying the set of additional variables in (4.49) so that the set of K additional regressors are replaced by KT additional regressors, Arellano (1993) obtains

$$\begin{pmatrix} y_i^+ \\ \bar{y}_i \end{pmatrix} = \begin{bmatrix} X_i^+ & 0 \\ \bar{X}_i' & X_i' \end{bmatrix} \begin{pmatrix} \beta \\ \lambda \end{pmatrix} + \begin{pmatrix} u_i^+ \\ \bar{u}_i \end{pmatrix} \tag{4.53}$$

where $X_i = (X_{i1}', \ldots, X_{iT}')'$ and λ is $KT \times 1$. Chamberlain (1982) test of correlated effects based on the reduced form approach turns out to be equivalent to testing for $\lambda = 0$ in (4.53). Once again this can be made robust to an arbitrary form of serial correlation and heteroskedasticity by using a Wald test for $\lambda = 0$ using White (1984) robust standard errors. This test is asymptotically distributed as χ^2_{TK}. Note that this clarifies the relationship between the Hausman specification test and Chamberlain omnibus goodness-of-fit test. In fact, both tests can be computed as Wald tests from the artificial regressions in (4.49) and (4.53). Hausman's test can be considered as a special case of the Chamberlain test for $\lambda_1 = \lambda_2 = .. = \lambda_T = \gamma/T$. Arellano (1993) extends this analysis to dynamic models and to the case where some of the explanatory variables are known to be uncorrelated or weakly correlated with the individual effects.

Ahn and Low (1996) showed that Hausman's test statistic can be obtained from the artificial regression of the GLS residuals $(y_{it}^* - X_{it}^{*\prime}\widehat{\beta}_{GLS})$ on $\widetilde{X}$ and $\bar{X}$, where $\widetilde{X}$ has typical element $\widetilde{X}_{it,k}$ and $\bar{X}$ is the matrix of regressors averaged over time. The test statistic is NT times the R^2 of this regression. Using (4.42), one can test H_0; $\gamma = 0$ by running the Gauss–Newton regression (GNR) evaluated at the restricted estimators under the null. Knowing θ, the restricted estimates yield $\widehat{\beta} = \widehat{\beta}_{GLS}$ and $\widehat{\gamma} = 0$. Therefore, the GNR on (4.42) regresses the GLS residuals $(y^*_{it} - X^{*\prime}_{it}\widehat{\beta}_{GLS})$ on the derivatives of the regression function with respect to β and γ evaluated at $\widehat{\beta}_{GLS}$ and $\widehat{\gamma} = 0$. These regressors are X^*_{it} and $\widetilde{X}_{it}$, respectively. But X^*_{it} and $\widetilde{X}_{it}$ span the same space as $\widetilde{X}_{it}$ and $\bar{X}_{i.}$. This follows immediately from the definition of X^*_{it} and $\widetilde{X}_{it}$ given above. Hence, this GNR yields the same regression sums of squares and therefore, the same Hausman test statistic as that proposed by Ahn and Low (1996); see problem 97.4.1 in *Econometric Theory* by Baltagi (1997).

Ahn and Low (1996) argue that Hausman's test can be generalized to test that each X_{it} is uncorrelated with μ_i and not simply that $\bar{X}_i$ is uncorrelated with μ_i. In this case, one computes NT times R^2 of the regression of GLS residuals $(y^*_{it} - X^{*\prime}_{it}\widehat{\beta}_{GLS})$ on $\widetilde{X}_{it}$ and $[X'_{i1}, \ldots, X'_{iT}]$. This LM statistic is identical to Arellano (1993) Wald statistic described earlier if the same estimates of the variance components are used. Ahn and Low (1996) argue that this test is recommended for testing the joint hypothesis of exogeneity of the regressors and the stability of the regression parameters over time. When the regression parameters are nonstationary over time, both $\widehat{\beta}_{GLS}$ and $\widetilde{\beta}_{Within}$ are inconsistent even though the regressors are exogenous. Monte Carlo experiments were performed that showed that both the Hausman test and the Ahn and Low (1996) test have good power in detecting endogeneity of the regressors. However, the latter test dominates if the coefficients of the regressors are nonstationary. For Ahn and Low (1996), rejection of the null does not necessarily favor the Within estimator since the latter estimator may be inconsistent. In this case, the authors recommend performing Chamberlain (1982) test or the equivalent test proposed by Angrist and Newey (1991).

Guggenberger (2010) shows that if one uses a Hausman pretest to decide between random and fixed effects inference in a panel data context, and then performs a second stage test on H_0; $\beta = \beta_0$, the size of the resulting two-stage test will be distorted. The paper recommends refraining from this pretesting in favor of inference based on the fixed effects estimator.

4.3.1 Example 1: Investment Equation

For the Grunfeld data, the Within estimates are given by $(\widetilde{\beta}_1, \widetilde{\beta}_2) = (0.1101238, 0.310065)$ with a variance–covariance matrix:

$$\text{var}(\widetilde{\beta}_{Within}) = \begin{bmatrix} 0.14058 & -0.077468 \\ & 0.3011788 \end{bmatrix} \times 10^{-3}$$

The Between estimates are given by $(0.1346461, 0.03203147)$ with variance–covariance matrix:

$$\text{var}(\widehat{\beta}_{Between}) = \begin{bmatrix} 0.82630142 & -3.7002477 \\ & 36.4572431 \end{bmatrix} \times 10^{-3}$$

The resulting Hausman test statistic based on (4.46) and (4.48) and labeled as m_3 yields an observed χ_2^2 statistic of 2.131. This is not significant at the 5% level, and we do not reject the null hypothesis of no correlation between the individual effects and the X_{it}. As a cautionary note, one should *not* use the Hausman command in Stata to perform the Hausman test based on a contrast between the fixed effects (FE) and Between (BE) estimators. This will automatically *subtract* the variance–covariance matrices of the two estimators, rather than *add* them as required in (4.48). However, the Hausman test statistic can be properly computed in Stata based upon the contrast between the RE (feasible GLS) estimator and fixed effects (FE). This is the Hausman statistic labeled as m_1 in (4.41) based on the contrast $\widehat{q}_1$ and var$(\widehat{q}_1)$ given in (4.40). Table 4.4 gives the Stata output using the Hausman command which computes (4.41). This yields an m_1 statistic of 2.33 which is distributed as χ_2^2. This does not reject the null hypothesis as obtained above using m_3. Note that the feasible GLS estimator in Stata is SWAR and is computed whenever the RE option is invoked. If one puts the option *sigmaless*, in the Hausman command, one is using the same estimate of σ_ν^2 (obtained from the consistent fixed effects estimates) in computing the variance–covariance matrix of both the consistent and efficient estimators. This yields an m_1 statistic of 2.13, which is exactly identical to m_3. One can also compute m_2 based on $\widehat{q}_2$ which is the contrast between the SWAR feasible GLS estimator and the Between estimator. Table 4.5 gives the Stata output that replicates this Hausman test yielding an m_2 statistic of 2.13. Note that the Stata comment in Table 4.5 that the Between estimator is consistent under the alternative is not necessarily true. Under the null however, the Between estimator is consistent, and this allows the computation of the variance of the difference as the difference in variances, as explained above. As expected, this statistic is not significant and does not reject the null hypothesis. The same result is obtained using m_1, m_2, and m_3. Hence, one does not reject the null hypothesis that the RE estimator is efficient. The augmented regression, given in (4.42) based on the SWAR feasible GLS estimate of θ, $(\widehat{\theta} = 0.861)$, yields the following estimates $\widehat{\beta} = (0.135, 0.032) = \widehat{\beta}_{Between}$, as derived in (4.45) and $\widehat{\gamma} = (-0.025, 0.278) = \widetilde{\beta}_{Within} - \widehat{\beta}_{Between}$, as derived in (4.46). The test for $H_0 : \gamma = 0$ yields an $F(2, 195)$ statistic of 1.07 with a p-value of 0.347. One can use the Stata user-written command *xtoverid* after running *xtreg, re*. This yields a χ_2^2 statistic of 2.131 which is identical to the Hausman test obtained from the three m statistics reported above. One can also get a robust Hausman test by running the augmented regression, given in (4.42) with the robust variance–covariance option and test for $\gamma = 0$. This yields the robust $F(2, 195)$ statistic of 1.58 with a p-value of 0.208. All Hausman statistics lead to not rejecting H_0.

Table 4.4 Grunfeld's data: Hausman test FE versus RE

```
. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
           F |    .1101238     .1097811        .0003427        .0055213
           C |    .3100653      .308113        .0019524        .0024516
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.33
                Prob>chi2 =      0.3119
```

Table 4.5 Grunfeld's data: Hausman test Between versus RE

```
. hausman be re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       be           re         Difference          S.E.
-------------+----------------------------------------------------------------
           F |    .1346461     .1097811        .0248649         .026762
           C |    .0320315      .308113       -.2760815        .1901633
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.13
                Prob>chi2 =      0.3445
```

4.3.2 Example 2: Gasoline Demand Equation

For the Baltagi and Griffin (1983) gasoline data, the Within estimates are given by $(\widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3) = (0.66128, -0.32130, -0.64015)$ with variance–covariance matrix

$$\text{var}(\widetilde{\beta}_{Within}) = \begin{bmatrix} 0.539 & 0.029 & -0.205 \\ & 0.194 & 0.009 \\ & & 0.088 \end{bmatrix} \times 10^{-2}$$

The Between estimates are given by $(0.96737, -0.96329, -0.79513)$ with variance–covariance matrix

$$\text{var}(\widehat{\beta}_{Between}) = \begin{bmatrix} 2.422 & -1.694 & -1.056 \\ & 1.766 & 0.883 \\ & & 0.680 \end{bmatrix} \times 10^{-2}$$

The resulting Hausman χ^2_3 test statistic is $m_3 = 26.507$ which is significant. Hence, we reject the null hypothesis of no correlation between the individual effects and the X_{it}, and we conclude that the random effects estimator is not consistent. One can similarly compute $m_2 = 27.45$, based on the contrast between the SWAR feasible GLS estimator and the between estimator, and $m_1 = 302.8$ based on the contrast between the SWAR feasible GLS estimator and the fixed effects estimator. These were obtained using Stata. If one puts the option *sigmaless*, in the Hausman command, one is using the same estimate of σ_ν^2 (obtained from the consistent fixed effects estimates) in computing the variance–covariance matrix of both the consistent and efficient estimators. This yields an m_1 statistic of 26.50, exactly identical to m_3. All three test statistics m_1, m_2, and m_3 lead to the same decision. The augmented regression, given in (4.42) based on the SWAR feasible GLS estimate of θ, ($\widehat{\theta} = 0.892$), yields the following estimates $\widehat{\beta} = (0.967, -0.963, -0.795) = \widehat{\beta}_{Between}$, as derived in (4.45) and $\widehat{\gamma} = (-0.306, 0.642, 0.155) = \widetilde{\beta}_{Within} - \widehat{\beta}_{Between}$, as derived in (4.46). The test for $H_0 : \gamma = 0$ yields an $F(3, 335)$ statistic of 8.83 with a p-value of 0.000. One can use the Stata user-written command *xtoverid* after running *xtreg, re*. This yields a χ_3^2 statistic of 26.495 which is identical to the Hausman test obtained from the three m statistics reported above. One can also get a robust Hausman test by running the augmented regression, given in (4.42) with the robust variance–covariance option and test for $\gamma = 0$. This yields the robust $F(3, 335)$ statistic of 14.91 with a p-value of 0.000. All Hausman statistics lead to rejecting H_0 and the RE estimator is not consistent.

4.3.3 Example 3: Canadian Manufacturing Industries

Owusu-Gyapong (1986) considered panel data on strike activity in 60 Canadian manufacturing industries for the period 1967–79. A one-way error component model is used and OLS, Within, and random effects GLS estimates are obtained. With $K' = 12$ regressors, $N = 60$ and $T = 13$, an F-test for the significance of industry-specific effects described in (2.12) yields an F-value of 5.56. This is distributed as $F_{59,709}$ under the null hypothesis of zero industry-specific effects. The null is soundly rejected and the Within estimator is preferred to the OLS estimator. Next, H_0; $\sigma_\mu^2 = 0$ is tested using the Breusch and Pagan (1980) two-sided LM test given as LM_1 in (4.23). This yields a χ^2 value of 21.4, which is distributed as χ_1^2 under the null hypothesis of no random effects. The null is soundly rejected and the GLS estimator is preferred to the OLS estimator. Finally, for a choice between the Within and GLS estimator, the author performs a Hausman (1978) type test to test H_0; $E(\mu_i/X_{it}) = 0$. This is based on the difference between the Within and GLS estimators as described in (4.41) and yields a χ^2 value equal to 3.84. This is distributed as χ_{11}^2 under the null and is not significant. The Hausman test was also run as an augmented regression-type test described in (4.42). This also did not reject the null of no correlation between the μ_i and the regressors. Based on these results, Owusu-Gyapong (1986) chose GLS as the preferred estimator.

4.3.4 Example 4: Sawmills in Washington State

Cardellichio (1990) estimated the production behavior of 1147 sawmills in the state of Washington observed biennially over the period 1972–84. A one-way error component model is used and OLS, Within, and random effects GLS estimates are presented. With $K' = 21$ regressors, $N = 1147$, and $T = 7$, an F-test for the stability of the slope parameters over time was performed which was not significant at the 5% level. In addition, an F -test for the significance of sawmill effects described in (2.12) was performed which rejected the null at the 1% significance level. Finally, a Hausman test was performed and it rejected the null at the 1% significance level. Cardellichio (1990) concluded that the regression slopes are stable over time, sawmill dummies should be included, and the Within estimator is preferable to OLS and GLS since the orthogonality assumption between the regressors and the sawmill effects is rejected.

4.3.5 Example 5: Mariage Premium

Cornwell and Rupert (1997) estimated the wage premium attributed to marriage using the 1971, 1976, 1978, and 1980 waves of the NLSY. They find that the Within estimates of the marriage premium are smaller than those obtained from feasible random effects GLS. A Hausman test based on the difference between these two estimators rejects the null hypothesis. This indicates the possibility of important omitted individual specific characteristics which are correlated with both marriage and the wage rate. They conclude that the marriage premium is purely an intercept shift and no more than 5–7%. They also cast doubt on the interpretation that marriage enhances productivity through specialization.

4.3.6 Example 6: Currency Union

Glick and Rose (2002) consider the question of whether leaving a currency union reduces international trade. Using annual data on bilateral trade among 217 countries from 1948 through 1997, they estimate an augmented gravity model controlling for several factors. These include real GDP, distance, land mass, common language, sharing a land border, whether they belong to the same regional trade agreement, land locked, island nations, common colonizer, current colony, ever a colony, and whether they remained part of the same nation. The focus variable is a binary variable which is unity if country i and country j use the same currency at time t. They apply OLS, FE, RE, and their preferred estimator is FE based on the Hausman test. They find that a pair of countries which joined/left a currency union experienced a near-doubling/halving of bilateral trade. The data set along with the Stata logs are available on Roses' website; see Problem 4.19.

4.3.7 Hausman's Test for the Two-Way Model

For the two-way error component model, Hausman (1978) test can still be based on the difference between the fixed effects estimator (with both time and individual dummies) and the two-way random effects GLS estimator. Also, the augmented regression, given in (4.42), can still be used as long as the Within and GLS transformations used are those for the two-way error component model. But, what about the equivalent tests described for the one-way model? Do they extend to the two-way model? Not quite. Kang (1985) showed that a similar equivalence for the Hausman test does not hold for the two-way error component model, since there would be two Between estimators, one Between time periods $\widehat{\beta}_T$ and one between cross-sections $\widehat{\beta}_C$. Also, $\widehat{\beta}_{GLS}$ is a weighted combination of $\widehat{\beta}_T$, $\widehat{\beta}_C$ and the Within estimator $\widetilde{\beta}_W$. Kang (1985) shows that the Hausman test based on $(\widetilde{\beta}_W - \widehat{\beta}_{GLS})$ is not equivalent to that based on $(\widehat{\beta}_C - \widehat{\beta}_{GLS})$ nor that based on $(\widehat{\beta}_T - \widehat{\beta}_{GLS})$. But there are other types of equivalencies (see Kang's Table 2). More importantly, Kang classifies five testable hypotheses:

(1) Assume that μ_i are fixed and test $E(\lambda_t/X_{it}) = 0$ based upon $\widetilde{\beta}_W - \widehat{\beta}_T$;
(2) Assume the μ_i are random and test $E(\lambda_t/X_{it}) = 0$ based upon $\widehat{\beta}_T - \widehat{\beta}_{GLS}$;
(3) Assume the λ_t are fixed and test $E(\mu_i/X_{it}) = 0$ based upon $\widetilde{\beta}_W - \widehat{\beta}_C$;
(4) Assume the λ_t are random and test $E(\mu_i/X_{it}) = 0$ based upon $\widehat{\beta}_C - \widehat{\beta}_{GLS}$;
(5) Compare two estimators, one which assumes both the μ_i and λ_T are fixed, and another that assumes both are random such that $E(\lambda_t/X_{it}) = E(\mu_i/X_{it}) = 0$. This test is based upon $\widehat{\beta}_{GLS} - \widetilde{\beta}_W$.

EViews computes two-way Hausman tests after running a two-way random effects model with both time period and individual effects. The two-way random effects model is set up as the comparison model, i.e., under the null, it is the efficient estimator. Table 4.6 reports the results of a two-way Hausman test for the Grunfeld data. Not shown is the two-way random effects model using the Wansbeek and Kapteyn option. This was reported in Table 3.4. Recall that the Swamy and Arora and Wallace and Hussain options yielded negative estimates of the variance component σ_λ^2; see Tables 3.2 and 3.3. The Hausman test based on two-way random versus two-way fixed yields a test statistic of 8.842 which is distributed as χ_2^2 under the null hypothesis. This has a p-value of .012 and is rejected at the 5% level. This is the same test as described in (5) by Kang (1985). The backup run reported by EViews is the two-way fixed effects model which was already given in Table 3.2 and is not reproduced here. This rejects the two-way random effects model estimator.

Note that EViews also reports two other Hausman tests. The first is based on the contrast between the two-way RE model and a *mixed* model with λ_t *random* and μ_i *fixed*. This assumes that under the null, the two-way RE model is efficient. The test statistic reported is 0.715 which is distributed as χ_2^2 under the null hypothesis and is not significant. The second Hausman test is based on the contrast between the two-way RE model and a mixed model with μ_i random and λ_t fixed. This again assumes

Table 4.6 Grunfeld's data: two-way Hausman tests
Correlated Random Effects - Hausman Test
Equation: EQFE
Test cross-section and period random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	0.715071	2	0.6994
Period random	7.208541	2	0.0272
Cross-section and period random	8.842195	2	0.0120

Correlated Random Effects - Hausman Test
Equation: EQFE
Test period random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Period random	8.291397	2	0.0158

Correlated Random Effects - Hausman Test
Equation: EQFE
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	2.963565	2	0.2272

that under the null, the two-way RE model is efficient. The test statistic reported is 7.209 which is distributed as χ^2_2 under the null hypothesis and is significant.

Having rejected the two-way RE model, one can check whether one effect is random given that the other effect is fixed. In this case, the comparison model will always be the two-way fixed effects model which is consistent rain or shine. The first Hausman test is based upon the contrast between a two-way *mixed* model with μ_i *random* and λ_t *fixed* versus a two-way fixed effects model. This yields a test statistic value of 2.964. This is distributed as χ^2_2 under the null hypothesis and is not significant. This does not reject the possibility that μ_i is random given that λ_t is fixed. Switching the mixed effects, one can run another Hausman test based upon the contrast between a two-way fixed effects model versus λ_t random and μ_i fixed. This yields a test statistic value of 8.291. This is distributed as χ^2_2 under the null hypothesis and is significant. Hence, the null is rejected. Note however, that all diagnostics in Table 4.1 have been indicating that time effects are not present and one should not employ a two-way random effects model, but rather a one-way random individual effects model.

4.4 Further Reading

Other tests for poolability include Ziemer and Wetzstein (1983) who suggest comparing pooled estimators (like $\widehat{\delta}_{OLS}$) with nonpooled estimators (like $\widehat{\delta}_{i,OLS}$) according to their forecast risk performance. Using a wilderness recreation demand model, they show that a Stein rule estimator gives a better forecast risk performance than the pooled or individual cross-section estimators. The Stein rule estimator for δ_i in (4.1) is given by

$$\widehat{\delta}_i^* = \widehat{\delta}_{OLS} + \left(1 - \frac{c}{F_{obs}}\right)(\widehat{\delta}_{i,OLS} - \widehat{\delta}_{OLS})$$

where $\widehat{\delta}_{i,OLS}$ is given in (4.6) and $\widehat{\delta}_{OLS}$ is given in (4.4). F_{obs} is the F-statistic to test H_0; $\delta_i = \delta$, given in (4.8), and the constant c is given by $c = ((N-1)K' - 2)/(NT - NK' + 2)$. Note that $\widehat{\delta}_i^*$ shrinks $\widehat{\delta}_{i,OLS}$ toward the pooled estimator $\widehat{\delta}_{OLS}$. Maddala et al. (1997) argued that shrinkage estimators appear to be better than either the pooled estimator or the individual cross-section estimators.

Baltagi, Hidalgo and Li (1996) derive a nonparametric test for poolability which is robust to functional form misspecification. In particular, they consider the following nonparametric panel data model

$$y_{it} = g_t(x_{it}) + \epsilon_{it} \qquad (i = 1, \ldots, N;\ t = 1, \ldots, T)$$

where $g_t(.)$ is an unspecified functional form that may vary over time. x_{it} is a $k \times 1$ column vector of predetermined explanatory variables with $(p \geq 1)$ variables being continuous and $k - p(\geq 0)$. Poolability of the data over time is equivalent to testing that $g_t(x) = g_s(x)$ almost everywhere for all t and $s = 1, 2, \ldots, T$; versus $g_t(x) \neq g_s(x)$ for some $t \neq s$ with probability greater than zero. The test statistic is shown to be consistent and asymptotically normal and is applied to an earnings equation using data from the PSID.

For cases where $N > T$, Pesaran, Smith and Im (1996) propose a Hausman (1978) type test for the poolability of the slopes based on the contrast of the FE estimator of β and an average of the individual OLS estimates for each β_i obtained in (4.6). The latter is called the *mean group* estimator $\widehat{\beta}_{MG} = \sum_i \widehat{\beta}_{i,ols}/N$. Under the null hypothesis, both estimators are consistent, but the FE estimator is more efficient than $\widehat{\beta}_{MG}$. However, Pesaran and Yamagata (2008) show that in case the model contains only strictly exogenous regressors, this Hausman test for slope homogeneity lacks power in all directions, if under the alternative hypothesis the slopes are random draws from the same distribution, i.e., a random coefficient model; see Swamy (1970).

For cases where $T > N$, Phillips and Sul (2003) propose an alternative type of Hausman test for slope homogeneity based on the difference between the individual

OLS estimates $\widehat{\beta}_{i,ols}$ stacked as an NK vector and the FE estimator $\widetilde{\beta}_{FE}$ repeated N times, i.e., $(\iota_N \otimes \widetilde{\beta}_{FE})$. Phillips and Sul consider a number of different consistent estimators of the variance of this contrast, including Andrew's (1993) median unbiased estimator and its extension to panels. But, as they note, all such estimators yield the same asymptotic covariance matrix as $T \to \infty$. Assuming the null of slope homogeneity holds, and some standard assumptions, Phillips and Sul show that the asymptotic distribution of this test statistic, for N fixed and $T \to \infty$, is $\chi_2^2(NK)$, as long as the variance of the contrast is a nonstochastic positive definite matrix.

For cases where $T > N$, Swamy (1970) suggest a test of slope homogeneity based on the dispersion of individual slope estimates $\widehat{\beta}_{i,ols}$ from a weighted pooled estimator allowing for cross-sectional heteroskedasticity. This is given by $\widehat{\beta}_S = (\sum_i X_i' E_T X_i/\widehat{\sigma}_i^2)^{-1}(\sum_i X_i' E_T y_i/\widehat{\sigma}_i^2)$ where $E_T = (I_T - \bar{J}_T)$ and $\widehat{\sigma}_i^2 = (y_i - X_i\widehat{\beta}_{i,ols})' E_T (y_i - X_i\widehat{\beta}_{i,ols})/(T - K - 1)$. The test statistic is given by $\widehat{S} = \sum_i (\widehat{\beta}_{i,ols} - \widehat{\beta}_S)'(X_i' E_T X_i/\widehat{\sigma}_i^2)(\widehat{\beta}_{i,ols} - \widehat{\beta}_S)$, and this is asymptotically distributed as $\chi_2^2((N-1)K)$ as T tends to infinity, with N fixed; see also Hsiao (2003, p. 149). Pesaran and Yamagata (2008) also consider $\widetilde{S}$ which replaces $\widehat{\sigma}_i^2$ everywhere by $\widetilde{\sigma}_i^2 = (y_i - X_i\widetilde{\beta}_{FE})' E_T (y_i - X_i\widetilde{\beta}_{FE})/(T-1)$. They also standardize the test statistics corresponding to $\widehat{S}$ and $\widetilde{S}$, denoting them by $\widehat{\Delta} = \sqrt{\frac{N}{2K}}(\frac{\widehat{S}}{N} - K)$ and $\widetilde{\Delta} = \sqrt{\frac{N}{2K}}(\frac{\widetilde{S}}{N} - K)$. They show that under a set of albeit restrictive assumptions, $\widehat{\Delta}$ and $\widetilde{\Delta}$ are asymptotically $N(0,1)$ as N and $T \to \infty$, as long as $\sqrt{N}/T \to 0$ for $\widehat{\Delta}$, and $\sqrt{N}/T^2 \to 0$ for $\widetilde{\Delta}$. Pesaran and Yamagata (2008) perform some Monte Carlo experiments to study the size and power of these tests. They find that the Hausman test has correct size, but no power irrespective of the sample size. On the other hand, Swamy's $\widehat{S}$ test has power, but tends to over-reject when T is small relative to N, with the extent of over-rejection diminishing only as T is increased relative to N. By contrast, the adjusted version of the dispersion test, $\widetilde{\Delta}_{adj} = \sqrt{\frac{N(T+1)}{2K(T-K-1)}}(\frac{\widetilde{S}}{N} - K)$, has the correct size for all combinations of sample sizes, even when T is very small relative to N.

The normality assumption on the error components disturbances may be untenable. Horowitz and Markatou (1996) show how to carry out nonparametric estimation of the densities of the error components. Using data from the Current Population Survey, they estimate an earnings model and show that the probability that individuals with low earnings will become high earners in the future is much lower than that obtained under the assumption of normality. One drawback of this nonparametric estimator is its slow convergence at a rate of $1/(\log N)$ where N is the number of individuals. Monte Carlo results suggest that this estimator should be used for N larger than 1000.

Blanchard and Mátyás (1996) perform Monte Carlo simulations to study the robustness of several tests for individual effects with respect to nonnormality of the disturbances. The alternative distributions considered are the exponential, lognormal, $t(5)$, and the Cauchy distributions. The main findings are that the F-test is robust against nonnormality while the one-sided and two-sided LM and LR tests are sensitive to nonnormality.

4.5 Notes

1. An elegant presentation of this F-statistic is given in Fisher (1970).
2. Critical values for the mixed χ^2_m are 7.289, 4.231, and 2.952 for $\alpha = 0.01, 0.05$, and 0.1, respectively.
3. Hausman (1978) tests $\gamma = 0$ from (4.42) using an F-statistic. The restricted regression yields OLS of y^* on X^*. This is the Fuller and Battese (1973) regression yielding GLS as described below (2.20). The unrestricted regression adds the matrix of Within regressors $\widetilde{X}$ as in (4.42). Baltagi and Liu (2007) showed that Hausman's test can be alternatively obtained by running the artificial regression of y^* on X^* and $\overline{X}$, and testing that the latter coefficients are zero, or running the artificial regression of y^* on X^* and X, and testing that the latter coefficients are zero; see problem 4.12.
4. For an important discussion of what null hypothesis is actually being tested using the Hausman test; see Holly (1982).

4.6 Problems

4.1 Verify the relationship between M and M^*, i.e., $MM^* = M^*$, given below (4.7). Hint: use the fact that $Z = Z^*I^*$ with $I^* = (\iota_N \otimes I_{K'})$.

4.2 Verify that $\dot{M}$ and $\dot{M}^*$ defined below (4.10) are both symmetric, idempotent, and satisfy $\dot{M}\dot{M}^* = \dot{M}^*$.

4.3 For *Grunfeld's data* given as Grunfeld.fil on the Springer website, verify the testing for the poolability results given in example 1, Sect. 4.1.3.

4.4 For the *gasoline data* given as Gasoline.dat on the Springer website, verify the testing for the poolability results given in example 2, Sect. 4.1.3.

4.5 Breusch and Pagan (1980) Lagrange multiplier test. Under normality of the disturbances, show that for the likelihood function given in (4.15),
(a) The information matrix is block-diagonal between $\theta' = (\sigma^2_\mu, \sigma^2_\lambda, \sigma^2_\nu)$ and δ.
(b) For H^c_0; $\sigma^2_\mu = \sigma^2_\lambda = 0$, verify (4.18), (4.20), and (4.22).

4.6 *Locally mean most powerful one-sided test*. Using the results of Baltagi, Chang and Li (1992), verify that the King and Wu (1997) LM test for H^c_0; $\sigma^2_\mu = \sigma^2_\lambda = 0$ is given by (4.30).

4.7 *Standardized LM tests*. For H^c_0; $\sigma^2_\mu = \sigma^2_\lambda = 0$, (a) Verify that the standardized Lagrange multiplier (SLM) test statistics derived by Honda (1991) is as described by (4.26) and (4.31).
(b) Also, verify that the King and Wu (1997) standardized test statistic is as described by (4.26) and (4.32).

4.8 Using the Monte Carlo setup for the two-way error component model described in Baltagi (1981),
(a) Compare the performance of the Chow F-test and the Roy–Zellner test for various values of the variance components.

(b) Compare the performance of the BP, KW, SLM, LR, GHM, and F-test statistics as done in Baltagi, Chang and Li (1992).

(c) Perform Hausman's specification test and discuss its size for the various experiments conducted.

4.9 For the *Grunfeld data*, replicate Table 4.1.

4.10 For the *gasoline data*, derive a similar table to test the hypotheses given in Table 4.1.

4.11 For the *public capital* data, derive a similar table to test the hypotheses given in Table 4.1.

4.12 Hausman (1978) *test based on an artificial regression*. Show that Hausman's test can be alternatively obtained from any one of the following artificial regressions:

$$y^* = X^*\beta + \widetilde{X}\gamma + w_1$$
$$y^* = X^*\beta + \bar{X}\gamma + w_2$$
$$y^* = X^*\beta + X\gamma + w_3$$

where $y^* = \sigma_\nu \Omega^{-1/2} y$, $X^* = \sigma_\nu \Omega^{-1/2} X$, $\widetilde{X} = QX$, and $\bar{X} = PX$; see (4.42). Hausman's test is equivalent to testing whether $\gamma = 0$ from any one of these three OLS regressions; see Baltagi and Liu (2007).

4.13 *Three contrasts yield the same Hausman test.* (a) Verify that m_2 is numerically exactly identical to m_1 and m_3, where $m_i = \widehat{q}_i' V_i^{-1} \widehat{q}_i$ defined below (4.48).

(b) Verify that these are also exactly numerically identical to $m_4 = \widehat{q}_4' V_4^{-1} \widehat{q}_4$ where $\widehat{q}_4 = \widehat{\beta}_{GLS} - \widehat{\beta}_{OLS}$ and $V_4 = \text{var}(\widehat{q}_4)$. Hint: see problem 89.3.3 in *Econometric Theory* by Baltagi (1989) and its solution by Koning (1990).

4.14 *Testing for correlated effects in panels.* This is based on problem 95.2.5 in *Econometric Theory* by Baltagi (1995a). This problem asks the reader to show that Hausman's test, studied in Sect. 4.3, can be derived from Arellano (1993) extended regression by using an alternative transformation of the data. In particular, consider the transformation given by $H = (C', \iota_T/T)'$ where C is the first $(T-1)$ rows of the Within transformation $E_T = I_T - \bar{J}_T$, I_T is an identity matrix of dimension T, and $\bar{J}_T = \iota_T \iota_T'/T$ with ι_T a vector of 1's of dimension T.

(a) Show that the matrix C satisfies the following properties: $C\iota_T = 0$, $C'(CC')^{-1}C = I_T - \bar{J}_T$; see Arellano and Bover (1995).

(b) For the transformed model $y_i^+ = Hy_i = (y_i^{*\prime}, \bar{y}_i)'$, where $y_i^* = Cy_i$ and $\bar{y}_i = \Sigma_{t=1}^T y_{it}/T$, the typical element of y_i^* is given by $y_{it}^* = [y_{it} - \bar{y}_i]$ for $t = 1, 2, \ldots, T-1$. Consider the extended regression similar to (4.49)

$$\begin{bmatrix} y_i^* \\ \bar{y}_i \end{bmatrix} = \begin{bmatrix} X_i^{*\prime} & 0 \\ \bar{X}_i' & \bar{X}_i' \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} u_i^* \\ \bar{u}_i \end{bmatrix}$$

and show that GLS on this extended regression yields $\widehat{\beta} = \widehat{\beta}_{Within}$ and $\widehat{\gamma} = \widehat{\beta}_{Between} - \widehat{\beta}_{Within}$, where $\widehat{\beta}_{Within}$ and $\widehat{\beta}_{Between}$ are the familiar panel data estimators. Conclude that Hausman's test for H_0: $E(\mu_i/X_i) = 0$ can be based on a test for $\gamma = 0$, as shown by Arellano (1993). See solution 95.2.5 in *Econometric Theory* by Xiong (1996).

4.15 For the *Grunfeld data*, replicate the Hausman test results given in example 1 of Sect. 4.3.

4.16 For the *gasoline data*, replicate the Hausman test results given in example 2 of Sect. 4.3.

4.17 Perform Hausman's test for the *public capital data*.

4.18 *The relative efficiency of the Between estimator with respect to the Within estimator*. This is based on problem 99.4.3 in *Econometric Theory* by Baltagi (1999). Consider the simple panel data regression model

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T \tag{1}$$

where α and β are scalars. Subtract the mean equation to get rid of the constant

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{..}) + u_{it} - \bar{u}_{..}, \tag{2}$$

where $\bar{x}_{..} = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}/NT$ and $\bar{y}_{..}$ and $\bar{u}_{..}$ are similarly defined. Add and subtract $\bar{x}_{i.}$ from the regressor in parentheses and rearrange

$$y_{it} - \bar{y}_{..} = \beta(x_{it} - \bar{x}_{i.}) + \beta(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{3}$$

where $\bar{x}_{i.} = \Sigma_{t=1}^{T} x_{it}/T$. Now run the unrestricted least squares regression

$$y_{it} - \bar{y}_{..} = \beta_w(x_{it} - \bar{x}_{i.}) + \beta_b(\bar{x}_{i.} - \bar{x}_{..}) + u_{it} - \bar{u}_{..} \tag{4}$$

where β_w is not necessarily equal to β_b.

(a) Show that the least squares estimator of β_w from (4) is the Within estimator and that of β_b is the Between estimator.

(b) Show that if $u_{it} = \mu_i + \nu_{it}$ where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and among themselves, then ordinary least squares (OLS) is equivalent to generalized least squares (GLS) on (4).

(c) Show that for model (1), the relative efficiency of the Between estimator with respect to the Within estimator is equal to $(B_{XX}/W_{XX})[(1-\rho)/(T\rho + (1-\rho))]$, where $W_{XX} = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T}(x_{it} - \bar{x}_{i.})^2$ denotes the Within variation and $B_{XX} = T\Sigma_{i=1}^{N}(\bar{x}_{i.} - \bar{x}_{..})^2$ denotes the Between variation. Also, $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ denotes the equicorrelation coefficient.

(d) Show that the square of the t-statistic used to test H_0; $\beta_w = \beta_b$ in (4) yields exactly Hausman (1978) specification test. See solution 99.4.3 in *Econometric Theory* by Gurmu (2000).

4.19 *Currency union and trade*. Using the Glick and Rose (2002) data set, downloadable from Roses' website at (http://haas.berkeley.edu)

(a) Replicate their results for the FE, RE, Between, and MLE estimators reported in Table 4 of their paper.

(b) Perform the Hausman test based on FE versus RE as well as Between versus RE using Stata.

4.20 *Investment and Tobin's q*. Schaller (1990) uses data based on financial statements of 188 large publicly traded US firms, over the period 1951–1985, to estimate an investment equation based on Tobin's q. The dependent variable is the ratio of investment to the capital stock (I/K). q is the ratio of the market value of the firm to the replacement cost of its assets. The data are available from the *Journal of Applied Econometrics* data archives.

(a) Replicate the descriptive statistics given in Table I of Schaller (1990, p. 313).

(b) Replicate the OLS, FE, and RE regressions for both the broad and narrow definitions of capital given in Tables II and III of Schaller (1990, p. 313).

(c) Perform the Hausman test for FE versus RE for Tables II and III.

References

Ahn, S.C., and S. Low. 1996. A reformulation of the Hausman test for regression models with pooled cross-section time-series data. *Journal of Econometrics* 71: 309–319.

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13.

Andrews, D.W.K. 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* 61: 139–165.

Angrist, J.D., and W.K. Newey. 1991. Over-identification tests in earnings functions with fixed effects. *Journal of Business and Economic Statistics* 9: 317–323.

Arellano, M. 1993. On the testing of correlated effects with panel data. *Journal of Econometrics* 59: 87–97.

Arellano, M., and O. Bover. 1995. Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics* 68: 29–51.

Baltagi, B.H. 1981. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error components model. *Journal of Econometrics* 17: 21–49.

Baltagi, B.H. 1989. The equivalence of the Boothe–MacKinnon and the Hausman specification tests in the context of panel data, Problem 89.3.3. *Econometric Theory* 5: 454.

Baltagi, B.H. 1995a. Testing for correlated effects in panels, Problem 95.2.5. *Econometric Theory* 11: 401–402.

Baltagi, B.H. 1995b. *Econometric Analysis of Panel Data*. Chichester: John Wiley.

Baltagi, B.H. 1997. Hausman's specification test as a Gauss–Newton regression, Problem 97.4.1. *Econometric Theory* 13: 757.

Baltagi, B.H. 1999. The relative efficiency of the between estimator with respect to the within estimator, Problem 99.4.3. *Econometric Theory* 15: 630–631.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2009. Testing the fixed effects restrictions? A monte carlo study of Chamberlain's minimum chi-squared test. *Statistics and Probability Letters* 79: 1358–1362.

Baltagi, B.H., Y.J. Chang, and Q. Li. 1992. Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics* 54: 95–120.

Baltagi, B.H., and J.M. Griffin. 1983. Gasoline demand in the OECD: An application of pooling and testing procedures. *European Economic Review* 22: 117–137.

Baltagi, B.H., J. Hidalgo, and Q. Li. 1996. A non-parametric test for poolability using panel data. *Journal of Econometrics* 75: 345–367.

Baltagi, B.H., and L. Liu. 2007. Alternative ways of obtaining Hausman's test using artificial regressions. *Statistics and Probability Letters* 77: 1413–1417.

Blanchard, P., and L. Mátyás. 1996. Robustness of tests for error components models to non-normality. *Economics Letters* 51: 161–167.
Breusch, T.S., and A.R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.
Cardellichio, P.A. 1990. Estimation of production behavior using pooled microdata. *Review of Economics and Statistics* 72: 11–18.
Carey, K. 1997. A panel data design for estimation of hospital cost functions. *Review of Economics and Statistics* 79: 443–453.
Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.
Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28: 591–605.
Cornwell, C., and P. Rupert. 1988. Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 3: 149–155.
Cornwell, C., and P. Rupert. 1997. Unobservable individual effects, marriage, and then the earnings of young men. *Economic Inquiry* 35: 285–294.
Evans, M.A., and M.L. King. 1985. Critical value approximations for tests of linear regression disturbances. *Australian Journal of Statistics* 27: 68–83.
Fisher, F.M. 1970. Tests of equality between sets of coefficients in two linear regressions: An expository note. *Econometrica* 38: 361–366.
Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.
Glick, R., and A.K. Rose. 2002. Does a currency union affect trade? *The time series evidence, European Economic Review* 46: 1125–1151.
Gourieroux, C., A. Holly, and A. Monfort. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50: 63–80.
Grunfeld, Y. 1958. The determinants of corporate investment, unpublished Ph.D. dissertation. Chicago: University of Chicago.
Gurmu, S. 2000. The relative efficiency of the between-estimator with respect to the within-estimator, Solution 99.4.3. *Econometric Theory* 16: 454–456.
Hartley, H.O., and J.N.K. Rao. 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54: 93–108.
Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320–340.
Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.
Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.
Hemmerle, W.J., and H.O. Hartley. 1973. Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics* 15: 819–831.
Holly, A. 1982. A remark on Hausman's specification test. *Econometrica* 50: 749–759.
Honda, Y. 1985. Testing the error components model with non-normal disturbances. *Review of Economic Studies* 52: 681–690.
Honda, Y. 1991. A standardized test for the error components model with the two-way layout. *Economics Letters* 37: 125–128.
Horowitz, J.L., and M. Markatou. 1996. Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63: 145–168.
Hsiao, C. 2003. *Analysis of Panel Data*. Cambridge: Cambridge University Press.
Kang, S. 1985. A note on the equivalence of specification tests in the two-factor multivariate variance components model. *Journal of Econometrics* 28: 193–203.
King, M.L., and P.X. Wu. 1997. Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* 16: 131–156.
Koning, R.H. 1990. The equivalence of the Boothe–MacKinnon and the Hausman specification tests in the context of panel data, Solution 89.3.3. *Econometric Theory* 6: 409.

Maddala, G.S., R.P. Trost, H. Li, and F. Joutz. 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics* 15: 90–100.

McElroy, M.B. 1977. Weaker MSE criteria and tests for linear restrictions in regression models with non-spherical disturbances. *Journal of Econometrics* 6: 389–394.

Moulton, B.R., and W.C. Randolph. 1989. Alternative tests of the error components model. *Econometrica* 57: 685–693.

Owusu-Gyapong, A. 1986. Alternative estimating techniques for panel data on strike activity. *Review of Economics and Statistics* 68: 526–531.

Pesaran, M.H., R. Smith and K.S. Im. 1996. Dynamic linear models for heterogenous panels, Chapter 8. In *The econometrics of panel data: A handbook of the theory with applications*, (eds.) L. Mátyás and P. Sevestre, Dordrecht: Kluwer Academic Publishers, 145–195.

Pesaran, M.H., and T. Yamagata. 2008. Testing slope homogeneity in large panels. *Journal of Econometrics* 142: 50–93.

Phillips, P.C.B., and D. Sul. 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6: 217–259.

Roy, S.N. 1957. *Some aspects of multivariate analysis*. New York: Wiley.

Schaller, H. 1990. A re-examination of the q theory of investment using U.S. firm data. *Journal of Applied Economics* 5: 209–325.

Swamy, P.A.V.B. 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38: 311–323.

Toro-Vizcarrondo, C., and T.D. Wallace. 1968. A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association* 63: 558–572.

Wallace, T.D. 1972. Weaker criteria and tests for linear restrictions in regression. *Econometrica* 40: 689–698.

Wallace, T.D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37: 55–72.

White, H. 1984. *Asymptotic theory for econometricians*. New York: Academic Press.

Xiong, W. 1996. Testing for correlated effects in panels, Solution 95.2.5. *Econometric Theory* 12: 405–406.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–368.

Ziemer, R.F., and M.E. Wetzstein. 1983. A Stein-rule method for pooling data. *Economics Letters* 11: 137–143.

Heteroskedasticity and Serial Correlation in the Error Component Model

5

5.1 Heteroskedasticity

The standard error component model given by Eqs. (2.1) and (2.2) assumes that the regression disturbances are homoskedastic with the same variance across time and individuals. This may be a restrictive assumption for panels, where the cross-sectional units may be of varying size and as a result may exhibit different variations. For example, when dealing with gasoline demand across OECD countries, steam electric generation across various size utilities, or estimating cost functions for various US airline firms, one should expect to find heteroskedasticity in the disturbance term. Assuming homoskedastic disturbances when heteroskedasticity is present will still result in consistent estimates of the regression coefficients, but these estimates will not be efficient. Also, the standard errors of these estimates will be biased and one should compute robust standard errors correcting for the possible presence of heteroskedasticity. In this section, we relax the assumption of homoskedasticity of the disturbances and introduce heteroskedasticity through the μ_i as first suggested by Mazodier and Trognon (1978). Next, we suggest an alternative heteroskedastic error component specification, where only the ν_{it} are heteroskedastic. We derive the true GLS transformation for these two models. We also consider two adaptive heteroskedastic estimators based on these models where the heteroskedasticity is of unknown form. These adaptive heteroskedastic estimators were suggested by Li and Stengos (1994) and Roy (2002).

Mazodier and Trognon (1978) generalized the homoskedastic error component model to the case where the μ_i are heteroskedastic, i.e., $\mu_i \sim (0, w_i^2)$ for $i = 1, \ldots, N$, but $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$. In vector form, $\mu \sim (0, \Sigma_\mu)$ where $\Sigma_\mu = \text{diag}[w_i^2]$ is a diagonal matrix of dimension $N \times N$, and $\nu \sim (0, \sigma_\nu^2 I_{NT})$. Therefore, using (2.4), one gets

$$\Omega = E(uu') = Z_\mu \Sigma_\mu Z_\mu' + \sigma_\nu^2 I_{NT} \tag{5.1}$$

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_5

This can be written as

$$\Omega = \text{diag}[w_i^2] \otimes J_T + \text{diag}[\sigma_\nu^2] \otimes I_T \tag{5.2}$$

where $\text{diag}[\sigma_\nu^2]$ is also of dimension $N \times N$. Using the Wansbeek and Kapteyn (1982) trick, Baltagi and Griffin (1988) derived the corresponding Fuller and Battese (1974) transformation as follows:

$$\Omega = \text{diag}[T w_i^2 + \sigma_\nu^2] \otimes \bar{J}_T + \text{diag}[\sigma_\nu^2] \otimes E_T$$

Therefore

$$\Omega^r = \text{diag}[(\tau_i^2)^r] \otimes \bar{J}_T + \text{diag}[(\sigma_\nu^2)^r] \otimes E_T \tag{5.3}$$

with $\tau_i^2 = T w_i^2 + \sigma_\nu^2$, and r is any arbitrary scalar. The Fuller–Battese transformation for the heteroskedastic case premultiplies the model by

$$\sigma_\nu \Omega^{-1/2} = \text{diag}[\sigma_\nu/\tau_i] \otimes \bar{J}_T + (I_N \otimes E_T) \tag{5.4}$$

Hence, $y^* = \sigma_\nu \Omega^{-1/2} y$ has a typical element $y^*_{it} = y_{it} - \theta_i \bar{y}_{i.}$ where $\theta_i = 1 - (\sigma_\nu/\tau_i)$ for $i = 1, \ldots, N$.

Baltagi and Griffin (1988) provided feasible GLS estimators including Rao's (1971a, b) MINQUE estimators for this model. Phillips (2003) argues that this model suffers from the incidental parameters problem and the variance estimates of μ_i (the ω_i^2) cannot be estimated consistently, so there is no guarantee that feasible GLS and true GLS will have the same asymptotic distributions. Instead, he suggests a stratified error component model where the variances change across strata and provide an EM algorithm to estimate it. It is important to note that Mazodier and Trognon (1978) had already suggested stratification in a two-way heteroskedastic error component model, and also that one can specify parametric variance functions which avoid the incidental parameter problem and then apply the GLS transformation described above. As in the cross-section heteroskedastic case, one has to know the variables that determine heteroskedasticity, but not necessarily the form. Adaptive estimation of heteroskedasticity of unknown form has been suggested for this model by Roy (2002). This follows a similar literature on adaptive estimation for cross-section data.

Alternatively, one could keep the μ_i homoskedastic with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and impose the heteroskedasticity on the ν_{it}, i.e., $\upsilon_{it} \sim (0, w_i^2)$ (see Problem 88.2.2 by Baltagi (1988) and its solution by Wansbeek (1989) in *Econometric Theory*). In this case, using (2.4) one obtains

$$\Omega = E(uu') = \text{diag}[\sigma_\mu^2] \otimes J_T + \text{diag}[w_i^2] \otimes I_T \tag{5.5}$$

Replacing J_T by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$, we get

$$\Omega = \text{diag}[T\sigma_\mu^2 + w_i^2] \otimes \bar{J}_T + \text{diag}[w_i^2] \otimes E_T$$

and

$$\Omega^r = \text{diag}[(\tau_i^2)^r] \otimes \bar{J}_T + \text{diag}[(w_i^2)^r] \otimes E_T \tag{5.6}$$

where $\tau_i^2 = T\sigma_\mu^2 + w_i^2$, and r is an arbitrary scalar. Therefore

$$\Omega^{-1/2} = \text{diag}[1/\tau_i] \otimes \bar{J}_T + \text{diag}[1/w_i] \otimes E_T \tag{5.7}$$

and $y^* = \Omega^{-1/2}y$ has a typical element

$$y_{it}^* = (\bar{y}_{i.}/\tau_i) + (y_{it} - \bar{y}_{i.})/w_i$$

Upon rearranging terms, we get

$$y_{it}^* = \frac{1}{w_i}(y_{it} - \theta_i \bar{y}_{i.}) \quad \text{where} \theta_i = 1 - (w_i/\tau_i)$$

One can argue that heteroskedasticity will contaminate both μ_i and ν_{it}, and it is hard to claim that it is in one component and not the other. Randolph (1988) gives the GLS transformation for a more general heteroskedastic model where *both* the μ_i and the ν_{it} are assumed heteroskedastic in the context of an unbalanced panel. In this case, the var$(\mu_i) = \sigma_i^2$ and $E(\nu\nu') = \text{diag}[\sigma_{it}^2]$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$. Li and Stengos (1994) considered the regression model given by (2.1) and (2.2) with $\mu_i \sim IID\left(0, \sigma_\mu^2\right)$ and $E\left[v_{it}|X_{it}'\right] = 0$ with $Var\left[v_{it}|X_{it}'\right] = \gamma\left(X_{it}'\right) \equiv \gamma_{it}$, so that the heteroskedasticity is on the remainder error term and it is of an unknown form.

Therefore $\sigma_{it}^2 = E\left[u_{it}^2|X_{it}\right] = \sigma_\mu^2 + \gamma_{it}$ and the proposed estimator of σ_μ^2 is given by

$$\widehat{\sigma}_\mu^2 = \frac{\sum_{i=1}^{N}\sum_{t\neq s}^{T} \widehat{u}_{it}\widehat{u}_{is}}{NT\,(T-1)}$$

where $\widehat{u}_{it}$ denotes the OLS residual. Also

$$\widehat{\gamma}_{it} = \frac{\sum_{j=1}^{N}\sum_{s=1}^{T} \widehat{u}_{js}^2 K_{it,js}}{\sum_{j=1}^{N}\sum_{s=1}^{T} K_{it,js}} - \widehat{\sigma}_\mu^2$$

where the kernel function is given by $K_{it,js} = K\left(\frac{X_{it}'-X_{js}'}{h}\right)$ and h is the smoothing parameter. These estimators of the variance components are used to construct a feasible adaptive GLS estimator of β which they denote by GLSAD. The computation of their feasible GLS estimator is simplified into an OLS regression using a recursive transformation that reduces the general heteroskedastic error component structure into classical errors; see Li and Stengos (1994) for details.

Roy (2002) considered the alternative heteroskedastic model $E\left[\mu_i|\overline{X}'_{i.}\right] = 0$ with

$$Var\left[\mu_i|\overline{X}'_{i.}\right] = \omega\left(\overline{X}'_{i.}\right) \equiv \omega_i$$

with $\overline{X}'_{i.} = \sum_{t=1}^{T} X'_{it}/T$ and $v_{it} \sim IID\left(0, \sigma_v^2\right)$, so that the heteroskedasticity is on the individual specific error component and it is of an unknown form. Roy (2002) used the usual estimator of σ_v^2 which is the MSE of the Within regression; see (2.24) and this can be written as

$$\widehat{\sigma}_v^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(y_{it} - \overline{y}_{i.}\right) - \left(X_{it} - \overline{X}_{i.}\right)'\widetilde{\beta}\right]^2}{N(T-1) - k}$$

where $\widetilde{\beta}$ is the fixed effects or Within estimator of β given in (2.7). Also

$$\widehat{\omega}_i = \frac{\sum_{j=1}^{N}\sum_{t=1}^{T}\widehat{u}_{jt}^2 K_{i.,j.}}{\sum_{j=1}^{N}\sum_{t=1}^{T} K_{i.,j.}} - \widehat{\sigma}_v^2$$

where the kernel function is given by

$$K_{i.,j.} = K\left(\frac{\overline{X}'_{i.} - \overline{X}'_{j.}}{h}\right)$$

Using these estimators of the variance components, Roy (2002) computed a feasible GLS estimator using the transformation derived by Baltagi and Griffin (1988), see (5.4). This was denoted by EGLS.

Both Li and Stengos (1994) and Roy (2002) performed Monte Carlo experiments based on the simple regression model given in (2.8). They compared the following estimators: (1) OLS; (2) Fixed effects or Within estimator (Within); (3) the conventional GLS estimator for the one-way error component model that assumes the error term u_{it} is homoskedastic (GLSH); and (4) their own adaptive heteroskedastic estimator denoted by (EGLS) for Roy (2002) and (GLSAD) for Li and Stengos (1994). Li and Stengos (1994) found that their adaptive estimator outperforms all the other estimators in terms of relative MSE with respect to true GLS for $N = 50, 100$, and $T = 3$ and for moderate to severe degrees of heteroskedasticity. Roy (2002) also found that her adaptive estimator performs well, although it was outperformed by fixed effects in some cases where there were moderate and severe degrees of heteroskedasticity. Baltagi, Bresson and Pirotte (2005) checked the sensitivity of the two proposed adaptive heteroskedastic estimators under misspecification of the form of heteroskedasticity. In particular, they ran Monte Carlo experiments using the heteroskedasticity setup of Li and Stengos (1994) to see how the misspecified Roy (2002) estimator performs. Next, they used the heteroskedasticity set up of Roy (2002) to see how the misspecified Li and Stengos (1994) estimator performs. They also checked the sensitivity of these results to the choice of the smoothing parameters, the sample size, and the degree of heteroskedasticity. Baltagi, Bresson and Pirotte (2005) found that in terms of loss in efficiency, misspecifying the adaptive form of heteroskedasticity can be costly when the Li and Stengos (1994) model is correct and the researcher performs the Roy (2002) estimator. This loss in efficiency is smaller when the true model is that of Roy (2002) and one performs the Li and Stengos (1994) estimator. The latter statement is true as long as the choice of bandwidth is not too small. Both papers also reported the 5% size performance of the estimated t-ratios of the slope coefficient. Li and Stengos (1994) found that only GLSAD had the correct size while OLS, GLSH and Within over-rejected the null hypothesis. Roy (2002) found that GLSH and EGLS had the correct size no matter what choice of h was used. Baltagi, Bresson and Pirotte (2005) found that OLS and GLSAD (small h) tend to over-reject the null when true no matter what form of adaptive heteroskedasticity. In contrast, GLSH, EGLS, and Within have size not significantly different from 5% when the true model is that of Roy (2002) and slightly over-reject (7–8%) when the true model is that of Li and Stengos (1994).

In Chap. 2, we pointed out that Arellano (1987) gave a neat way of obtaining standard errors for the fixed effects estimator that are robust to heteroskedasticity and serial correlation of arbitrary form; see Eq. (2.16). In Chap. 4, we discussed how Arellano (1993) suggested a Hausman (1978) test as well as a Chamberlain (1982) omnibus goodness of fit test that are robust to heteroskedasticity and serial correlation of arbitrary form; see Eqs. (4.49) and (4.53). Li and Stengos (1994) suggested a modified Breusch and Pagan test for significance of the random individual effects, i.e., H_0; $\sigma_\mu^2 = 0$, which is robust to heteroskedasticity of unknown form in the remainder error term.

5.1.1 Testing for Homoskedasticity in an Error Component Model

Verbon (1980) derived a Lagrange multiplier test for the null hypothesis of homoskedasticity against the heteroskedastic alternative $\mu_i \sim \left(0, \sigma^2_{\mu_i}\right)$ and $\nu_{it} \sim \left(0, \sigma^2_{\nu_i}\right)$. In Verbon's model, however, $\sigma^2_{\mu_i}$ and $\sigma^2_{\nu_i}$ are, up to a multiplicative constant, *identical* parametric functions of time-invariant exogenous variables Z_i, i.e., $\sigma^2_{\mu_i} = \sigma^2_\mu f\,(Z_i\theta_2)$ and $\sigma^2_{\nu_i} = \sigma^2_\nu f\,(Z_i\theta_1)$. Lejeune (2006), on the other hand, dealt with maximum likelihood estimation and Lagrange multiplier testing of a general heteroskedastic one-way error component regression model assuming that $\mu_i \sim \left(0, \sigma^2_{\mu_i}\right)$ and $\nu_{it} \sim \left(0, \sigma^2_{\nu_{it}}\right)$ where $\sigma^2_{\mu_i}$ and $\sigma^2_{\nu_{it}}$ are *distinct* parametric functions of exogenous variables Z_{it} and F_i, i.e., $\sigma^2_{\nu_{it}} = \sigma^2_\nu h_\nu\,(Z_{it}\theta_1)$ and $\sigma^2_{\mu_i} = \sigma^2_\mu h_\mu\,(F_i\theta_2)$. In the context of incomplete panels, Lejeune (2006) derived two joint LM tests for no individual effects and homoskedasticity in the remainder error term. The first LM test considers a random effects one-way error component model with $\mu_i \sim IIN\left(0, \sigma_\mu^2\right)$ and a remainder error term that is heteroskedastic, $v_{it} \sim N\left(0, \sigma_{v_{it}}^2\right)$ with $\sigma_{v_{it}}^2 = \sigma_v^2 h_v\left(Z_{it}\theta_1\right)$. The joint hypothesis $H_0;\ \theta_1 = \sigma_\mu^2 = 0$ renders OLS the restricted MLE. Lejeune argued that there is no need to consider a variance function for μ_i since one is testing σ_μ^2 equal to zero. While the computation of the LM test statistic is simplified under this assumption, i.e., $\mu_i \sim IIN\left(0, \sigma_\mu^2\right)$, this is not in the original spirit of Lejeune's ML estimation where both μ_i and v_{it} have general variance functions. Lejeune's second LM test considers a fixed effects one-way error component model where μ_i is a fixed parameter to be estimated, and the remainder error term is heteroskedastic with $v_{it} \sim N\left(0, \sigma_{v_{it}}^2\right)$ and $\sigma_{v_{it}}^2 = \sigma_v^2 h_v\left(Z_{it}\theta_1\right)$. The joint hypothesis is $H_0;\ \mu_i = \theta_1 = 0$ for all $i = 1, 2, \ldots, N$. This renders OLS the restricted MLE.

Holly and Gardiol (2000) derived a score test for homoskedasticity in a one-way error component model where the alternative model is that the μ_i's are independent and distributed as $N(0, \sigma_{\mu_i}^2)$ where $\sigma_{\mu_i}^2 = \sigma_\mu^2 h_\mu\left(F_i\theta_2\right)$. Here, F_i is a vector of p explanatory variables such that $F_i\theta_2$ does not contain a constant term and h_μ is a strictly positive twice differentiable function satisfying $h_\mu(0) = 1$ with $h'_\mu(0) \neq 0$ and $h''_\mu(0) \neq 0$. The score test statistic for $H_0;\ \theta_2 = 0$ turns out to be one half the explained sum of squares of the OLS regression of $(\hat{s}/\bar{s}) - \iota_N$ against the p regressors in F as in the Breusch and Pagan test for homoskedasticity. Here $\hat{s}_i = \hat{u}_i'\bar{J}_T\hat{u}_i$ and $\bar{s} = \sum_{i=1}^{N} \hat{s}_i/N$ where $\widehat{u}$ denote the maximum likelihood residuals from the restricted model under $H_0;\ \theta_2 = 0$. This is a one-way homoskedastic error component model with $\mu_i \sim N(0, \sigma_\mu^2)$. The reader is asked to verify this result in Problem 5.3.

In the spirit of the general heteroskedastic model of Randolph (1988) and Lejeune (2006), Baltagi, Bresson and Pirotte (2006) derived a *joint* Lagrange multiplier test for homoskedasticity, i.e., $H_0;\ \theta_1 = \theta_2 = 0$. Under the null hypothesis, the model is a homoskedastic one-way error component regression model. Note that this is different from Lejeune (2006), where under his null, $\sigma_\mu^2 = 0$, so that the restricted MLE is OLS and not MLE on a one-way homoskedastic error component model. Allowing for $\sigma_\mu^2 > 0$ is more likely to be the case in panel data where heterogeneity across the individuals is likely to be present even if heteroskedasticity is not. The model under the null is exactly that of Holly and Gardiol (2000) but it is more general under the alternative since it does not assume a homoskedastic remainder error term. Next, Baltagi, Bresson and Pirotte (2006) derived an LM test for the null hypothesis of homoskedasticity of the individual random effects assuming homoskedasticity of the remainder error term, i.e., $\theta_2 = 0 \mid \theta_1 = 0$. Not surprisingly, they get the Holly and Gardiol (2000) LM test. Last but not least, Baltagi, Bresson and Pirotte (2006) derived an LM test for the null hypothesis of homoskedasticity of the remainder error term assuming homoskedasticity of the individual effects, i.e., $\theta_1 = 0 \mid \theta_2 = 0$. The details for the derivations and the resulting statistics are not provided here and the reader is referred to their paper. Monte Carlo experiments showed that the joint LM test performed well when both error components were heteroskedastic, and performed second best when one of the components was homoskedastic while the other was not. In contrast, the marginal LM tests performed best when heteroskedasticity was present in the right error component. They yielded misleading results if heteroskedasticity was present in the wrong error component.

5.2 Serial Correlation

The classical error component disturbances given by (2.2) assume that the only correlation over time is due to the presence of the same individual across the panel. In Chap. 2, this equicorrelation coefficient was shown to be correl$(u_{it}, u_{is}) = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ for $t \neq s$. Note that it is the same no matter how far t is from s. This may be a restrictive assumption for economic relationships, like investment or consumption, where an unobserved shock this period may affect the behavioral relationship for at least the next few periods. This type of serial correlation is not allowed for in the simple error component model. Ignoring serial correlation when it is present results in consistent but inefficient estimates of the regression coefficients and biased standard errors. This section introduces serial correlation in the ν_{it}, first as an autoregressive process of order one AR(1), as in the Lillard and Willis (1978) study on earnings, next, as a second-order autoregressive process AR(2), also as a special fourth-order autoregressive process AR(4) for quarterly data, and finally as a first-order moving average MA(1) process. For all these serial correlation specifications, a simple generalization of the Fuller and Battese (1973) transformation is derived and the implications for predictions are given. Testing for individual effects and serial correlation is taken up in the last subsection.

5.2.1 The AR(1) Process

Lillard and Willis (1978) generalized the error component model to the serially correlated case, by assuming that the remainder disturbances (the ν_{it}) follow an AR(1) process. In this case $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, whereas

$$\nu_{it} = \rho \nu_{i,t-1} + \epsilon_{it} \tag{5.8}$$

$|\rho| < 1$, and $\epsilon_{it} \sim \text{IID}(0, \sigma_\epsilon^2)$. The μ_i are independent of the ν_{it} and $\nu_{i0} \sim (0, \sigma_\epsilon^2/(1-\rho^2))$. Baltagi and Li (1991a) derived the corresponding Fuller and Battese (1974) transformation for this model. First, one applies the Prais–Winsten (PM) transformation matrix

$$C = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ . & . & . & \cdots & . & . & . \\ . & . & . & \cdots & . & . & . \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

to transform the remainder AR(1) disturbances into serially uncorrelated classical errors. For panel data, this has to be applied for N individuals. The transformed regression disturbances are in vector form

$$u^* = (I_N \otimes C)u = (I_N \otimes C\iota_T)\mu + (I_N \otimes C)\nu \tag{5.9}$$

Using the fact that $C\iota_T = (1-\rho)\iota_T^\alpha$, where $\iota_T^{\alpha\prime} = (\alpha, \iota_{T-1}^\prime)$ and $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, one can rewrite (5.9) as

$$u^* = (1-\rho)(I_N \otimes \iota_T^\alpha)\mu + (I_N \otimes C)\nu \tag{5.10}$$

Therefore, the variance–covariance matrix of the transformed disturbances is

$$\Omega^* = E(u^* u^{*\prime}) = \sigma_\mu^2 (1-\rho)^2 [I_N \otimes \iota_T^\alpha \iota_T^{\alpha\prime}] + \sigma_\epsilon^2 (I_N \otimes I_T)$$

since $(I_N \otimes C)E(\nu\nu^\prime)(I_N \otimes C^\prime) = \sigma_\epsilon^2 (I_N \otimes I_T)$. Alternatively, this can be rewritten as

$$\Omega^* = d^2 \sigma_\mu^2 (1-\rho)^2 [I_N \otimes \iota_T^\alpha \iota_T^{\alpha\prime}/d^2] + \sigma_\epsilon^2 (I_N \otimes I_T) \tag{5.11}$$

where $d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = \alpha^2 + (T-1)$. This replaces $J_T^\alpha = \iota_T^\alpha \iota_T^{\alpha\prime}$ by $d^2 \bar{J}_T^\alpha$, its idempotent counterpart, where $\bar{J}_T^\alpha = \iota_T^\alpha \iota_T^{\alpha\prime}/d^2$. Extending the Wansbeek and Kapteyn trick, we replace I_T by $E_T^\alpha + \bar{J}_T^\alpha$, where $E_T^\alpha = I_T - \bar{J}_T^\alpha$. Collecting terms with the same matrices, one obtains the spectral decomposition of Ω^*.

$$\Omega^* = \sigma_\alpha^2 (I_N \otimes \bar{J}_T^\alpha) + \sigma_\epsilon^2 (I_N \otimes E_T^\alpha) \tag{5.12}$$

where $\sigma_\alpha^2 = d^2 \sigma_\mu^2 (1-\rho)^2 + \sigma_\epsilon^2$. Therefore

$$\sigma_\epsilon \Omega^{*-1/2} = (\sigma_\epsilon/\sigma_\alpha)(I_N \otimes \bar{J}_T^\alpha) + (I_N \otimes E_T^\alpha) = I_N \otimes I_T - \theta_\alpha (I_N \otimes \bar{J}_T^\alpha) \tag{5.13}$$

where $\theta_\alpha = 1 - (\sigma_\epsilon/\sigma_\alpha)$.

Premultiplying the PW transformed observations $y^* = (I_N \otimes C)y$ by $\sigma_\epsilon \Omega^{*-1/2}$, one gets $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$. The typical elements of $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ are given by

$$(y_{i1}^* - \theta_\alpha \alpha b_i, y_{i2}^* - \theta_\alpha b_i, \ldots, y_{iT}^* - \theta_\alpha b_i)^\prime \tag{5.14}$$

where $b_i = [(\alpha y_{i1}^* + \sum_2^T y_{it}^*)/d^2]$ for $i = 1, \ldots, N$.[1] The first observation gets special attention in the AR(1) error component model. First, the PW transformation gives it a special weight $\sqrt{1-\rho^2}$ in y^*. Second, the Fuller and Battese transformation gives it a special weight $\alpha = \sqrt{(1+\rho)/(1-\rho)}$ in computing the weighted average b_i and the pseudo-difference in (5.14). Note that (i) if $\rho = 0$, then $\alpha = 1$, $d^2 = T$, $\sigma_\alpha^2 = \sigma_1^2$, and $\theta_\alpha = \theta$. Therefore, the typical element of y_{it}^{**} reverts to the familiar $(y_{it} - \theta \bar{y}_{i.})$ transformation for the one-way error component model with no serial correlation. (ii) If $\sigma_\mu^2 = 0$, then $\sigma_\alpha^2 = \sigma_\epsilon^2$ and $\theta_\alpha = 0$. Therefore, the typical element of y_{it}^{**} reverts to the PW transformation y_{it}^*.

The BQU estimators of the variance components arise naturally from the spectral decomposition of Ω^*. In fact, $(I_N \otimes E_T^\alpha)u^* \sim (0, \sigma_\epsilon^2[I_N \otimes E_T^\alpha])$ and $(I_N \otimes \bar{J}_T^\alpha)u^* \sim (0, \sigma_\alpha^2[I_N \otimes \bar{J}_T^\alpha])$ and

$$\widehat{\sigma}_\epsilon^2 = u^{*\prime}(I_N \otimes E_T^\alpha)u^*/N(T-1) \text{ and } \widehat{\sigma}_\alpha^2 = u^{*\prime}(I_N \otimes \bar{J}_T^\alpha)u^*/N \tag{5.15}$$

provide the BQU estimators of σ_ϵ^2 and σ_α^2, respectively. Baltagi and Li (1991a) suggest estimating ρ from Within residuals $\widetilde{\nu}_{it}$ as $\widetilde{\rho} = \sum_{i=1}^N \sum_{t=1}^T \widetilde{\nu}_{it}\widetilde{\nu}_{i,t-1}/\sum_{i=1}^N \sum_{t=2}^T \widetilde{\nu}_{i,t-1}^2$. Then, $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ are estimated from (5.15) by substituting OLS residuals $\hat{u}^*$ from the PW transformed equation using $\widetilde{\rho}$. Using Monte Carlo experiments, Baltagi and Li (1997) found that $\widetilde{\rho}$ performs poorly for small T and recommended an alternative estimator of ρ which is based on the autocovariance function $Q_s = E(u_{it}u_{i,t-s})$. For the AR(1) model given in (5.8), it is easy to show that $Q_s = \sigma_\mu^2 + \sigma_\nu^2\rho^s$. From Q_0, Q_1, and Q_2, one can easily show that $\rho + 1 = (Q_0 - Q_2)/(Q_0 - Q_1)$. Hence, a consistent estimator of ρ (for large N) is given by

$$\widehat{\rho} = \frac{\widetilde{Q}_0 - \widetilde{Q}_2}{\widetilde{Q}_0 - \widetilde{Q}_1} - 1 = \frac{\widetilde{Q}_1 - \widetilde{Q}_2}{\widetilde{Q}_0 - \widetilde{Q}_1}$$

where $\widetilde{Q}_s = \sum_{i=1}^N \sum_{t=s+1}^T \widehat{u}_{it}\widehat{u}_{i,t-s}/N(T-s)$ and $\widehat{u}_{it}$ denotes the OLS residuals from (2.1). $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ are estimated from (5.15) by substituting OLS residuals $\hat{u}^*$ from the PW transformed equation using $\widehat{\rho}$ rather than $\widetilde{\rho}$.

Therefore, the estimation of an AR(1) serially correlated error component model is considerably simplified by (i) applying the PW transformation in the first step, as is usually done in the time-series literature, and (ii) subtracting a pseudo-average from these transformed data as in (5.14) in the second step.

5.2.2 The AR(2) Process

This simple transformation can be extended to allow for an AR(2) process on the ν_{it}, i.e.,

$$\nu_{it} = \rho_1\nu_{i,t-1} + \rho_2\nu_{i,t-2} + \epsilon_{it} \tag{5.16}$$

where $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$, $|\rho_2| < 1$ and $|\rho_1| < (1-\rho_2)$. Let $E(\nu_i\nu_i') = \sigma_\epsilon^2 V$, where $\nu_i' = (\nu_{i1}, \ldots, \nu_{iT})$ and note that V is invariant to $i = 1, \ldots, N$. The unique $T \times T$ lower triangular matrix C with positive diagonal elements which satisfies $CVC' = I_T$ is given by

$$C = \begin{bmatrix} \gamma_0 & 0 & 0\,0 \ldots & 0 & 0 & 0 & 0 \\ -\gamma_2 & \gamma_1 & 0\,0 \ldots & 0 & 0 & 0 & 0 \\ -\rho_2 & -\rho_1 & 1\,0 \ldots & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot\,\cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot\,\cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0\,0 \ldots & -\rho_2 & -\rho_1 & 1 & 0 \\ 0 & 0 & 0\,0 \ldots & 0 & -\rho_2 & -\rho_1 & 1 \end{bmatrix}$$

where $\gamma_0 = \sigma_\epsilon/\sigma_\nu$, $\gamma_1 = \sqrt{1-\rho_2^2}$, $\gamma_2 = \gamma_1[\rho_1/(1-\rho_2)]$, and $\sigma_\nu^2 = \sigma_\epsilon^2(1-\rho_2)/(1+\rho_2)[(1-\rho_2)^2 - \rho_1^2]$. The transformed disturbances are given by

$$u^* = (I_N \otimes C)u = (1-\rho_1-\rho_2)(I_N \otimes \iota_T^\alpha)\mu + (I_N \otimes C)\nu \tag{5.17}$$

This uses the fact that $C\iota_T = (1-\rho_1-\rho_2)\times$ (the new ι_T^α) where $\iota_T^{\alpha\prime} = (\alpha_1, \alpha_2, \iota_{T-2}^\prime)$, $\alpha_1 = \sigma_\epsilon/\sigma_\nu(1-\rho_1-\rho_2)$, and $\alpha_2 = \sqrt{(1+\rho_2)/(1-\rho_2)}$.

Similarly, one can define

$$d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = \alpha_1^2 + \alpha_2^2 + (T-2),\ J_{T,}^\alpha E_T^\alpha,\ \text{etc.}$$

as in Sect. 5.2.1, to obtain

$$\Omega^* = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2[I_N \otimes \bar{J}_T^\alpha] + \sigma_\epsilon^2[I_N \otimes I_T] \tag{5.18}$$

as in (5.11). The only difference is that $(1-\rho_1-\rho_2)$ replaces $(1-\rho)$ and ι_T^α is defined in terms of α_1 and α_2 rather than α. Similarly, one can obtain $\sigma_\epsilon\Omega^{*-1/2}$ as in (5.13) with $\sigma_\alpha^2 = d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2 + \sigma_\epsilon^2$. The typical elements of $y^{**} = \sigma_\epsilon\Omega^{*-1/2}y^*$ are given by

$$(y_{i1}^* - \theta_\alpha\alpha_1 b_i,\ y_{i2}^* - \theta_\alpha\alpha_2 b_i,\ y_{i3}^* - \theta_\alpha b_i, \ldots, y_{iT}^* - \theta_\alpha b_i) \tag{5.19}$$

where $b_i = [(\alpha_1 y_{i1}^* + \alpha_2 y_{i2}^* + \sum_3^T y_{it}^*)/d^2]$. The first two observations get special attention in the AR(2) error component model, first in the matrix C defined above (5.17) and second in computing the average b_i and the Fuller and Battese transformation in (5.19). Therefore, one can obtain GLS on this model by (i) transforming the data as in the time-series literature by the C matrix defined above (5.17) and (ii) subtracting a pseudo-average in the second step as in (5.19).

5.2.3 The AR(4) Process for Quarterly Data

Consider the specialized AR(4) process for quarterly data, i.e., $\nu_{it} = \rho\nu_{i,t-4} + \epsilon_{it}$, where $|\ \rho\ | < 1$ and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$. The C matrix for this process can be defined as follows: $u_i^* = Cu_i$ where

$$\begin{aligned} u_{it}^* &= \sqrt{1-\rho^2}\ u_{it} \text{for } t = 1,2,3,4 \\ &= u_{it} - \rho u_{i,t-4} \text{ for } t = 5,6,\ldots,Ts \end{aligned} \tag{5.20}$$

This means that the μ_i component of u_{it} gets transformed as $\sqrt{1-\rho^2}\ \mu_i$ for $t = 1,2,3,4$ and as $(1-\rho)\mu_i$ for $t = 5,6,\ldots,T$. This can be rewritten as $\alpha(1-\rho)\mu_i$ for $t = 1,2,3,4$ where $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, and $(1-\rho)\mu_i$ for $t = 5,\ldots,T$, so that $u^* = (I_N \otimes C)u$ is given by (5.9) with a new C, the same α, but $\iota_T^{\alpha\prime} = (\alpha, \alpha, \alpha, \alpha, \iota_{T-4}^\prime)$, $d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = 4\alpha^2 + (T-4)$, and the derivations Ω^* and $\sigma_\epsilon\Omega^{*-1/2}$ in (5.12) and (5.13) are the same. The typical elements of $y^{**} = \sigma_\epsilon\Omega^{*-1/2}y^*$ are given by

$$(y_{i1}^* - \theta_\alpha \alpha b_i, \ldots, y_{i4}^* - \theta_\alpha b_i, y_{i5}^* - \theta_\alpha b_i, \ldots, y_{iT}^* - \theta_\alpha b_i) \tag{5.21}$$

where $b_i = [(\alpha(\sum_{t=1}^4 y_{it}^*) + \sum_{t=5}^T y_{it}^*)/d^2]$. Once again, GLS can be easily computed by applying (5.20) to the data in the first step and (5.21) in the second step.

5.2.4 The MA(1) Process

For the MA(1) model, defined by

$$\nu_{it} = \epsilon_{it} + \lambda \epsilon_{i,t-1} \tag{5.22}$$

where $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$ and $|\lambda| < 1$, Balestra (1980) gives the following C matrix, $C = D^{-1/2} P$ where $D = \text{diag}\{a_t a_{t-1}\}$ for $t = 1, \ldots, T$,

$$P = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \lambda & a_1 & 0 & \ldots & 0 \\ \lambda^2 & a_1\lambda & a_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda^{T-1} & a_1\lambda^{T-2} & a_2\lambda^{T-3} & \ldots & a_{T-1} \end{bmatrix}$$

and $a_t = 1 + \lambda^2 + \ldots + \lambda^{2t}$ with $a_0 = 1$. For this C matrix, one can show that the new $\iota_T^\alpha = C\iota_T = (\alpha_1, \alpha_2, \ldots, \alpha_T)'$ where these α_t can be solved for recursively as follows:

$$\begin{aligned} \alpha_1 &= (a_0/a_1)^{1/2} \\ \alpha_t &= \lambda(a_{t-2}/a_t)^{1/2}\alpha_{t-1} + (a_{t-1}/a_t)^{1/2} t = 2, \ldots, T \end{aligned} \tag{5.23}$$

Therefore, $d^2 = \iota_T^{\alpha\prime}\iota_T^\alpha = \sum_{t=1}^T \alpha_t^2$, $\sigma_\alpha^2 = d^2\sigma_\mu^2 + \sigma_\epsilon^2$ and the spectral decomposition of Ω^* is the same as that given in (5.12), with the newly defined ι_T^α and σ_α^2. The typical elements of $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ are given by

$$(y_{i1}^* - \theta_\alpha \alpha_1 b_i, \ldots, y_{iT}^* - \theta_\alpha \alpha_T b_i) \tag{5.24}$$

where $b_i = [\sum_{t=1}^T \alpha_t y_{it}^*/d^2]$. Therefore, for an MA(1) error component model, one applies the recursive transformation given in (5.23) in the first step and subtracts a pseudo-average described in (5.24) in the second step; see Baltagi and Li (1991a) for more details. In order to implement the estimation of an error component model with MA(1) remainder errors, Baltagi and Li (1997) proposed an alternative transformation that is simple to compute and requires only least squares. This can be summarized as follows.

Let $\gamma_s = E(\nu_{it}\nu_{i,t-s})$ denote the autocovariance function of ν_{it} and $r = \gamma_1/\gamma_0$. Note that when ν_{it} follows an MA(1) process; we have $Q_s = \sigma_\mu^2 + \gamma_s$ for $s = 0, 1$ and $Q_s = \sigma_\mu^2$ for $s > 1$. Hence we have $\gamma_\tau = Q_\tau - Q_s (\tau = 0, 1)$ for some $s > 1$.

Step 1: Compute $y_{i1}^* = y_{i1}/\sqrt{g_1}$ and $y_{it}^* = [y_{it} - (ry_{i,t-1}^*)/\sqrt{g_{t-1}}]/\sqrt{g_t}$ for $t = 2, \ldots, T$, where $g_1 = 1$ and $g_t = 1 - r^2/g_{t-1}$ for $t = 2, \ldots, T$. Note that this transformation depends only on r, which can be estimated by $\widehat{r} = \widehat{\gamma}_1/\widehat{\gamma}_0 = (\widetilde{Q}_1 - \widetilde{Q}_s)/(\widetilde{Q}_0 - \widetilde{Q}_s)$ for some $s > 1$.

Step 2: Compute y^{**} using the result that $\iota_T^\alpha = C\iota_T = (\alpha_1, \ldots, \alpha_T)'$ with $\alpha_1 = 1$ and $\alpha_t = [1 - r/\sqrt{g_{t-1}}]/\sqrt{g_t}$ for $t = 2, \ldots, T$. Note that in this case $\sigma^2 = \gamma_0$. The estimators of σ_α^2 and σ^2 are simply given by $\widehat{\sigma}_\alpha^2 = (\sum_{t=1}^T \widehat{\alpha}_t^2)\widehat{\sigma}_\mu^2 + \widehat{\sigma}^2$, and $\widehat{\sigma}^2 = \widehat{\gamma}_0 = \widetilde{Q}_0 - \widetilde{Q}_s$ for some $s > 1$ with $\widehat{\sigma}_\mu^2 = \widetilde{Q}_s$ for some $s > 1$. Finally $\widehat{\delta} = 1 - \sqrt{\widehat{\gamma}_0/\widehat{\sigma}_\alpha^2}$. Again, the OLS estimator on the (**) transformed equation is equivalent to GLS on (2.1).

In summary, a simple transformation for the one-way error component model with serial correlation, can be easily generalized to any error process generating the remainder disturbances ν_{it} as long as there exists a *simple* $(T \times T)$ matrix C such that the transformation $(I_N \otimes C)\nu$ has zero mean and variance $\sigma^2 I_{NT}$.

Step 1: Perform the C transformation on the observations of each individual $y_i' = (y_{i1}, \ldots, y_{iT})$ to obtain $y_i^* = Cy_i$ free of serial correlation.

Step 2: Perform another transformation on the y_{it}^*'s, obtained in step 1, which subtracts from y_{it}^* a fraction of a weighted average of observations on y_{it}^*, i.e.,

$$y_{it}^{**} = y_{it}^* - \theta_\alpha \alpha_t (\Sigma_{s=1}^T \alpha_s y_{is}^*)/(\Sigma_{s=1}^T \alpha_s^2)$$

where the α_t's are the elements of $\iota_T^\alpha = C\iota_T \equiv (\alpha_1, \alpha_2, \ldots, \alpha_T)'$ and $\theta_\alpha = 1 - (\sigma/\sigma_\alpha)$ with $\sigma_\alpha^2 = \sigma_\mu^2(\Sigma_{t=1}^T \alpha_t^2) + \sigma^2$. See Baltagi and Li (1994) for an extension to the MA(q) case and Galbraith and Zinde-Walsh (1995) for an extension to the ARMA(p, q) case.

5.2.5 Unequally Spaced Panels with AR(1) Disturbances

Some panel data sets cannot be collected in every period due to lack of resources or cut in funding. Instead, these panels are collected over unequally spaced time intervals. For example, a panel of households could be collected over unequally spaced years rather than annually. This is also likely when collecting data on countries, states, or firms where in certain years, the data are not recorded, are hard to obtain, or are simply missing. Other common examples are panel data sets using daily data from the stock market, including stock prices, commodity prices, futures, etc. These panel data sets are unequally spaced when the market closes on weekends and holidays. This is also common for housing resale data where the pattern of resales for each house occurs at different time periods, and the panel is unbalanced because we observe different number of resales for each house. Baltagi and Wu (1999) extend the Baltagi and Li (1991a) results to the estimation of an unequally spaced panel data regression model with AR(1) remainder disturbances. A feasible generalized least squares procedure is proposed as weighted least squares that can handle a wide range of unequally spaced panel data patterns. This procedure is simple to compute and provides natural estimates of the serial correlation and variance components parameters. Baltagi and Wu (1999) also provide a locally best invariant (*LBI*) test for zero first-order serial correlation against positive or negative serial correlation in case of unequally spaced panel data. Details are given in that paper. This is programmed in Stata under the (*xtregar, re lbi*) command. Table 5.1 gives the Stata output for Grunfeld's investment equation, given in (2.40), with random effects and an AR(1) remainder disturbance term. The bottom of Table 5.1 produces the Baltagi–Wu *LBI* statistic of 0.956 and the Bhargava, Franzini and Narendranathan (1982) Durbin–Watson statistic for zero first-order serial correlation described in (5.44) below. Both tests reject the null hypothesis of no first-order serial correlation. The estimate of ρ for the AR(1) remainder disturbances is 0.67 while $\widehat{\sigma}_\mu = 74.52$ and $\widehat{\sigma}_\nu = 41.48$. Note that $\widehat{\beta}_1$ in (2.41) drops from 0.110 for a typical random effects GLS estimator reported in Table 2.1 to 0.095 for the random effects GLS estimator with AR(1) remainder disturbances in Table 5.1. This is contrasted to an increase in $\widehat{\beta}_2$ from 0.308 in Table 2.1 to 0.320 in Table 2.5. Note that if we have missing data on say 1951 and 1952, Stata computes this unequally spaced panel estimation for the random effects with AR(1) disturbances. Table 5.2 reproduces this output. Note that it is based on 180 observations, due to the loss of two years of data for all 10 firms. The Baltagi–Wu *LBI* statistic is 1.139 and the Bhargava, Franzini and Narendranathan (1982) Durbin–Watson statistic is 0.807, exactly as reported in Table 1 of Baltagi and Wu (1999, p. 822). Both test statistics reject the null hypothesis of no first-order serial correlation. Problem 5.19 asks the reader to replicate these results for other patterns of missing observations.

5.2.6 Prediction

In Sect. 2.5, we derived Goldberger (1962) BLUP of $y_{i,T+S}$ for the one-way error component model without serial correlation. For ease of reference, we reproduce Eq. (2.37) for predicting one period ahead for the ith individual

$$\widehat{y}_{i,T+1} = Z'_{i,T+1}\widehat{\delta}_{GLS} + w'\Omega^{-1}\widehat{u}_{GLS} \tag{5.25}$$

where $\widehat{u}_{GLS} = y - Z\widehat{\delta}_{GLS}$ and $w = E(u_{i,T+1}u)$. For the AR(1) model with no error components, a standard result is that the last term in (5.25) reduces to $\rho\widehat{u}_{i,T}$, where $\widehat{u}_{i,T}$ is the Tth GLS residual for the ith individual. For the one-way error component model without serial correlation (see Taub 1979, or Sect. 2.5), the last term of (5.25) reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2 + \sigma_\nu^2)]\overline{\widehat{u}}_{i.}$, where $\overline{\widehat{u}}_{i.} = \sum_{t=1}^{T}\widehat{u}_{it}/T$ is the average of the ith individual's GLS residuals. This section summarizes the Baltagi and Li (1992) derivation of the last term of (5.25) when *both* error components and

Table 5.1 Grunfeld's data. Random effects and AR(1) remainder disturbances

```
. xtregar I F C , re lbi

RE GLS regression with AR(1) disturbances       Number of obs      =       200
Group variable (i): fn                          Number of groups   =        10

R-sq:  within  = 0.7649                         Obs per group: min =        20
       between = 0.8068                                        avg =      20.0
       overall = 0.7967                                        max =        20

                                                Wald chi2(3)       =    360.31
corr(u_i, Xb)      = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .0949215   .0082168    11.55   0.000     .0788168    .1110262
           C |   .3196589   .0258618    12.36   0.000     .2689707    .3703471
       _cons |  -44.38123   26.97525    -1.65   0.100    -97.25175    8.489292
-------------+----------------------------------------------------------------
      rho_ar |  .67210608   (estimated autocorrelation coefficient)
     sigma_u |  74.517098
     sigma_e |  41.482494
     rho_fov |   .7634186   (fraction of variance due to u_i)
       theta |  .67315699
------------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = .6844797
Baltagi-Wu LBI = .95635623
```

Table 5.2 Grunfeld's data. Unequally spaced panel

```
.  xtregar I F C if yr!=1951 & yr!= 1952 , re lbi

RE GLS regression with AR(1) disturbances       Number of obs      =       180
Group variable (i): fn                          Number of groups   =        10

R-sq:  within  = 0.7766                         Obs per group: min =        18
       between = 0.8112                                        avg =      18.0
       overall = 0.8024                                        max =        18

                                                Wald chi2(3)       =    341.38
corr(u_i, Xb)      = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .0919986   .0083459    11.02   0.000     .0756409    .1083563
           C |   .3243706   .0266376    12.18   0.000     .2721618    .3765793
       _cons |  -43.01923   27.05662    -1.59   0.112    -96.04924    10.01077
-------------+----------------------------------------------------------------
      rho_ar |  .68934342   (estimated autocorrelation coefficient)
     sigma_u |  74.002133
     sigma_e |  41.535675
     rho_fov |  .76043802   (fraction of variance due to u_i)
       theta |   .6551959
------------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = .80652308
Baltagi-Wu LBI = 1.1394026
```

serial correlation are present. This provides the applied researcher with a simple way of augmenting the GLS predictions obtained from the Fuller and Battese (1973) transformation described above.

For the one-way error component model with AR(1) remainder disturbances, considered in Sect. 5.2.1, Baltagi and Li (1992) find that

$$w'\Omega^{-1}\widehat{u}_{GLS} = \rho\widehat{u}_{i,T} + \left(\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\alpha^2}\right)\left[\alpha\hat{u}_{i1}^* + \sum_{t=2}^{T}\hat{u}_{it}^*\right] \tag{5.26}$$

Note that the first PW-transformed GLS residual receives an α weight in averaging across the ith individual's residuals in (5.26). (i) If $\sigma_\mu^2 = 0$, so that only serial correlation is present, (5.26) reduces to $\rho\widehat{u}_{i,T}$. Similarly, (ii) if $\rho = 0$, so that only error components are present, (5.26) reduces to $[T\sigma_\mu^2/(T\sigma_\mu^2+\sigma_\nu^2)]\bar{\widehat{u}}_{i.}$.

For the one-way error component model with remainder disturbances following an AR(2) process, considered in Sect. 5.2.2, Baltagi and Li (1992) find that

$$\begin{aligned} w'\Omega^{-1}\widehat{u}_{GLS} &= \rho_1\widehat{u}_{i,T-1} + \rho_2\widehat{u}_{i,T-2} \\ &\quad + \left[\frac{(1-\rho_1-\rho_2)^2\sigma_\mu^2}{\sigma_\alpha^2}\right]\left[\alpha_1\hat{u}_{i1}^* + \alpha_2\hat{u}_{i2}^* + \sum_{t=3}^{T}\hat{u}_{it}^*\right] \end{aligned} \tag{5.27}$$

where

$$\begin{aligned} \alpha_1 &= \sigma_\epsilon/\sigma_\nu(1-\rho_1-\rho_2)\alpha_2 = \sqrt{(1+\rho_2)/(1-\rho_2)} \\ \sigma_\alpha^2 &= d^2\sigma_\mu^2(1-\rho_1-\rho_2)^2 + \sigma_\epsilon^2 \\ d^2 &= \alpha_1^2 + \alpha_2^2 + (T-2) \end{aligned}$$

and

$$\begin{aligned} \hat{u}_{i1}^* &= (\sigma_\epsilon/\sigma_\nu)\widehat{u}_{i1} \\ \hat{u}_{i2}^* &= \sqrt{1-\rho_2^2}\,[\widehat{u}_{i2} - (\rho_1/(1-\rho_2))\widehat{u}_{i1}] \\ \hat{u}_{it}^* &= \widehat{u}_{it} - \rho_1\widehat{u}_{i,t-1} - \rho_2\widehat{u}_{i,t-2} \quad \text{for } t = 3,\ldots,T \end{aligned}$$

Note that if $\rho_2 = 0$, this predictor reduces to (5.26). Also, note that for this predictor, the first two residuals are weighted differently when averaging across the ith individual's residuals in (5.27).

For the one-way error component model with remainder disturbances following the specialized AR(4) process for quarterly data, considered in Sect. 5.2.3, Baltagi and Li (1992) find that

$$w'\Omega^{-1}\widehat{u}_{GLS} = \rho\widehat{u}_{i,T-3} + \left[\frac{(1-\rho)^2\sigma_\mu^2}{\sigma_\alpha^2}\right]\left[\alpha\sum_{t=1}^{4}\hat{u}_{it}^* + \sum_{t=5}^{T}\hat{u}_{it}^*\right] \tag{5.28}$$

where $\alpha = \sqrt{(1+\rho)/(1-\rho)}$, $\sigma_\alpha^2 = d^2(1-\rho)^2\sigma_\mu^2 + \sigma_\epsilon^2$, $d^2 = 4\alpha^2 + (T-4)$, and

$$\begin{aligned} u_{it}^* &= \sqrt{1-\rho^2}\, u_{it} \text{ for } t = 1, 2, 3, 4 \\ &= u_{it} - \rho u_{i,t-4} \quad \text{for } t = 5, 6, \ldots, T \end{aligned}$$

Note, for this predictor, that the first four quarterly residuals are weighted by α when averaging across the ith individual's residuals in (5.28).

Finally, for the one-way error component model with remainder disturbances following an MA(1) process, considered in Sect. 5.2.4, Baltagi and Li (1992) find that

$$w'\Omega^{-1}\widehat{u}_{GLS} = -\lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\hat{u}_{iT}^* + \left[1 + \lambda\left(\frac{a_{T-1}}{a_T}\right)^{1/2}\alpha_T\right]\left(\frac{\sigma_\mu^2}{\sigma_\alpha^2}\right)\left[\sum_{t=1}^{T}\alpha_t\hat{u}_{it}^*\right] \tag{5.29}$$

where the $\hat{u}_{it}^*$ can be solved for recursively as follows:

$$\begin{aligned} \hat{u}_{i1}^* &= (a_0/a_1)^{1/2}\hat{u}_{i1} \\ \hat{u}_{it}^* &= \lambda(a_{t-2}/a_t)^{1/2}\hat{u}_{i,t-1}^* + (a_{t-1}/a_t)^{1/2}\hat{u}_{i,t} \quad t = 2, \ldots, T \end{aligned}$$

If $\lambda = 0$, then from (5.23), $a_t = \alpha_t = 1$ for all t, and (5.29) reduces to the predictor for the error component model with no serial correlation. If $\sigma_\mu^2 = 0$, the second term in (5.29) drops out and the predictor reduces to that of the MA(1) process.

Frees and Miller (2004) forecast the sale of state lottery tickets using panel data from 50 postal (ZIP) codes in Wisconsin observed over 40 weeks. The first 35 weeks of data are used to estimate the model and the remaining five weeks are used to assess the validity of model forecasts. Using the mean absolute error criteria and the mean absolute percentage error criteria, the best forecasts were given by the error component model with AR(1) disturbances followed by the fixed effects model with AR(1) disturbances.

5.2.7 Testing for Serial Correlation and Individual Effects

In this section, we address the problem of *jointly* testing for serial correlation and individual effects. Baltagi and Li (1995) derived three LM statistics for an error component model with first-order serially correlated errors. The first LM statistic jointly tests for zero first-order serial correlation and random individual effects. The second LM statistic tests for zero first-order serial correlation assuming fixed individual effects, and the third LM statistic tests for zero first-order serial correlation assuming random individual effects. In all three cases, Baltagi and Li (1995) showed that the corresponding LM statistic is the *same* whether the alternative is AR(1) or MA(1). Also, Baltagi and Li (1995) derived two extensions of the Burke, Godfrey and Termayne (1990) AR(1) versus MA(1) test from the time-series to the panel data literature. The first extension tests the null of AR(1) disturbances against MA(1) disturbances, and the second the null of MA(1) disturbances against AR(1) disturbances in an error component model. These tests are computationally simple requiring only OLS or Within residuals. In what follows, we briefly review the basic ideas behind these tests.

Consider the panel data regression given in (2.3)

$$y_{it} = Z'_{it}\delta + u_{it} \quad i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T \tag{5.30}$$

where δ is a $(K + 1) \times 1$ vector of regression coefficients including the intercept. The disturbance follows a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{5.31}$$

where $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and the remainder disturbance follow a stationary AR(1) process, $\nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$ with $|\ \rho\ | < 1$, or an MA(1) Process, $\nu_{it} = \epsilon_{it} + \lambda\epsilon_{i,t-1}$ with $|\ \lambda\ | < 1$, and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$. In what follows, we will show that the joint LM test statistic for H_1^a: $\sigma_\mu^2 = 0; \lambda = 0$ is the same as that for H_1^b: $\sigma_\mu^2 = 0; \rho = 0$.

A Joint LM Test for Serial Correlation and Random Individual Effects

Let us consider the joint LM test for the error component model where the remainder disturbances follow an MA(1) process. In this case, the variance–covariance matrix of the disturbances is given by

$$\Omega = E(uu') = \sigma_\mu^2 I_N \otimes J_T + \sigma_\epsilon^2 I_N \otimes V_\lambda \tag{5.32}$$

where

$$V_\lambda = \begin{pmatrix} 1+\lambda^2 & \lambda & 0 \ldots & 0 \\ \lambda & 1+\lambda^2 & \lambda \ldots & 0 \\ \vdots & \vdots & \vdots \ddots & \vdots \\ 0 & 0 & 0 \ldots & 1+\lambda^2 \end{pmatrix} \tag{5.33}$$

and the log-likelihood function is given by $L(\delta, \theta)$ in (4.15) with $\theta = (\lambda, \sigma_\mu^2, \sigma_\epsilon^2)'$. In order to construct the LM test statistic for $H_1^a : \sigma_\mu^2 = 0; \lambda = 0$, one needs $D(\theta) = \partial L(\theta)/\partial\theta$ and the information matrix $J(\theta) = E[\partial^2 L(\theta)/\partial\theta\partial\theta']$ evaluated at the restricted maximum likelihood estimator $\widehat{\theta}$. Note that under the null hypothesis $\Omega^{-1} = (1/\sigma_\epsilon^2)I_{NT}$. Using the general Hemmerle and Hartley (1973) formula given in (4.17), one gets the scores

$$\begin{aligned} \partial L(\theta)/\partial\lambda &= NT \sum_{i=1}^{N}\sum_{t=2}^{T} \widehat{u}_{it}\widehat{u}_{i,t-1} \Big/ \sum_{i=1}^{N}\sum_{t=2}^{T} \widehat{u}_{it}^2 \equiv NT(\widehat{u}'\widehat{u}_{-1}/\widehat{u}'\widehat{u}) \\ \partial L(\theta)/\partial\sigma_\mu^2 &= -(NT/2\widehat{\sigma}_\epsilon^2)[1 - \widehat{u}'(I_N \otimes J_T)\widehat{u}/(\widehat{u}'\widehat{u})] \end{aligned} \tag{5.34}$$

where $\widehat{u}$ denotes the OLS residuals and $\widehat{\sigma}_\epsilon^2 = \widehat{u}'\widehat{u}/NT$. Using (4.19), see Harville (1977), one gets the information matrix

$$\widehat{J} = (NT/2\widehat{\sigma}_\epsilon^4)\begin{pmatrix} T & 2(T-1)\widehat{\sigma}_\epsilon^2/T & 1 \\ 2(T-1)\widehat{\sigma}_\epsilon^2/T & 2\widehat{\sigma}_\epsilon^4(T-1)/T & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{5.35}$$

Hence the LM statistic for the null hypothesis H_1^a: $\sigma_\mu^2 = 0$; $\lambda = 0$ is given by

$$LM_1 = \widehat{D}'\widehat{J}^{-1}\widehat{D} = \frac{NT^2}{2(T-1)(T-2)}[A^2 - 4AB + 2TB^2] \tag{5.36}$$

where $A = [\widehat{u}'(I_N \otimes J_T)\widehat{u}/(\widehat{u}'\widehat{u})] - 1$ and $B = (\widehat{u}'\widehat{u}_{-1}/\widehat{u}'\widehat{u})$. This is asymptotically distributed (for large N) as χ_2^2 under H_1^a.

It remains to show that LM_1 is exactly the same as the joint test statistic for $H_1^b : \sigma_\mu^2 = 0$; $\rho = 0$, where the remainder disturbances follow an AR(1) process (see Baltagi and Li 1991b). In fact, if we repeat the derivation given in (5.32)–(5.36), the only difference is to replace the V_λ matrix by its AR(1) counterpart

$$V_\rho = \begin{pmatrix} 1 & \rho & \ldots & \rho^{T-1} \\ \rho & 1 & \ldots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \ldots & 1 \end{pmatrix}$$

Note that under the null hypothesis, we have $(V_\rho)_{\rho=0} = I_T = (V_\lambda)_{\lambda=0}$ and

$$(\partial V_\rho/\partial\rho)_{\rho=0} = G = (\partial V_\lambda/\partial\lambda)_{\lambda=0}$$

where G is the bidiagonal matrix with bidiagonal elements all equal to one. Using these results, Problem 5.14 asks the reader to verify that the resulting joint LM test statistic is the same whether the residual disturbances follow an AR(1) or an MA(1) process. Hence, the joint LM test statistic for random individual effects and first-order serial correlation is independent of the form of serial correlation, whether it is AR(1) or MA(1). This extends the Breusch and Godfrey (1981) result from time-series regression to a panel data regression using an error component model.

Note that the A^2 term is the basis for the LM test statistic for H_2: $\sigma_\mu^2 = 0$ assuming there is no serial correlation (see Breusch and Pagan 1980). In fact, $\text{LM}_2 = \sqrt{NT/2(T-1)}A$ is asymptotically distributed (for large N) as $N(0, 1)$ under H_2 against the one-sided alternative H_2'; $\sigma_\mu^2 > 0$; see (4.25). Also, the B^2 term is the basis for the LM test statistic for $H_3 : \rho = 0$ (or $\lambda = 0$) assuming there are no individual effects (see Breusch and Godfrey 1981). In fact, $\text{LM}_3 = \sqrt{NT^2/(T-1)}B$ is asymptotically distributed (for large N) as $N(0, 1)$ under H_3 against the one-sided alternative H_3'; ρ (or λ) > 0. The presence of an interaction term in the joint LM test statistic, given in (5.36), emphasizes the importance of the joint test when both serial correlation and random individual effects are suspected. However, when T is large the interaction term becomes negligible.

Note that all the LM tests considered assume that the underlying null hypothesis is that of white noise disturbances. However, in panel data applications, especially with large labor panels, one is concerned with individual effects and is guaranteed their existence. In this case, it is inappropriate to test for serial correlation assuming no individual effects as is done in H_3. In fact, if one uses LM_3 to test for serial correlation, one is very likely to reject the null hypothesis of H_3 even if the null is true. This is because the μ_i are correlated for the same individual across time, and this will contribute to rejecting the null of no serial correlation.

An LM Test for First-order Serial Correlation in a Random Effects Model
Baltagi and Li (1995) also derived a conditional LM test for first-order serial correlation given the existence of random individual effects. In case of an AR(1) model, the null hypothesis is $H_4^b : \rho = 0$ (given $\sigma_\mu^2 > 0$) versus $H_4^{b\prime} : \rho \neq 0$ (given $\sigma_\mu^2 > 0$). The variance–covariance matrix (under the alternative) is

$$\Omega_1 = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\nu^2 (I_N \otimes V_\rho) \tag{5.37}$$

Under the null hypothesis H_4^b, we have

$$\begin{aligned}
(\Omega_1^{-1})_{\rho=0} &= (1/\sigma_\epsilon^2) I_N \otimes E_T + (1/\sigma_1^2) I_N \otimes \bar{J}_T \\
(\partial\Omega_1/\partial\rho) \mid_{\rho=0} &= \sigma_\epsilon^2 (I_N \otimes G) \\
(\partial\Omega_1/\partial\sigma_\mu^2) \mid_{\rho=0} &= (I_N \otimes J_T) \\
(\partial\Omega_1/\partial\sigma_\epsilon^2) \mid_{\rho=0} &= (I_N \otimes I_T)
\end{aligned}$$

where $\bar{J}_T = \iota_T \iota_T'/T$, $E_T = I_T - \bar{J}_T$, G is a bidiagonal matrix with bidiagonal elements all equal to one, and $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\epsilon^2$.

When the first-order serial correlation is of the MA(1) type, the null hypothesis becomes H_4^a, $\lambda = 0$ (given that $\sigma_\mu^2 > 0$) versus $H_4^{a\prime} : \lambda \neq 0$ (given that $\sigma_\mu^2 > 0$). In this case, the variance–covariance matrix is

$$\Omega_2 = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\epsilon^2 (I_N \otimes V_\lambda) \tag{5.38}$$

and under the null hypothesis H_4^a,

$$\begin{aligned}
(\Omega_2^{-1})_{\lambda=0} &= (1/\sigma_\epsilon^2)(I_N \otimes E_T) + (1/\sigma_1^2)(I_N \otimes \bar{J}_T) = (\Omega_1^{-1})_{\rho=0} \\
(\partial\Omega_2/\partial\lambda) \mid_{\lambda=0} &= \sigma_\epsilon^2 (I_N \otimes G) = (\partial\Omega_1/\partial\rho) \mid_{\rho=0} \\
(\partial\Omega_2/\partial\sigma_\mu^2) \mid_{\lambda=0} &= (I_N \otimes J_T) = (\partial\Omega_1/\partial\sigma_\mu^2) \mid_{\rho=0} \\
(\partial\Omega_2/\partial\sigma_\epsilon^2) \mid_{\lambda=0} &= (I_N \otimes I_T) = (\partial\Omega_1/\partial\sigma_\epsilon^2) \mid_{\rho=0}
\end{aligned}$$

Using these results, Problem 5.15 asks the reader to verify that the test statistic for H_4^a is the same as that for H_4^b. This conditional LM statistic, call it LM_4, is not given here but is derived in Baltagi and Li (1995).

To summarize, the conditional LM test statistics for testing first-order serial correlation, assuming random individual effects, are invariant to the form of serial correlation (i.e., whether it is AR(1) or MA(1)). Also, these conditional LM tests require restricted mle of a one-way error component model with random individual effects rather than OLS estimates as is usual with LM tests.

Bera, Sosa-Escudero and Yoon (2001) criticize this loss of simplicity in computation of LM tests that use OLS residuals and suggest an adjustment of these LM tests that are robust to local misspecification. Instead of $LM_\mu = NTA^2/2(T-1) = LM_2^2$ for testing $H_2; \sigma_\mu^2 = 0$ which ignores the possible presence of serial correlation, they suggest computing

$$LM_\mu^* = \frac{NT(2B-A)^2}{2(T-1)(1-(2/T))}$$

This test essentially modifies LM_μ by correcting the mean and variance of the score $\partial L/\partial \sigma_\mu^2$ for its asymptotic correlation with $\partial L/\partial \rho$. Under the null hypothesis, LM_μ^* is asymptotically distributed as χ_1^2. Under local misspecification, this adjusted test statistic is equivalent to Neyman's $C(\alpha)$ test and shares its optimality properties. Similarly, they suggest computing

$$LM_\rho^* = \frac{NT^2[B-(A/T)]^2}{(T-1)(1-(2/T))}$$

instead of $LM_\rho = NT^2B^2/(T-1) = LM_3^2$ to test $H_3; \rho = 0$, against the alternative that $\rho \neq 0$, ignoring the presence of random individual effects. They also show that

$$LM_\mu^* + LM_\rho = LM_\rho^* + LM_\mu = LM_1$$

where LM_1 is the joint LM test given in (5.36). In other words, the two-directional LM test for σ_μ^2 and ρ can be decomposed into the sum of the *adjusted* one-directional test of one type of alternative and the *unadjusted* form of the other hypothesis. Bera, Sosa-Escudero and Yoon (2001) argue that these tests use only OLS residuals and are easier to compute than the conditional LM tests derived by Baltagi and Li (1995). Bera, Sosa-Escudero and Yoon (2001) perform Monte Carlo experiments that show the usefulness of these modified Rao's Score tests in guarding against *local* misspecification.

For the Grunfeld data, we computed the Breusch and Pagan (1980) test for random effects H_2: $\sigma_\mu^2 = 0$. This yielded $LM_\mu = 798.162$ (in Table 4.2) using *xttest0* after running (*xtreg, re*). Using the Stata command *xttest1* after performing (*xtreg, re*), one can generate $LM_\rho = 143.523$,to test for H_3; $\rho = 0$. The joint LM_1 statistic for $H_1^b : \sigma_\mu^2 = 0; \rho = 0$ is 808.471. This is done in Table 5.3. The Bera, Sosa-Escudero and Yoon (2001) LM tests that are robust to local misspecifications LM_μ^* and LM_ρ^* are also reported. The joint test rejects the null of no first- order serial correlation and no random firm effects. The one-directional tests LM_ρ and LM_ρ^* reject the null of no first-order serial correlation, while the one-directional tests LM_μ and LM_μ^* reject the null of no random firm effects.

Table 5.3 Grunfeld's data. Joint test for random effects and AR(1) remainder disturbances

```
. xttest1, unadjusted

Tests for the error component model:

        I[fn,t] = Xb + u[fn] + v[fn,t]
           v[fn,t] = lambda v[fn,(t-1)] + e[fn,t]

        Estimated results:
                          |       Var      sd = sqrt(Var)
                 ---------+-----------------------------
                        I |   47034.89       216.8753
                        e |   2784.458       52.767964
                        u |     7089.8       84.20095

        Tests:
           Random Effects, Two Sided:
           LM(Var(u)=0)           =   798.16 Pr>chi2(1) =   0.0000
           ALM(Var(u)=0)          =   664.95 Pr>chi2(1) =   0.0000

           Random Effects, One Sided:
           LM(Var(u)=0)           =    28.25 Pr>N(0,1)  =   0.0000
           ALM(Var(u)=0)          =    25.79 Pr>N(0,1)  =   0.0000

           Serial Correlation:
           LM(lambda=0)           =   143.52 Pr>chi2(1) =   0.0000
           ALM(lambda=0)          =    10.31 Pr>chi2(1) =   0.0013

           Joint Test:
           LM(Var(u)=0,lambda=0)  =   808.47 Pr>chi2(2) =   0.0000
```

An LM Test for First-order Serial Correlation in a Fixed Effects Model

The model is the same as (5.30), and the null hypothesis is H_5^b; $\rho = 0$ given that the μ_i are fixed parameters. Writing each individual's variables in a $T \times 1$ vector form, we have

$$y_i = Z_i\delta + \mu_i\iota_T + \nu_i \tag{5.39}$$

where $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, Z_i is $T \times (K+1)$ and ν_i is $T \times 1$. $\nu_i \sim N(0, \Omega_\rho)$ where $\Omega_\rho = \sigma_\epsilon^2 V_\rho$ for the AR(1) disturbances. The log-likelihood function is

$$L(\delta, \rho, \mu, \sigma_\epsilon^2) = \text{ constant } - \frac{1}{2}\log \mid \Omega \mid \tag{5.40}$$

$$-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N}[(y_i - Z_i\delta - \mu_i\iota_T)'V_\rho^{-1}(y_i - Z_i\delta - \mu_i\iota_T)]$$

where $\Omega = I_N \otimes \Omega_\rho$ is the variance–covariance matrix of $\nu' = (\nu_1', \ldots, \nu_N')$. One can easily check that the maximum likelihood estimator of μ_i is given by $\widehat{\mu}_i = \{(\iota_T' V_\rho^{-1}\iota_T)^{-1}[\iota_T' V_\rho^{-1}(y_i - Z_i\widehat{\delta})]\}_{\rho=0} = \bar{y}_{i.} - \bar{Z}_{i.}'\widehat{\delta}$, where $\widehat{\delta}$ is the maximum likelihood estimator of δ, $\bar{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$, and $\bar{Z}_{i.}$ is a $(K+1) \times 1$ vector of averages of Z_{it} across time.

Write the log-likelihood function in vector form of ν as

$$L(\delta, \mu, \theta) = \text{constant} - \frac{1}{2}\log \mid \Omega \mid -\frac{1}{2}\nu'\Omega^{-1}\nu \tag{5.41}$$

where $\theta' = (\rho, \sigma_\epsilon^2)$. Now (5.41) has a similar form to (4.15). By following a similar derivation as that given earlier, one can easily verify that the LM test statistic for testing H_5^b is

$$LM = [NT^2/(T-1)](\widehat{\nu}'\widehat{\nu}_{-1}/\widehat{\nu}'\widehat{\nu})^2 \tag{5.42}$$

which is asymptotically distributed (for large T) as χ_1^2 under the null hypothesis H_5^b. Note that $\widehat{\nu}_{it} = y_{it} - Z'_{it}\widehat{\delta} - \widehat{\mu}_i = (\widetilde{y}_{it} - \widetilde{Z}'_{it}\widehat{\delta}) + (\bar{y}_{i.} - \bar{Z}'_{i.}\widehat{\delta} - \widehat{\mu}_i)$ where $\widetilde{y}_{it} = y_{it} - \bar{y}_{i.}$ is the usual Within transformation. Under the null of $\rho = 0$, the last term in parentheses is zero since $\{\widehat{\mu}_i\}_{\rho=0} = \bar{y}_{i.} - \bar{Z}'_{i.}\widehat{\delta}$ and $\{\widehat{\nu}_{it}\}_{\rho=0} = \widetilde{y}_{it} - \widetilde{Z}_{it}\widehat{\delta} = \widetilde{\nu}_{it}$. Therefore, the LM statistic given in (5.42) can be expressed in terms of the usual Within residuals (the $\widetilde{\nu}$), and the one-sided test for H_5^b (corresponding to the alternative $\rho > 0$) is

$$LM_5 = \sqrt{NT^2/(T-1)}(\widetilde{\nu}'\widetilde{\nu}_{-1}/\widetilde{\nu}'\widetilde{\nu}) \tag{5.43}$$

This is asymptotically distributed (for large T) as $N(0, 1)$.

By a similar argument, one can show that the LM test statistic for $H_5^a : \lambda = 0$, in a fixed effects model with MA(1) residual disturbances, is identical to LM_5.

Note also that LM_5 differs from LM_3 only by the fact that the Within residuals $\widetilde{\nu}$ (in LM_5) replace the OLS residuals $\widehat{u}$ (in LM_3). Since the Within transformation wipes out the individual effects whether fixed or random, one can also use (5.43) to test for serial correlation in the random effects models.

The Durbin–Watson Statistic for Panel Data

For the fixed effects model described in (5.39) with ν_{it} following an AR(1) process, Bhargava, Franzini and Narendranathan (1982), hereafter BFN, suggested testing for H_0; $\rho = 0$ against the alternative that $|\rho| < 1$, using the Durbin–Watson statistic only based on the Within residuals (the $\widetilde{\nu}_{it}$) rather than OLS residuals:

$$d_p = \sum_{i=1}^{N}\sum_{t=2}^{T}(\widetilde{\nu}_{it} - \widetilde{\nu}_{i,t-1})^2 / \sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\nu}_{it}^2 \tag{5.44}$$

BFN showed that for arbitrary regressors, d_p is a locally most powerful invariant test in the neighborhood of $\rho = 0$. They argued that exact critical values using the Imhof routine are both impractical and unnecessary for panel data since they involve the computation of the nonzero eigenvalues of a large $NT \times NT$ matrix. Instead, BFN tabulate upper and lower bounds of d_p at the 5% levels for $N = 50, 100, 150, 250, 500, 1000$, $T = 6, 10$, and $k = 1, 3, 5, 7, 9, 11, 13, 15$. BFN remark that d_p would be rarely inconclusive since the bounds will be very tight even for moderate values of N. Also, for very large N, BFN argue that it is not necessary to compute these bounds, but simply test whether d_p is less than two when testing against positive serial correlation. Stata computes this DW statistic when using the *xtregar* command as demonstrated in Sect. 5.2.1 and Tables 5.1 and 5.2. However, no critical value or p-value is provided and these are not tabulated for unbalanced panels.

Inoue and Solon (2006) propose a portmanteau test for serially correlated errors in a fixed effects model. This tests the null hypothesis of no serial correlation between any two periods against a general alternative that at least some of the autocorrelations are nonzero. This test is attractive for short T, but will lack power as T gets large because the dimension of the null hypothesis is $(T-1)(T-2)/2$ which grows with T^2. This test needs N to be much larger than $T^2/2$. This test can be implemented with Stata with the command *xtistest*.

Testing AR(1) Against MA(1) in an Error Component Model

Testing AR(1) against MA(1) has been extensively studied in the time-series literature; see King and McAleer (1987) for a Monte Carlo comparison of non-nested, approximate point optimal, as well as LM tests.[3] In fact, King and McAleer (1987) found that the non-nested tests perform poorly in small samples, while King (1983) points optimal test performs the best. Burke, Godfrey and Termayne (1990) (hereafter BGT) derived a simple test to distinguish between AR(1) and MA(1) processes. Baltagi and Li (1995) proposed two extensions of the BGT test to the error component model. These tests are simple to implement requiring Within or OLS residuals.

The basic idea of the BGT test is as follows: under the null hypothesis of an AR(1) process, the remainder error term υ_{it} satisfies

$$\text{Correl}\ (\nu_{it}, \nu_{i,t-\tau}) = \rho_\tau = (\rho_1)^\tau \quad \tau = 1, 2, \ldots \tag{5.45}$$

Therefore, under the null hypothesis

$$\rho_2 - (\rho_1)^2 = 0 \tag{5.46}$$

Under the alternative hypothesis of an MA(1) process on ν_{it}, $\rho_2 = 0$ and hence $\rho_2 - (\rho_1)^2 < 0$. Therefore, BGT recommend a test statistic based on (5.46) using estimates of ρ obtained from OLS residuals. One problem remains. King (1983) suggests that any "good" test should have a size which tends to zero, asymptotically, for $\rho > 0.5$. The test based on (5.46) does not guarantee this property. To remedy this, BGT proposed supplementing (5.46) with the decision to accept the null hypothesis of AR(1) if $\widehat{\rho}_1 > \frac{1}{2} + 1/\sqrt{T}$.

In an error component model, the Within transformation wipes out the individual effects, and one can use the Within residuals of $\widetilde{u}_{it} (= \widetilde{\nu}_{it})$ instead of OLS residuals $\widehat{u}_{it}$ to construct the BGT test. Let

$$(\widetilde{\rho}_1)_i = \sum_{t=2}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-1} / \sum_{t=1}^{T} \widetilde{u}_{it}^2 \tag{5.47}$$

and

$$(\widetilde{\rho}_2)_i = \sum_{t=3}^{T} \widetilde{u}_{it}\widetilde{u}_{i,t-2} / \sum_{t=1}^{T} \widetilde{u}_{it}^2 \quad \text{for } i = 1, \ldots, N \tag{5.48}$$

The following test statistic, based on (5.48),

$$\widetilde{\gamma}_i = \sqrt{T}[(\widetilde{\rho}_2)_i - (\widetilde{\rho}_1^2)_i]/[1 - (\widetilde{\rho}_2)_i] \tag{5.49}$$

is asymptotically distributed (for large T) as $N(0, 1)$ under the null hypothesis of an AR(1). Using the data on all N individuals, we can construct a generalized BGT test statistic for the error component model

$$\widetilde{\gamma} = \sqrt{N}\left(\sum_{i=1}^{N}\widetilde{\gamma}_i/N\right) = \sqrt{NT}\sum_{i=1}^{N}\left[\frac{(\widetilde{\rho}_2)_i - (\widetilde{\rho}_1^2)_i}{1-(\widetilde{\rho}_2)_i}\right]/N \tag{5.50}$$

$\widetilde{\gamma}_i$ are independent for different i since the $\widetilde{u}_i$ are independent. Hence $\widetilde{\gamma}$ is also asymptotically distributed (for large T) as $N(0, 1)$ under the null hypothesis of an AR(1) process. The test statistic (5.50) is supplemented by

$$\widetilde{r}_1 = \sum_{i=1}^{N}(\widetilde{r}_1)_i/N \equiv \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=2}^{T}\widetilde{u}_{it}\widetilde{u}_{i,t-1}/\sum_{t=1}^{T}\widetilde{u}_{it}^2\right] \tag{5.51}$$

and the Baltagi and Li (1995) proposed BGT_1 test can be summarized as follows:

(1) Use the Within residuals $\widetilde{u}_{it}$ to calculate $\widetilde{\gamma}$ and $\widetilde{r}_1$ from (5.50) and (5.51).
(2) Accept the AR(1) model if $\widetilde{\gamma} > c_\alpha$, or $\widetilde{r}_1 > \frac{1}{2} + 1/\sqrt{T}$, where $\Pr[N(0, 1) \leqslant c_\alpha] = \alpha$.

The bias in estimating $\rho_s (s = 1, 2)$ by using Within residuals is of $O(1/T)$ as $N \to \infty$ (see Nickell 1981). Therefore, BGT_1 may not perform well for small T. Since for typical labor panels, N is large and T is small, it would be desirable if an alternative simple test can be derived which performs well for large N rather than large T. In the next section we will give such a test.

An Alternative BGT Type Test for Testing AR(1) versus MA(1)

Let the null hypothesis be H_7: $\nu_{it} = \epsilon_{it} + \lambda\epsilon_{i,t-1}$ and the alternative be H_7' : $\nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$, where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$. Note that this test differs from the BGT_1 test in that the null hypothesis is MA(1) rather than AR(1). The alternative BGT type test uses autocorrelation estimates derived from OLS residuals and can be motivated as follows. Let

$$Q_0 = \frac{\sum\sum u_{it}^2}{NT} = u'u/NT \text{ and}$$

$$Q_s = \frac{\sum\sum u_{it}u_{i,t-s}}{N(T-s)} = u'(I_N \otimes G_s)u/N(T-s) \text{ for } s = 1, \ldots, S$$

where $G_s = \frac{1}{2}\text{Toeplitz}(\iota_s)$, ι_s is a vector of zeros with the $(s+1)$th element being one. $s = 1, \ldots, S$ with $S \leqslant (T-1)$ and S is finite.[2] Given the true residuals (the u), and assuming

$$\left[\frac{u'Au}{n} - E\left(\frac{u'Au}{n}\right)\right] \xrightarrow{P} 0$$

where $n = NT$ and A is an arbitrary symmetric matrix, Baltagi and Li (1995) proved the following results, as $N \to \infty$:
(1) For the MA(1) model

$$\begin{aligned}
\text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\nu^2 = \sigma_\mu^2 + \sigma_\epsilon^2(1+\lambda^2) \\
\text{plim } Q_1 &= \sigma_\mu^2 + \lambda\sigma_\epsilon^2 \\
\text{plim } Q_s &= \sigma_\mu^2 \quad \text{for } s = 2, \ldots, S
\end{aligned} \tag{5.52}$$

(2) For the AR(1) model

$$\begin{aligned}
\text{plim } Q_0 &= \sigma_\mu^2 + \sigma_\nu^2 \\
\text{plim } Q_s &= \sigma_\mu^2 + \rho^s\sigma_\nu^2 \quad \text{for } s = 1, \ldots, S
\end{aligned} \tag{5.53}$$

See Problem 5.17. Baltagi and Li (1995) showed that for large N, one can distinguish the AR(1) process from the MA(1) process based on the information obtained from $Q_s - Q_{s+l}$, for $s \geqslant 2$ and $l \geqslant 1$. To see this, note that $\text{plim}(Q_s - Q_{s+l}) = 0$ for the MA(1) process and $\text{plim}(Q_s - Q_{s+l}) = \sigma_\nu^2\rho^s(1-\rho^l) > 0$ for the AR(1) process.

Hence, Baltagi and Li (1995) suggest an asymptotic test of H_7 against H_7' based upon

$$\gamma = \sqrt{N/V}(Q_2 - Q_3) \tag{5.54}$$

where $V = 2tr\{[(\sigma_\mu^2 J_T + \sigma_\epsilon^2 V_\lambda)(G_2/(T-2) - G_3/(T-3))]^2\}$. Under some regularity conditions, γ is asymptotically distributed (for large N) as $N(0,1)$ under the null hypothesis of an MA(1) process.[3] In order to calculate V, we note that for the MA(1) process, $\sigma_\nu^2 = \sigma_\epsilon^2(1+\lambda^2)$ and $\sigma_\epsilon^2 V_\lambda = \sigma_\upsilon^2 I_T + \sigma_\epsilon^2\lambda G$. Therefore, we do not need to estimate λ in order to compute the test statistic γ; all we need to get are some consistent estimators for σ_ν^2, $\lambda\sigma_\epsilon^2$, and σ_μ^2. These are obtained as follows:

$$\begin{aligned}
\widehat{\sigma}_\upsilon^2 &= \widehat{Q}_0 - \widehat{Q}_2 \\
\lambda\widehat{\sigma}_\epsilon^2 &= \widehat{Q}_0 - \widehat{Q}_1 \\
\widehat{\sigma}_\mu^2 &= \widehat{Q}_2
\end{aligned}$$

where $\widehat{Q}_s$ are obtained from Q_s by replacing u_{it} by the OLS residuals $\widehat{u}_{it}$. Substituting these consistent estimators into V we get $\widehat{V}$, and the test statistic γ becomes

$$\widehat{\gamma} = \sqrt{N/\widehat{V}}(\widehat{Q}_2 - \widehat{Q}_3) \tag{5.55}$$

where

$$(\widehat{Q}_2 - \widehat{Q}_3) = \sum_{i=1}^{N}\sum_{t=3}^{N} \widehat{u}_{it}\widehat{u}_{i,t-2}/N(T-2) - \sum_{i=1}^{N}\sum_{t=4}^{T} \widehat{u}_{it}\widehat{u}_{i,t-3}/N(T-3)$$

and

$$\widehat{V} = 2\text{tr}\{[(\widehat{\sigma}_\mu^2 J_T + \widehat{\sigma}_\nu^2 I_T + \sigma_\epsilon^2 \widehat{\lambda} G)/(G_2/(T-2) + G_3/(T-3))]^2\}$$

$\widehat{\gamma}$ is asymptotically distributed (for large N) as $N(0, 1)$ under the null hypothesis H_7 and is referred to as the BGT_2 test.

Baltagi and Li (1995) perform extensive Monte Carlo experiments using the regression model setup considered in Chap. 4. However, the remainder disturbances are now allowed to follow the AR(1) or MA(1) process. Table 5.4 gives a summary of all tests considered. Their main results can be summarized as follows.

(1) The joint LM_1 test performs well in testing the null of H_1: $\rho = \sigma_\mu^2 = 0$. Its estimated size is not statistically different from its nominal size. Let $\omega = \sigma_\mu^2/\sigma^2$ denote the proportion of the total variance that is due to individual effects. Baltagi and Li (1995) find that in the presence of large individual effects ($\omega > 0.2$), or high serial correlation, ρ (or λ) > 0.2, LM_1 has high power rejecting the null in $99 - 100\%$ of the cases. It only has low power when $\omega = 0$ and ρ (or λ) $= 0.2$, or when $\omega = 0.2$ and ρ (or λ) $= 0$.

(2) The test statistic LM_2 for testing H_2: $\sigma_\mu^2 = 0$ implicitly assumes that ρ (or λ) $= 0$. When ρ is indeed equal to zero, this test performs well. However, as ρ moves away from zero and increases, this test tends to be biased in favor of rejecting the null. This is because a large serial correlation coefficient (i.e., large ρ) contributes to a large correlation among the individuals in the sample, even though $\sigma_\mu^2 = 0$. For example, when the null is true ($\sigma_\mu^2 = 0$) but $\rho = 0.9$, LM_2 rejects in 100% of the cases. Similar results are obtained in case ν_{it} follows an MA(1) process. In general, the presence of positive serial correlation tends to bias the case in favor of finding nonzero individual effects.

(3) Similarly, the LM_3 test for testing H_3: $\rho = 0$ implicitly assumes $\sigma_\mu^2 = 0$. This test performs well when $\sigma_\mu^2 = 0$. However, as σ_μ^2 increases, the performance of this test deteriorates. For example, when the null is true ($\rho = 0$) but $\omega = 0.9$, LM_3 rejects the null hypothesis in 100% of the cases. The large correlation among the μ_i contributes to the rejection of null hypothesis of no serial correlation. These results strongly indicate that one should not ignore the individual effects when testing for serial correlation.

(4) In contrast to LM_3, both LM_4 and LM_5 take into account the presence of individual effects. For large values of ρ or λ (greater than 0.4), both LM_4 and LM_5 have high power, rejecting the null more than 99% of the time. However, the estimated size of LM_4 is closer to the 5% nominal value than that of LM_5. In addition, Baltagi and Li (1995) show that Bhargava, Franzini and Narendranathan (1982) modified Durbin–Watson performs better than LM_5 and is recommended.

(5) The BGT_1 test, which uses Within residuals and tests the null of an AR(1) against the alternative of an MA(1) performs well if $T \geqslant 60$ and $T > N$. However,

Table 5.4 Testing for serial correlation and individual effects

	Null hypothesis H_0	Alternative hypothesis H_A	Test statistics	Asymptotic distribution under H_0
1a.	$H_1^a : \sigma_\mu^2 = 0; \lambda = 0$	σ_μ^2 or $\lambda \neq 0$	LM_1	χ_2^2
1b.	$H_1^b : \sigma_\mu^2 = 0; \rho = 0$	σ_μ^2 or $\rho \neq 0$	LM_1	χ_2^2
2.	$H_2 : \sigma_\mu^2 = 0$	$\sigma_\mu^2 > 0$	LM_2	$N(0, 1)$
3a.	$H_3^a : \lambda = 0$	$\lambda > 0$	LM_3	$N(0, 1)$
3b.	$H_3^b : \rho = 0$	$\rho > 0$	LM_3	$N(0, 1)$
4a.	$H_4^a : \lambda = 0\,(\sigma_\mu^2 > 0)$	$\lambda > 0\,(\sigma_\mu^2 > 0)$	LM_4	$N(0, 1)$
4b.	$H_4^b : \rho = 0\,(\sigma_\mu^2 > 0)$	$\rho > 0\,(\sigma_\mu^2 > 0)$	LM_4	$N(0, 1)$
5a.	$H_5^a : \lambda = 0\,(\mu_i \text{ fixed})$	$\lambda > 0\,(\mu_i \text{ fixed})$	LM_5	$N(0, 1)$
5b.	$H_5^b : \rho = 0\,(\mu_i \text{ fixed})$	$\rho > 0\,(\mu_i \text{ fixed})$	LM_5	$N(0, 1)$
6	H_6 : AR(1)	MA(1)	BGT_1	$N(0, 1)$
7	H_7 : MA(1)	AR(1)	BGT_2	$N(0, 1)$

Source Baltagi and Li (1995). Reproduced by permission of Elsevier Science Publishers BN (North Holland)

when T is small, or T is of moderate size but N is large, BGT_1 will tend to over-reject the null hypothesis. Therefore BGT_1 is not recommended for these cases. For typical labor panels, N is large and T is small. For these cases, Baltagi and Li (1995) recommend the BGT_2 test, which uses OLS residuals and tests the null of an MA(1) against the alternative of an AR(1). This test performs well when N is large and does not rely on T to achieve its asymptotic distribution. The Monte Carlo results show that BGT_2's performance improves as either N or T increases.

Baltagi and Li (1997) perform Monte Carlo experiments to compare the finite sample relative efficiency of a number of pure and *pretest* estimators for an error component model with remainder disturbances that are generated by an AR(1) or an MA(1) process. These estimators are (1) Ordinary Least Squares (OLS); (2) the Within estimator; (3) Conventional GLS which ignores the serial correlation in the remainder disturbances but accounts for the random error components structure. This is denoted by CGLS. (4) GLS assuming random error components with the remainder disturbances following an MA(1) process. This is denoted by GLSM. (5) GLS assuming random error components with the remainder disturbances following an AR(1) process. This is denoted by GLSA. (6) A pretest estimator which is based on the results of two tests. This is denoted by PRE. The first test is LM_4 which tests for the presence of serial correlation given the existence of random individual effects. If the null is not rejected, this estimator reduces to conventional GLS. In case serial correlation is found, the BGT_2 test is performed to distinguish between the AR(1) and MA(1) process and GLSA or GLSM is performed. (7) A Generalized Method of Moments (GMM) estimator, where the error component structure of the disturbances is ignored and a general variance–covariance matrix is estimated across the time dimension. Finally, (8) True GLS which is denoted by TGLS is obtained for comparison purposes. In fact, the relative efficiency of each estimator is obtained by dividing its MSE by that of TGLS. It is important to emphasize that all the estimators considered are consistent as long as the explanatory variables and the disturbances are uncorrelated, as $N \to \infty$, with T fixed. The primary concern here is with their small sample properties. The results show that the correct GLS procedure is always the best, but the researcher does not have perfect foresight on which one it is: GLSA for an AR(1) process, or GLSM for an MA(1) process. In this case, the pretest estimator is a viable alternative given that its performance is a close second to correct GLS whether the true serial correlation process is AR(1) or MA(1).

5.3 Time-Wise Autocorrelated and Cross-Sectionally Heteroskedastic Panel Regression

An alternative method for dealing with time-wise autocorrelated and cross-sectionally heteroskedastic disturbances in a panel regression model is described in Kmenta (1986). The basic idea is to allow for first-order autoregressive disturbances in Eq. (2.1) that follow a simple AR(1) process

$$u_{it} = \rho_i u_{i,t-1} + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{5.56}$$

where the autoregressive parameter can vary across cross-sections with $\mid \rho_i \mid < 1$. Also, the remainder error ϵ_{it} is assumed to be normal with zero mean and a general variance–covariance matrix that allows for possible heteroskedasticity as well as correlation across cross-sections, i.e.,

$$E(\epsilon\epsilon') = \Sigma \otimes I_T \quad \text{where} \epsilon' = (\epsilon_{11}, \ldots, \epsilon_{1T}, \ldots, \epsilon_{N1}, \ldots, \epsilon_{NT}) \tag{5.57}$$

and Σ is $N \times N$. The initial values of the disturbances are assumed to have the following properties:

$$u_{i0} \sim N\left(0, \frac{\sigma_{ii}}{1-\rho_i^2}\right) \quad \text{and } E(u_{i0}u_{j0}) = \frac{\sigma_{ij}}{1-\rho_i\rho_j} \qquad i, j = 1, 2, \ldots, N$$

Two special cases of this general specification are also considered. The first special case assumes that there is no correlation across different cross-sections (i.e., $\sigma_{ij} = 0$ for $i \neq j$), but there is heteroskedasticity (i.e., Σ is diagonal). The second special case assumes that Σ is diagonal but uses the additional restriction that *all* the ρ_i are equal to ρ for $i = 1, 2, \ldots, N$. The exogeneity assumption on the regressors renders OLS unbiased and consistent for this model. Hence, the OLS estimates can be used to estimate the ρ_i's and Σ. In fact, Beck and Katz (1995) use least squares residuals to obtain robust estimates of the variance–covariance matrix of OLS. This can be done using the *xtpcse* command in Stata. For the general variance–covariance structure given in (5.57), Kmenta (1986) describes how to obtain feasible GLS estimators of the regression coefficients. In the first step, OLS residuals are used to get consistent estimates of the ρ_i. Next, a Prais–Winsten transformation is applied using the estimated $\widehat{\rho}_i$ to get a consistent estimate of Σ from the resulting residuals. In the last step, GLS is applied to the Prais–Winsten transformed model using the consistent estimate of Σ. This can be done using the *xtgls* command in Stata. This

Table 5.5 Common rho and heteroskedastic AR(1) for Grunfeld data

```
. xtgls I F C, corr(ar1) panels(heteroskedastic)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   common AR(1) coefficient for all panels  (0.9261)

Estimated covariances      =        10          Number of obs      =       200
Estimated autocorrelations =         1          Number of groups   =        10
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    107.43
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .0715306   .0087269     8.20   0.000     .0544262     .088635
           C |   .1405652   .0314945     4.46   0.000     .0788371    .2022933
       _cons |  -1.979683   6.781349    -0.29   0.770    -15.27088    11.31152
------------------------------------------------------------------------------
```

may be a suitable pooling method for N small and T very large, but for typical labor or consumer panels where N is large and T is small it may be *infeasible*. In fact, for $N > T$, estimate of Σ will be *singular*. Note that the number of extra parameters to be estimated for this model is $N(N+1)/2$ corresponding to the elements of Σ plus N distinct ρ_i. This is in contrast to the simple one-way error component model with N extra parameters to estimate for the fixed effects model or two extra variance components to estimate for the random effects model. For example, even for a small $N = 50$, the number of extra parameters to estimate for the Kmenta technique is 1325 compared to 50 for fixed effects and two for the random effects model. Baltagi (1986) discusses the advantages and disadvantages of the Kmenta and the error components methods and compares their performance using Monte Carlo experiments. For typical panels with N large and T small, the error component model is parsimonious in its estimation of variance–covariance parameters compared to the time-wise autocorrelated, cross-sectionally heteroskedastic specification and is found to be more robust to misspecification.

Table 5.5 gives the estimation results for the Grunfeld investment equation given in (2.40) using the command *xtgls* in Stata. This performs AR(1) estimation with *common rho*, allowing for *heteroskedasticity* across firms, but no cross-firm correlation. One common rho, and 10 different variances are estimated, one for each firm. Table 5.6 performs AR(1) estimation with *a different rho* for each firm, allowing for *heteroskedasticity* across firms, but no cross-firm correlation. Ten different rhos, and 10 different variances are estimated, one for each firm. Table 5.7 performs AR(1) estimation with *a different rho* for each firm, allowing for *heteroskedasticity* across firms, *and cross-firm correlation*. Ten different rhos, and 55 elements of the variance–covariance matrix are estimated. This is not a parsimonious model and should *only* be used if N is very small compared to T. The estimates and their standard errors do change when compared to OLS in Table 2.1.

Table 5.6 Varying rhos and heteroskedastic AR(1) for Grunfeld data

```
. xtgls I F C, corr(psar1) panels(heteroskedastic)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =        10          Number of obs      =       200
Estimated autocorrelations =        10          Number of groups   =        10
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    100.13
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |    .071577     .00818     8.75   0.000     .0555444    .0876096
           C |    .165713   .0336642     4.92   0.000     .0997325    .2316935
       _cons |   5.663732   9.870044     0.57   0.566     -13.6812    25.00866
------------------------------------------------------------------------------
```

Table 5.7 Varying rhos and cross-section dependence AR(1) for Grunfeld data

```
. xtgls I F C, corr(psar1) panels(correlated)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   panel-specific AR(1)

Estimated covariances      =        55          Number of obs      =       200
Estimated autocorrelations =        10          Number of groups   =        10
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    326.73
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .0846195   .0056835    14.89   0.000       .07348     .095759
           C |    .245775   .0214979    11.43   0.000       .20364    .2879101
       _cons |    -8.8195   6.304894    -1.40   0.162    -21.17686    3.537865
------------------------------------------------------------------------------
```

Empirical Example: Nickell, Nunziata and Ochel (2005) estimate a dynamic unemployment rate equation using a panel of 20 OECD countries over the period 1961–95. Their results indicate that broad movements in unemployment across the OECD can be explained by shifts in labor market institutions. The dependent variable is ur_{it}, the unemployment rate in country i at time t. This is regressed on the lagged unemployment rate $ur_{i,t-1}$ as well as measures of labor institutions, like the employment protection legislation index (ep), the unemployment benefit replacement rate (br), unemployment benefit duration (bd), labor union density (ud), coordination in wage setting (co), and the tax wedge (tw). Also, the economic shocks are represented by the labor demand shock (lds), the TFP shock ($tfphpc$), the import price shock (tts), the acceleration in money supply ($d2ms$), and the real interest rate

($rirl$). All regressions included *year* and *country* dummies as well as country-specific time trends. In addition, these regressions included interaction terms where the variables were expressed as deviations from their overall means. Table 5.8 replicates column 1 of Table 5 of Nickell, Nunziata and Ochel (2005, p. 14) using iterative generalized least squares allowing for heteroskedastic errors and country-specific first-order serial correlation. Strictly speaking, this approach applies to static panel models with no lagged dependent variable. Yet the authors use it for the estimation of a dynamic panel model with serial correlation. See Chap. 8 for the proper estimation of a dynamic panel data model. The data set and variable definitions as well as Stata program output are provided on the LSE website as an attached ZIP file for an earlier working paper version of this paper (WP0502) http://cep.lse.ac.uk/_new/publications/series.asp?prog=CEP.

5.4 Further Reading

Hansen (2007) suggests a GLS estimator for a fixed effects panel model where the disturbances follow an AR(p) process. Baltagi and Li (1994) study the MA(q) case on the remainder disturbances, while MaCurdy (1982) and Galbraith and Zinde-Walsh (1995) study the autoregressive moving average ARMA(p, q) case. For an extension to the two-way model with serially correlated disturbances, see Revankar (1979) who considers the case where the λ_t follow an AR(1) process. Also, see Karlsson and Skoglund (2004) for the two-way error component model with an ARMA process on the time-specific effects. They derive the maximum likelihood estimator under normality of the disturbances and propose LM tests for serial correlation and for the choice between the AR(1) and MA(1) specification for the time- specific effects following Baltagi and Li (1995). Magnus and Woodland (1988) generalize this Revankar (1979) model to the multivariate error component model case and derive the corresponding maximum likelihood estimator. Chamberlain (1982, 1984) allows for arbitrary serial correlation and heteroskedastic patterns by viewing each time period as an equation and treating the panel as a multivariate setup. Baltagi, Song and Jung (2010) derive a joint LM test for homoskedasticity and no first-order serial correlation for a panel data error components regression model. In econometrics, when one tests for heteroskedasticity, serial correlation is ignored, and when one tests for serial correlation, heteroskedasticity is ignored. Baltagi, Song and Jung (2010) derive a conditional LM test for homoskedasticity given serial correlation, as well as, a conditional LM test for no first-order serial correlation given heteroskedasticity, all in the context of a random effects panel data model. Monte Carlo results show that these tests along with their likelihood ratio alternatives have good size and power under various forms of heteroskedasticity including exponential and quadratic functional forms.

Drukker (2003) provides an alternative test for serial correlation for panel data which one can find in Wooldridge (2010), using the *xtserial* command in Stata. The basic idea is to difference the model and hence get rid of the individual effects, the μ_i's, whether fixed or random, and test that the correlation of the resulting differenced

Table 5.8 Explaining OECD unemployment, 1961–95

.xtgls ur ur1 ep br bd gd_bd_br c_ud co gd_co_ud tw gd_co_tw lds tfphpc tts d2ms rirl t66-t95 tre1-tre20 id2-id20 , p(hetero) corr(psar1) rhotype(theil) igls

```
Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =        20         Number of obs      =       599
Estimated autocorrelations =        20         Number of groups   =        20
Estimated coefficients     =        85         Obs per group: min =        12
                                                              avg =     29.95
                                                              max =        33
                                               Wald chi2(84)      =  47272.06
                                               Prob > chi2        =    0.0000
```

ur	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ur1	.8569874	.0177229	48.35	0.000	.8222511	.8917237
ep	.1024391	.1615809	0.63	0.526	-.2142536	.4191317
br	2.664894	.4153647	6.42	0.000	1.850795	3.478994
bd	.865036	.2104416	4.11	0.000	.452578	1.277494
gd_bd_br	3.97257	.9477118	4.19	0.000	2.115089	5.830051
c_ud	7.378327	2.277199	3.24	0.001	2.915099	11.84155
co	-1.021815	.2968673	-3.44	0.001	-1.603664	-.4399658
gd_co_ud	-6.958332	1.172198	-5.94	0.000	-9.255797	-4.660867
tw	.7077332	.8891361	0.80	0.426	-1.034942	2.450408
gd_co_tw	-3.605739	1.057825	-3.41	0.001	-5.679038	-1.532441
lds	-23.97486	2.25323	-10.64	0.000	-28.39111	-19.55861
tfphpc	-17.65527	1.270991	-13.89	0.000	-20.14637	-15.16417
tts	6.616593	1.771568	3.73	0.000	3.144384	10.0888
d2ms	.4200588	.2506539	1.68	0.094	-.0712139	.9113314
rirl	2.052865	1.183245	1.73	0.083	-.2662521	4.371981

Time and country dummies as well as country specific time trends are not shown to save space.

error $(\nu_{it} - \nu_{i,t-1})$ one period apart is -0.5. This will be the case if the original remainder error ν_{it} is not serially correlated. Chapter 8 on dynamic panels will actually difference the model this way and the remainder error will be MA(1) unit root with correlation -0.5. In fact Arellano and Bond (1991) provide a test for the remainder error to be zero first-order serially correlated; see Chap. 8. For the Grunfeld data, the differenced regression and the Wooldridge test for serial correlation are shown in Table 5.9. Note that this runs the differenced regression, obtains the estimated residuals, then runs these residuals on their lagged value, and tests that the coefficient is -0.5 using robust standard errors, clustering on the firms. The test rejects zero first-

Table 5.9 Wooldridge test for serial correlation using Grunfeld's data

```
. xtserial I F C, output

Linear regression                                      Number of obs =     190
                                                       F(  2,     9) =   47.80
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4288
                                                       Root MSE      =  42.896

                                     (Std. Err. adjusted for 10 clusters in fn)
------------------------------------------------------------------------------
             |               Robust
         D.I |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         D1. |   .0890628   .0145088     6.14   0.000     .0562416    .1218841
             |
           C |
         D1. |    .278694    .138404     2.01   0.075    -.0343977    .5917856
------------------------------------------------------------------------------

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,       9) =    263.592
           Prob > F =      0.0000
```

order serial correlation in the remainder error ν_{it}. Born and Breitung (2016) compare the performance of this Wooldridge test for serial correlation with the Lagrange Multiplier (LM) test suggested by Baltagi and Li (1995), see (5.43), and the Bhargava, Franzini and Narendranathan (1982) modification of the classical Durbin–Watson statistic for the fixed effects model given in (5.44). Under the null hypothesis of no serial correlation, all tests possess a standard normal limiting distribution as $N \to \infty$ and T is fixed. Analyzing the local power of the tests, they find that the LM statistic of Baltagi and Li (1995) has superior power properties. Born and Breitung (2016) also propose a generalization to test for autocorrelation up to some given lag order and a test statistic that is robust against time- dependent heteroskedasticity. These can be implemented with Stata using the commands *xtqptest* and *xthrtest*. In fact, after performing (*xtreg, fe*) on the Grunfeld data, one issues the command *(xtqptest, lags(1))* which yields a χ_1^2 statistic of 7.77 with a p-value of 0.005 rejecting no serial correlation up to order 1. Similarly, issuing *(xtqptest, lags(2))* yields a χ_2^2 statistic of 9.38 with a p-value of 0.009 rejecting no serial correlation up to order 2. On the other hand, if one issues the command *(xtqptest, order(1))* this yields an LM statistic distributed as $N(0, 1)$ of 2.79 with a p-value of 0.005 rejecting no serial correlation of order 1. Similarly, issuing *(xtqptest, order(2))* yields an LM statistic distributed as $N(0, 1)$ of 2.59 with a p-value of 0.010 rejecting no serial correlation of order 2. Finally, issuing the command *xthrtest* yields a heteroskedasticity robust $N(0, 1)$ statistic of 0.93 with a p-value of 0.351 not rejecting first-order serial correlation. So except for the heteroskedasticity robust Born and Breitung (2016) test, serial correlation is not rejected for the Grunfeld data.

5.5 Notes

1. Bhargava, Franzini and Narendranathan (1982) derive the corresponding transformation for the one-way error component model with fixed effects and first-order autoregressive disturbances.
2. Let $a = (a_1, a_2, \ldots, a_n)'$ denote an arbitrary $n \times 1$ vector, then Toeplitz (a) is an $n \times n$ symmetric matrix generated from the $n \times 1$ vector a with the diagonal elements all equal to a_1 second diagonal elements equal to a_2, etc.
3. Obviously, there are many different ways to construct such a test. For example, we can use $Q_2 + Q_3 - 2Q_4$ instead of $Q_2 - Q_3$ to define the γ test. In this case,

$$V = 2\text{tr}\{[(\sigma_\mu^2 J_T + \sigma_\epsilon^2 V_\lambda)(G_2/(T-2) + G_3/(T-3) - 2G_4/(T-4))]^2\}$$

5.6 Problems

5.1 *Heteroskedastic individual effects*. (a) For the one-way error component model with heteroskedastic μ_i, i.e., $\mu_i \sim (0, w_i^2)$, verify that $\Omega = E(uu')$ is given by (5.1) and (5.2).
(b) Using the Wansbeek and Kapteyn (1982) trick show that Ω can also be written as in (5.3).

5.2 (a) Using (5.3) and (5.4), verify that $\Omega\Omega^{-1} = I$ and that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.
(b) Show that $y^* = \sigma_\nu \Omega^{-1/2} y$ has a typical element $y_{it}^* = y_{it} - \theta_i \bar{y}_{i.}$ where $\theta_i = 1 - (\sigma_\nu/\tau_i)$ and $\tau_i^2 = Tw_i^2 + \sigma_\nu^2$ for $i = 1, \ldots, N$.

5.3 *An LM test for heteroskedasticity in a one-way error component model.* Holly and Gardiol (2000) derived a score test for homoskedasticity in a one-way error component model where the alternative model is that the μ_i's are independent and distributed as $N(0, \sigma_{\mu_i}^2)$ where $\sigma_{\mu_i}^2 = \sigma_\mu^2 h_\mu(F_i\theta_2)$. Here, F_i is a vector of p explanatory variables such that $F_i\theta_2$ does not contain a constant term and h_μ is a strictly positive twice differentiable function satisfying $h_\mu(0) = 1$ with $h'_\mu(0) \neq 0$ and $h''_\mu(0) \neq 0$. Show that the score test statistic for H_0; $\theta_2 = 0$ is equal to one half of the explained sum of squares of the OLS regression of $(\hat{s}/\bar{s}) - \iota_N$ against the p regressors in F as in the Breusch and Pagan test for homoskedasticity. Here $\hat{s}_i = \hat{u}_i' \bar{J}_T \hat{u}_i$ and $\bar{s} = \sum_{i=1}^N \hat{s}_i/N$ where $\widehat{u}$ denote the maximum likelihood residuals from the restricted model under H_0; $\theta_2 = 0$.

5.4 *An alternative heteroskedastic error component model.* (a) For the one-way error component model with heteroskedastic remainder disturbances, i.e., $\nu_{it} \sim (0, w_i^2)$, verify that $\Omega = E(uu')$ is given by (5.5).
(b) Using the Wansbeek and Kapteyn (1982) trick show that Ω can also be written as in (5.6).

5.5 (a) Using (5.6) and (5.7), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.
(b) Show that $y^* = \Omega^{-1/2} y$ has a typical element $y_{it}^* = (y_{it} - \theta_i \bar{y}_{i.})/w_i$ where $\theta_i = 1 - (w_i/\tau_i)$ and $\tau_i^2 = T\sigma_\mu^2 + w_i^2$ for $i = 1, \ldots, N$.

5.6 *AR(1) process*. (a) For the one-way error component model with remainder disturbances ν_{it} following a stationary AR(1) process as in (5.8), verify that $\Omega^* = E(u^* u^{*\prime})$ is that given by (5.11).
(b) Using the Wansbeek and Kapteyn (1982) trick, show that Ω^* can be written as in (5.12).

5.7 (a) Using (5.12) and (5.13), verify that $\Omega^* \Omega^{*-1} = I$ and $\Omega^{*-1/2} \Omega^{*-1/2} = \Omega^{*-1}$.
(b) Show that $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ has a typical element given by (5.14).
(c) Show that for $\rho = 0$, (5.14) reduces to $(y_{it} - \theta \bar{y}_{i.})$.
(d) Show that for $\sigma_\mu^2 = 0$, (5.14) reduces to y_{it}^*.

5.8 *Unbiased estimates of the variance components under the AR(1) model*. Prove that $\widehat{\sigma}_\epsilon^2$ and $\widehat{\sigma}_\alpha^2$ given by (5.15) are unbiased for σ_ϵ^2 and σ_α^2, respectively.

5.9 *AR(2) process*. (a) For the one-way error component model with remainder disturbances ν_{it} following a stationary AR(2) process as in (5.16), verify that $\Omega^* = E(u^* u^{*\prime})$ is that given by (5.18).
(b) Show that $y^{**} = \sigma_\epsilon \Omega^{*-1/2} y^*$ has a typical element given by (5.19).

5.10 *AR(4) process for quarterly data*. For the one-way error component model with remainder disturbances ν_{it} following a specialized AR(4) process $\nu_{it} = \rho \nu_{i,t-4} + \epsilon_{it}$ with $|\rho| < 1$ and $\epsilon_{it} \sim \text{IIN}(0, \sigma_\epsilon^2)$, verify that $y^{**} = \sigma_\epsilon \Omega^{-1/2} y^*$ is given by (5.21).

5.11 *MA(1) process*. For the one-way error component model with remainder disturbances v_{it} following an MA(1) process given by (5.22), verify that $y^{**} = \sigma_\epsilon \Omega^{-1/2} y^*$ is given by (5.24).

5.12 *Prediction in the serially correlated error component model*. For the BLU predictor of $y_{i,T+1}$ given in (5.25), show that when v_{it} follows
(a) the AR(1) process, the GLS predictor is corrected by the term in (5.26);
(b) the AR(2) process, the GLS predictor is corrected by the term given in (5.27);
(c) the specialized AR(4) process, the GLS predictor is corrected by the term given in (5.28);
(d) the MA(1) process, the GLS predictor is corrected by the term given in (5.29).

5.13 *A joint LM test for serial correlation and random individual effects*. Using (4.17) and (4.19), verify (5.34) and (5.35) and derive the LM_1 statistic given in (5.36).

5.14 (a) Verify that $(\partial V_\rho / \partial \rho)_{\rho=0} = G = (\partial V_\lambda / \partial \lambda)_{\lambda=0}$ where G is the bidiagonal matrix with bidiagonal elements all equal to one.
(b) Using this result verify that the joint LM statistic given in (5.36) is the same whether the residual disturbances follow an AR(1) or an MA(1) process, i.e., the joint LM test statistic for H_1^a: $\sigma_\mu^2 = 0$; $\lambda = 0$ is the same as that for H_1^b: $\sigma_\mu^2 = 0$; $\rho = 0$.

5.15 *Conditional LM test for serial correlation assuming random individual effects*. For $H_4^b : \rho = 0$ (given $\sigma_\mu^2 > 0$),
(a) Derive the score, the information matrix, and the LM statistic for H_4^b.

(b) Verify that for H_4^a; $\lambda = 0$ (given $\sigma_\mu^2 > 0$), one obtains the same LM statistic as in part (a).

5.16 *An LM test for first-order serial correlation in a fixed effects model.* For H_5^b; $\rho = 0$ (given the μ_i are fixed),
(a) Verify that the likelihood is given by (5.40) and derive the MLE of the μ_i.
(b) Using (5.34) and (5.35), verify that the LM statistic for H_5^b is given by (5.42).
(c) Verify that for H_5^a; $\lambda = 0$ (given the μ_i are fixed), one obtains the same LM statistic as in (5.42).

5.17 (a) Verify (5.52) for the MA(1) model. Hint: Use the fact that $\lim E(u'u)/(NT) = \lim \text{tr}(\Omega)/(NT)$ for deriving $\text{plim} Q_0$. Similarly, use the fact that

$$\lim\ E(u'(I_N \otimes G_1)u)/N(T-1) = \lim\ \text{tr}[\Omega(I_N \otimes G_1)]/N(T-1)$$

for deriving $\text{plim} Q_1$. Also,

$$\lim\ E(u'(I_N \otimes G_s)u)/N(T-s) = \lim\ \text{tr}[\Omega(I_N \otimes G_s)]/N(T-s)$$

for deriving $\text{plim} Q_s$ for $s = 2, \ldots, S$.
(b) Verify (5.53) for the AR(1) model.

5.18 Using the Monte Carlo setup in Baltagi and Li (1995), study the performance of the tests proposed in Table 5.4.

5.19 For the *Grunfeld data*,
(a) Perform the tests described in Table 5.4.
(b) Using the unbalanced patterns described in Table 1 of Baltagi and Wu (1999), replicate the Baltagi–Wu *LBI* and Bhargava, Franzini and Narendranathan (1982) Durbin–Watson test statistics reported in that table. This can be easily done using (*xtregar, re lbi*) command in Stata.

5.20 For the *gasoline data* given on the Springer website, perform the tests described in Table 5.4.

5.21 For the *public capital data*, given on the Springer website, perform the tests described in Table 5.4.

5.22 Using the Grunfeld investment equation in (2.40),
(a) Replicate Table 5.5 using the AR(1) estimation with common ρ and heteroskedastic variances.
(b) Replicate Table 5.6 using the AR(1) estimation with varying ρ_i and heteroskedastic variances across firms.
(c) Replicate Table 5.7 using the AR(1) estimation with varying ρ_i, heteroskedastic variances, and cross-firm dependence in the variance–covariance matrix. Compare with the error component estimates obtained in Table 2.1.

5.23 *Time-wise autocorrelated and cross-sectionally heteroskedastic disturbances.* Using the Nickell, Nunziata and Ochel (2005) data set which is the basis for the empirical example in Sect. 5.3 explaining dynamic unemployment using a panel of 20 OECD countries over the period 1961–95,
(a) Replicate Table 5.8 using iterative generalized least squares allowing for heteroskedastic errors and country- specific first-order serial correlation.
(b) Replicate the rest of Table 5 in Nickell, Nunziata and Ochel (2005, p. 14).

References

Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49: 431–434.

Arellano, M. 1993. On the testing of correlated effects with panel data. *Journal of Econometrics* 59: 87–97.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Balestra, P. 1980. A note on the exact transformation associated with the first-order moving average process. *Journal of Econometrics* 14: 381–394.

Baltagi, B.H. 1986. Pooling under misspecification: Some Monte Carlo evidence on the Kmenta and the error components techniques. *Econometric Theory* 2: 429–440.

Baltagi, B.H. 1988. An alternative heteroscedastic error components model, Problem 88.2.2. *Econometric Theory* 4: 349–350.

Baltagi, B.H., and J.M. Griffin. 1988. A generalized error component model with heteroscedastic disturbances. *International Economic Review* 29: 745–753.

Baltagi, B.H., and Q. Li. 1991a. A transformation that will circumvent the problem of autocorrelation in an error component model. *Journal of Econometrics* 48: 385–393.

Baltagi, B.H., and Q. Li. 1991b. A joint test for serial correlation and random individual effects. *Statistics and Probability Letters* 11: 277–280.

Baltagi, B.H., and Q. Li. 1992. Prediction in the one-way error component model with serial correlation. *Journal of Forecasting* 11: 561–567.

Baltagi, B.H., and Q. Li. 1994. Estimating error component models with general MA (q) disturbances. *Econometric Theory* 10: 396–408.

Baltagi, B.H., and Q. Li. 1995. Testing AR (1) against MA (1) disturbances in an error component model. *Journal of Econometrics* 68: 133–151.

Baltagi, B.H., and Q. Li. 1997. Monte Carlo results on pure and pretest estimators of an error component model with autocorrelated disturbances. *Annales D'Économie et de Statistique* 48: 69–82.

Baltagi, B.H., and P.X. Wu. 1999. Unequally spaced panel data regressions with AR (1) disturbances. *Econometric Theory* 15: 814–823.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2005. Adaptive estimation of heteroskedastic error component models. *Econometric Reviews* 24: 39–58.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2006. Joint LM test for heteroskedasticity in a one-way error component model. *Journal of Econometrics* 134: 401–417.

Baltagi, B.H., S.H. Song, and B.C. Jung. 2010. Testing for heteroskedasticity and serial correlation in a random effects panel data model. *Journal of Econometrics* 154: 122–124.

Beck, N., and J. Katz. 1995. What to do (and not to do) with time-series-cross-section data in comparative politics. *American Political Science Review* 89: 634–647.

Bera, A.K., W. Sosa-Escudero, and M. Yoon. 2001. Tests for the error component model in the presence of local misspecification. *Journal of Econometrics* 101: 1–23.

Bhargava, A., L. Franzini, and W. Narendranathan. 1982. Serial correlation and fixed effects model. *Review of Economic Studies* 49: 533–549.

Born, B., and J. Breitung. 2016. Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews* 35: 1290–1316.

Breusch, T.S., and L.G. Godfrey. 1981. A review of recent work on testing for autocorrelation in dynamic simultaneous models. In *Macroeconomic analysis, essays in macroeconomics and economics*, (eds.) D.A. Currie, R. Nobay, and D. Peel, 63–100. London: Croom Helm.

Breusch, T.S., and A.R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Burke, S.P., L.G. Godfrey, and A.R. Termayne. 1990. Testing AR(1) against MA(1) disturbances in the linear regression model: An alternative procedure. *Review of Economic Studies* 57: 135–145.

Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.

Chamberlain, G. 1984. Panel data, Chapter 22. In *Handbook of Econometrics*, (eds.) Z. Griliches, and M. Intrilligator, 1247–1318. Amsterdam: North-Holland.

Drukker, D.M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3 (2): 168–177.

Frees, E.W., and T.W. Miller. 2004. Sales forecasting using longitudinal data models. *International Journal of Forecasting* 20: 99–114.

Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.

Fuller, W.A., and G.E. Battese. 1974. Estimation of linear models with cross-error structure. *Journal of Econometrics* 2: 67–78.

Galbraith, J.W., and V. Zinde-Walsh. 1995. Transforming the error-component model for estimation with general ARMA disturbances. *Journal of Econometrics* 66: 349–355.

Goldberger, A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57: 369–375.

Hansen, C.B. 2007. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics* 140: 670–694.

Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320–340.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hemmerle, W.J., and H.O. Hartley. 1973. Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics* 15: 819–831.

Holly, A., and L. Gardiol. 2000. A score test for individual heteroscedasticity in a one-way error components model, Chapter 10. In *Panel data econometrics: Future directions*, (eds.) J. Krishnakumar, and E. Ronchetti, 199–211. Amsterdam: North-Holland.

Inoue, A., and G. Solon. 2006. A portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory* 22: 835–851.

Karlsson, S., and J. Skoglund. 2004. Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects. *Empirical Economics* 29: 79–88.

King, M.L. 1983. Testing for autoregressive against moving average errors in the linear regression model. *Journal of Econometrics* 21: 35–51.

King, M.L., and M. McAleer. 1987. Further results on testing AR(1) against MA(1) disturbances in the linear regression model. *Review of Economic Studies* 54: 649–663.

Kmenta, J. 1986. *Elements of econometrics*. New York: MacMillan.

Lejeune, B. 2006. A full heteroscedastic one-way error components model: Pseudo maximum likelihood estimation and specification testing, Chapter 2. In *Panel data econometrics: Theoretical contributions and empirical applications*, (ed.) B. Baltagi, 31–66. Amsterdam: Elsevier Science.

Li, Q., and T. Stengos. 1994. Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. *International Economic Review* 35: 981–1000.

Lillard, L.A., and R.J. Willis. 1978. Dynamic aspects of earning mobility. *Econometrica* 46: 985–1012.

MaCurdy, T.E. 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18: 83–114.

Magnus, J.R., and A.D. Woodland. 1988. On the maximum likelihood estimation of multivariate regression models containing serially correlated error components. *International Economic Review* 29: 707–725.

Mazodier, P., and A. Trognon. 1978. Heteroskedasticity and stratification in error components models. *Annales de l'INSEE* 30–31: 451–482.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

Nickell, S., L. Nunziata, and W. Ochel. 2005. Unemployment in the OECD since the 1960s. What do we know? *Economic Journal* 115: 1–27.

Phillips, R.L. 2003. Estimation of a stratified error components model. *International Economic Review* 44: 501–521.

Randolph, W.C. 1988. A transformation for heteroscedastic error components regression models. *Economics Letters* 27: 349–354.

Rao, C.R. 1971a. Estimation of variance and covariance components–MINQUE theory. *Journal of Multivariate Analysis* 1: 257–275.

Rao, C.R. 1971b. Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1: 445–456.

Revankar, N.S. 1979. Error component models with serial correlated time effects. *Journal of the Indian Statistical Association* 17: 137–160.

Roy, N. 2002. Is adaptive estimation useful for panel models with heteroscedasticity in the individual specific error component? Some Monte Carlo evidence. *Econometric Reviews* 21: 189–203.

Taub, A.J. 1979. Prediction in the context of the variance-components model. *Journal of Econometrics* 10: 103–108.

Verbon, H.A.A. 1980. Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components (SUREVC). *Economics Letters* 5: 149–153.

Wansbeek, T.J. 1989. An alternative heteroscedastic error components model, Solution 88.1.1. *Econometric Theory* 5: 326.

Wansbeek, T.J., and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.

Wooldridge, J.M. 2010. *Econometric analysis of cross-section and panel data*. Massachusetts: MIT Press.

6 Seemingly Unrelated Regressions with Error Components

6.1 The One-Way Model

In several instances in economics, one is confronted with estimating a set of equations. This could be a set of demand equations across different sectors, industries, or regions. Other examples include the estimation of a translog cost function along with the corresponding cost share equations. In these cases, Zellner's (1962) seemingly unrelated regressions (SUR) approach is popular since it captures the efficiency due to the correlation of the disturbances across equations. Applications of the SUR procedure with time-series or cross-section data are too numerous to cite. In this chapter, we focus on the estimation of a set of SUR equations with panel data.

Avery (1977) seems to be the first to consider the SUR model with error component disturbances. In this case, we have a set of M equations

$$y_j = Z_j\delta_j + u_j \quad (j = 1, \ldots, M) \tag{6.1}$$

where y_j is $NT \times 1$, Z_j is $NT \times k'_j$, $\delta'_j = (\alpha_j, \beta'_j)$, β_j is $k_j \times 1$ and $k'_j = k_j + 1$ with

$$u_j = Z_\mu \mu_j + \nu_j \quad (j = 1, \ldots, M) \tag{6.2}$$

where $Z_\mu = (I_N \otimes \iota_T)$ and $\mu'_j = (\mu_{1j}, \mu_{2j}, \ldots, \mu_{Nj})$ and $\nu'_j = (\nu_{11j}, \ldots, \nu_{1Tj}, \ldots, \nu_{N1j}, \ldots, \nu_{NTj})$ are random vectors with zero means and covariance matrix

$$E\begin{pmatrix} \mu_j \\ \nu_j \end{pmatrix}(\mu'_l, \nu'_l) = \begin{bmatrix} \sigma^2_{\mu_{jl}} I_N & 0 \\ 0 & \sigma^2_{\nu_{jl}} I_{NT} \end{bmatrix} \tag{6.3}$$

for $j, l = 1, 2, \ldots, M$. This can be justified as follows: $\mu \sim (0, \Sigma_\mu \otimes I_N)$ and $\nu \sim (0, \Sigma_\nu \otimes I_{NT})$ where $\mu' = (\mu'_1, \mu'_2, \ldots, \mu'_M)$, $\nu' = (\nu'_1, \nu'_2, \ldots, \nu'_M)$, $\Sigma_\mu = [\sigma^2_{\mu_{jl}}]$ and $\Sigma_\nu = [\sigma^2_{\nu_{jl}}]$ for $j, l = 1, 2, \ldots, M$. In other words, *each* error component follows the same standard Zellner (1962) SUR assumptions imposed on classical disturbances. Using (6.2), it follows that

$$\Omega_{jl} = E(u_j u'_l) = \sigma^2_{\mu_{jl}}(I_N \otimes J_T) + \sigma^2_{\nu_{jl}}(I_N \otimes I_T) \tag{6.4}$$

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_6

In this case, the covariance matrix between the disturbances of different equations has the same one-way error component form. Except now, there are additional *cross-equations* variance components to be estimated. The variance–covariance matrix for the set of M equations is given by

$$\Omega = E(uu') = \Sigma_\mu \otimes (I_N \otimes J_T) + \Sigma_\nu \otimes \ (I_N \otimes I_T) \tag{6.5}$$

where $u' = (u_1', u_2', \ldots, u_M')$ is a $1 \times MNT$ vector of disturbances with u_j defined in (6.2) for $j = 1, 2, \ldots, M$. $\Sigma_\mu = [\sigma^2_{\mu jl}]$ and $\Sigma_\nu = [\sigma^2_{\nu jl}]$ are both $M \times M$ matrices. Replacing J_T by $T\bar{J}_T$ and I_T by $E_T + \bar{J}_T$, and collecting terms one gets

$$\begin{aligned}\Omega &= (T\Sigma_\mu + \Sigma_\nu) \otimes (I_N \otimes \bar{J}_T) + \Sigma_\nu \otimes (I_N \otimes E_T) \\ &= \Sigma_1 \otimes P + \Sigma_\nu \otimes Q\end{aligned} \tag{6.6}$$

where $\Sigma_1 = T\Sigma_\mu + \Sigma_\nu$. Also, $P = I_N \otimes \bar{J}_T$ and $Q = I_{NT} - P$ were defined (2.4) (6.6) is the spectral decomposition of Ω derived by Baltagi (1980), which means that

$$\Omega^r = \Sigma_1^r \otimes P + \Sigma_\nu^r \otimes Q \tag{6.7}$$

where r is an arbitrary scalar (see also Magnus 1982). For $r = -1$, one gets the inverse Ω^{-1} and for $r = -\frac{1}{2}$ one gets

$$\Omega^{-1/2} = \Sigma_1^{-1/2} \otimes P + \Sigma_\nu^{-1/2} \otimes Q \tag{6.8}$$

Kinal and Lahiri (1990) suggest obtaining the Cholesky decomposition of Σ_ν and Σ_1 in (6.8) to reduce the computation and simplify the transformation of the system.

One can estimate Σ_ν by $\widehat{\Sigma}_\nu = U'QU/N(T-1)$ and Σ_1 by $\widehat{\Sigma}_1 = U'PU/N$ where $U = [u_1, \ldots, u_M]$ is the $NT \times M$ matrix of disturbances for all M equations. Problem 6.7 asks the reader to verify that knowing U, $\widehat{\Sigma}_\nu$, and $\widehat{\Sigma}_1$ are unbiased estimates of Σ_ν and Σ_1, respectively. For feasible GLS estimates of the variance components, Avery (1977) following Wallace and Hussain (1969) in the single equation case recommends replacing U by OLS residuals, while Baltagi (1980) following Amemiya's (1971) suggestion for the single equation case recommends replacing U by Within-type residuals.

For this model, a block-diagonal Ω makes GLS on the whole system equivalent to GLS on each equation separately; see Problem 6.3. However, when the same X appear in each equation, GLS on the whole system is not equivalent to GLS on each equation separately (see Avery 1977). As in the single equation case, if N and $T \to \infty$, then the Within estimator of this system is asymptotically efficient and has the same asymptotic variance–covariance matrix as the GLS estimator.

6.2 The Two-Way Model

It is easy to extend the analysis to a two-way error component structure across the system of equations. In this case, (6.2) becomes

$$u_j = Z_\mu \mu_j + Z_\lambda \lambda_j + \nu_j \quad (j = 1, \ldots, M) \tag{6.9}$$

where $\lambda'_j = (\lambda_{1j}, \ldots, \lambda_{Tj})$ is a random vector with zero mean and covariance matrix given by the following:

$$E\begin{pmatrix}\mu_j\\ \lambda_j\\ \nu_j\end{pmatrix}(\mu'_l, \lambda'_l, \nu'_l) = \begin{bmatrix}\sigma^2_{\mu_{jl}} I_N & 0 & 0\\ 0 & \sigma^2_{\lambda_{jl}} I_T & 0\\ 0 & 0 & \sigma^2_{\nu_{jl}} I_{NT}\end{bmatrix} \quad (6.10)$$

for $j, l = 1, 2, \ldots, M$. In this case, $\lambda \sim (0, \Sigma_\lambda \otimes I_T)$ where $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_T)$ and $\Sigma_\lambda = [\sigma^2_{\lambda_{jl}}]$ is $M \times M$. Like μ and v, the λ follow a standard Zellner SUR-type assumption. Therefore,

$$\Omega_{jl} = E(u_j u'_l) = \sigma^2_{\mu_{jl}}(I_N \otimes J_T) + \sigma^2_{\lambda_{jl}}(J_N \otimes I_T) + \sigma^2_{\nu_{jl}}(I_N \otimes I_T) \quad (6.11)$$

As in the one-way SUR model, the covariance between the disturbances of different equations has the same two-way error component form. Except now, there are additional cross-equations variance components to be estimated. The variance–covariance matrix of the system of M equations is given by

$$\Omega = E(uu') = \Sigma_\mu \otimes (I_N \otimes J_T) + \Sigma_\lambda \otimes (J_N \otimes I_T) + \Sigma_\nu \otimes (I_N \otimes I_T) \quad (6.12)$$

where $u' = (u'_1, u'_2, \ldots, u'_M)$ with u_j defined in (6.9). Using the Wansbeek and Kapteyn (1982) trick, one gets (see Problem 6.5):

$$\Omega = \sum_{i=1}^{4} \Lambda_i \otimes Q_i \quad (6.13)$$

where $\Lambda_1 = \Sigma_\nu$, $\Lambda_2 = T\Sigma_\mu + \Sigma_\nu$, $\Lambda_3 = N\Sigma_\lambda + \Sigma_\nu$, and $\Lambda_4 = T\Sigma_\mu + N\Sigma_\lambda + \Sigma_\nu$, with Q_i defined (3.13). This is the spectral decomposition of Ω (see Baltagi 1980), with

$$\Omega^r = \sum_{i=1}^{4} \Lambda_i^r \otimes Q_i \quad (6.14)$$

for r an arbitrary scalar. When $r = -1$ one gets the inverse Ω^{-1} and when $r = -\frac{1}{2}$ one gets

$$\Omega^{-1/2} = \sum_{i=1}^{4} \Lambda_i^{-1/2} \otimes Q_i \quad (6.15)$$

Once again, the Cholesky decompositions of the Λ_i can be obtained in (6.15) to reduce the computation and simplify the transformation of the system (see Kinal and Lahiri 1990). Knowing the true disturbances U, quadratic unbiased estimates of the variance components are obtained from

$$\widehat{\Sigma}_v = \frac{U'Q_1U}{(N-1)(T-1)} \quad \widehat{\Lambda}_2 = \frac{U'Q_2U}{(N-1)} \quad \text{and} \quad \widehat{\Lambda}_3 = \frac{U'Q_3U}{(T-1)} \quad (6.16)$$

see Problem 6.7. Feasible estimates of (6.16) are obtained by replacing U by OLS residuals or Within-type residuals. One should check for positive definite estimates of Σ_μ and Σ_λ before proceeding. The Within estimator has the same asymptotic variance–covariance matrix as GLS when N and $T \to \infty$. Also, as long as the estimate of Σ_ν is consistent and the estimates of Σ_μ and Σ_λ have a finite positive definite probability limit, the corresponding feasible SUR-GLS estimate of the regression coefficients is asymptotically efficient.

6.3 Applications and Extensions

Verbon (1980) applies the SUR procedure with one-way error components to a set of four labor demand equations, using data from the Netherlands on 18 industries over 10 semiannual periods covering the period 1972–79. Verbon (1980) extends the above error component specification to allow for heteroskedasticity in the individual effects modeled as a simple function of p time-invariant variables. He applies a Breusch and Pagan (1979) LM test to check for the existence of heteroskedasticity.

Beierlein, Dunn and McConnon (1981) estimated the demand for electricity and natural gas in the northeastern United States using an SUR model with two-way error component disturbances. The data were collected for nine states comprising the Census Bureau's northeastern region of the USA for the period 1967–77. Six equations were considered corresponding to the various sectors considered. These were residential gas, residential electric, commercial gas, commercial electric, industrial gas, and industrial electric. Comparison of the error components SUR estimates with those obtained from OLS and single equation error component procedures showed substantial improvement in the estimates and a sizable reduction in the empirical standard errors.

Magnus (1982) derives the maximum likelihood estimator for the linear and nonlinear multivariate error component model under various assumptions on the errors. Sickles (1985) applies Magnus' multivariate nonlinear error components analysis to model the technology and specific factor productivity growth in the US airline industry.

Baltagi, Griffin and Rich (1995) estimate a SUR model consisting of a translog variable cost function and its corresponding input share equations for labor, fuel, and material. The panel data consists of 24 U.S. airlines over the period 1971–1986. Firm and time dummies are included in the variable cost equation and symmetry as well as adding-up restrictions on the share equations are imposed. A general Solow type index of technical change is estimated and its determinants are in turn analyzed. One of the main findings of this study is that despite the slowing of productivity growth in the 1980s, deregulation does appear to have stimulated technical change due to more efficient route structure.

Biorn (2004) considers the problem of estimating a system of regression equations with random individual effects from unbalanced panel data. The unbalancedness is due to random attrition. Biorn (2004) shows that GLS on this system can be interpreted as a matrix weighted average of group specific GLS estimators with weights equal to the inverse of their respective variance–covariance matrices. The grouping of individuals in the panel is according to the number of times they are observed (not necessarily the same period and not necessarily consecutive periods). Biorn also derives a stepwise algorithm for obtaining the MLE under normality of the disturbances.

Platoni, Sckokai and Moro (2012) extend Biorn's (2004) estimation of the unbalanced seemingly unrelated regressions (SUR) from the one-way to the two-way error components case. Once again, the GLS estimator can be interpreted as a matrix weighted average of the group specific GLS estimators with weights equal to the inverse of their respective covariance matrices.

Baltagi and Rich (2005) utilize the National Bureau of Economic Research (NBER) manufacturing productivity database file which provides annual data on 459 manufacturing industries at the SIC 4-digit level. They estimate a SUR model consisting of a translog cost function and its corresponding input share equations for production workers, nonproduction workers, energy, materials, and capital. Industry and time dummies are included in the cost equation and symmetry as well as adding-up restrictions on the share equations are imposed. Using the general index approach of Baltagi and Griffin (1988), they establish an explicit and unconstrained time path for nonneutral technical change between production and nonproduction labor in US manufacturing industries over the 1959–1996 period. Their findings confirm the prevailing interpretation in the labor economics literature that substantial reductions in the relative share of production labor are attributable to a sustained period of nonneutral technical change. However, they find that skill-biased technical change effects are most evident prior to 1983. This predates the diffusion of personal computer technologies in the workplace and the dramatic wage structure changes associated with the 1980s. This confirms previous findings that historically, biased technological change has been an important source of increased (relative) demand for skilled labor, and that one should avoid exaggerating the uniqueness of the computer revolution.

6.4 Problems

6.1 *Seemingly unrelated regressions with one-way error component disturbances*. Using the one-way error component structure on the disturbances of the jth equation given in (6.2) and (6.3), verify that Ω_{jl}, the variance–covariance matrix between the jth and lth equation disturbances, is given by (6.4).

6.2 Using (6.6) and (6.7), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

6.3 *Special cases of the SUR model with error component disturbances*. Consider a set of two equations with one-way error components disturbances.

(a) Show that if the variance–covariance matrix between the equations is *block-diagonal*, then GLS on the system is equivalent to GLS on each equation separately (see Avery 1977; Baltagi 1980).

(b) Show that if the explanatory variables are the *same* across the two equations, GLS on the system does not necessarily revert to GLS on each equation separately (seeAvery 1977; Baltagi 1980).

(c) Does your answer to parts (a) and (b) change if the disturbances followed a two-way error component model?

6.4 *Seemingly unrelated regressions with two-way error component disturbances*. Using the two-way error component structure on the disturbances of the jth equation given in (6.9) and (6.10), verify that Ω_{jl}, the variance–covariance matrix between the jth and lth equation disturbances, is given by (6.11).

6.5 Using the form of Ω given in (6.12) and the Wansbeek and Kapteyn (1982) trick verify (6.13).

6.6 Using (6.13) and (6.14), verify that $\Omega\Omega^{-1} = I$ and $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$.

6.7 *Unbiased estimates of the variance components of the one-way SUR model*.
(a) Using (6.6), verify that $\widehat{\Sigma}_\nu = U'QU/N(T-1)$ and $\widehat{\Sigma}_1 = U'PU/N$ yield unbiased estimates of Σ_ν and Σ_1, respectively.
(b) Using (6.13), verify that (6.16) results in unbiased estimates of Σ_ν, Λ_2, and Λ_3, respectively.

References

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13.

Avery, R.B. 1977. Error components and seemingly unrelated regressions. *Econometrica* 45: 199–209.

Baltagi, B.H. 1980. On seemingly unrelated regressions with error components. *Econometrica* 48: 1547–1551.

Baltagi, B.H., and J.M. Griffin. 1988. A general index of technical change. *Journal of Political Economy* 96: 20–41.

Baltagi, B.H., J.M. Griffin, and D. Rich. 1995. Airline deregulation: The cost pieces of the puzzle. *International Economic Review* 36: 245–258.

Baltagi, B.H., and D. Rich. 2005. Skilled-biased technical change in U.S. manufacturing: A general index approach. *Journal of Econometrics* 126: 549–570.

Beierlein, J.G., J.W. Dunn, and J.C. McConnon Jr. 1981. The demand for electricity and natural gas in the Northeastern United States. *Review of Economics and Statistics* 63: 403–408.

Biorn, E. 2004. Regression systems for unbalanced panel data: A stepwise maximum likelihood procedure. *Journal of Econometrics* 122: 281–291.

Breusch, T.S., and A.R. Pagan. 1979. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47: 1287–1294.

Kinal, T., and K. Lahiri. 1990. A computational algorithm for multiple equation models with panel data. *Economics Letters* 34: 143–146.

Magnus, J.R. 1982. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics* 19: 239–285.

Platoni, S., P. Sckokai, and D. Moro. 2012. A note on two-way ECM estimation of SUR systems on unbalanced panel data. *Econometric Reviews* 31: 119–141.

Sickles, R.C. 1985. A nonlinear multivariate error components analysis of technology and specific factor productivity growth with an application to U.S. airlines. *Journal of Econometrics* 27: 61–78.

Verbon, H.A.A. 1980. Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components (SUREVC). *Economics Letters* 5: 149–153.

Wallace, T.D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37: 55–72.

Wansbeek, T.J., and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.
Zellner, A. 1962. An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–368.

Simultaneous Equations with Error Components

7

7.1 Single Equation Estimation

Endogeneity of the right-hand regressors is a serious problem in econometrics. By endogeneity, we mean the correlation of the right-hand-side regressors and the disturbances. This may be due to the omission of relevant variables, measurement error, sample selectivity, self-selection, or other reasons. Endogeneity causes inconsistency of the usual OLS estimates and requires instrumental variable (IV) methods like two-stage least squares (2SLS) to obtain consistent parameter estimates. The applied literature is full of examples of endogeneity: demand and supply equations for labor, money, goods and commodities to mention a few. Also, behavioral relationships like consumption, production, investment, import, and export are just a few more examples in economics where endogeneity is suspected. We assume that the reader is familiar with the identification and estimation of a single equation and a system of simultaneous equations. In this chapter, we focus on the estimation of simultaneous equations using panel data.

Consider the following first structural equation of a simultaneous equation model

$$y_1 = Z_1\delta_1 + u_1 \tag{7.1}$$

where $Z_1 = [Y_1, X_1]$ and $\delta_1' = (\gamma_1', \beta_1')$. As in the standard simultaneous equation literature, Y_1 is the set of g_1 right-hand-side endogenous variables, and X_1 is the set of k_1 included exogenous variables. Let $X = [X_1, X_2]$ be the set of all exogenous variables in the system. This equation is identified with k_2 the number of excluded exogenous variables from the first equation (X_2) being larger than or equal to g_1.

Throughout this chapter, we will focus on the one-way error component model

$$u_1 = Z_\mu \mu_1 + \nu_1 \tag{7.2}$$

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_7

where $Z_\mu = (I_N \otimes \iota_T)$ and $\mu_1' = (\mu_{11}, \ldots, \mu_{N1})$ and $\nu_1' = (\nu_{111}, \ldots, \nu_{NT1})$ are random vectors with zero means and covariance matrix

$$E\begin{pmatrix}\mu_1\\ \nu_1\end{pmatrix}(\mu_1', \nu_1') = \begin{bmatrix}\sigma^2_{\mu_{11}} I_N & 0\\ 0 & \sigma^2_{\nu_{11}} I_{NT}\end{bmatrix} \tag{7.3}$$

This differs from the SUR setup in Chap. 6 only in the fact that there are right-hand-side endogenous variables in Z_1.[1] In this case,

$$E(u_1 u_1') = \Omega_{11} = \sigma^2_{\nu_{11}} I_{NT} + \sigma^2_{\mu_{11}}(I_N \otimes J_T) \tag{7.4}$$

In other words, the first structural equation has the typical variance–covariance matrix of a one-way error component model described in Chap. 2. The only difference is that now a double subscript (1, 1) is attached to the variance components to specify that this is the first equation. One can transform (7.1) by $Q = I_{NT} - P$ with $P = I_N \otimes \bar{J}_T$, to get

$$Qy_1 = QZ_1\delta_1 + Qu_1 \tag{7.5}$$

Let $\widetilde{y}_1 = Qy_1$ and $\widetilde{Z}_1 = QZ_1$. Performing 2SLS on (7.5) with $\widetilde{X} = QX$ as the set of instruments, one gets Within 2SLS (or Fixed Effects 2SLS)

$$\widetilde{\delta}_{1,W2SLS} = (\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1)^{-1} \widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1 \tag{7.6}$$

with $\mathrm{var}(\widehat{\widetilde{\delta}}_{1,W2SLS}) = \sigma^2_{\nu_{11}} (\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1)^{-1}$. This can be obtained using the Stata command (*xtivreg, fe*) specifying the endogenous variables Y_1 and the set of instruments X. Within 2SLS can also be obtained as GLS on

$$\widetilde{X}'\widetilde{y}_1 = \widetilde{X}'\widetilde{Z}_1\delta_1 + \widetilde{X}'\widetilde{u}_1 \tag{7.7}$$

see problem 7.1. Similarly, if we let $\bar{y}_1 = Py_1$ and $\bar{Z}_1 = PZ_1$, we can transform (7.1) by P and perform 2SLS with $\bar{X} = PX$ as the set of instruments. In this case, we get the Between 2SLS estimator of δ_1

$$\widehat{\delta}_{1,B2SLS} = (\bar{Z}_1' P_{\bar{X}} \bar{Z}_1)^{-1} \bar{Z}_1' P_{\bar{X}} \bar{y}_1 \tag{7.8}$$

with $\mathrm{var}(\widehat{\delta}_{1,B2SLS}) = \sigma^2_{1_{11}} (\bar{Z}_1' P_{\bar{X}} \bar{Z}_1)^{-1}$ where $\sigma^2_{1_{11}} = T\sigma^2_{\mu_{11}} + \sigma^2_{\nu_{11}}$. This can also be obtained using the Stata command (*xtivreg, be*) specifying the endogenous variables Y_1 and the set of instruments X. Between 2SLS can also be obtained as GLS on

$$\bar{X}'\bar{y}_1 = \bar{X}'\bar{Z}_1\delta_1 + \bar{X}'\bar{u}_1 \tag{7.9}$$

Stacking these two transformed equations in (7.7) and (7.9) as a system, as in (2.28) and noting that δ_1 is the same for these two transformed equations, one gets

$$\begin{pmatrix}\widetilde{X}'\widetilde{y}_1\\ \bar{X}'\bar{y}_1\end{pmatrix} = \begin{pmatrix}\widetilde{X}'\widetilde{Z}_1\\ \bar{X}'\bar{Z}_1\end{pmatrix}\delta_1 + \begin{pmatrix}\widetilde{X}'\widetilde{u}_1\\ \bar{X}'\bar{u}_1\end{pmatrix} \tag{7.10}$$

where

$$E\begin{pmatrix}\widetilde{X}'\widetilde{u}_1\\ \bar{X}'\bar{u}_1\end{pmatrix} = 0 \quad \text{and} \quad \mathrm{var}\begin{pmatrix}\widetilde{X}'\widetilde{u}_1\\ \bar{X}'\bar{u}_1\end{pmatrix} = \begin{bmatrix}\sigma^2_{v_{11}}\widetilde{X}'\widetilde{X} & 0\\ 0 & \sigma^2_{1_{11}}\bar{X}'\bar{X}\end{bmatrix}$$

Performing GLS on (7.10) yields the error component two-stage least squares (EC2SLS) estimator of δ_1 derived by Baltagi (1981b)

$$\widehat{\delta}_{1,EC2SLS} = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]^{-1} \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{y}_1}{\sigma^2_{1_{11}}}\right] \quad (7.11)$$

with $\mathrm{var}(\widehat{\delta}_{1,EC2SLS})$ given by the first inverted bracket in (7.11); see problem 7.2. Note that $\widehat{\delta}_{1,EC2SLS}$ can also be written as a matrix-weighted average of $\widetilde{\delta}_{1,W2SLS}$ and $\widehat{\delta}_{1,B2SLS}$ with the weights depending on their respective variance–covariance matrices:

$$\widehat{\delta}_{1,EC2SLS} = W_1 \widehat{\delta}_{1,W2SLS} + W_2 \widehat{\delta}_{1,B2SLS} \quad (7.12)$$

with

$$W_1 = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]^{-1} \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}}\right]$$

and

$$W_2 = \left[\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma^2_{\nu_{11}}} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]^{-1} \left[\frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma^2_{1_{11}}}\right]$$

Consistent estimates of $\sigma^2_{\nu_{11}}$ and $\sigma^2_{1_{11}}$ can be obtained from W2SLS and B2SLS residuals, respectively. In fact

$$\widehat{\sigma}^2_{\nu_{11}} = (y_1 - Z_1 \widetilde{\delta}_{1,W2SLS})' Q (y_1 - Z_1 \widetilde{\delta}_{1,W2SLS}) / N(T-1) \quad (7.13)$$

$$\widehat{\sigma}^2_{1_{11}} = (y_1 - Z_1 \widehat{\delta}_{1,B2SLS})' P (y_1 - Z_1 \widehat{\delta}_{1,B2SLS}) / N \quad (7.14)$$

Substituting these variance-components estimates in (7.11), one gets a feasible estimate of EC2SLS. Note that unlike the usual 2SLS procedure, EC2SLS requires estimates of the variance components. One can correct for degrees of freedom in (7.13) and (7.14) especially for small samples, but the panel is assumed to have large N. Also, one should check that $\widehat{\sigma}^2_{\mu_{11}} = (\widehat{\sigma}^2_{1_{11}} - \widehat{\sigma}^2_{\nu_{11}})/T$ is positive.

Alternatively, one can premultiply (7.1) by $\Omega_{11}^{-1/2}$ where Ω_{11} is given in (7.4), to get

$$y_1^* = Z_1^* \delta_1 + u_1^* \quad (7.15)$$

with $y_1^* = \Omega_{11}^{-1/2} y_1$, $Z_1^* = \Omega_{11}^{-1/2} Z_1$ and $u_1^* = \Omega_{11}^{-1/2} u_1$. In this case, $\Omega_{11}^{-1/2}$ is given by (2.20) with the additional subscripts (1, 1) for the variance components, i.e.,

$$\Omega_{11}^{-1/2} = (P/\sigma_{1_{11}}) + (Q/\sigma_{\nu_{11}}) \quad (7.16)$$

Therefore, the typical element of y_1^* is $y^*_{1_{it}} = (y_{1_{it}} - \theta_1 \bar{y}_{1_{i.}})/\sigma_{\nu_{11}}$ where $\theta_1 = 1 - (\sigma_{\nu_{11}}/\sigma_{1_{11}})$ and $\bar{y}_{1_{i.}} = \sum_{t=1}^T y_{1_{it}}/T$.

Given a set of instruments A, 2SLS on (7.15) using A gives

$$\widehat{\delta}_{1,2SLS} = (Z_1^{*\prime} P_A Z_1^*)^{-1} Z_1^{*\prime} P_A y_1^* \quad (7.17)$$

where $P_A = A(A'A)^{-1}A'$. Using the results in White (1986), the optimal set of instrumental variables in (7.15) is

$$X^* = \Omega_{11}^{-1/2} X = \frac{QX}{\sigma_{\nu_{11}}} + \frac{PX}{\sigma_{1_{11}}} = \frac{\widetilde{X}}{\sigma_{\nu_{11}}} + \frac{\bar{X}}{\sigma_{1_{11}}}$$

Using $A = X^*$, one gets the Balestra and Varadharajan-Krishnakumar (1987) generalized two-stage least squares (G2SLS):

$$\widehat{\delta}_{1,G2SLS} = (Z_1^{*\prime} P_{X^*} Z_1^*)^{-1} Z_1^{*\prime} P_{X^*} y_1^* \tag{7.18}$$

Cornwell, Schmidt and Wyhowski (1992) showed that Baltagi's (1981b) EC2SLS can be obtained from (7.17), i.e., using a 2SLS package on the transformed equation (7.15) with the set of instruments $A = [QX, PX] = [\widetilde{X}, \bar{X}]$. In fact, QX is orthogonal to PX and $P_A = P_{\widetilde{X}} + P_{\bar{X}}$. This also means that

$$\begin{aligned} P_A Z_1^* &= (P_{\widetilde{X}} + P_{\bar{X}})[\Omega_{11}^{-1/2} Z_1] \\ &= (P_{\widetilde{X}} + P_{\bar{X}}) \left[\frac{Q}{\sigma_{\nu_{11}}} + \frac{P}{\sigma_{1_{11}}} \right] Z_1 = \frac{P_{\widetilde{X}} \widetilde{Z}_1}{\sigma_{\nu_{11}}} + \frac{P_{\bar{X}} \bar{Z}_1}{\sigma_{1_{11}}} \end{aligned} \tag{7.19}$$

with

$$Z_1^{*\prime} P_A Z_1^* = \left(\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{Z}_1}{\sigma_{\nu_{11}}^2} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{Z}_1}{\sigma_{1_{11}}^2} \right)$$

and

$$Z_1^{*\prime} P_A y_1^* = \left(\frac{\widetilde{Z}_1' P_{\widetilde{X}} \widetilde{y}_1}{\sigma_{\nu_{11}}^2} + \frac{\bar{Z}_1' P_{\bar{X}} \bar{y}_1}{\sigma_{1_{11}}^2} \right)$$

Therefore, $\widehat{\delta}_{1,EC2SLS}$ given by (7.11) is the same as (7.17) with $A = [\widetilde{X}, \bar{X}]$.

So, how is Baltagi's (1981b) EC2SLS given by (7.11) different from the Balestra and Varadharajan-Krishnakumar (1987) G2SLS given by (7.18)? It should be clear to the reader that the set of instruments used by Baltagi (1981b), i.e., $A = [\widetilde{X}, \bar{X}]$ spans the set of instruments used by Balestra and Varadharajan-Krishnakumar (1987), i.e., $X^* = [\widetilde{X}/\sigma_{\nu_{11}} + \bar{X}/\sigma_{1_{11}}]$. In fact, one can show that $A = [\widetilde{X}, \bar{X}]$, $B = [X^*, \widetilde{X}]$ and $C = [X^*, \bar{X}]$ yield the same projection, and therefore the same 2SLS estimator given by EC2SLS (see problem 7.3). Without going into proofs, we note that Baltagi and Li (1992) showed that $\widehat{\delta}_{1,G2SLS}$ and $\widehat{\delta}_{1,EC2SLS}$ yield the same asymptotic variance–covariance matrix. Therefore, using White's (1986) terminology, $\widetilde{X}$ in B and $\bar{X}$ in C are *redundant* with respect to X^*. Redundant instruments can be interpreted loosely as additional sets of instruments that do not yield extra gains in asymptotic efficiency; see White (1986) for the strict definition and Baltagi and Li (1992) for the proof in this context.

For applications, it is easy to obtain EC2SLS using a standard 2SLS package:

Step 1: Run W2SLS, and B2SLS using a standard 2SLS package, i.e., run 2SLS of $\widetilde{y}$ on $\widetilde{Z}$ using $\widetilde{X}$ as instruments, and run 2SLS of $\bar{y}$ on $\bar{Z}$ using $\bar{X}$ as instruments. This yields (7.6) and (7.8), respectively.[2] Alternatively, this can be computed using the (*xtivreg, fe*) and (*xtivreg, be*) commands in Stata, specifying the endogenous variables and the set of instruments.

Step 2: Compute $\widehat{\sigma}^2_{\nu_{11}}$ and $\widehat{\sigma}^2_{1_{11}}$ from (7.13) and (7.14) and obtain y_1^*, Z_1^*, and X^* as described below (7.15). This transforms (7.1) by $\widehat{\Omega}_{11}^{-1/2}$ as in (7.15).
Step 3: Run 2SLS on this transformed equation (7.15) using the instrument set $A = X^*$ or $A = [QX, PX]$ as suggested above, i.e., run 2SLS of y_1^* on Z_1^* using X^* as instruments to get G2SLS, or $[\widetilde{X}, \bar{X}]$ as instruments to get EC2SLS. This yields (7.18) and (7.11), respectively. These computations are easy. They involve simple transformations on the data and the application of 2SLS three times. Alternatively, this can be done with Stata using the (*xtivreg, re*) command to get G2SLS, and (*xtivreg, ec2sls*) to get EC2SLS.

7.2 Empirical Example: Crime in North Carolina

Cornwell and Trumbull (1994), hereafter (CT), estimated an economic model of crime using panel data on 90 counties in North Carolina over the period 1981–1987. Table 7.1 replicates the Between and fixed effects estimates of CT using Stata. The empirical model relates the crime rate (which is an FBI index measuring the number of crimes divided by the county population) to a set of explanatory variables which include deterrent variables as well as variables measuring returns to legal opportunities. All variables are in logs except for the regional and time dummies. The explanatory variables consist of the probability of arrest (which is measured by the ratio of arrests to offenses), probability of conviction given arrest (which is measured by the ratio of convictions to arrests), probability of a prison sentence given a conviction (measured by the proportion of total convictions resulting in prison sentences); average prison sentence in days as a proxy for sanction severity. The number of police per capita as a measure of the county's ability to detect crime, the population density which is the county population divided by county land area, a dummy variable indicating whether the county is in the SMSA with population larger than 50,000. Percent minority, which is the proportion of the county's population that is minority or non-white. Percent young male which is the proportion of the county's population that is male and between the ages of 15 and 24. Regional dummies for western and central counties. Opportunities in the legal sector are captured by the average weekly wage in the county by industry. These industries are construction; transportation, utilities and communication; wholesale and retail trade; finance, insurance and real estate; services; manufacturing; and federal, state and local government.

Fixed effects results show that the probability of arrest, the probability of conviction given arrest, and the probability of imprisonment given conviction all have a negative and significant effect on the crime rate with estimated elasticities of −0.355, −0.282, and −0.173, respectively. The sentence severity has a negative but insignificant effect on the crime rate. The greater the number of police per capita, the greater the number of reported crimes per capita. The estimated elasticity is 0.413 and it is significant. This could be explained by the fact that the larger the police force, the larger the reported crime. Alternatively, this could be an endogeneity problem

Table 7.1 Economics of crime estimates for North Carolina, 1981–1987 (standard errors in parentheses)

lcrmrte	Between	Fixed Effects	FE2SLS	BE2SLS	EC2SLS
lprbarr	-0.648 (0.088)	-0.355 (0.032)	-0.576 (0.802)	-0.503 (0.241)	-0.413 (0.097)
lprbconv	-0.528 (0.067)	-0.282 (0.021)	-0.423 (0.502)	-0.525 (0.100)	-0.323 (0.054)
lprbpris	0.297 (0.231)	-0.173 (0.032)	-0.250 (0.279)	0.187 (0.318)	-0.186 (0.042)
lavgsen	-0.236 (0.174)	-0.002 (0.026)	0.009 (0.049)	-0.227 (0.179)	-0.010 (0.027)
lpolpc	0.364 (0.060)	0.413 (0.027)	0.658 (0.847)	0.408 (0.193)	0.435 (0.090)
ldensity	0.168 (0.077)	0.414 (0.283)	0.139 (1.021)	0.226 (0.102)	0.429 (0.055)
lwcon	0.195 (0.210)	-0.038 (0.039)	-0.029 (0.054)	0.314 (0.259)	-0.007 (0.040)
lwtuc	-0.196 (0.170)	0.046 (0.019)	0.039 (0.031)	-0.199 (0.197)	0.045 (0.020)
lwtrd	0.129 (0.278)	-0.021 (0.040)	-0.018 (0.045)	0.054 (0.296)	-0.008 (0.041)
lwfir	0.113 (0.220)	-0.004 (0.028)	-0.009 (0.037)	0.042 (0.306)	-0.004 (0.029)
lwser	-0.106 (0.163)	0.009 (0.019)	0.019 (0.039)	-0.135 (0.174)	0.006 (0.020)
lwmfg	-0.025 (0.134)	-0.360 (0.112)	-0.243 (0.420)	-0.042 (0.156)	-0.204 (0.080)
lwfed	0.156 (0.287)	-0.309 (0.176)	-0.451 (0.527)	0.148 (0.326)	-0.164 (0.159)
lwsta	-0.284 (0.256)	0.053 (0.114)	-0.019 (0.281)	-0.203 (0.298)	-0.054 (0.106)
lwloc	0.010 (0.463)	0.182 (0.118)	0.263 (0.312)	0.044 (0.494)	0.163 (0.120)
lpctmle	-0.095 (0.158)	0.627 (0.364)	0.351 (1.011)	-0.095 (0.192)	-0.108 (0.140)
lpctmin	0.148 (0.049)	--	--	0.169 (0.053)	0.189 (0.041)
west	-0.230 (0.108)	--	--	-0.205 (0.114)	-0.227 (0.100)
central	-0.164 (0.064)	--	--	-0.173 (0.067)	-0.194 (0.060)
urban	-0.035 (0.132)	--	--	-0.080 (0.144)	-0.225 (0.116)
_cons	-2.097 (2.822)	--	--	-1.977 (4.001)	-0.954 (1.284)

Time dummies were included except for Between and BE2SLS. The number of observations is 630. The F-statistics for significance of county dummies in fixed effects is F(89,518) = 36.38. The corresponding F-statistic using FE2SLS is 29.66. Both are significant. Hausman's test for (fixed effects – random effects) is $\chi^2(22) = 49.4$ with p-value of 0.0007. The corresponding Hausman test for (FE2SLS – EC2SLS) is $\chi^2(22) = 19.5$ with a p-value of 0.614.

with more crime resulting in the hiring of more police. The higher the density of the population the higher the crime rate, but this is insignificant. Returns to legal activity are insignificant except for wages in the manufacturing sector and wages in the transportation, utilities and communication sector. The manufacturing wage has a negative and significant effect on crime with an estimated elasticity of -0.36, while the transportation, utilities and communication sector wages have a positive and significant effect on crime with an estimated elasticity of 0.046. Percent young male is insignificant at the 5% level.

Cornwell and Trumbull (1994) argue that the Between estimates do not control for county effects and yield much higher deterrent effects than the fixed effects estimates. These Between estimates as well as the random effects estimates are rejected as inconsistent by a Hausman (1978) test. CT worried about the endogeneity of police per capita and the probability of arrest. They used as instruments two additional variables. Offense mix which is the ratio of crimes involving face-to-face contact (such as robbery, assault, and rape) to those that do not. The rationale for using this variable is that arrest is facilitated by positive identification of the offender. The second instrument is per capita tax revenue. This is justified on the basis that counties with preferences for law enforcement will vote for higher taxes to fund a larger police force. The fixed effects 2SLS estimates are reported in Table 7.1. These results are based on the replication by Baltagi (2006a) using Stata, but they do not match those in Table 3 of Cornwell and Trumbull (1994). All deterrent variables had insignificant t-statistics. These include the probability of arrest, the probability of conviction given arrest as well as the probability of imprisonment given conviction. Also insignificant were sentence severity and police per capita. Manufacturing wage, which was significant using the fixed effects 2SLS estimates of Cornwell and Trumbull (1994), turn out to be insignificant in our replication. In fact, CT find that all variables were insignificant using fixed effects 2SLS except for the percent young male and the manufacturing wage. CT also report 2SLS estimates ignoring the heterogeneity in the county effects for comparison. However, they warn against using biased and inconsistent estimates that ignore county effects. In fact, county effects were always significant; see the F-statisitcs reported in Table 7.1.

An alternative to dealing with the endogeneity problem is to run a random effects 2SLS estimator that allows for the endogeneity of police per capita and the probability of arrest. This estimator is a matrix-weighted average of Between 2SLS and fixed effects 2SLS and was denoted by EC2SLS in (7.11). The Stata output for EC2SLS is given in Table 7.2 using (*xtivreg, ec2sls*) and the results are summarized in Table 7.1. All the deterrent variables are significant with slightly higher elasticities than fixed effects. The sentence severity variable is still insignificant and police per capita is still positive and significant. Manufacturing wage is negative and significant and percent minority is positive and significant. Obtaining an estimate of the last coefficient is an advantage of EC2SLS over the fixed effects estimators, because it allows us to recapture estimates of variables that were invariant across time and wiped out by the fixed effects transformation; see also Hausman and Taylor (1981) and Sect. 7.4. Table 7.3 gives the random effects (G2SLS) estimator described in (7.18) using (*xtivreg, re*). G2SLS gives basically the same results as EC2SLS but the stan-

Table 7.2 EC2SLS estimates for the crime data

```
. xtivreg  lcrmrte lprbconv lprbpris lavgsen  ldensity lwcon lwtuc lwtrd lwfir lw
> ser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban d82 d83 d84 d85
>  d86  d87 ( lprbarr lpolpc= ltaxpc lmix), ec2sls

EC2SLS Random-effects regression               Number of obs      =       630
Group variable: county                         Number of groups   =        90

R-sq:  within  = 0.4521                        Obs per group: min =         7
       between = 0.8158                                       avg =       7.0
       overall = 0.7840                                       max =         7

                                               Wald chi2(26)      =    575.74
corr(u_i, X)        = 0 (assumed)              Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.4129261    .097402    -4.24   0.000    -.6038305   -.2220218
      lpolpc |   .4347492    .089695     4.85   0.000     .2589502    .6105482
    lprbconv |  -.3228872   .0535517    -6.03   0.000    -.4278465   -.2179279
    lprbpris |  -.1863195   .0419382    -4.44   0.000    -.2685169   -.1041222
     lavgsen |  -.0101765   .0270231    -0.38   0.706    -.0631408    .0427877
    ldensity |   .4290282   .0548483     7.82   0.000     .3215275     .536529
       lwcon |  -.0074751   .0395775    -0.19   0.850    -.0850455    .0700954
       lwtuc |    .045445   .0197926     2.30   0.022     .0066522    .0842379
       lwtrd |  -.0081412   .0413828    -0.20   0.844    -.0892499    .0729676
       lwfir |  -.0036395   .0289238    -0.13   0.900    -.0603292    .0530502
       lwser |   .0056098   .0201259     0.28   0.780    -.0338361    .0450557
       lwmfg |  -.2041398   .0804393    -2.54   0.011     -.361798   -.0464816
       lwfed |  -.1635108   .1594496    -1.03   0.305    -.4760263    .1490047
       lwsta |  -.0540503   .1056769    -0.51   0.609    -.2611732    .1530727
       lwloc |   .1630523    .119638     1.36   0.173    -.0714339    .3975384
    lpctymle |  -.1081057   .1396949    -0.77   0.439    -.3819026    .1656912
     lpctmin |    .189037   .0414988     4.56   0.000     .1077009    .2703731
        west |  -.2268433   .0995913    -2.28   0.023    -.4220387   -.0316479
     central |  -.1940428   .0598241    -3.24   0.001    -.3112958   -.0767898
       urban |  -.2251539   .1156302    -1.95   0.052    -.4517851    .0014772
         d82 |   .0107452   .0257969     0.42   0.677    -.0398158    .0613062
         d83 |  -.0837944   .0307088    -2.73   0.006    -.1439825   -.0236063
         d84 |  -.1034997   .0370885    -2.79   0.005    -.1761918   -.0308076
         d85 |  -.0957017   .0494502    -1.94   0.053    -.1926223    .0012189
         d86 |  -.0688982   .0595956    -1.16   0.248    -.1857036    .0479071
         d87 |  -.0314071   .0705197    -0.45   0.656    -.1696232    .1068091
       _cons |  -.9538032   1.283966    -0.74   0.458    -3.470331    1.562725
-------------+----------------------------------------------------------------
     sigma_u |  .21455964
     sigma_e |  .14923892
         rho |  .67394413   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   lprbarr lpolpc
Instruments:    lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser
                lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban d82
                d83 d84 d85 d86 d87 ltaxpc lmix
```

dard errors are higher. Remember that EC2SLS uses more instruments than G2SLS. Problem 04.1.1 in *Econometric Theory* by Baltagi (2004) suggests a Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS. For the crime data, this yields a Hausman statistic of 19.50 which is distributed as $\chi^2(22)$ and is insignificant with a p-value of 0.614. This does not reject the null hypothesis that EC2SLS yields a consistent estimator. This can be computed using the Hausman command after storing the EC2SLS and FE2SLS estimates. Recall that the random effects estimator was rejected by Cornwell and Trumbull (1994) based on the standard Hausman (1978) test. The latter was based on the contrast between fixed effects and random effects assuming that the endogeneity comes entirely from the correlation between the county effects and the explanatory variables. This does not account

Table 7.3 Random effects 2SLS for crime data (G2SLS)

```
 . xtivreg  lcrmrte lprbconv lprbpris lavgsen  ldensity lwcon lwtuc lwtrd lwfir lw
> ser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban d82 d83 d84 d85
>  d86  d87 ( lprbarr lpolpc= ltaxpc lmix), re

G2SLS Random-effects regression                 Number of obs      =       630
Group variable: county                          Number of groups   =        90

R-sq:  within  = 0.4521                         Obs per group: min =         7
       between = 0.8036                                        avg =       7.0
       overall = 0.7725                                        max =         7

                                                Wald chi2(26)      =    542.48
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lcrmrte |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.4141383   .2210496    -1.87   0.061    -.8473875    .0191109
      lpolpc |   .5049461   .2277778     2.22   0.027     .0585098    .9513824
    lprbconv |  -.3432506   .1324648    -2.59   0.010    -.6028768   -.0836244
    lprbpris |  -.1900467   .0733392    -2.59   0.010     -.333789   -.0463045
     lavgsen |  -.0064389   .0289407    -0.22   0.824    -.0631617    .0502838
    ldensity |   .4343449   .0711496     6.10   0.000     .2948943    .5737956
       lwcon |  -.0042958   .0414226    -0.10   0.917    -.0854826    .0768911
       lwtuc |   .0444589   .0215448     2.06   0.039     .0022318    .0866859
       lwtrd |  -.0085579   .0419829    -0.20   0.838    -.0908428     .073727
       lwfir |  -.0040305   .0294569    -0.14   0.891    -.0617649    .0537038
       lwser |   .0105602   .0215823     0.49   0.625    -.0317403    .0528608
       lwmfg |   -.201802   .0839373    -2.40   0.016    -.3663161   -.0372878
       lwfed |  -.2134579   .2151046    -0.99   0.321    -.6350551    .2081393
       lwsta |  -.0601232   .1203149    -0.50   0.617     -.295936    .1756896
       lwloc |   .1835363   .1396775     1.31   0.189    -.0902265    .4572992
    lpctymle |  -.1458703   .2268086    -0.64   0.520    -.5904071    .2986664
     lpctmin |   .1948763   .0459385     4.24   0.000     .1048384    .2849141
        west |  -.2281821    .101026    -2.26   0.024    -.4261894   -.0301747
     central |  -.1987703   .0607475    -3.27   0.001    -.3178332   -.0797075
       urban |  -.2595451   .1499718    -1.73   0.084    -.5534844    .0343942
         d82 |   .0132147   .0299924     0.44   0.660    -.0455692    .0719987
         d83 |  -.0847693    .032001    -2.65   0.008    -.1474901   -.0220485
         d84 |  -.1062027   .0387893    -2.74   0.006    -.1822284   -.0301769
         d85 |  -.0977457   .0511681    -1.91   0.056    -.1980334     .002542
         d86 |  -.0719451   .0605819    -1.19   0.235    -.1906835    .0467933
         d87 |  -.0396595   .0758531    -0.52   0.601    -.1883289    .1090099
       _cons |  -.4538501   1.702983    -0.27   0.790    -3.791636    2.883935
-------------+----------------------------------------------------------------
     sigma_u |  .21455964
     sigma_e |  .14923892
         rho |  .67394413   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   lprbarr lpolpc
Instruments:    lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser
                lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban d82
                d83 d84 d85 d86 d87 ltaxpc lmix
```

for the endogeneity of the conventional simultaneous equation type between police per capita and the probability of arrest and the crime rate. This alternative Hausman test based on the contrast between fixed effects 2SLS and EC2SLS failed to reject the null hypothesis. Accounting for this endogeneity, the random effects 2SLS becomes a viable estimator whose consistency cannot be rejected. We also ran the first stage regressions to check for weak instruments. For the probability of arrest, the F-statistic of the fixed effects first-stage regression was 15.6 as compared to 4.62 for the Between first-stage regression. Similarly, for the police per capita, the F-statistic of the fixed effects first-stage regression was 9.27 as compared to 2.60 for the Between first-stage regression. This indicates that these instruments may be weaker in the Between first-stage regressions (for Between 2SLS) than in the fixed effects first stage regressions (for fixed effects 2SLS). This example confirms the Cornwell and Trumbull (1994) conclusion that county effects cannot be ignored in estimating an economic model of crime using panel data in North Carolina. It also shows that the usual Hausman test based on the difference between fixed effects and random effects may lead to misleading inference if there are endogenous regressors of the conventional simultaneous equation type. An alternative Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS did not reject the consistency of random effects 2SLS, an estimator which yields plausible estimates of the crime equation.

7.3 System Estimation

Consider the system of identified equations:

$$y = Z\delta + u \tag{7.20}$$

where $y' = (y'_1, \ldots, y'_M)$, $Z = \text{diag}[Z_j]$, $\delta' = (\delta'_1, \ldots, \delta'_M)$, and $u' = (u'_1, \ldots, u'_M)$ with $Z_j = [Y_j, X_j]$ of dimension $NT \times (g_j + k_j)$, for $j = 1, \ldots, M$. In this case, there are g_j included right-hand side Y_j and k_j included right-hand side X_j. This differs from the SUR model only in the fact that there are right-hand-side endogenous variables in the system of equations. For the one-way error component model, the disturbance of the jth equation u_j is given by (6.2) and $\Omega_{jl} = E(u_j u'_l)$ is given by (6.4) as in the SUR case. Once again, the covariance matrix between the disturbances of different equations has the same error component form. Except now, there are additional cross-equations variance components to be estimated. The variance–covariance matrix of the set of M structural equations $\Omega = E(uu')$ is given by (6.5) and $\Omega^{-1/2}$ is given by (6.8). Premultiplying (7.20) by $(I_M \otimes Q)$ yields

$$\widetilde{y} = \widetilde{Z}\delta + \widetilde{u} \tag{7.21}$$

where $\widetilde{y} = (I_M \otimes Q)y$, $\widetilde{Z} = (I_M \otimes Q)Z$ and $\widetilde{u} = (I_M \otimes Q)u$. Performing 3SLS on (7.21) with $(I_M \otimes \widetilde{X})$ as the set of instruments, where $\widetilde{X} = QX$, one gets the Within 3SLS estimator:

$$\widetilde{\delta}_{W3SLS} = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z}]^{-1}[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{y}] \tag{7.22}$$

Similarly, transforming (7.20) by$(I_M \otimes P)$ yields

$$\bar{y} = \bar{Z}\delta + \bar{u} \tag{7.23}$$

where $\bar{y} = (I_M \otimes P)y$, $\bar{Z} = (I_M \otimes P)Z$ and $\bar{u} = (I_M \otimes P)u$. Performing 3SLS on the transformed system (7.23) using $(I_M \otimes \bar{X})$ as the set of instruments, where $\bar{X} = PX$, one gets the Between 3SLS estimator:

$$\widehat{\delta}_{B3SLS} = [\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{y}] \tag{7.24}$$

Next, we stack the two transformed systems given in (7.21) and (7.23) after premultiplying by $(I_M \otimes \widetilde{X}')$ and $(I_M \otimes \bar{X}')$, respectively. Then, we perform GLS noting that δ is the same for these two transformed systems (see problem 7.5). The resulting estimator of δ is the error components three-stage least squares (EC3SLS) given by Baltagi (1981b)

$$\begin{aligned}\widehat{\delta}_{EC3SLS} = {} & [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1} \\ & \times[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{y} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{y}] \end{aligned} \tag{7.25}$$

Note that $\widehat{\delta}_{EC3SLS}$ can also be written as a matrix-weighted average of $\widehat{\delta}_{W3SLS}$ and $\widehat{\delta}_{B3SLS}$ as follows:

$$\widehat{\delta}_{EC3SLS} = W_1\widehat{\delta}_{W3SLS} + W_2\widehat{\delta}_{B3SLS} \tag{7.26}$$

with

$$W_1 = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z}]$$

and

$$W_2 = [\widetilde{Z}'(\Sigma_\nu^{-1} \otimes P_{\widetilde{X}})\widetilde{Z} + \bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]^{-1}[\bar{Z}'(\Sigma_1^{-1} \otimes P_{\bar{X}})\bar{Z}]$$

Consistent estimates of Σ_ν and Σ_1 can be obtained as in (7.13) and (7.14) using W2SLS and B2SLS residuals with

$$\widehat{\sigma}^2_{\nu_{jl}} = (y_j - Z_j\widetilde{\delta}_{j,W2SLS})'Q(y_l - Z_l\widetilde{\delta}_{l,W2SLS})/N(T-1) \tag{7.27}$$

$$\widehat{\sigma}^2_{1_{jl}} = (y_j - Z_j\widehat{\delta}_{j,B2SLS})'P(y_l - Z_l\widehat{\delta}_{l,B2SLS})/N \tag{7.28}$$

One should check whether $\widehat{\Sigma}_\mu = (\widehat{\Sigma}_1 - \widehat{\Sigma}_\nu)/T$ is positive definite.

Using $\Omega^{-1/2}$ from (6.8), one can transform (7.20) to get

$$y^* = Z^*\delta + u^* \tag{7.29}$$

with $y^* = \Omega^{-1/2}y$, $Z^* = \Omega^{-1/2}Z$ and $u^* = \Omega^{-1/2}u$. For an arbitrary set of instruments A, the 3SLS estimator of (7.29) becomes

$$\widehat{\delta}_{3SLS} = (Z^{*\prime}P_A Z^*)^{-1}Z^{*\prime}P_A y^* \tag{7.30}$$

Using the results of White (1986), the optimal set of instruments is

$$X^* = \Omega^{-1/2}(I_M \otimes X) = (\Sigma_\nu^{-1/2} \otimes QX) + (\Sigma_1^{-1/2} \otimes PX)$$

Substituting $A = X^*$ in (7.30), one gets the efficient three-stage least squares (E3SLS) estimator:

$$\widehat{\delta}_{E3SLS} = (Z^{*\prime}P_{X^*}Z^*)^{-1}Z^{*\prime}P_{X^*}y^* \tag{7.31}$$

This is not the G3SLS estimator suggested by Balestra and Varadharajan-Krishnakumar (1987). In fact, Balestra and Varadharajan-Krishnakumar (1987) suggest using

$$\begin{aligned} A &= \Omega^{1/2}\,\mathrm{diag}[\Omega_{jj}^{-1}](I_M \otimes X) \\ &= \Sigma_\nu^{1/2}\,\mathrm{diag}\left(\frac{1}{\sigma^2_{\nu_{jj}}}\right) \otimes \widetilde{X} + \Sigma_1^{1/2}\,\mathrm{diag}\left(\frac{1}{\sigma^2_{1_{jj}}}\right) \otimes \bar{X} \end{aligned} \tag{7.32}$$

Substituting this A in (7.30) yields the G3SLS estimator of δ. So, how are G3SLS, EC3SLS, and E3SLS related? Baltagi and Li (1992) show that Baltagi's (1981b) EC3SLS estimator can be obtained from (7.30) with $A = [I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$. From this it is clear that the set of instruments $[I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$ used by Baltagi (1981b) spans the set of instruments $[\Sigma_\nu^{-1/2} \otimes \widetilde{X} + \Sigma_1^{-1/2} \otimes \bar{X}]$ needed for E3SLS. In addition, we note without proof that Baltagi and Li (1992) show that $\widehat{\delta}_{EC3SLS}$ and $\widehat{\delta}_{E3SLS}$ yield the same asymptotic variance–covariance matrix. Problem 7.6 shows that Baltagi's (1981b) EC3SLS estimator has redundant instruments with respect to those used by the E3SLS estimator. Therefore, using White's (1984) terminology, the extra instruments used by Baltagi (1981b) do not yield extra gains in asymptotic efficiency. However, Baltagi and Li (1992) also show that both EC3SLS and E3SLS are asymptotically more efficient than the G3SLS estimator corresponding to the set of instruments given by (7.32). In applications, it is easy to obtain EC3SLS using a standard 3SLS package:

Step 1: Obtain W2SLS and B2SLS estimates of each structural equation as described in the first step of computing EC2SLS.
Step 2: Compute estimates of $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_\nu$ as described in (7.27) and (7.28).
Step 3: Obtain the Cholesky decomposition of $\widehat{\Sigma}_1^{-1}$ and $\widehat{\Sigma}_\nu^{-1}$ and use those instead of $\widehat{\Sigma}_1^{-1/2}$ and $\widehat{\Sigma}_\nu^{-1/2}$ in the transformation described in (7.29), i.e., obtain y^*, Z^*, and X^* as described below (7.30).
Step 4: Apply 3SLS to this transformed system (7.29) using as a set of instruments $A = X^*$ or $A = [I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$, i.e., run 3SLS of y^* on Z^* using as instruments X^* or $[I_M \otimes \widetilde{X}, I_M \otimes \bar{X}]$. These yield (7.31) and (7.25), respectively. The computations are again easy, requiring simple transformations and a 3SLS package.

Baltagi (1981b) shows that EC3SLS reduces to EC2SLS when the disturbances of the different structural equations are uncorrelated with each other, but not necessarily when all the structural equations are just-identified. This is different from the analogous conditions between 2SLS and 3SLS in the classical simultaneous equations model (see problem 7.7).

Baltagi (1984) also performs Monte Carlo experiments on a two-equation simultaneous model with error components and demonstrates the efficiency gains in terms of mean squared error in performing EC2SLS and EC3SLS over the standard simultaneous equation counterparts, 2SLS and 3SLS. EC2SLS and EC3SLS also performed better than a two- or three-stage variance-components method where right-hand-side endogenous variables are replaced by their predicted values from the reduced form and the standard error component GLS is performed in the second step. Also, Baltagi (1984) demonstrates that better estimates of the variance components do not necessarily imply better estimates of the structural or reduced form parameters.[3]

Baltagi and Blien (1998) use FE-2SLS to estimate wage curves for Germany using data for 142 labor market regions over the period 1981-90. Briefly, the wage curve describes the negative relationship between the local unemployment rate and the level of wages. Baltagi and Blien (1998) find that ignoring endogeneity of the local employment rate yields results in favor of the wage curve only for younger and less qualified workers. Accounting for endogeneity of the unemployment rate yields evidence in favor of the wage curve across all types of workers. In particular, the wages of less qualified workers are more responsive to local unemployment rates than the wages of more qualified workers. Also, the wages of men are slightly more responsive to local unemployment rates than the wages of women. Applications of EC2SLS and EC3SLS include (i) an econometric rational-expectations macroeconomic model for developing countries with capital controls (see Haque, Lahiri and Montiel 1993), and (ii) an econometric model measuring income and price elasticities of foreign trade for developing countries (see Kinal and Lahiri 1990).

Empirical Example: *Economic growth and foreign aid*. Bruckner (2013) investigates the simultaneity problem between per capita GDP growth and foreign aid. He utilizes an unbalanced panel of 47 least developed countries (LDCs) observed over the period 1960–2000. Using fixed effects 2sls with time and country effects, Bruckner shows that a 1% point increase in GDP per capita growth decreased foreign aid by over 4%. The endogenous variables are the change in log GDP (D_lgdp) and the change in log foreign aid (D_laid). The instrumental variables used for this equation were the change in log rainfall (D_rain), its square (D_rain_sq), and the log-changes in the commodity price index (p_index). The data can be downloaded from the *Journal of Applied Econometrics* website. Accounting for the endogeneity of foreign aid, Bruckner then estimates that a 1% increase in foreign aid increased real per capita GDP growth by around 0.1% points. Problem 7.16 asks the reader to replicate Table I of Bruckner (2013) relating the effect of economic growth on foreign aid. Here, in the spirit of Bruckner, we estimate a system of two equations

Table 7.4 Fixed Effects 3SLS for Economic Growth and Foreign Aid

```
. reg3 (D.lgdp D.laid p_index D.rain D.rain_sq time* Iccode*) (D.laid D.lgdp
D.polity2 war time* Iccode*)
Three-stage least-squares regression
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lgdp       |
        laid |
         D1. |   .1196744   .1493232     0.80   0.423    -.1729936    .4123425
             |
     p_index |   .4760824   .2910518     1.64   0.102    -.0943686    1.046533
             |
        rain |
         D1. |   .2732289   .1245043     2.19   0.028      .029205    .5172529
             |
     rain_sq |
         D1. |  -.0187174   .0086531    -2.16   0.031    -.0356772   -.0017576
             |
       _cons |   -.011686    .027273    -0.43   0.668    -.0651401     .041768
-------------+----------------------------------------------------------------
D_laid       |
        lgdp |
         D1. |  -3.800951   1.900492    -2.00   0.046    -7.525848   -.0760546
             |
     polity2 |
         D1. |    .016687   .0083164     2.01   0.045     .0003871     .032987
             |
         war |  -.0735758   .0778736    -0.94   0.345    -.2262053    .0790536
------------------------------------------------------------------------------
```

with FE-3SLS. The first equation relates the change in log GDP per capita to the change in log foreign aid, the log-changes in the commodity price index, log rain and its square, time and country fixed effects. The second equation relates the change in log foreign aid to the change in log GDP, changes in political institutions (D.polity2), the incidence of civil war (war) along with time and country fixed effects. Table 7.4 shows these results using *reg3* in Stata. This is 3SLS with dummy variables for countries and time. The results corroborate the significant negative effect of 1% change in log GDP per capita on log of foreign aid, although this effect is smaller 3.8%. Also, it corroborates the effect of 1% in log of foreign aid on economic growth is 0.1%, but this is insignificant. Note that the time and country dummy variables are not shown to conserve space.

7.4 The Hausman and Taylor Estimator

Let us reconsider the single equation estimation case but now focus on endogeneity occurring through the unobserved individual effects. Examples where μ_i and the explanatory variables may be correlated include an earnings equation, where the unobserved individual ability may be correlated with schooling and experience; also a production function, where managerial ability may be correlated with the inputs. Mundlak (1978) considered the one-way error component regression model in (2.5) but with the additional auxiliary regression:

$$\mu_i = \bar{X}'_{i.}\pi + \epsilon_i \tag{7.33}$$

where $\epsilon_i \sim \text{IIN}(0, \sigma_\epsilon^2)$ and $\bar{X}'_{i.}$ is $1 \times K$ vector of observations on the explanatory variables averaged over time. In other words, Mundlak assumed that the individual effects are a linear function of the averages of *all* the explanatory variables across time. These effects are uncorrelated with the explanatory variables if and only if $\pi = 0$. Mundlak (1978) assumed, without loss of generality, that the X are deviations from their sample mean. In vector form, one can write (7.33) as

$$\mu = Z'_\mu X\pi/T + \epsilon \tag{7.34}$$

where $\mu' = (\mu_1, \ldots, \mu_N)$, $Z_\mu = I_N \otimes \iota_T$ and $\epsilon' = (\epsilon_1, \ldots, \epsilon_N)$. Substituting (7.34) in (2.5), with no constant, one gets

$$y = X\beta + PX\pi + (Z_\mu\epsilon + \nu) \tag{7.35}$$

where $P = I_N \otimes \bar{J}_T$. Using the fact that the ϵ and the υ are uncorrelated, the new error in (7.35) has zero mean and variance–covariance matrix

$$V = E(Z_\mu\epsilon + \nu)(Z_\mu\epsilon + \nu)' = \sigma_\epsilon^2(I_N \otimes J_T) + \sigma_\nu^2 I_{NT} \tag{7.36}$$

Using partitioned inverse, one can verify (see problem 7.8) that GLS on (7.35) yields

$$\widehat{\beta}_{GLS} = \widetilde{\beta}_{Within} = (X'QX)^{-1}X'Qy \tag{7.37}$$

and

$$\widehat{\pi}_{GLS} = \widehat{\beta}_{Between} - \widetilde{\beta}_{Within} = (X'PX)^{-1}X'Py - (X'QX)^{-1}X'Qy \tag{7.38}$$

with

$$\begin{aligned} \operatorname{var}(\widehat{\pi}_{GLS}) &= \operatorname{var}(\widehat{\beta}_{Between}) + \operatorname{var}(\widetilde{\beta}_{Within}) \\ &= (T\sigma_\epsilon^2 + \sigma_\nu^2)(X'PX)^{-1} + \sigma_\nu^2(X'QX)^{-1} \end{aligned} \tag{7.39}$$

Problem 7.8 also shows that OLS and GLS on (7.35) are equivalent. Therefore, Mundlak (1978) showed that the best linear unbiased estimator of (2.5) becomes the *fixed effects* (Within) estimator once these individual effects are modeled as a linear function of *all* the X_{it} as in (7.33). The random effects estimator on the other hand is biased because it ignores (7.33). Note that Hausman's test based on the Between minus Within estimators is basically a test for H_0; $\pi = 0$ and this turns out to be another natural derivation for the test considered in Chap. 4, namely,

$$\widehat{\pi}'_{GLS}(\operatorname{var}(\widehat{\pi}'_{GLS}))^{-1}\widehat{\pi}_{GLS} \overset{H_0}{\rightarrow} \chi^2_K$$

Mundlak's (1978) formulation in (7.35) assumes that *all* the explanatory variables are related to the individual effects. The random effects model on the other hand assumes *no* correlation between the explanatory variables and the individual effects. The random effects model generates the GLS estimator, whereas Mundlak's formulation produces the Within estimator. Instead of this " *all* or *nothing*" correlation among the X and the μ_i, Hausman and Taylor (1981) consider a model where some of the explanatory variables are related to the μ_i. In particular, they consider the following model:

$$y_{it} = X_{it}\beta + Z_i\gamma + \mu_i + \nu_{it} \tag{7.40}$$

where the Z_i are cross-sectional time-invariant variables. Hausman and Taylor (1981), hereafter HT, split X and Z into two sets of variables: $X = [X_1; X_2]$ and $Z = [Z_1; Z_2]$ where X_1 is $n \times k_1$, X_2 is $n \times k_2$, Z_1 is $n \times g_1$, Z_2 is $n \times g_2$ and $n = NT$. X_1 and Z_1 are assumed exogenous in that they are not correlated with μ_i, and ν_{it} while X_2 and Z_2 are endogenous because they are correlated with the μ_i, but not the ν_{it}. The Within transformation would sweep the μ_i and remove the bias, but in the process it would also remove the Z_i and hence the Within estimator will not give an estimate of γ. To get around that HT suggest premultiplying the model by $\Omega^{-1/2}$ and using the following set of instruments; $A_0 = [Q, X_1, Z_1]$, where $Q = I_{NT} - P$ and $P = (I_N \otimes \bar{J}_T)$. Breusch, Mizon and Schmidt (1989), hereafter BMS, show that this set of instruments yields the same projection and is therefore equivalent to another set, namely $A_{HT} = [QX_1, QX_2, PX_1, Z_1]$. The latter set of instruments A_{HT} is feasible, whereas A_0 is not. The order condition for identification gives the result that k_1 the number of variables in X_1 must be at least as large as g_2 the number of variables in Z_2. Note that $\widetilde{X}_1 = QX_1$, $\widetilde{X}_2 = QX_2$, $\bar{X}_1 = PX_1$ and Z_1 are used as instruments. Therefore, X_1 is used twice, once as averages and another time as deviations from these averages. This is an advantage of panel data allowing instruments from *within* the model. Note that the Within transformation wipes out the Z_i

and does not allow the estimation of γ. In order to get consistent estimates of γ, HT propose obtaining the Within residuals and averaging them over time

$$\widehat{d}_i = \bar{y}_{i.} - \bar{X}'_{i.}\widetilde{\beta}_W \tag{7.41}$$

Then, (7.40) averaged over time can be estimated by running 2SLS of $\widehat{d}_i$ on Z_i with the set of instruments $A = [X_1, Z_1]$. This yields

$$\widehat{\gamma}_{2SLS} = (Z'P_A Z)^{-1}Z'P_A\widehat{d} \tag{7.42}$$

where $P_A = A(A'A)^{-1}A'$. It is clear that the order condition has to hold ($k_1 \geqslant g_2$) for $(Z'P_A Z)$ to be nonsingular. Next, the variance-components estimates are obtained as follows:

$$\widetilde{\sigma}^2_\nu = \widetilde{y}'\bar{P}_{\widetilde{X}}\widetilde{y}/N(T-1) \tag{7.43}$$

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$, $\bar{P}_A = I - P_A$ and

$$\widetilde{\sigma}^2_1 = \frac{(y_{it} - X_{it}\widetilde{\beta}_W - Z_i\widehat{\gamma}_{2SLS})'P(y_{it} - X_{it}\widetilde{\beta}_W - Z_i\widehat{\gamma}_{2SLS})}{N} \tag{7.44}$$

This last estimate is based upon an NT vector of residuals. Once the variance-components estimates are obtained, the model in (7.40) is transformed using $\widehat{\Omega}^{-1/2}$ as follows:

$$\widehat{\Omega}^{-1/2}y_{it} = \widehat{\Omega}^{-1/2}X_{it}\beta + \widehat{\Omega}^{-1/2}Z_i\gamma + \widehat{\Omega}^{-1/2}u_{it} \tag{7.45}$$

The HT estimator is basically 2SLS on (7.45) using $A_{HT} = [\widetilde{X}, \bar{X}_1, Z_1]$ as a set of instruments.

(1) If $k_1 < g_2$, then the equation is under-identified. In this case, $\widehat{\beta}_{HT} = \widetilde{\beta}_W$ and $\widehat{\gamma}_{HT}$ does not exist.
(2) If $k_1 = g_2$, then the equation is just-identified. In this case, $\widehat{\beta}_{HT} = \widetilde{\beta}_W$ and $\widehat{\gamma}_{HT} = \widehat{\gamma}_{2SLS}$ given by (7.42).
(3) If $k_1 > g_2$, then the equation is over-identified and the HT estimator obtained from (7.45) is more efficient than the Within estimator.

A test for over-identification is obtained by computing

$$\widehat{m} = \hat{q}'[\text{var}(\widetilde{\beta}_W) - \text{var}(\widehat{\beta}_{HT})]^{-}\widehat{q} \tag{7.46}$$

with $\widehat{q} = \widehat{\beta}_{HT} - \widetilde{\beta}_W$ and $\widehat{\sigma}^2_\nu\widehat{m} \overset{H_0}{\rightarrow} \chi^2_l$ where $l = \min[k_1 - g_2, NT - K]$.

Note that $y^* = \widehat{\sigma}_\nu\widehat{\Omega}^{-1/2}y$ has a typical element $y^*_{it} = y_{it} - \widehat{\theta}\bar{y}_{i.}$ where $\widehat{\theta} = 1 - \widehat{\sigma}_\nu/\widehat{\sigma}_1$ and similar terms exist for X^*_{it} and Z^*_i. In this case, 2SLS on (7.45) yields

$$\begin{pmatrix}\widehat{\beta}\\ \widehat{\gamma}\end{pmatrix} = \left[\begin{pmatrix}X^{*\prime}\\ Z^{*\prime}\end{pmatrix}P_A(X^*, Z^*)\right]^{-1}\begin{pmatrix}X^{*\prime}\\ Z^{*\prime}\end{pmatrix}P_A y^* \tag{7.47}$$

where P_A is the projection matrix on $A_{HT} = [\widetilde{X}, \bar{X}_1, Z_1]$.

Amemiya and MaCurdy (1986), hereafter AM, suggest a more efficient set of instruments $A_{AM} = [QX_1, QX_2, X^*_1, Z_1]$ where $X^*_1 = X^0_1 \otimes \iota_T$ and

$$X_1^0 = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1T} \\ \vdots & \vdots & \dots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{NT} \end{bmatrix} \tag{7.48}$$

is an $N \times k_1 T$ matrix. So X_1 is used $(T+1)$ times, once as $\widetilde{X}_1$ and T times as X_1^*. The order condition for identification is now more likely to be satisfied $(Tk_1 > g_2)$. However, this set of instruments requires a stronger exogeneity assumption than that of Hausman and Taylor (1981). The latter requires only uncorrelatedness of the mean of X_1 from the μ_i, i.e.,

$$\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\bar{X}_{1i.}\mu_i\right) = 0$$

while Hausman and Taylor (1981) require

$$\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^{N}X_{1it}\mu_i\right) = 0 \quad \text{for} \quad t = 1, \dots, T$$

i.e., uncorrelatedness at each point in time. Breusch, Mizon and Schmidt (1989) suggest yet a more efficient set of instruments

$$A_{BMS} = [\widetilde{X}, \bar{X}_1, (\widetilde{X})^*, Z_1]$$

so that X_1 is used $(T+1)$ times and X_2 is used T times. This requires even more exogeneity assumptions, i.e., $\widetilde{X}_2 = QX_2$ should be uncorrelated with the μ_i effects. The BMS order condition becomes $Tk_1 + (T-1)k_2 \geqslant g_2$.

Computational Note: The number of instruments used by the AM and BMS procedures can increase rapidly as T and the number of variables in the equation get large. For large N panels, small T and reasonable k, this should not be a problem. However, even for $T = 7$, $k_1 = 4$, and $k_2 = 5$ as in the empirical illustration used in the next section, the number of additional instruments used by HT are 4 as compared to 28 for AM and 58 for BMS.[4]

7.5 Empirical Example: Earnings Equation Using PSID Data

Cornwell and Rupert (1988) apply these three instrumental variable (IV) methods to a returns to schooling example based on a panel of 595 individuals observed over the period 1976-82 and drawn from the Panel Study of Income Dynamics (PSID). A description of the data is given in Cornwell and Rupert (1988) and is put on the Springer website as Wage.xls. In particular, log wage is regressed on years of education (ED), weeks worked (WKS), years of full-time work experience (EXP), occupation (OCC = 1, if the individual is in a blue-collar occupation), residence

Table 7.5 Mincer Wage Equation. Dependent Variable: Log Wage*

	GLS	*Within*	*HT*	*AM*
Constant	4.264 (0.098)	–	2.913 (0.284)	2.927 (0.275)
WKS	0.0010 (0.0008)	0.0008 (0.0006)	0.0008 (0.0006)	0.0008 (0.0006)
SOUTH	−0.017 (0.027)	−0.002 (0.034)	0.007 (0.032)	0.007 (0.032)
SMSA	−0.014 (0.020)	−0.042 (0.019)	−0.042 (0.019)	−0.042 (0.019)
MS	−0.075 (0.023)	−0.030 (0.019)	−0.030 (0.019)	−0.030 (0.019
EXP	0.082 (0.003)	0.113 (0.002)	0.113 (0.002)	0.113 (0.002)
EXP^2	−0.0008 (0.00006)	−0.0004 (0.00005)	−0.0004 (0.00005)	−0.0004 (0.00005)
OCC	−0.050 (0.017)	−0.021 (0.014)	−0.021 (0.014)	−0.021 (0.014)
IND	0.004 (0.017)	0.019 (0.015)	0.014 (0.015)	0.014 (0.015)
UNION	0.063 (0.017)	0.033 (0.015)	0.033 (0.015)	0.032 (0.015)
FEM	−0.339 (0.051)	–	−0.131 (0.127)	−0.132 (0.127)
BLK	−0.210 (0.058)	–	−0.286 (0.156)	−0.286 (0.155)
ED	0.100 (0.006)	–	0.138 (0.021)	0.137 (0.021)
		$\chi_9^2 = 5075$	$\chi_3^2 = 5.26$	$\chi_{13}^2 = 14.74$

* X_2 = (OCC, SOUTH, SMSA, IND), Z_1 = (FEM, BLK)
Source Baltagi and Khanti-Akom (1990). Reproduced by permission of John Wiley & Sons Ltd

(SOUTH = 1, SMSA = 1, if the individual resides in the South, or in a standard metropolitan statistical area), industry (IND = 1, if the individual works in a manufacturing industry), marital status (MS = 1, if the individual is married), sex and race (FEM = 1, BLK = 1, if the individual is female or black), union coverage (UNION = 1, if the individual's wage is set by a union contract) and time dummies to capture productivity and price level effects. Baltagi and Khanti-Akom (1990) replicate this study, and some of their results in Table II are reproduced in Table 7.5. The conventional GLS indicates that an additional year of schooling produces a 10% wage gain. But conventional GLS does not account for the possible correlation of the explanatory variables with the individual effects. The Within transformation eliminates the

Table 7.6 Hausman and Taylor estimates of a mincer wage equation

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog (exp exp2
wks ms union ed)

Hausman-Taylor estimation                       Number of obs      =      4165
Group variable (i): id                          Number of groups   =       595

                                                Obs per group: min =         7
                                                               avg =         7
                                                               max =         7

Random effects u_i ~ i.i.d.                     Wald chi2(12)      =   6891.87
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         occ |  -.0207047   .0137809    -1.50   0.133    -.0477149    .0063055
       south |   .0074398    .031955     0.23   0.816    -.0551908    .0700705
        smsa |  -.0418334   .0189581    -2.21   0.027    -.0789906   -.0046761
         ind |   .0136039   .0152374     0.89   0.372    -.0162608    .0434686
TVendogenous |
         exp |   .1131328    .002471    45.79   0.000     .1082898    .1179758
        exp2 |  -.0004189   .0000546    -7.67   0.000    -.0005259   -.0003119
         wks |   .0008374   .0005997     1.40   0.163    -.0003381    .0020129
          ms |  -.0298508     .01898    -1.57   0.116    -.0670508    .0073493
       union |   .0327714   .0149084     2.20   0.028     .0035514    .0619914
TIexogenous  |
         fem |  -.1309236    .126659    -1.03   0.301    -.3791707    .1173234
         blk |  -.2857479   .1557019    -1.84   0.066    -.5909179    .0194221
TIendogenous |
          ed |    .137944   .0212485     6.49   0.000     .0962977    .1795902
             |
       _cons |   2.912726   .2836522    10.27   0.000     2.356778    3.468674
-------------+----------------------------------------------------------------
     sigma_u |  .94180304
     sigma_e |  .15180273
         rho |  .97467788   (fraction of variance due to u_i)
------------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.
```

individual effects and all the Z_i variables, and the resulting Within estimator is consistent even if the individual effects are correlated with the explanatory variables. The Within estimates are quite different from those of GLS, and the Hausman test based on the difference between these two estimates yields $\chi^2_9 = 5075$ which is significant. This rejects the hypothesis of no correlation between the individual effects and the explanatory variables. This justifies the use of the IV methods represented as HT and AM in Table 7.5. We let X_1 = (OCC, SOUTH, SMSA, IND), X_2 = (EXP, EXP^2, WKS, MS, UNION), Z_1 = (FEM, BLK), and Z_2 = (ED). Table 7.6 reproduces the Hausman and Taylor (1981) estimates using the (*xthtaylor*) command in Stata. The coefficient of ED is estimated as 13.8%, thirty-eight percent higher than the estimate obtained using GLS (10%). A Hausman test based on the difference between HT and the Within estimator yields $\chi^2_3 = 5.26$ which is not significant at the 5% level. There are three degrees of freedom since there are three over-identifying conditions (the number of X_1 variables minus the number of Z_2 variables).

Therefore, we cannot reject that the set of instruments X_1 and Z_1 chosen are legitimate. Table 7.7 reproduces the Amemiya and MaCurdy (1986) estimates using

Table 7.7 Amemiya and MaCurdy estimates of a mincer wage equation

```
. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog
(exp exp2 wks ms union ed) amacurdy

Amemiya-MaCurdy estimation                       Number of obs      =      4165
Group variable (i): id                           Number of groups   =       595

                                                 Obs per group: min =         7
                                                                avg =         7
                                                                max =         7

Random effects u_i ~ i.i.d.                      Wald chi2(12)      =   6879.20
                                                 Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         occ |  -.0208498   .0137653    -1.51   0.130    -.0478292    .0061297
       south |   .0072818   .0319365     0.23   0.820    -.0553126    .0698761
        smsa |  -.0419507   .0189471    -2.21   0.027    -.0790864   -.0048149
         ind |   .0136289    .015229     0.89   0.371    -.0162194    .0434771
TVendogenous |
         exp |   .1129704   .0024688    45.76   0.000     .1081316    .1178093
        exp2 |  -.0004214   .0000546    -7.72   0.000    -.0005283   -.0003145
         wks |   .0008381   .0005995     1.40   0.162    -.0003368     .002013
          ms |  -.0300894   .0189674    -1.59   0.113    -.0672649    .0070861
       union |   .0324752   .0148939     2.18   0.029     .0032837    .0616667
TIexogenous  |
         fem |   -.132008   .1266039    -1.04   0.297     -.380147    .1161311
         blk |  -.2859004   .1554857    -1.84   0.066    -.5906468    .0188459
TIendogenous |
          ed |   .1372049   .0205695     6.67   0.000     .0968894    .1775205
             |
       _cons |   2.927338   .2751274    10.64   0.000     2.388098    3.466578
-------------+----------------------------------------------------------------
     sigma_u |  .94180304
     sigma_e |  .15180273
         rho |  .97467788   (fraction of variance due to u_i)
------------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.
```

the (*xthtaylor*) command in Stata with the (*amacurdy*) option. These estimates are close to the HT estimates. The additional exogeneity assumptions needed for the AM estimator are not rejected using a Hausman test based on the difference between the HT and AM estimators. This yields $\chi^2_{13} = 14.74$ which is not significant at the 5 level. The BMS estimates (not reported here but available in Baltagi and Khanti-Akom (1990)) are similar to those of AM. Again, the additional exogeneity assumptions needed for the BMS estimator are not rejected using a Hausman test based on the difference between the AM and BMS estimators. This yields a $\chi^2_{13} = 9.59$ which is not significant at the 5% level.

Other applications of the Hausman–Taylor estimator include the following:

(1) Contoyannis and Rice (2001) for a study of the impact of health on wages. In particular, this paper considers the effect of self-assessed general and psychological health on hourly wages using longitudinal data from the six waves of the British Household Panel Survey. Contoyannis and Rice show that reduced psychological health reduces the hourly wage for males, while excellent self-assessed health increases the hourly wage for females.

(2) Egger and Pfaffermayr (2004) for a study of the effects of distance as a common determinant of exports and foreign direct investment (FDI) in a three-factor New

Trade Theory model. They used industry-level data of bilateral outward FDI stocks and exports of the U.S. and Germany to other countries between 1989 and 1999. They find that distance exerts a positive and significant impact on bilateral stocks of outward FDI of both the U.S. and Germany. However, the effect of distance on exports is much smaller in absolute size and significantly negative for the U.S. but insignificant for Germany. Problem 7.17 asks the reader to replicate the results of this paper.

(3) Serlenga and Shin (2007) who apply the Hausman–Taylor estimation methodology to the gravity equation of bilateral trade flows among 15 European countries over the period 1960–2001. Among their findings is that the impact of country-specific variables, like distance, common language and common border can be recovered using this approach in addition to allowing such variables to be endogenous. In particular, they argue that common language is a proxy for cultural, historical, and linguistic proximity, and this in turn is highly correlated with country-specific effects. Problem 7.15 asks the reader to replicate the results of this paper.

7.6 Further Reading

Cornwell, Schmidt and Wyhowski (1992) consider a simultaneous equation model with error components that distinguishes between two types of exogenous variables, namely *singly exogenous* and *doubly exogenous* variables. A singly exogenous variable is correlated with the individual effects but not with the remainder noise. These are given the subscript (2). On the other hand, a doubly exogenous variable is uncorrelated with both the effects and the remainder disturbance term. These are given the subscript (1). Cornwell, Schmidt and Wyhowski extend the results of HT, AM, and BMS by transforming each structural equation by its $\Omega^{-1/2}$ and applying 2SLS on the transformed equation using $A = [QX, PB]$ as the set of instruments in (7.47). B is defined as follows:

(1) $B_{HT} = [X_{(1)}, Z_{(1)}]$ for the Hausman and Taylor (1981) type estimator. This B_{HT} is the set of all *doubly* exogenous variables in the system.
(2) $B_{AM} = [X^*_{(1)}, Z_{(1)}]$ for the Amemiya and MaCurdy (1986) type estimator. The (*) notation has been defined in (7.48).
(3) $B_{BMS} = [X^*_{(1)}, Z_{(1)}, (QX_{(2)})^*]$ for the Breusch, Mizon and Schmidt (1989) type estimator. Cornwell, Schmidt and Wyhowski (1992) also derive a similar set of instruments for the 3SLS analogue and give a generalized method of moments interpretation to these estimators. Finally, they consider the possibility of a different set of instruments for each equation, say $A_j = [QX, PB_j]$ for the jth equation, where for the HT type estimator, B_j consists of all doubly exogenous variables of equation j (i.e., exogenous variables that are uncorrelated with the individual effects in equation j). Wyhowski (1994) extends the HT, AM, and BMS approaches to the two-way error component model and gives the appropriate set of instruments.

Baltagi and Chang (2000) compare the performance of several single and system estimators of a two-equation simultaneous model with unbalanced panel data. The Monte Carlo design varies the degree of unbalancedness in the data and the variance components ratio due to the individual effects. Many of the results obtained for the simultaneous equation error component model with balanced data carry over to the unbalanced case. For example, both feasible EC2SLS estimators considered performed reasonably well and it is hard to choose between them. Simple ANOVA methods can still be used to obtain good estimates of the structural and reduced form parameters even in the unbalanced panel data case. Replacing negative estimates of the variance components by zero did not seriously affect the performance of the corresponding structural or reduced form estimates. Better estimates of the structural variance components do not necessarily imply better estimates of the structural coefficients. Finally, do not make the data balanced to simplify the computations. The loss in root mean squared error can be huge.

Most applied work in economics have made the choice between the RE and FE estimators based upon the standard Hausman (1978) test. This is based upon a contrast between the FE and RE estimators. If this standard Hausman test rejects the null hypothesis that the conditional mean of the disturbances given the regressors is zero, the applied researcher reports the FE estimator. Otherwise, the researcher reports the RE estimator; see the discussion in Chap. 4 and the two empirical applications by Owusu-Gyapong (1986) and Cardellichio (1990). Baltagi, Bresson and Pirotte (2003) suggest an alternative *pretest* estimator based on the Hausman and Taylor (1981) model. This pretest estimator reverts to the RE estimator if the standard Hausman test based on the FE versus the RE estimators is not rejected. It reverts to the HT estimator if the choice of strictly exogenous regressors is not rejected by a second Hausman test based on the difference between the FE and HT estimators. Otherwise, this pretest estimator reverts to the FE estimator. In other words, this pretest alternative suggests that the researcher consider a Hausman–Taylor model where some of the variables, but not all, may be correlated with the individual effects. Monte Carlo experiments were performed to compare the performance of this pretest estimator with the standard panel data estimators under various designs. The estimators considered were ordinary least squares (OLS), fixed effects (FE), random effects (RE), and the Hausman–Taylor (HT) estimators. In one-design, some regressors were correlated with the individual effects, i.e., a Hausman–Taylor world. In another design, the regressors were not allowed to be correlated with the individual effects, i.e., a RE world. Results showed that the pretest estimator is a viable estimator and is always second best to the efficient estimator. It is second in RMSE performance to the RE estimator in a RE world and second to the HT estimator in an HT world. The FE estimator is a consistent estimator under both designs but its disadvantage is that it does not allow the estimation of the coefficients of the time-invariant regressors. When there is endogeneity among the regressors, Baltagi, Bresson and Pirotte (2003) show that there is substantial bias in OLS and the RE estimators and both yield misleading inference. Even when there is no correlation between the individual effects and the regressors, i.e., in a RE world, inference based on OLS can be seriously misleading. This last result was emphasized by Moulton (1986).

Baltagi and Liu (2012) extend the HT estimator to allow for serial correlation of the AR(1) type in the remainder disturbances. They demonstrate the gains in efficiency of this estimator versus the standard panel data estimators that ignore serial correlation using Monte Carlo experiments. This estimator computes a fixed effects Prais–Winsten (PW) GLS estimator in the first step, rather than the usual FE estimator used by Hausman and Taylor (1981) in the absence of serial correlation. Averaging the residuals from this regression over time as in HT but now weighting the initial period differently from the rest as in Baltagi and Li (1991), one performs 2sls as in the second step of the HT estimator to retrieve consistent estimates of the time-invariant coefficients. With consistent estimates of the residuals, new estimates of the variance components are computed as in HT and the final IV-GLS step is performed on the PW transformed model.

Baltagi and Bresson (2012) apply the useful robust panel data methods suggested by Bramati and Croux (2007) and Wagenvoort and Waldmann (2002) to the Hausman and Taylor (1981) estimator. They demonstrate using Monte Carlo experiments the substantial gains in efficiency as measured by MSE of this robust HT estimator over its classical counterpart. The magnitude of the gains in MSE depend upon the *type* and *degree* of contamination of the observations. They illustrate this robust HT method by applying it to the classical Mincer wage equation using the empirical study of Cornwell and Rupert (1988). For this empirical study, the returns to education seem to be robust to outliers, while the magnitude and significance of the female coefficient is sensitive to robustification of the HT estimator.

7.7 Notes

1. The analysis in this chapter can be easily extended to the two-way error component model; see the problems at the end of this chapter and Baltagi (1981b).
2. As in the classical regression case, the variances of W2SLS have to be adjusted by the factor $(NT - k_1 - g_1 + 1)/[N(T-1) - k_1 - g_1 + 1]$, whenever 2SLS is performed on the Within transformed equation. Note also that the set of instruments is $\widetilde{X}$ and not X as emphasized in (7.6).
3. This is analogous to the result found in the single equation error component literature by Taylor (1980) and Baltagi (1981a).
4. Im et al. (1999) point out that for panel data models, the exogeneity assumptions imply many more moment conditions than the standard random and fixed effects estimators use. Im et al. (1999) provide the assumptions under which the efficient GMM estimator based on the entire set of available moment conditions reduces to these simpler estimators. In other words, the efficiency of the simple estimators is established by showing the redundancy of the moment conditions that they do not use.

7.8 Problems

7.1 *Within 2SLS and Between 2SLS*. Verify that GLS on (7.7) yields (7.6) and GLS on (7.9) yields (7.8), the Within 2SLS and Between 2SLS estimators of δ_1, respectively.

7.2 *Error component two-stage least squares*. Verify that GLS on (7.10) yields the EC2SLS estimator of δ_1 given in (7.11) (see Baltagi (1981b)).

7.3 *Equivalence of several EC2SLS estimators*. (a) Show that $A = [\widetilde{X}, \bar{X}]$; $B = [X^*, \widetilde{X}]$ and $C = [X^*, \bar{X}]$ yield the same projection, i.e., $P_A = P_B = P_C$ and hence the same EC2SLS estimator given by (7.11) (see Baltagi and Li (1992)).
(b) Show that $P_A P_{X^*} = P_{X^*}$, and that $P_A - P_{X^*}$ is idempotent. Use this result to prove that the $avar\left(\sqrt{n}\widehat{\beta}_{G2SLS}\right) - avar\left(\sqrt{n}\widehat{\beta}_{EC2SLS}\right)$ is positive semi-definite, where $avar$ denotes asymptotic variance and $n = NT$. While this result may not be of consequence asymptotically since both estimators are asymptotically efficient. It may lead to smaller standard errors in finite samples. The intuition comes from the fact that extra instruments will in general lead to more efficient estimators, and in small samples, to lower standard errors. See problem 7.11 for an empirical illustration.

7.4 *Within 3SLS and Between 3SLS*. Verify that 3SLS on (7.21) with $(I_M \otimes \widetilde{X})$ as the set of instruments yields (7.22). Similarly, verify that 3SLS on (7.23) with $(I_M \otimes \bar{X})$ as the set of instruments yields (7.24). These are the Within 3SLS and Between 3SLS estimators of δ_1, respectively.

7.5 *Error component three-stage least squares*. Verify that GLS on the stacked system (7.21) and (7.23) each premultiplied by $(I_M \otimes \widetilde{X}')$ and $(I_M \otimes \bar{X}')$, respectively, yields the EC3SLS estimator of δ given in (7.25) (see Baltagi (1981b)).

7.6 *Equivalence of several EC3SLS estimators*. (a) Prove that $A = (I_M \otimes \widetilde{X}, I_M \otimes \bar{X})$ yields the same projection as $B = (H \otimes \widetilde{X}, G \otimes \bar{X})$ or $C = [(H \otimes \widetilde{X} + G \otimes \bar{X}), H \otimes \widetilde{X}]$ or $D = [H \otimes \widetilde{X} + G \otimes \bar{X}), G \otimes \bar{X}]$ where H and G are nonsingular $M \times M$ matrices (see Baltagi and Li 1992). Conclude that these sets of instruments yield the same EC3SLS estimator of δ given by (7.25).
(b) Let $H = \Sigma_\nu^{-1/2}$ and $G = \Sigma_1^{-1/2}$, and note that A is the set of instruments proposed by Baltagi (1981b) while $B = (\Sigma_\nu^{-1/2} \otimes \widetilde{X}, \Sigma_1^{-1/2} \otimes \bar{X})$ is the optimal set of instruments X^* defined (7.30). Conclude that $H \otimes \widetilde{X}$ is redundant in C and $G \otimes \bar{X}$ is redundant in D with respect to the optimal set of instruments X^*.
(c) Show that $P_A = P_{I_M \otimes \tilde{X}} + P_{I_M \otimes \bar{X}}$, and that $P_A P_B = P_B$ where $B = (\Sigma_\nu^{-1/2} \otimes \widetilde{X}, \Sigma_1^{-1/2} \otimes \bar{X})$. Use this result to prove that the $avar\left(\sqrt{n}\hat{\delta}_{E3SLS}\right) - avar\left(\sqrt{n}\hat{\delta}_{EC3SLS}\right)$ is positive semi-definite.

7.7 *Special cases of the simultaneous equations model with one-way error component disturbances*. (a) Consider a system of two structural equations with one-way error component disturbances. Show that if the disturbances between the two equations are *uncorrelated*, then EC3SLS is equivalent to EC2SLS (see Baltagi (1981b)).

(b) Show that if this system of two equations with one-way error component disturbances is *just-identified*, then EC3SLS does not necessarily reduce to EC2SLS (see Baltagi (1981b)).

7.8 *Mundlak's fixed effects result*. (a) Using partitioned inverse, show that GLS on (7.35) yields $\widehat{\beta}_{GLS} = \widetilde{\beta}_{Within}$ and $\widehat{\pi}_{GLS} = \widehat{\beta}_{Between} - \widetilde{\beta}_{Within}$ as given in (7.37) and (7.38).
(b) Verify that $\text{var}(\widehat{\pi}_{GLS}) = \text{var}(\widehat{\beta}_{Between}) + \text{var}(\widetilde{\beta}_{Within})$ as given in (7.39).
(c) Show that Mundlak's result can also be obtained using *system estimation* without invoking partitioned inversion. Premultiply (7.35) by P

$$Py = PX(\beta + \pi) + P\eta$$

Here, we define $\eta = Z_\mu \epsilon + \nu$, and we use the fact that $P^2 = P$ and $PZ_\mu = Z_\mu$. Note that OLS or GLS on this equation yields $\widehat{(\beta + \pi)} = (X'PX)^{-1}X'Py$ which is the Between estimator. Similarly, premultiplying (7.35) by Q one gets

$$Qy = QX\beta + Q\nu$$

since $QP = 0$. OLS or GLS on this equation yields $\widetilde{\beta}_{Within} = (X'QX)^{-1}X'Qy$ which is the usual Within or Fixed Effects estimator. Stacking the system of equations, we get

$$\begin{pmatrix} Py \\ Qy \end{pmatrix} = \begin{pmatrix} PX \\ QX \end{pmatrix}\beta + \begin{pmatrix} PX \\ 0 \end{pmatrix}\pi + \begin{pmatrix} P\eta \\ Q\nu \end{pmatrix}$$

and the system error vector has mean 0 and variance–covariance matrix given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 P & 0 \\ 0 & \sigma_\nu^2 Q \end{pmatrix}$$

Show that OLS or GLS on this system yields the same results that Mundlak found by applying GLS to (7.35).
(d) Prove that the Zyskind (1967) necessary and sufficient condition for OLS to be equivalent to GLS on the system of equations given in part (c) is satisfied. This calls for $P_Z\Sigma = \Sigma P_Z$, where $Z = \begin{pmatrix} PX & PX \\ QX & 0 \end{pmatrix}$ is the matrix of regressors and Σ is the variance–covariance matrix of its disturbances. See the solution in Baltagi (2006b).
(e) Show that Mundlak's (1978) result could have been obtained with OLS rather than GLS on (7.35). Prove that Zyskind's (1967) necessary and sufficient condition for OLS to be equivalent to GLS holds for (7.35). (Hint: rewrite X as $(P + Q)X$ in (7.35) and collect like terms before applying OLS).

7.9 *EC2SLS and EC3SLS for the two-way error component Model*. Consider the two-way error component model given in (6.9) and the covariance matrix Ω_{jl} between the jth and lth equation disturbances given in (6.11):
(a) Derive the EC2SLS estimator for δ_1 in (7.1).
(b) Derive the EC3SLS estimator for δ in (7.20) (Hint: See Baltagi (1981b).
(c) Repeat problem 7.7 parts (a) and (b) for the two-way error component EC2SLS and EC3SLS.

7.10 Using the Monte Carlo setup for a two-equation simultaneous model with error component disturbances, given in Baltagi (1984), compare EC2SLS and EC3SLS with the usual 2SLS and 3SLS estimators that ignore the error component structure.

7.11 *Crime in North Carolina*. Using the Cornwell and Trumbull (1994) panel data set described in the empirical example in Sect. 7.1 and given on the Springer website as crime.dat, replicate Table 7.1 and the associated test statistics. See also the replication by Baltagi (2006a). Note that the standard errors of EC2SLS are smaller than those of G2SLS.

7.12 *Mincer wage equation*. Using the Cornwell and Rupert (1988) panel data set described in the empirical example in Sect. 7.4 and given on the Springer website as wage.xls, replicate Table 7.6 and the associated test statistics.

7.13 *A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares*. This is based on Problem 04.1.1 in *Econometric Theory* by Baltagi (2004). Consider the first structural equation of a simultaneous equation panel data model given in (7.1). Hausman (1978) suggests comparing the FE and RE estimators in the classic panel data regression. With endogenous right-hand-side regressors like Y_1 this test can be generalized to test H_0; $E(u_1 \mid Z_1) = 0$ based on $\widehat{q}_1 = \widetilde{\delta}_{1,FE2SLS} - \widehat{\delta}_{1,EC2SLS}$ where $\widetilde{\delta}_{1,FE2SLS}$ is defined in (7.6) and $\widehat{\delta}_{1,EC2SLS}$ is defined in (7.11).

(a) Show that under H_0: $E(u_1 \mid Z_1) = 0$, plim$\widehat{q}_1 = 0$ and the asymptotic cov $(\widehat{q}_1, \widehat{\delta}_{1,EC2SLS}) = 0$.

(b) Conclude that $\text{var}(\widehat{q}_1) = \text{var}(\widetilde{\delta}_{1,FE2SLS}) - \text{var}(\widehat{\delta}_{1,EC2SLS})$, where var denotes the asymptotic variance. This is used in computing the Hausman test statistic given by $m_1 = \hat{q}_1' [\text{var}(\widehat{q}_1)]^{-1} \widehat{q}_1$. Under H_0, m_1 is asymptotically distributed as χ^2_r, where r denotes the dimension of the slope vector of the time varying variables in Z_1. This can be easily implemented using Stata.

(c) Compute the usual Hausman test based on FE and RE and this alternative Hausman test based on FE2SLS and EC2SLS for the crime data considered in problem 7.11. What do you conclude?

(d) Show that Hausman's test based on the contrast between FE2SLS and EC2SLS can be alternatively obtained from any one of the following artificial 2SLS regressions with instruments $A = \left[\tilde{X}, \bar{X}\right]$:

$$y_1^* = Z_1^* \delta_1 + \tilde{Z}_1 \gamma_1 + \omega_1$$
$$y_1^* = Z_1^* \delta_1 + \bar{Z}_1 \gamma_1 + \omega_2$$
$$y_1^* = Z_1^* \delta_1 + Z_1 \gamma_1 + \omega_3$$

Here, $Z_1^* = \Omega_{11}^{-1/2} Z_1$, $\widetilde{Z}_1 = Q Z_1$, and $\bar{Z}_1 = P Z_1$, where $\Omega_{11}^{-1/2}$ is defined in (7.16). $\tilde{X}$, $\bar{X}$ and y^* are similarly defined; see (7.15). Show that Hausman's test is equivalent to testing whether $\gamma_1 = 0$ in any one of these three 2SLS regressions; see Baltagi and Liu (2007).

7.14 *Fixed Effects v.s. Omission of time-invariant variables.* This is based on Oaxaca and Geisler (2003). Consider the case where the fixed effects $\mu_i = Z_i'\pi$ are just omitted time-invariant variables made up of $Z_i's$ which are of dimension 1xg. Note that this may include a constant, but unlike the Hausman and Taylor (1981) or Mundlak (1978) reduced form model, this does not include an error term. The true regression model in this case is $y_{it} = X_{it}'\beta + Z_i'\pi + \nu$ with $\nu \sim \text{IIN}(0, \sigma_\nu^2 I_{NT})$. For this model, pooled OLS is BLUE, but the researcher omits the $Z_i's$, and runs instead fixed effects to estimate β. Oaxaca and Geisler (2003) show that the pooled OLS estimate of π can be obtained as a GLS of fixed effects residuals averaged over time (as in the Hausman–Taylor procedure, see (7.41)), i.e., $\widehat{d}_i = \bar{y}_{i.} - \bar{X}_{i.}'\widetilde{\beta}_{FE}$ on Z_i.

(a) Prove this result, and show that the resulting standard errors from the two regressions will be different due to the different estimates of the remainder variance σ_ν^2.

(b) Oaxaca and Geisler (2003) suggest a Chow F-test based on the difference between the restricted sum of squares from pooled OLS imposing the restriction H_0^a; $\mu_i = Z_i'\pi$, versus the unrestricted sum of squares residuals obtained from the fixed effects regression. Note that this is different from the test for H_0^b; $\pi = 0$, which can also be obtained from a Chow F-test based on the difference between the restricted OLS residual sum of squares of y on X, versus the unrestricted OLS residual sum of squares of y on X and Z. It is also different from the usual F-test for fixed effects, i.e., H_0^c; $\mu_i = 0$ which was introduced in Chap. 2. The latter is based on the difference between the restricted sum of squares residuals from OLS of y on X,versus the unrestricted sum of squares residuals obtained from the fixed effects regression.

Comment: The crucial assumption here is that μ_i is a *known function* of the $Z_i's$, which is restrictive, and unlike the Hausman–Taylor model, μ_i can not be a function of some of the time varying variables, the $X_{it}'s$, nor can it have a stochastic error term. This is why the second step in the Hausman–Taylor procedure is 2SLS and not GLS. In the HT model, the $Z_i's$ are correlated with the individual effects, i.e., the $\mu_i's$, because both of them appear in the regression model; see (7.40). Here, GLS is performed in the second step because it is assumed that the $\mu_i's$ have been replaced by their known functional form $Z_i'\pi$.

7.15 *Gravity models of intra-EU trade*. Using the Serlenga and Shin (2007) panel data set, which can be downloaded from the *Journal of Applied Econometrics* website, replicate Tables II and III of that paper and the associated test statistics. In particular, you are asked to apply the Hausman–Taylor estimation methodology to the gravity equation of bilateral trade flows among 15 European countries over the period 1960–2001. The general model regresses bilateral trade (Trade) on GDP, similarity in relative size (Sim), differences in relative factor endowments between trading partners (Rlf), real exchange rate (Rer), a dummy variable which is 1 when both countries belong to the European community (Cee), a dummy variable which is 1 when both countries adopt a common currency (Emu); distance between capital cities (Dist); common border (Bor); and common language (Lan). (a) replicate the FE results in Table II for the full model; (b) replicate the HT estimation results in Table II; (c) perform the Hausman test for over-identification when GDP, RLF, and RER are used as instruments for Lan in the HT estimation; (d) What happens if you perform the Amemiya and MaCurdy (1986) estimator?

7.16 *Economic growth and foreign aid*. This is the empirical example in Sect. 7.3 based on Bruckner (2013) who investigates the simultaneity problem between per capita GDP growth and foreign aid. The data can be downloaded from the *Journal of Applied Econometrics* website.

(a) Replicate Table I of Bruckner (2013) relating the effect of economic growth on foreign aid.

(b) For the FE-2SLS with both country and time dummies. Show that for this differenced model, the country dummies are jointly insignificant, while the time dummies are jointly significant.

(c) Replicate Table 7.4 in this chapter applying FE-3SLS using *reg3* with time and country dummies in both equations.

7.17 *Distance, Trade, and FDI*. Egger and Pfaffermayr (2004) study the effects of *distance* as a common determinant of exports and foreign direct investment (FDI) in a three-factor New Trade Theory model. They use industry-level data of bilateral outward FDI stocks and exports of the U.S. and Germany to other countries between 1989 and 1999. They find that distance exerts a positive and significant impact on bilateral stocks of outward FDI of both the U.S. and Germany. However, the effect of distance on exports is much smaller in absolute size and significantly negative for the U.S. but insignificant for Germany. Using the Egger and Pfaffermayr (2004) panel data set, which can be downloaded from the *Journal of Applied Econometrics* website, replicate Table II on pages 236–237 of that paper and the associated test statistics. In particular replicate the Within and Hausman–Taylor estimation for real bilateral exports and real bilateral FDI. Do that for US and Germany.

7.18 *Twin Crises*. Hutchison and Noy (2005) investigate the output effects of banking and currency crises in emerging markets, focusing on whether "twin crises" entail large losses. Using a panel data set of 24 emerging market economies over 1975–97, they find that currency (banking) crises are very costly, reducing output by about 5–8% (8–10%) over a 2–4 year period. They find more than 51 currency and 33 banking crises episodes over the past 25 years in their emerging markets sample and 20 occurrences of "twin crises"—currency and banking crises that occurred simultaneously. (a) Replicate their Table 5, column (1), p.743, which runs a Hausman–Taylor model regressing Real GDP growth (dlrgdp) on its lagged value (dlrgdpl), the change in budget surplus to real GDP ratio at period (t-1) (dlbsurgd), credit growth (dlcrdtl), external growth rate (fgrowth1), real exchange rate overvaluation (t-1) (rxrdev5), openness (open1), and banking crisis (crisis). (b) Replicate Table 5, column 3, page 743, which adds currency crisis dummy at times (t) and (t-1) (xrp_n, xrp_nwl) as well as a twin crisis dummy (twin2).

References

Amemiya, T., and T.E. MaCurdy. 1986. Instrumental-variable estimation of an error components model. *Econometrica* 54: 869–881.

Balestra, P., and J. Varadharajan-Krishnakumar. 1987. Full information estimations of a system of simultaneous equations with error components structure. *Econometric Theory* 3: 223–246.

Baltagi, B.H. 1981a. Pooling: An experimental study of alternative testing and estimation procedures in a two-way error components model. *Journal of Econometrics* 17: 21–49.

Baltagi, B.H. 1981b. Simultaneous equations with error components. *Journal of Econometrics* 17: 189–200.

Baltagi, B.H. 1984. A Monte Carlo study for pooling time-series of cross-section data in the simultaneous equations model. *International Economic Review* 25: 603–624.

Baltagi, B.H. 2004. A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares, Problem 04.1.1. *Econometric Theory* 20: 223–224.

Baltagi, B.H. 2006a. Estimating an economic model of crime using panel data from North Carolina. *Journal of Applied Econometrics* 21: 543–547.

Baltagi, B.H. 2006b. An alternative derivation of Mundlak's fixed effects results using system estimation. *Econometric Theory* 22: 1191–1194.

Baltagi, B.H., and U. Blien. 1998. The German wage curve: Evidence from the IAB employment sample. *Economics Letters* 61: 135–142.

Baltagi, B.H., and G. Bresson. 2012. A robust Hausman-Taylor estimator. *Advances in Econometrics* 29: 175–218.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2003. Fixed effects, random effects or Hausman-Taylor? *A pretest estimator, Economics Letters* 79: 361–369.

Baltagi, B.H., and Y.J. Chang. 2000. Simultaneous equations with incomplete panels. *Econometric Theory* 16: 269–279.

Baltagi, B.H., and S. Khanti-Akom. 1990. On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 5: 401–406.

Baltagi, B.H., and Q. Li. 1991. A transformation that will circumvent the problem of autocorrelation in an error component model. *Journal of Econometrics* 48: 385–393.

Baltagi, B.H., and Q. Li. 1992. A note on the estimation of simultaneous equations with error components. *Econometric Theory* 8: 113–119.

Baltagi, B.H., and L. Liu. 2007. Alternative ways of obtaining Hausman's test using artificial regressions. *Statistics and Probability Letters* 77: 1413–1417.

Baltagi, B.H., and L. Liu. 2012. The Hausman-Taylor panel data model with serial correlation. *Statistics and Probability Letters* 82: 1401–1406.

Bramati, M.C., and C. Croux. 2007. Robust estimators for the fixed effects panel data model. *Econometrics Journal* 10: 521–540.

Breusch, T.S., G.E. Mizon, and P. Schmidt. 1989. Efficient estimation using panel data. *Econometrica* 57: 695–700.

Bruckner, M. 2013. On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics* 28: 126–150.

Cardellichio, P.A. 1990. Estimation of production behavior using pooled microdata. *Review of Economics and Statistics* 72: 11–18.

Contoyannis, P., and N. Rice. 2001. The impact of health on wages: Evidence from the British Household Panel Survey. *Empirical Economics* 26: 599–622.

Cornwell, C., and P. Rupert. 1988. Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 3: 149–155.

Cornwell, C., and W.N. Trumbull. 1994. Estimating the economic model of crime with panel data. *Review of Economics and Statistics* 76: 360–366.

Cornwell, C., P. Schmidt, and D. Wyhowski. 1992. Simultaneous equations and panel data. *Journal of Econometrics* 51: 151–181.

Egger, P., and M. Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19: 227–246.

Haque, N.U., K. Lahiri, and P. Montiel. 1993. Estimation of a macroeconomic model with rational expectations and capital controls for developing countries. *Journal of Development Economics* 42: 337–356.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.

Hutchison, M.M., and I. Noy. 2005. How bad are twins? Output costs of currency and banking crises. *Journal of Money, Credit, and Banking* 37: 725–752.

Im, K.S., S.C. Ahn, P. Schmidt, and J.M. Wooldridge. 1999. Efficient estimation of panel data models with strictly exogenous explanatory variables. *Journal of Econometrics* 93: 177–201.

Kinal, T., and K. Lahiri. 1990. A computational algorithm for multiple equation models with panel data. *Economics Letters* 34: 143–146.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Mundlak, Y. 1978. On the pooling of time series and cross-section data. *Econometrica* 46: 69–85.

Oaxaca, R.L., and I. Geisler. 2003. Fixed effects models with time invariant variables: a theoretical note. *Economics Letters* 80: 373–377.

Owusu-Gyapong, A. 1986. Alternative estimating techniques for panel data on strike activity. *Review of Economics and Statistics* 68: 526–531.

Serlenga, L., and Y. Shin. 2007. Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics* 22: 361–381.

Taylor, W.E. 1980. Small sample considerations in estimation from panel data. *Journal of Econometrics* 13: 203–223.

Wagenvoort, R., and R. Waldmann. 2002. On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics* 106: 297–324.

White, H. 1984. *Asymptotic theory for econometricians*. New York: Academic Press.

White, H. 1986. Instrumental variables analogs of generalized least squares estimators. In *Advances in statistical analysis and statistical computing*, (ed.) R.S. Mariano, Vol. 1, 173–277. New York: JAI Press.

Wyhowski, D.J. 1994. Estimation of a panel data model in the presence of correlation between regressors and a two-way error component. *Econometric Theory* 10: 130–139.

Zyskind, G. 1967. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics* 36: 1092–1109.

Dynamic Panel Data Models

8

8.1 Introduction

Many economic relationships are dynamic in nature, and one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment. See, for example, Baltagi and Levin (1986) on dynamic demand for an addictive commodity like cigarettes, Arellano and Bond (1991) on a dynamic model of employment, Blundell et al. (1992) on a dynamic model of company investment, Ziliak (1997) on a dynamic life cycle labor supply model, and Acemoglu et al. (2005) on a dynamic model relating democracy to education. These dynamic relationships are characterized by the presence of a lagged dependent variable among the regressors, i.e.,

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + u_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \tag{8.1}$$

where δ is a scalar, x'_{it} is $1 \times K$ and β is $K \times 1$. We will assume that the u_{it} follow a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{8.2}$$

where $\mu_i \sim \text{IID}(0, \sigma^2_\mu)$ and $\nu_{it} \sim \text{IID}(0, \sigma^2_\nu)$ independent of each other and among themselves. The dynamic panel data regression described in (8.1) and (8.2) is characterized by two sources of persistence over time. Autocorrelation is due to the presence of a lagged dependent variable among the regressors and individual effects characterizing the heterogeneity among the individuals. In this chapter, we review some of the econometric studies that propose estimation and testing procedures for this model.

Let us start with some of the basic problems introduced by the inclusion of a lagged dependent variable. Since y_{it} is a function of μ_i, it immediately follows that $y_{i,t-1}$ is also a function of μ_i. Therefore, $y_{i,t-1}$, a right-hand regressor in (8.1), is correlated with the error term. This renders the OLS estimator biased and inconsistent

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_8

even if the ν_{it} are not serially correlated. For the fixed effects (FE) estimator, the Within transformation wipes out the μ_i (see Chap. 2), but $(y_{i,t-1} - \bar{y}_{i.-1})$ where $\bar{y}_{i.-1} = \sum_{t=2}^{T} y_{i,t-1}/(T-1)$ will still be correlated with $(\nu_{it} - \bar{\nu}_{i.})$ even if the ν_{it} are not serially correlated. This is because $y_{i,t-1}$ is correlated with $\bar{\nu}_{i.}$ by construction. The latter average contains $\nu_{i,t-1}$ which is obviously correlated with $y_{i,t-1}$. Also, ν_{it} is correlated with $\bar{y}_{i.-1}$ because the latter average contains y_{it}. These are the leading terms causing the correlation and they are both of order $(T-1)$. This result was discovered by Nickell (1981) who showed that the Within estimator is biased of $O(1/T)$. This bias does not vanish as the number of individuals increase, so the Within estimator is inconsistent for N large and T small. However, as T gets large the fixed effects estimator becomes consistent. Several suggestions to correct for the bias of the popular FE estimator have been proposed. Most notably, Kiviet (1995) who derives an approximation for the bias of the Within estimator in a dynamic panel data model with serially uncorrelated disturbances and strongly exogenous regressors. He then proposes a bias-corrected FE estimator that subtracts a consistent estimator of this bias from the original FE estimator.

For typical micro-panels where N is large and T is short and fixed, the Within estimator is biased and inconsistent, and it is worth emphasizing that only if $T \to \infty$ will the Within estimator of δ and β be consistent for the dynamic error component model. For macro-panels, studying, for example, long-run growth, the data covers a large number of countries N over a moderate size T. In this case, T is not very small relative to N. Hence, some researchers may still favor the Within estimator arguing that its bias may not be large. Judson and Owen (1999) performed some Monte Carlo experiments for $N = 20$ or 100 and $T = 5$, 10, 20 and 30 and found that the bias in the Within estimator can be sizeable, even when $T = 30$. This bias increases with δ and decreases with T. But even for $T = 30$, this bias could be as much as 20% of the true value of the coefficient of interest.

An alternative transformation that wipes out the individual effects is the first difference (FD) transformation. In this case, correlation between the predetermined explanatory variables and the remainder error is easier to handle. In fact, Anderson and Hsiao (1982) suggested first differencing the model to get rid of the μ_i and then using $\Delta y_{i,t-2} = (y_{i,t-2} - y_{i,t-3})$ or simply $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1} = (y_{i,t-1} - y_{i,t-2})$. These instruments will not be correlated with $\Delta\nu_{it} = \nu_{i,t} - \nu_{i,t-1}$, as long as the ν_{it} themselves are not serially correlated. This instrumental variable (IV) estimation method leads to consistent but not necessarily efficient estimates of the parameters in the model because it does not make use of all the available moment conditions (see Ahn and Schmidt 1995), and it does not take into account the differenced structure on the residual disturbances $(\Delta\nu_{it})$. Arellano (1989) finds that for simple dynamic error components models, the estimator that uses differences $\Delta y_{i,t-2}$ rather than levels $y_{i,t-2}$ for instruments has a singularity point and very large variances over a significant range of parameter values. In contrast, the estimator that uses instruments in levels, i.e., $y_{i,t-2}$, has no singularities and much smaller variances and is therefore recommended.

Dessi and Robertson (2003) estimate dynamic panel regressions relating debt and Tobin's Q using a panel of $N = 557$ UK firms observed over the period 1967–89 ($T = 23$). They find that firm fixed effects are highly significant concluding that unobserved firm characteristics are important determinants of both capital structure and expected performance (as measured by Tobin's Q). Applying the Anderson and Hsiao (1982) estimator, they find highly significant dynamic effects in the determination of debt and Tobin's Q. Hence, emphasizing the importance of capturing firm *heterogeneity* and *dynamics*, two of the main advantages of applying panel data methods.

Arellano and Bond (1991) proposed a generalized method of moments (GMM) procedure that is more efficient than the Anderson and Hsiao (1982) estimator. While Ahn and Schmidt (1995) derived additional nonlinear moment restrictions not exploited by the Arellano and Bond (1991) GMM estimator. This literature is generalized and extended by Arellano and Bover (1995) and Blundell and Bond (1998) to mention a few. In addition, an alternative method of estimation of the dynamic panel data model is proposed by Keane and Runkle (1992). This is based on the forward filtering idea in time-series analysis. We focus on these studies and describe their respective contributions to the estimation and testing of dynamic panel data models. This chapter concludes with some applications and suggested readings. [1]

8.2 The Arellano and Bond Estimator

Arellano and Bond (1991) argue that additional instruments can be obtained in a dynamic panel data model if one utilizes the orthogonality conditions that exist between lagged values of y_{it} and the disturbances ν_{it}. Let us illustrate this with the simple autoregressive model with no regressors:

$$y_{it} = \delta y_{i,t-1} + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{8.3}$$

where $u_{it} = \mu_i + \nu_{it}$ with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$, independent of each other and among themselves. In order to get a consistent estimate of δ as $N \to \infty$ with T fixed, we first difference (8.3) to eliminate the individual effects

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (\nu_{it} - \nu_{i,t-1}) \tag{8.4}$$

and note that $(\nu_{it} - \nu_{i,t-1})$ is MA(1) with unit root. For $t = 3$, the first period we observe this relationship, we have

$$y_{i3} - y_{i2} = \delta(y_{i2} - y_{i1}) + (v_{i3} - v_{i2})$$

In this case, y_{i1} is a valid instrument, since it is highly correlated with $(y_{i2} - y_{i1})$ and not correlated with $(\nu_{i3} - \nu_{i2})$ as long as the ν_{it} are not serially correlated. But note what happens for $t = 4$, the second period we observe (8.4)

$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (\nu_{i4} - \nu_{i3})$$

In this case, y_{i2} as well as y_{i1} are valid instruments for $(y_{i3} - y_{i2})$, since both y_{i2} and y_{i1} are not correlated with $(\nu_{i4} - \nu_{i3})$. One can continue in this fashion, adding an extra valid instrument with each forward period, so that for period T, the set of valid instruments becomes $(y_{i1}, y_{i2}, \ldots, y_{i,T-2})$.

This instrumental variable procedure still does not account for the differenced error term in (8.4). In fact,

$$E(\Delta\nu_i \; \Delta\nu_i') = \sigma_\nu^2 G \tag{8.5}$$

where $\Delta\nu_i' = (\nu_{i3} - \nu_{i2}, \ldots, \nu_{iT} - \nu_{i,T-1})$ and

$$G = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

is $(T-2) \times (T-2)$, since $\Delta\nu_i$ is MA(1) with unit root. Define

$$W_i = \begin{bmatrix} [y_{i1}] & & & 0 \\ & [y_{i1}, y_{i2}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, \ldots, y_{i,T-2}] \end{bmatrix} \tag{8.6}$$

Then, the matrix of instruments is $W = [W_1', \ldots, W_N']'$ and the moment equations described above are given by $E(W_i'\Delta\nu_i) = 0$. These moment conditions have also been pointed out by Holtz-Eakin, Newey and Rosen (1988) and Ahn and Schmidt (1995). Premultiplying the differenced equation (8.4) in vector form by W', one gets

$$W'\Delta y = W'(\Delta y_{-1})\delta + W'\Delta\nu \tag{8.7}$$

Performing GLS on (8.7) one gets the Arellano and Bond (1991) preliminary one-step consistent estimator

$$\begin{aligned} \widehat{\delta}_1 = {} & [(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y_{-1})]^{-1} \\ & \times[(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y)] \end{aligned} \tag{8.8}$$

The optimal generalized method of moments (GMM) estimator of δ à la Hansen (1982) for $N \to \infty$ and T fixed using only the above moment restrictions yields the same expression as in (8.8) except that

$$W'(I_N \otimes G)W = \sum_{i=1}^{N} W_i' G W_i$$

is replaced by

$$V_N = \sum_{i=1}^{N} W_i'(\Delta\nu_i)(\Delta\nu_i)'W_i$$

This GMM estimator requires no knowledge concerning the initial conditions or the distributions of ν_i and μ_i. To operationalize this estimator, $\Delta\nu$ is replaced by differenced residuals obtained from the preliminary consistent estimator $\widehat{\delta}_1$. The resulting estimator is the *two-step* Arellano and Bond (1991) GMM estimator:

$$\widehat{\delta}_2 = [(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y_{-1})]^{-1}[(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y)] \tag{8.9}$$

A consistent estimate of the asymptotic var$(\widehat{\delta}_2)$ is given by the first term in (8.9),

$$\widehat{\text{var}}(\widehat{\delta}_2) = [(\Delta y_{-1})'W\widehat{V}_N^{-1}W'(\Delta y_{-1})]^{-1} \tag{8.10}$$

Note that $\widehat{\delta}_1$ and $\widehat{\delta}_2$ are asymptotically equivalent if the ν_{it} are IID$(0, \sigma_\nu^2)$. The *one-step* Arellano and Bond estimator, given in (8.8), is the *default* option in Stata's command *xtabond*. The option *two-step* gives the estimator in (8.9).

8.2.1 Testing for Over-Identification Restrictions and Serial Correlation in Dynamic Panel Models

The basic idea of the test for over-identification restrictions can be explained using the simple autoregressive model given in (8.3). Assume there are only four periods, i.e., $T = 4$. Then Arellano and Bond (1991) give us three moment conditions to identify one parameter:

$$E[(y_{i,1}(u_{i,3} - u_{i,2})] = 0 \tag{8.11a}$$
$$E[(y_{i,1}(u_{i,4} - u_{i,3})] = 0 \tag{8.11b}$$
$$E[(y_{i,2}(u_{i,4} - u_{i,3})] = 0 \tag{8.11c}$$

Any of these moment conditions can be used to estimate δ. The remainig two are over-identification restrictions. For the general case, given by the moment conditions $E(W_i'\Delta\nu_i) = 0$ with W_i defined by (8.6), Arellano and Bond (1991) suggest the following *Sargan test* for over-identifying restrictions:

$$m = \Delta\widehat{\nu}'W\left[\sum_{i=1}^{N} W_i'(\Delta\widehat{\nu}_i)(\Delta\widehat{\nu}_i)'W_i\right]^{-1} W'(\Delta\widehat{\nu}) \sim \chi^2_{p-K-1}$$

where p refers to the number of columns of W and $\Delta\widehat{\nu}$ denote the residuals from the two-step Arellano and Bond estimator. This statistic can be obtained with Stata using the command *estat sargan*. Other tests suggested are Sargan's difference statistic to test nested hypotheses concerning serial correlation in a sequential way, or a Griliches and Hausman (1986) type test based on the difference between the two-step GMM estimators assuming the disturbances in levels are MA(0) and MA(1), respectively. These are described in more detail in Arellano and Bond (1991) (p. 283).

Using Monte Carlo experiments, Bowsher (2002) finds that the use of too many moment conditions causes the Sargan test for over-identifying restrictions to be undersized and have extremely low power. Fixing N at 100, and letting T increase over the range (5, 7, 9, 11, 13, 15), the performance of the Sargan's test using the full set of Arellano–Bond moment conditions is examined for $\delta = 0.4$. For T = 5, the Monte Carlo mean of the Sargan χ^2_5 statistic is 5.12 when it should be 5, and its Monte Carlo variance is 9.84 when it should be 10. The size of the test is 0.052 at the 5% level and the power under the alternative is 0.742. For T = 15, the Sargan χ^2_{90} statistic has a Monte Carlo mean of 91.3 when its theoretical mean is 90. However, its Monte Carlo variance is 13.7 when it should be 180. This underestimation of the theoretical variance results in zero rejection rate under the null and alternative. In general, the Monte Carlo mean is a good estimator of the mean of the asymptotic χ^2 statistic. However, the Monte Carlo variance is much smaller than its asymptotic counterpart when T is large. The Sargan test never rejects when T is too large for a given N. Zero rejection rates under the null and alternative were also observed for the following (N, T) pairs (125, 16), (85, 13), and (40, 10). This is attributed to poor estimates of the weighting matrix in GMM rather than to weak instruments.

Additionally, Arellano and Bond (1991) propose a test for the hypothesis that there is *no second-order serial correlation* for the disturbances of the first-differenced equation. This test is important because the *consistency* of the GMM estimator relies upon the fact that $E[\Delta\nu_{it}\Delta\nu_{i,t-2}] = 0$. The test statistic is given in equation (8) of Arellano and Bond (1991) (p. 282) and will not be reproduced here, This hypothesis is true if the ν_{it} are not serially correlated or follow a random walk. Under the latter situation, both OLS and GMM of the first-differenced version of (8.1) are consistent, and Arellano and Bond (1991) suggest a Hausman-type test based on the difference between the two estimators. A test for *first-order* and *second-order* serial correlation can be obtained with Stata using the command *estat abond*.[2]

To summarize, dynamic panel data estimation of equations (8.1) and (8.2) with fixed effects suffers from the Nickell (1981) bias which disappears only if T tends to infinity. For fixed T and large N, the recommended estimator in this case is GMM suggested by Arellano and Bond (1991) which basically differences the model to get rid of the individual specific effects and along with it any time-invariant regressor. This also gets rid of any endogeneity that may be due to the correlation of these individual effects and the right-hand side regressors. The moment conditions utilize the orthogonality conditions between the differenced errors and lagged values of the dependent variable. This assumes that the original disturbances in (8.1) and (8.2) are serially uncorrelated and that the differenced error is MA(1) with unit root. In fact, two diagnostics are computed using the Arellano and Bond GMM procedure to test for first-order and second-order serial correlation in the disturbances. One should reject the null of the absence of first-order serial correlation and not reject the absence of second-order serial correlation. A special feature of dynamic panel data GMM estimation is that the number of moment conditions increases with T. Therefore, a Sargan test is performed to test the over-identification restrictions. The next two subsections discuss two weaknesses of the Arellano and Bond two-step estimator. Section 8.2.2 argues that poor estimation of the weight matrix leads to biased asymptotic standard errors and weak inference, while Sect. 8.2.3 discusses the bias/efficiency trade-off in using a subset rather than all the moment conditions available.

8.2.2 Downward Bias of the Estimated Asymptotic Standard Errors

A limited Monte Carlo study was performed by Arellano and Bond (1991) based on 100 replications from a simple autoregressive model with one regressor and no constant, i.e., $y_{it} = \delta y_{i,t-1} + \beta x_{it} + \mu_i + \nu_{it}$ with $N = 100$ and $T = 7$. The results showed that the GMM estimators have negligible finite sample biases and substantially smaller variances than those associated with simpler IV estimators à la Anderson and Hsiao (1982). However, the estimated standard error of the two-step GMM estimator was found to be *downward biased*. The tests proposed above also performed reasonably well. These estimation and testing methods were applied to a model of employment using a panel of 140 quoted UK companies for the period 1979–84. This is the benchmark data set used in Stata and EViews to obtain the one-step and two-step estimators described in (8.8) and (8.9); see Problem 8.8.

Windmeijer (2005) attributes the small sample downward bias of the estimated asymptotic standard errors of the two-step efficient GMM estimator to the estimation of the *weight matrix*. He suggests a correction term based on a Taylor series expansion that accounts for the estimation of this *weight matrix*. He shows that this correction term provides a more accurate approximation in finite samples when all the moment conditions are linear. These corrected standard errors are available using Stata commands with the option *vce (robust)*.

The asymptotic standard errors of the two-step GMM estimator in dynamic panel data models underestimate the variability of this estimator in small samples. This in turn renders the Wald tests for parameter restrictions oversized. Bond, Bowsher and Windmeijer (2001) suggest using *criterion-based inference* for test of hypotheses in dynamic panel data models rather than Wald tests based on two-step GMM. This criterion-based statistic is computed as the difference between the standard GMM tests of over-identifying restrictions in the restricted and unrestricted models. Monte Carlo experiments show that this outperforms conventional Wald tests. It also has similar size and power properties as the computationally burdensome alternatives based on continuously updated GMM or exponential tilting.

8.2.3 Too Many Moment Conditions and the Bias Efficiency Trade-Off

Ziliak (1997) asked the question whether the bias/efficiency trade-off for the GMM estimator considered by Tauchen (1986) for the time-series case is still binding in panel data where the sample size is normally larger than 500. For time-series data, Tauchen (1986) showed that even for $T = 50$ or 75 there is a bias/efficiency trade-off as the number of moment conditions increase. Therefore, Tauchen recommended the use of sub-optimal instruments in small samples. This problem becomes more pronounced with panel data since the number of moment conditions increases dramatically as the number of strictly exogenous variables and the number of time-series observations increase. Even though it is desirable from an asymptotic efficiency point of view to include as many moment conditions as possible, it may be infeasible or impractical to do so in many cases. For example, for $T = 10$ and five strictly exogenous regressors, this generates 500 moment conditions for GMM. Ziliak (1997) performed an extensive set of Monte Carlo experiments for a dynamic panel data model and found that the same trade-off between bias and efficiency exists for GMM as the number of moment conditions increases, and that one is better off with *sub-optimal instruments*. In fact, Ziliak found that GMM performed well with sub-optimal instruments, but is not recommended when all the moments are exploited for estimation. Ziliak demonstrated this with a life cycle labor supply model under uncertainty based on 532 men observed over 10 years of data (1978–87) from the panel study of income dynamics. The sample was restricted to continuously married, continuously working prime age men aged 22–51 in 1978. These men were paid an hourly wage or salaried and could not be piece-rate workers or self-employed. Ziliak found that the downward bias of GMM was quite severe as the number of moment conditions increased, outweighing the gains in efficiency. Ziliak reported estimates of the intertemporal substitution elasticity. This measures the intertemporal changes in hours of work due to an anticipated change in the real wage. For GMM, this estimate changed from 0.519 to 0.093 when the number of moment conditions used in GMM was increased from 9 to 212. The standard error of this estimate dropped from 0.36 to 0.07. Ziliak attributed this bias to the correlation between the sample moments used in estimation and the estimated weight matrix. Roodman (2009) discusses the problem of too many instruments and develops *xtabond2* in Stata that has many advantages in implementing the Arellano and Bond (1991) GMM estimator including a collapse option that reduces the number of moments conditions by collapsing the W_i matrix in (8.6). This is in the spirit of Ziliak's call for not using all moments to trade-off bias for efficiency. We will use *xtabond2* in the empirical examples.

8.3 The Arellano and Bover Estimator

Arellano and Bover (1995) develop a unifying GMM framework for looking at efficient IV estimators for dynamic panel data models. They do that in the context of the Hausman and Taylor (1981) model given in (7.40), which in static form is reproduced here for convenience:

$$y_{it} = x'_{it}\beta + Z'_i\gamma + u_{it} \tag{8.12}$$

where β is $K \times 1$ and γ is $g \times 1$. The Z_i are time-invariant variables, whereas the x_{it} vary over individuals and time. In vector form, (8.12) can be written as

$$y_i = W_i\eta + u_i \tag{8.13}$$

with the disturbances following a one-way error component model

$$u_i = \mu_i\iota_T + \nu_i \tag{8.14}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $u_i = (u_{i1}, \ldots, u_{iT})'$, $\eta' = (\beta', \gamma')$, $W_i = [X_i, \iota_T Z'_i]$, $X_i = (x_{i1}, \ldots, x_{iT})'$ and ι_T is a vector of ones of dimension T. In general, $E(u_i u'_i/w_i)$ will be unrestricted depending on $w_i = (x'_i, Z'_i)'$ where $x_i = (x'_{i1}, \ldots, x'_{iT})'$. However, the literature emphasizes two cases with cross-sectional homoskedasticity.

Case 1 $E(u_i u'_i) = \Omega$ independent of w_i, but general to allow for arbitrary Ω as long as it is the same across individuals, i.e., Ω is the same for $i = 1, \ldots, N$.

Case 2 the traditional error component model where $\Omega = \sigma^2_\nu I_T + \sigma^2_\mu \iota_T\iota'_T$.

Arellano and Bover transform the system of T equations in (8.13) using the nonsingular transformation

$$H = \begin{bmatrix} C \\ \iota'_T/T \end{bmatrix} \tag{8.15}$$

where C is any $(T-1) \times T$ matrix of rank $(T-1)$ such that $C\iota_T = 0$. For example, C could be the first $(T-1)$ rows of the Within group operator or the first difference operator.[3] Note that the transformed disturbances

$$u^+_i = Hu_i = \begin{bmatrix} Cu_i \\ \bar{u}_i \end{bmatrix} \tag{8.16}$$

have the first $(T-1)$ transformed errors free of μ_i. Hence, all exogenous variables are valid instruments for these first $(T-1)$ equations. Let m_i denote the subset of variables of w_i assumed to be uncorrelated in levels with μ_i and such that the dimension of m_i is greater than or equal to the dimension of η. In the Hausman and Taylor study, $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$ where X_1 and Z_1 are exogenous of dimension $NT \times k_1$ and $N \times g_1$. X_2 and Z_2 are correlated with the individual effects and are of dimension $NT \times k_2$ and $N \times g_2$. In this case, m_i includes the set of X_1 and Z_1 variables and m_i would be based on $(Z'_{1,i}, x'_{1,i1}, \ldots, x'_{1iT})'$. Therefore, a valid IV matrix for the complete transformed system is

$$M_i = \begin{bmatrix} w_i' & & & 0 \\ & \ddots & & \\ & & w_i' & \\ 0 & & & m_i' \end{bmatrix} \tag{8.17}$$

and the moment conditions are given by

$$E(M_i' H u_i) = 0 \tag{8.18}$$

Defining $W = (W_1', \ldots, W_N')'$, $y = (y_1', \ldots, y_N')'$, $M = (M_1', \ldots, M_N')'$, $\bar{H} = I_N \otimes H$ and $\bar{\Omega} = I_N \otimes \Omega$, and premultiplying (8.13) in vector form by $M'\bar{H}$ one gets

$$M'\bar{H}y = M'\bar{H}W\eta + M'\bar{H}u \tag{8.19}$$

Performing GLS on (8.19) one gets the Arellano and Bover (1995) estimator

$$\widehat{\eta} = [W'\bar{H}'M(M'\bar{H}\bar{\Omega}\bar{H}'M)^{-1}M'\bar{H}W]^{-1}W'\bar{H}'M(M'\bar{H}\bar{\Omega}\bar{H}'M)^{-1}M'\bar{H}y \tag{8.20}$$

In practice, the covariance matrix of the transformed system $\Omega^+ = H\Omega H'$ is replaced by a consistent estimator, usually

$$\widehat{\Omega}^+ = \sum_{i=1}^{N} \widehat{u}_i^+ \widehat{u}_i^{+\prime}/N \tag{8.21}$$

where $\widehat{u}_i^+$ are residuals based on consistent preliminary estimates. The resulting $\widehat{\eta}$ is the optimal GMM estimator of η with constant Ω based on the above moment restrictions. Further, efficiency can be achieved using Chamberlain (1982) or Hansen (1982) GMM type estimator which replaces $(\sum_i M_i'\Omega^+ M_i)$ in (8.20) by $(\sum_i M_i'\widehat{u}_i^+\widehat{u}_i^{+\prime} M_i)$. For the error component model, $\widetilde{\Omega}^+ = H\widetilde{\Omega}H'$ with $\widetilde{\Omega} = \widetilde{\sigma}_\nu^2 I_T + \widetilde{\sigma}_\mu^2 \iota_T \iota_T'$, where $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_\mu^2$ denote consistent estimates σ_ν^2 and σ_μ^2.

The Hausman and Taylor (1981) (HT) estimator, given in Sect. 7.3, is $\widehat{\eta}$ with $\widetilde{\Omega}^+$ and $m_i = (Z_{1,i}', \bar{x}_{1,i}')'$ where $\bar{x}_i' = \iota_T' X_i/T = (\bar{x}_{1,i}', \bar{x}_{2,i}')$. The Amemiya and MaCurdy (1986) (AM) estimator is $\widehat{\eta}$ with $\widetilde{\Omega}^+$ and $m_i = (Z_{1i}', x_{1,i1}', \ldots, x_{1,iT}')'$. The Breusch, Mizon and Schmidt (1989) (BMS) estimator exploits the additional moment restrictions that the correlation between $x_{2,it}$, and μ_i is constant over time. In this case, $\widetilde{x}_{2,it} = x_{2,it} - \bar{x}_{2,i}$ are valid instruments for the last equation of the transformed system. Hence, BMS is $\widehat{\eta}$ with $\widetilde{\Omega}^+$ and $m_i = (Z_{1,i}', x_{1,i1}', \ldots, x_{1,iT}', \widetilde{x}_{2,i1}', \ldots, \widetilde{x}_{2,iT}')'$.

Because the set of instruments M_i is block-diagonal, Arellano and Bover show that $\widehat{\eta}$ is invariant to the choice of C. Another advantage of their representation is that the form of $\Omega^{-1/2}$ need not be known. Hence, this approach generalizes the HT, AM, BMS type estimators to a more general form of Ω than that of error components, and it easily extends to the dynamic panel data case as can be seen next.

Let us now introduce a lagged dependent variable into the right-hand side of (8.12):

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + Z_i'\gamma + u_{it} \tag{8.22}$$

Assuming that t=0 is observed, we redefine η'=$(\delta, \beta', \gamma')$ and W_i=$[y_{i(-1)}, X_i, \iota_T Z_i']$ with $y_{i(-1)} = (y_{i,0}, \ldots, y_{i,T-1})'$. Provided there are enough valid instruments to ensure identification, the GMM estimator defined in (8.20) remains consistent for this model. The matrix of instruments M_i is the same as before adjusting for the fact that $t = 0$ is now the first period observed, so that $w_i = [x_{i0}', ..., x_{iT}', Z_i']'$. In this case $y_{i(-1)}$ is excluded despite its presence in W_i. The same range of choices for m_i are available, for example, $m_i = (Z_{1i}', x_{1i}', \widetilde{x}_{2,i1}', ...\widetilde{x}_{2,iT}')$ is the BMS-type estimator. However, for this choice of m_i the rows of CX_i are linear combinations of m_i. This means that the same instrument set is valid for all equations, and we can use $M_i = I_T \otimes m_i'$ without altering the estimator. The consequence is that the transformation is unnecessary and the estimator can be obtained by applying 3SLS to the original system of equations using m_i as the vector of instruments for all equations:

$$\begin{aligned}\widehat{\eta} = \left[\sum_i (W_i \otimes m_i)' \left(\widehat{\Omega} \otimes \sum_i m_i m_i'\right)^{-1} \sum_i (W_i \otimes m_i)\right]^{-1} \sum_i (W_i \otimes m_i)' \\ \times \left(\widehat{\Omega} \otimes \sum_i m_i m_i'\right)^{-1} \sum_i (y_i \otimes m_i)\end{aligned} \tag{8.23}$$

Arellano and Bover (1995) prove that this 3SLS estimator is asymptotically equivalent to the limited information maximum likelihood procedure with Ω unrestricted developed by Bhargava and Sargan (1983). The latter estimator can be applied to short T linear dynamic panel data models using Stata's *xtdpdqml* command; see Kripfganz (2016). In fact, the *xtdpdqml* command also performs quasi–maximum likelihood (QML) estimation of linear dynamic short T panel data models sugested by Hsiao, Pesaran and Tahmiscioglu (2002). Bhargava and Sargan (1983) assume a random effects dynamic panel, while Hsiao, Pesaran and Tahmiscioglu (2002) estimate a differenced model. Both estimators condition on an endogenous initial y_{i0} and specify its functional form. Kripfganz (2016) illustrates this using the Arellano and Bond (1991) dynamic employment empirical application. This is restricted to one lag on the dependent variable when Arellano and Bond (1991) used two lags, and no distributed lag on the exogenous variables when Arellano and Bond (1991) used up to two lags on the exogenous variables. The illustration also omits the industry output variable arguing that Arellano and Bond (1991) found it insignificant. You are asked to replicate these results in Problem 8.13.

Kripfganz and Schwarz (2018) proposed a two-step estimation procedure a la Hausman and Taylor (1981) to identify the coefficients of time-invariant regressors in a dynamic panel data model. In the first step, they estimate the coefficients of the time-varying regressors using GMM or QML. Subsequently, they regress the first-stage residuals on the time-invariant regressors and adjust the second-step standard errors to account for the first-step estimation error. This can be done using *xtseqreg* in Stata. One advantage of the two-step approach is the invariance of the first-step estimates to incorrect assumptions needed to identify the coefficients of the time-invariant regressors. They illustrate their two-step estimator using a dynamic version of the gravity model for the foreign direct investment (FDI) study by Egger and Pfaffermayr (2004) which is considered in Problem 7.17. Kripfganz and Schwarz (2018) find that the lagged real bilateral stock of US outward FDI is significant. This suggests that neglecting the dynamic nature of the model results in a sizable overestimation of the effect of the time-invariant geographical distance variable. You are asked to replicate these results in Problem 8.14. Note that this two-step procedure is an alternative to the system estimation procedure of Arellano and Bover (1995) that allows one to obtain estimates of time-invariant variables in a dynamic panel data model.

8.4 The Ahn and Schmidt Moment Conditions

Ahn and Schmidt (1995) show that under the standard assumptions used in a dynamic panel data model, there are additional *nonlinear* moment conditions that were ignored by the Arellano and Bond (1991) estimator. In this section, we explain how these additional restrictions arise for the simple dynamic model and show how they can be utilized in a GMM framework.

Consider the simple dynamic model with no regressors given in (8.3), and assume that y_{i0}, ..., y_{iT} are observable. In vector form, this is given by

$$y_i = \delta y_{i-1} + u_i \tag{8.24}$$

where $y_i' = (y_{i1}, \ldots, y_{iT})$, $y_{i-1}' = (y_{i0}, \ldots, y_{i,T-1})$ and $u_i' = (u_{i1}, \ldots, u_{iT})$. The standard assumptions on the dynamic model (8.24) are that
(A.1) For all i, ν_{it} is uncorrelated with y_{i0} for all t.
(A.2) For all i, ν_{it} is uncorrelated with μ_i for all t.
(A.3) For all i, the ν_{it} are mutually uncorrelated.
Ahn and Schmidt (1995) argue that these assumptions on the initial value y_{i0} are weaker than those often made in the literature (see Bhargava and Sargan 1983).

Under these assumptions, one obtains the following $T(T-1)/2$ moment conditions:

$$E(y_{is}\Delta u_{it}) = 0 \quad t = 2, \ldots, T \quad s = 0, \ldots, t-2 \tag{8.25}$$

These are the same moment restrictions given below (8.6) and exploited by Arellano and Bond (1991). However, Ahn and Schmidt (1995) find $T-2$ additional moment conditions not implied by (8.25). These are given by

$$E(u_{iT}\Delta u_{it}) = 0 \quad t = 2, \ldots, T-1 \tag{8.26}$$

Therefore, (8.25) and (8.26) imply a set of $T(T-1)/2 + (T-2)$ moment conditions which represent all of the moment conditions implied by the assumptions that the v_{it} are mutually uncorrelated among themselves and with μ_i and y_{i0}. More formally, the standard assumptions impose restrictions on the following covariance matrix:

$$\Sigma = \operatorname{cov}\begin{bmatrix} \nu_{i1} \\ \nu_{i1} \\ \vdots \\ \nu_{iT} \\ y_{i0} \\ \mu_i \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1T} & \sigma_{10} & \sigma_{1\mu} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2T} & \sigma_{20} & \sigma_{2\mu} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \ldots & \sigma_{TT} & \sigma_{T0} & \sigma_{T\mu} \\ \sigma_{01} & \sigma_{02} & \ldots & \sigma_{0T} & \sigma_{00} & \sigma_{0\mu} \\ \sigma_{\mu 1} & \sigma_{\mu 2} & \ldots & \sigma_{\mu T} & \sigma_{\mu 0} & \sigma_{\mu\mu} \end{bmatrix} \tag{8.27}$$

But, we do not observe μ_i and ν_{it}, only their sum $u_{it} = \mu_i + \nu_{it}$ which can be written in terms of the data and δ. Hence to get observable moment restrictions, we have to look at the following covariance matrix:

$$\Lambda = \operatorname{cov}\begin{bmatrix} \mu_i + v_{i1} \\ \mu_i + v_{i2} \\ \vdots \\ \mu_i + v_{iT} \\ y_{i0} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1T} & \lambda_{10} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2T} & \lambda_{20} \\ \vdots & \vdots & & \vdots & \vdots \\ \lambda_{T1} & \lambda_{T2} & \ldots & \lambda_{TT} & \lambda_{T0} \\ \lambda_{01} & \lambda_{02} & \ldots & \lambda_{0T} & \lambda_{00} \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_{\mu\mu} + \sigma_{11} + 2\sigma_{\mu 1}) & (\sigma_{\mu\mu} + \sigma_{12} + \sigma_{\mu 1} + \sigma_{\mu 2}) & \ldots \\ (\sigma_{\mu\mu} + \sigma_{12} + \sigma_{\mu 1} + \sigma_{\mu 2}) & (\sigma_{\mu\mu} + \sigma_{22} + 2\sigma_{\mu 2}) & \ldots \\ \vdots & \vdots & \ldots \\ (\sigma_{\mu\mu} + \sigma_{1T} + \sigma_{\mu 1} + \sigma_{\mu T}) & (\sigma_{\mu\mu} + \sigma_{2T} + \sigma_{\mu 2} + \sigma_{\mu T}) & \ldots \\ (\sigma_{0\mu} + \sigma_{01}) & (\sigma_{0\mu} + \mu_{02}) & \ldots \end{bmatrix} \tag{8.28}$$

$$\begin{matrix} (\sigma_{\mu\mu} + \sigma_{1T} + \sigma_{\mu 1} + \sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{01}) \\ (\sigma_{\mu\mu} + \sigma_{2T} + \sigma_{\mu 2} + \sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{02}) \\ \vdots & \vdots \\ (\sigma_{\mu\mu} + \sigma_{TT} + 2\sigma_{\mu T}) & (\sigma_{0\mu} + \sigma_{0T}) \\ (\sigma_{0\mu} + \sigma_{0T}) & \sigma_{00} \end{matrix}\Bigg]$$

Under the standard assumptions (A.1)–(A.3), we have $\sigma_{ts} = 0$ for all $t \neq s$, and $\sigma_{\mu t} = \sigma_{0t} = 0$ for all t. Then Λ simplifies as follows:

$$\Delta = \begin{bmatrix} (\sigma_{\mu\mu} + \sigma_{11}) & \sigma_{\mu\mu} & \ldots & \sigma_{\mu\mu} & \sigma_{0\mu} \\ \sigma_{\mu\mu} & (\sigma_{\mu\mu} + \sigma_{22}) & \ldots & \sigma_{\mu\mu} & \sigma_{0\mu} \\ \vdots & \vdots & & \vdots & \vdots \\ \sigma_{\mu\mu} & \sigma_{\mu\mu} & \ldots & (\sigma_{\mu\mu} + \sigma_{TT}) & \sigma_{0\mu} \\ \sigma_{0\mu} & \sigma_{0\mu} & \ldots & \sigma_{0\mu} & \sigma_{00} \end{bmatrix} \tag{8.29}$$

There are $T - 1$ restrictions, that $\lambda_{0t} = E(y_{i0}u_{it})$ is the same for $t = 1, \ldots, T$; and $[T(T-1)/2] - 1$ restrictions, that $\lambda_{ts} = E(u_{is}u_{it})$ is the same for $t, s = 1, \ldots, T$, $t \neq s$. Adding the number of restrictions, we get $T(T-1)/2 + (T-2)$.

In order to see how these additional moment restrictions are utilized, consider our simple dynamic model in differenced form along with the last period's observation in levels:

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta u_{it} \quad t = 2, 3, \ldots .T \tag{8.30}$$

$$y_{iT} = \delta y_{i,T-1} + u_{iT} \tag{8.31}$$

The usual IV estimator, utilizing the restrictions in (8.25), amounts to estimating the first-differenced equations (8.30) using three-stage least squares, imposing the restriction that δ is the same in every equation, where the instrument set is y_{i0} for $t = 2$; (y_{i0}, y_{i1}) for $t = 3$; $\ldots$; $(y_{i0}, \ldots, y_{i,T-2})$ for $t = T$ (see Sect. 8.2). Even though there are no legitimate observable instruments for the levels equation (8.31), Ahn and Schmidt argue that (8.31) is still useful in estimation because of the additional covariance restrictions implied by (8.26), i.e., that u_{iT} is uncorrelated with Δu_{it} for $t = 2, ..., T-1$. Ahn and Schmidt show that any additional covariance restrictions besides (8.26) are redundant and implied by the basic moment conditions given by (8.25). Ahn and Schmidt also point out that the moment conditions (8.25) and (8.26) hold under weaker conditions than those implied by the standard assumptions (A.1)–(A.3). In fact, one only needs
(B.1) $\text{cov}(\nu_{it}, y_{i0})$ is the same for all i and t instead of $\text{cov}(\nu_{it}, y_{i0}) = 0$, as in (A.1);
(B.2) $\text{cov}(\nu_{it}, \mu_i)$ is the same for all i and t instead of $\text{cov}(\nu_{it}, \mu_i) = 0$, as in (A.2);
(B.3) $\text{cov}(\nu_{it}, \upsilon_{is})$ is the same for all i and $t \neq s$, instead of $\text{cov}(\nu_{it}, \nu_{is}) = 0$, as in (A.3).

Ahn and Schmidt (1995) show that GMM based on (8.25) and (8.26) is asymptotically equivalent to Chamberlain (1982, 1984) optimal minimum distance estimator, and that it reaches the semiparametric efficiency bound. Ahn and Schmidt also explore additional moment restrictions obtained from assuming the ν_{it} homoskedastic for all i and t and the stationarity assumption of Arellano and Bover (1995) that $E(y_{it}\mu_i)$ is the same for all t. The reader is referred to their paper for more details. For specific parameter values, Ahn and Schmidt compute asymptotic covariance matrices and show that the extra moment conditions lead to substantial gains in asymptotic efficiency.

Ahn and Schmidt also consider the dynamic version of the Hausman and Taylor (1981) model studied in Sect. 8.3 and show how one can make efficient use of exogenous variables as instruments. In particular, they show that the strong exogeneity assumption implies more orthogonality conditions which lie in the deviations from mean space. These are irrelevant in the static Hausman–Taylor model but are relevant for the dynamic version of that model. For more details on these conditions; see Schmidt, Ahn and Wyhowski (1992) and Ahn and Schmidt (1995). In a follow-up paper, Ahn and Schmidt (1997) proposed a linearized GMM estimator that is asymptotically as efficient as the nonlinear GMM estimator. They also provided simple moment tests for the validity of these nonlinear restrictions. In addition, they investigated the circumstances under which the optimal GMM estimator is equivalent to a linear instrumental variable estimator. They found that these circumstances were quite restrictive and go beyond uncorrelatedness and homoskedasticity of the errors.

Ahn and Schmidt (1995) provided some evidence on the efficiency gains from the nonlinear moment conditions which in turn provided support for their use in practice. By employing all these conditions, the resulting GMM estimator is asymptotically efficient and has the same asymptotic variance as the MLE under normality.

8.5 The Blundell and Bond System GMM Estimator

Blundell and Bond (1998) revisit the importance of exploiting the initial condition in generating efficient estimators of the dynamic panel data model when T is small. They consider a simple autoregressive panel data model with no exogenous regressors

$$y_{it} = \delta y_{i,t-1} + \mu_i + \nu_{it} \tag{8.32}$$

with $E(\mu_i)=0$; $E(\nu_{it}) = 0$; and $E(\mu_i \nu_{it}) = 0$ for $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$. Blundell and Bond (1998) focus on the case where $T = 3$ and therefore there is only one orthogonality condition given by $E(y_{i1}\Delta\nu_{i3}) = 0$, so that δ is just-identified. In this case, the first-stage IV regression is obtained by running Δy_{i2} on y_{i1}. Note that this regression can be obtained from (8.32) evaluated at $t = 2$ by subtracting y_{i1} from both sides of this equation, i.e.,

$$\Delta y_{i2} = (\delta - 1)y_{i,1} + \mu_i + \nu_{i2} \tag{8.33}$$

Since we expect $E(y_{i1}\mu_i) > 0$, $(\delta - 1)$ will be biased upwards with

$$\text{plim}(\widehat{\delta} - 1) = (\delta - 1)\frac{c}{c + (\sigma_\mu^2/\sigma_u^2)} \tag{8.34}$$

where $c = (1 - \delta)/(1 + \delta)$. The bias term effectively scales the estimated coefficient on the instrumental variable y_{i1} toward zero. They also find that the F-statistic of the first-stage IV regression converges to χ_1^2 with noncentrality parameter

$$\tau = \frac{(\sigma_u^2 c)^2}{\sigma_\mu^2 + \sigma_u^2 c} \to 0 \text{ as } \delta \to 1 \tag{8.35}$$

As $\tau \to 0$, the instrumental variable estimator performs poorly. Hence, Blundell and Bond attribute the bias and the poor precision of the first difference GMM estimator to the problem of *weak instruments* and characterize this by its concentration parameter τ.

Next, Blundell and Bond (1998) show that an additional *mild stationarity restriction* on the initial conditions process allows the use of an extended system GMM estimator that uses lagged differences of y_{it} as instruments for equations in levels, in addition to lagged levels of y_{it} as instruments for equations in first differences. More specifically, this stationarity condition on y_{i1} requires

$$E[(y_{i1} - \frac{\mu_i}{1-\delta})\mu_i] = 0, \tag{8.36}$$

so that y_{it} converges toward its mean $\frac{\mu_i}{1-\delta}$ for each individual from period $t = 2$ onwards. This in turn yields the condition $E[\Delta y_{i,t-1}\mu_i] = 0$ for $i = 1, 2, .., N$. Using the usual mild assumption that $E[\Delta \nu_{it}\mu_i] = 0$ for $i = 1, 2, .., N$ and $t = 3, 4, \ldots, T$, we get the additional $T - 2$ non-redundant linear moment conditions $E[\Delta y_{i,t-1}(\mu_i + \nu_{it})] = 0$ for $t = 3, 4, \ldots, T$, see also Ahn and Schmidt (1995). Together with the Arellano and Bond (1991) conditions on the first differenced equation, these moment conditions on equations in levels yield the system GMM estimator. This can be applied in Stata using *xtabond2* and *xtdpd*. Blundell and Bond (1998) show that this system GMM estimator produces *dramatic efficiency gains* over the basic first difference GMM as $\delta \to 1$ and $(\sigma^2_\mu/\sigma^2_u)$ increases. In fact, for $T = 4$ and $(\sigma^2_\mu/\sigma^2_u) = 1$, the asymptotic variance ratio of the first difference GMM estimator to this system GMM estimator is 1.75 for $\delta = 0$ and increases to 3.26 for $\delta = 0.5$ and 55.4 for $\delta = 0.9$. This clearly demonstrates that the *levels restrictions* remain informative in cases where first differenced instruments become weak. Things improve for first difference GMM as T increases. However, with short T and persistent series, the Blundell and Bond findings support the use of the extra moment conditions. Blundell and Bond (2000) revisit the estimates of the capital and labor coefficients in a Cobb–Douglas production function considered by Griliches and Mairesse (1998). Using data on 509 R&D performing US manufacturing companies observed over 8 years (1982–89), the standard GMM estimator that uses moment conditions on the first differenced model finds a low estimate of the capital coefficient and low precision for all coefficients estimated. However, the system GMM estimator gives reasonable and more precise estimates of the capital coefficient and constant returns to scale is not rejected. They conclude that "a careful examination of the original series and consideration of the system GMM estimator can usefully overcome many of the disappointing features of the standard GMM estimator for dynamic panel models".

Bun and Windmeijer (2010) show that there is still a weak instrument problem for the system GMM estimator for the covariance stationary panel data AR(1) model when the variances of the individual heterogeneity and idiosyncratic errors are the same. In fact, they show that the Blundell and Bond (1998) SYS GMM estimator has indeed a smaller bias and rmse than the Arellano and Bond (1991) DIF GMM when the series are persistent, but that this bias increases with increasing variance ratio of heterogeneity to idiosyncratic error and can become substantial. The Wald test can be severely size distorted for both DIF and SYS GMM with persistent data, but the SYS Wald test size properties deteriorate further with increasing variance ratio.

It is important to note that the *initial condition* assumption is very important in dynamic models. This is emphasized by Anderson and Hsiao (1982) and Hayakawa (2009). The latter paper shows that if the initial condition renders the dependent variable *mean nonstationary*, and the variance of the individual effects is significantly larger than that of the remainder disturbances, the FD-GMM estimator performs quite well. In fact, it does not suffer from a weak instrument problem as the correlation between the lagged dependent variable and instruments gets large owing to the unremoved individual effects even if the data is persistent. This is in contrast to the

Blundell and Bond (1998) result pointing out the weak instruments problem of the FD-GMM estimator under the assumption of mean-stationarity. The Blundell and Bond (1998) estimator is implemented with Stata using *xtabond2* or *xtdpd*. This is illustrated in the empirical example in Sect. 8.8.

8.6 The Keane and Runkle Estimator

Let $y = X\beta + u$ be our panel data model with X containing a lagged dependent variable. We consider the case where $E(u_{it}/X_{it}) \neq 0$, and there exists a set of predetermined instruments W such that $E(u_{it}/W_{is}) = 0$ for $s \leqslant t$, but $E(u_{it}/W_{is}) \neq 0$ for $s > t$. In other words, W may contain lagged values of y_{it}. For this model, the 2SLS estimator will provide a consistent estimator for β. Now consider the random effects model or any other kind of serial correlation which is invariant across individuals, $\Omega_{TS} = E(uu') = I_N \otimes \Sigma_{TS}$. In this case, 2SLS will not be efficient. Keane and Runkle (1992), henceforth KR, suggest an alternative more efficient algorithm that takes into account this more general variance–covariance structure for the disturbances based on the forward filtering idea from the time-series literature. This method of estimation eliminates the general serial correlation pattern in the data, while preserving the use of predetermined instruments in obtaining consistent parameter estimates. First, one gets a consistent estimate of Σ_{TS}^{-1} and its corresponding Cholesky's decomposition $\widehat{P}_{TS}$. Next, one premultiplies the model by $\widehat{Q}_{TS} = (I_N \otimes \widehat{P}_{TS})$ and estimates the model by 2SLS using the original instruments. In this case

$$\widehat{\beta}_{KR} = [X'\widehat{Q}'_{TS} P_W \widehat{Q}_{TS} X]^{-1} X'\widehat{Q}'_{TS} P_W \widehat{Q}_{TS} y \tag{8.37}$$

where $P_W = W(W'W)^{-1}W'$ is the projection matrix for the set of instruments W. Note that this allows for a general covariance matrix Σ_{TS}, and its distinct elements $T(T+1)/2$ have to be much smaller than N. This is usually the case for large consumer or labor panels where N is very large and T is very small. Using the consistent 2SLS residuals, say $\widehat{u}_i$ for the ith individual, where $\widehat{u}_i$ is of dimension $(T \times 1)$, one can form

$$\widehat{\Sigma}_{TS} = \widehat{U}'\widehat{U}/N = \sum_{i=1}^{N} \widehat{u}_i \widehat{u}'_i/N$$

where $\widehat{U}' = [\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_N]$ is of dimension $(T \times N)$.[4]

First differencing is also used in dynamic panel data models to get rid of individual specific effects. The resulting first-differenced errors are serially correlated of an MA(1) type with unit root if the original ν_{it} are classical errors. In this case, there will be gain in efficiency in performing the KR procedure on the first-differenced (FD) model. Get $\widehat{\Sigma}_{FD}$ from FD-2SLS residuals and obtain $\widehat{Q}_{FD} = I_N \otimes \widehat{P}_{FD}$, then estimate the transformed equation by 2SLS using the original instruments.

Underlying this estimation procedure are two important hypotheses that are testable. The first is H_A; the set of instruments W are *strictly exogenous*. In order to test H_A, KR propose a test based on the difference between fixed effects 2SLS (FE-2SLS) and first-difference 2SLS (FD-2SLS). FE-2SLS is consistent only if H_A is true. In fact if the W are predetermined rather than strictly exogenous, then $E(W_{it}\bar{\nu}_{i.}) \neq 0$ and our estimator would not be consistent. In contrast, FD-2SLS is consistent whether H_A is true or not, i.e., $E(W_{it}\Delta\nu_{it}) = 0$ rain or shine. An example of this is when $y_{i,t-2}$ is a member of $W_{it,}$ then $y_{i,t-2}$ is predetermined and not correlated with $\Delta\nu_{it}$ as long as the ν_{it} are not serially correlated. However, $y_{i,t-2}$ is correlated with $\bar{\nu}_{i.}$ because this last average contains $\nu_{i,t-2}$. If H_A is not rejected, one should check whether the individual effects are correlated with the set of instruments. In this case, the usual Hausman (1978) test applies. This is based on the difference between the FE and GLS estimator of the regression model. The FE estimator would be consistent rain or shine since it wipes out the individual effects. However, the GLS estimator would be consistent and efficient only if $E(\mu_i/W_{it}) = 0$, and inconsistent otherwise. If H_A is rejected, the instruments are predetermined and the Hausman test is inappropriate. The test for H_B; $E(\mu_i/W_{it}) = 0$ will now be based on the difference between FD-2SLS and 2SLS. Under H_B, both estimators are consistent, but if H_B is not true, FD-2SLS remains consistent while 2SLS does not.

These tests are Hausman (1978) type tests except that

$$\begin{aligned}\text{var}(\widehat{\beta}_{FE-2SLS} - \widehat{\beta}_{FD-2SLS}) = &(\widetilde{X}'P_W\widetilde{X})^{-1}(\widetilde{X}'P_W\widetilde{\Omega}_{FE-2SLS}P_W\widetilde{X})(\widetilde{X}'P_W\widetilde{X})^{-1}\\ &-(\widetilde{X}'P_W\widetilde{X})^{-1}(\widetilde{X}'P_W\widetilde{\Omega}_{FEFD}P_WX_{FD})(X'_{FD}P_WX_{FD})^{-1}\\ &-(X'_{FD}P_WX_{FD})^{-1}(X'_{FD}P_W\widetilde{\Omega}_{FEFD}P_W\widetilde{X})(\widetilde{X}'P_W\widetilde{X})^{-1}\\ &+(X'_{FD}P_WX_{FD})^{-1}(X'_{FD}P_W\widehat{\Omega}_{FD-2SLS}P_WX_{FD})\\ &\times(X'_{FD}P_WX_{FD})^{-1}\end{aligned} \tag{8.38}$$

where $\widetilde{\Sigma}_{FE-2SLS} = \widetilde{U}'_{FE}\widetilde{U}_{FE}/N$, $\widehat{\Sigma}_{FD-2SLS} = \widehat{U}'_{FD}\widehat{U}_{FD}/N$ and $\widehat{\Sigma}_{FEFD} = \widetilde{U}'_{FE}\widehat{U}_{FD}/N$. As described above, $\widetilde{U}'_{FE} = [\widetilde{u}_1, \ldots, \widetilde{u}_N]_{FE}$ denotes the FE-2SLS residuals and $\widetilde{U}'_{FD} = [\widetilde{u}_1, \ldots, \widetilde{u}_N]_{FD}$ denotes the FD-2SLS residuals. Recall that for the Keane–Runkle approach, $\Omega = I_N \otimes \Sigma$.

Similarly, the var$(\widehat{\beta}_{2SLS} - \widehat{\beta}_{FD-2SLS})$ is computed as above with $\widetilde{X}$ being replaced by X, $\widetilde{\Omega}_{FE-2SLS}$ by $\widehat{\Omega}_{2SLS}$ and $\widetilde{\Omega}_{FEFD}$ by $\widehat{\Omega}_{2SLSFD}$. Also, $\widehat{\Sigma}_{2SLS} = \widehat{U}'_{2SLS}\widehat{U}_{2SLS}/N$ and $\widehat{\Sigma}_{2SLSFD} = \widehat{U}'_{2SLS}\widehat{U}_{FD}/N$.

The variances are complicated because KR do not use the efficient estimator under the null as required by a Hausman-type test (see Schmidt, Ahn and Wyhowski (1992)). Keane and Runkle (1992) apply their testing and estimation procedures to a simple version of the rational expectations life-cycle consumption model. Based on a sample of 627 households surveyed between 1972 and 1982 by the Michigan Panel Study on Income Dynamics (PSID), KR reject the strong exogeneity of the instruments. This means that the Within estimator is inconsistent and the standard Hausman test based on the difference between the standard Within and GLS estimators is inappropriate. KR also fail to reject the null hypothesis of no correlation between the individual effects and the instruments. This means that there is no need to first difference to get rid of the individual effects. Based on the KR-2SLS estimates, the authors cannot reject the simple life-cycle model. However, they show that if one uses the inconsistent Within estimates for inference one would get misleading evidence against the life-cycle model.

In the Monte Carlo experiments performed by Ziliak (1997), he recommended the Keane and Runkle (1992) estimator which performed better in terms of the bias/efficiency trade-off than the Arellano and Bond estimator. Forward filtering eliminates all forms of serial correlation while still maintaining orthogonality with the initial instrument set. Schmidt, Ahn and Wyhowski (1992) argued that filtering is irrelevant if one exploits all sample moments during estimation. However, in practice, the number of moment conditions increases with the number of time periods T and the number of regressors K and can become computationally intractable. In fact for $T = 15$ and $K = 10$, the number of moment conditions for Schmidt, Ahn and Wyhowski (1992) is $T(T-1)K/2$ which is 1040 restrictions, highlighting the computational burden of this approach. In addition, Ziliak argued that the over-identifying restrictions are less likely to be satisfied possibly due to the weak correlation between the instruments and the endogenous regressors. In this case, the Keane and Runkle (1992) estimator is desirable yielding less bias than GMM and sizeable gains in efficiency. In fact, for the life cycle labor example, the Keane and Runkle estimate of the intertemporal substitution elasticity was 0.135 for 9 moment conditions compared to 0.296 for 212 moment conditions. The standard error of these estimates dropped from 0.32 to 0.09. Keane and Neal (2016) programmed the Keane and Runkle (1992) estimator Stata using the command *xtkr*. This will be illustrated for the empirical example on Cigarette demand in Sect. 8.8.1.

8.7 Limited Information Maximum Likelihood

The dynamic panel model generates many over-identifying restrictions even for moderate values of T. Also, the number of instruments increases with T, but the quality of these instruments is often *poor* because they tend to be only weakly correlated with first-differenced endogenous variables that appear in the equation. Limited information maximum likelihood (LIML) is strongly preferred to 2SLS if the number of instruments gets large as the sample size tends to infinity. Alternative normalization rules adopted by LIML and 2SLS are at the root of their different sampling behavior. Alonso-Borrego and Arellano (1999) derive a symmetrically normalized GMM (SNM) and compare it with ordinary GMM and LIML analogues by means of simulations. Monte Carlo and empirical results show that GMM can exhibit large biases when the instruments are poor, while LIML and SNM remain essentially unbiased. However, LIML and SNM always had a larger interquartile range than GMM. For $T = 4$, $N = 100$, $\sigma_\mu^2 = 0.2$ and $\sigma_\nu^2 = 1$, the bias for $\delta = 0.5$ was 6.9% for GMM, 1.7% for SNM and 1.7% for LIML. This bias increases to 17.8% for GMM, 3.7% for SNM, and 4.1% for LIML for $\delta = 0.8$.

Alvarez and Arellano (2003) studied the asymptotic properties of FE, one-step GMM and non-robust LIML for a first-order autoregressive model when both N and T tend to infinity with $(N/T) \to c$ for $0 \leq c < 2$. For this autoregressive model, the

FE estimator is inconsistent for T fixed and N large, but becomes consistent as T gets large. GMM is consistent for fixed T, but the number of orthogonality conditions increases with T. GMM estimators that use the full set of moments available can be severely biased, especially when the instruments are weak and the number of moment conditions is large relative to N. Alvarez and Arellano show that for $T < N$, GMM bias is always smaller than FE bias and LIML bias is smaller than the other two. In fixed T framework, GMM and LIML are asymptotically equivalent, but as T increases, LIML has a smaller asymptotic bias than GMM. These results provide some theoretical support for LIML over GMM.[5] Alvarez and Arellano (2003) derive the asymptotic properties of the FE, GMM, and LIML estimators of a dynamic model with random effects. When both T and N $\to \infty$. GMM and LIML are *consistent* and asymptotically equivalent to the FE estimator. When (T/N $\to$ 0), the fixed T results for GMM and LIML remain valid, but FE, although consistent, still exhibits an asymptotic bias term in its asymptotic distribution. When T/N $\to c$, where $0 < c \leq 2$, all three estimators are consistent. The basic intuition behind this result is that contrary to the structural equation setting where too many instruments produce over-fitting and undesirable closeness to OLS; here, a larger number of instruments is associated with larger values of T and closeness to FE is desirable since the endogeneity bias $\to$ 0 as T$\to \infty$. Nevertheless, FE, GMM, and LIML exhibit a bias term in their asymptotic distributions; the biases are of order 1/T, 1/N, and 1/(2N-T), respectively. Provided T<N, the asymptotic bias of GMM is always smaller than the FE bias, and the LIML bias is smaller than the other two. When T = N, the asymptotic bias is the same for all three estimators.

Alvarez and Arellano (2003) also consider a random effects MLE that leaves the mean and variance of the initial conditions unrestricted but enforces time-series homoskedasticity. This estimator has no asymptotic bias because it does not entail incidental parameters in the N and T dimensions, and it becomes robust to heteroskedasticity as T$\to \infty$. For the simple autoregressive model in (8.32) with $|\delta| < 1$, ν_{it} being *iid* across time and individuals and independent of μ_i and y_{i0}. Alvarez and Arellano (2003) find that as T$\to \infty$, regardless of whether N is fixed or tends to ∞, provided N/T$^3 \to 0$,

$$\sqrt{NT}\left[\widetilde{\delta}_{FE} - (\delta - \frac{1}{T}(1+\delta))\right] \to N(0, 1-\delta^2) \tag{8.39}$$

Also, as N, T$\to \infty$ such that $(\log T^2)/N \to 0$, $\widehat{\delta}_{GMM} \to \delta$. Moreover, provided T/N$\to c$, $0 < c < \infty$,

$$\sqrt{NT}\left[\widehat{\delta}_{GMM} - (\delta - \frac{1}{N}(1+\delta))\right] \to N(0, 1-\delta^2) \tag{8.40}$$

when T$\to \infty$, the number of GMM orthogonality conditions T(T-1)/2 $\to \infty$. In spite of this fact, $\widehat{\delta}_{GMM} \to \delta$. Also, as N, T$\to \infty$ provided T/N$\to c$, $0 \leq c \leq 2$, $\widehat{\delta}_{LIML} \to \delta$. Moreover,

$$\sqrt{NT}\left[\widehat{\delta}_{LIML} - (\delta - \frac{1}{2N-T}(1+\delta))\right] \to N(0, 1-\delta^2) \tag{8.41}$$

LIML like GMM is consistent for δ despite T$\rightarrow\infty$ and T/N$\rightarrow c$. Provided T<N, the bias of LIML< bias of GMM<bias of FE. In fact, for $\delta = 0.2$, T = 11, N = 100, the median of 1000 Monte Carlo replications yield 0.063 for FE, 0.188 for GMM and 0.196 for LIML. For $\delta = 0.8$, T = 11, N = 100, the median of 1000 Monte Carlo replications yield 0.554 for FE, 0.763 for GMM and 0.792 for LIML. When we increase T to 51, N = 100, and $\delta = 0.8$, the median of 1000 Monte Carlo replications yield 0.760 for FE, 0.779 for GMM and 0.789 for LIML.

Wansbeek and Knapp (1999) consider a simple dynamic panel data model with heterogeneous coefficients on the lagged dependent variable and the time trend, i.e.,

$$y_{it} = \delta_i y_{i,t-1} + \xi_i t + \mu_i + u_{it}. \tag{8.42}$$

They show that double differencing gets rid of the individual country effects (μ_i) on the first round of differencing and the heterogeneous coefficient on the time trend (ξ_i) on the second round of differencing. Modified OLS, IV, and GMM methods are adapted to this model, and LIML is suggested as a viable alternative to GMM to guard against the small sample bias of GMM. Simulations show that LIML is the superior estimator for $T \geqslant 10$ and $N \geqslant 50$. Macro-economic data are subject to measurement error, and Wansbeek and Knapp (1999) show how these estimators can be modified to account for measurement error that is white noise. For example, GMM is modified so that it discards the orthogonality conditions that rely on the absence of measurement error.

8.8 Empirical Examples

8.8.1 Example 1: Dynamic Demand for Cigarettes

Baltagi and Levin (1986) estimate a dynamic demand model for cigarettes based on panel data from 46 American states. This data, updated from 1963–92, is available on the Springer web site as cigar.txt. The estimated equation is

$$\ln C_{it} = \alpha + \beta_1 \ln C_{i,t-1} + \beta_2 \ln P_{i,t} + \beta_3 \ln Y_{it} + \beta_4 \ln Pn_{it} + u_{it} \tag{8.43}$$

where the subscript i denotes the ith state ($i = 1, \ldots, 46$), and the subscript t denotes the tth year ($t = 1, \ldots, 30$). C_{it} is real per capita sales of cigarettes by persons of smoking age (16 years and older). This is measured in packs of cigarettes per head. P_{it} is the average retail price of a pack of cigarettes measured in real terms. Y_{it} is real per capita disposable income. Pn_{it} denotes the minimum real price of cigarettes in any neighboring state. This last variable is a proxy for the casual smuggling effect across state borders. It acts as a substitute price attracting consumers from high-tax states like Massachusetts with 26¢ per pack to cross over to New Hampshire where the tax is only 12¢ per pack. The disturbance term is specified as a two-way error component model:

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \quad i = 1, \ldots, 46 \quad t = 1, \ldots, 30 \tag{8.44}$$

where μ_i denotes a state-specific effect, and λ_t denotes a year-specific effect. The time-period effects (the λ_t) are assumed fixed parameters to be estimated as coefficients of time dummies for each year in the sample. This can be justified given the numerous policy interventions as well as health warnings and Surgeon General's reports. For example:

(1) the imposition of warning labels by the Federal Trade Commission effective January 1965;
(2) the application of the Fairness Doctrine Act to cigarette advertising in June 1967, which subsidized antismoking messages from 1968 to 1970;
(3) the Congressional ban on broadcast advertising of cigarettes effective January 1971.

The μ_i are state-specific effects which can represent any state-specific characteristic including the following:

(1) States with Indian reservations like Montana, New Mexico, and Arizona are among the biggest losers in tax revenues from non-Indians purchasing tax-exempt cigarettes from the reservations.
(2) Florida, Texas, Washington, and Georgia are among the biggest losers of revenues due to the purchasing of cigarettes from tax-exempt military bases in these states.
(3) Utah, which has a high percentage of Mormon population (a religion which forbids smoking), has a per capita sales of cigarettes in 1988 of 55 packs, a little less than half the national average of 113 packs.
(4) Nevada, which is a highly touristic state, has a per capita sales of cigarettes of 142 packs in 1988, 29 more packs than the national average.

These state-specific effects may be assumed fixed, in which case one includes state dummy variables in equation (8.43). The resulting estimator is the Within estimator reported in Table 8.1. Note that OLS, which ignores the state and time effects, yields a low short-run price elasticity of -0.09. However, the coefficient of lagged consumption is 0.97 which implies a high long-run price elasticity of -2.98. The Within estimator with both state and time effects yields a higher short-run price elasticity of -0.30, but a lower long-run price elasticity of -1.79. Both state and time dummies were jointly significant with an observed F-statistic of 7.39 and a p-value of 0.0001. The observed F-statistic for the significance of state dummies (given the existence of time dummies) is 4.16 with a p-value of 0.0001. The observed F-statistic for the significance of time dummies (given the existence of state dummies) is 16.05 with a p-value of 0.0001. These results emphasize the importance of including state and time effects in the cigarette demand equation. This is a dynamic equation and the OLS and Within estimators do not take into account the endogeneity of the lagged dependent variable. Hence, we report 2SLS and Within-2SLS using as instruments the lagged exogenous regressors. These give lower estimates of lagged consumption and higher estimates of own price elasticities. The Hausman-type test based on the

Table 8.1 Pooled estimation results. *Cigarette Demand

	$lnC_{i,t-1}$	lnP_{it}	$lnPn_{it}$	lnY_{it}
OLS	0.97 (157.7)	−0.090 (6.2)	0.024 (1.8)	−0.03 (5.1)
Within	0.83 (66.3)	−0.299 (12.7)	0.034 (1.2)	0.10 (4.2)
2SLS	0.85 (25.3)	−0.205 (5.8)	0.052 (3.1)	−0.02 (2.2)
2SLS-KR	0.71 (22.7)	−0.311 (13.9)	0.071 (3.7)	−0.02 (1.5)
Within-2SLS	0.60 (17.0)	−0.496 (13.0)	−0.106 (0.5)	0.19 (6.4)
FD-2SLS	0.51 (9.5)	−0.348 (12.3)	0.112 (3.5)	0.10 (2.9)
FD-2SLS-KR	0.40 (13.3)	−0.341 (18.4)	0.083 (4.2)	0.21 (10.0)
GMM-two-step	0.70 (10.2)	−0.396 (6.0)	−0.105 (1.3)	0.13 (3.5)
System GMM	0.70 (8.8)	−0.415 (4.3)	−0.003 (1.0)	0.09 (3.4)

*Numbers in parentheses are t-statistics. All regressions except OLS and 2SLS include time dummies

Source Some of the results in this table are reported in Baltagi, Griffin and Xiong (2000)

difference between Within-2SLS and FD-2SLS and discussed in Sect. 8.6, yields a χ^2_4 statistic $= 118.6$. This rejects the consistency of the Within-2SLS estimator. The Hausman-type test based on the difference between 2SLS and FD-2SLS yields a χ^2_4 statistic $= 96.6$. This rejects the consistency of 2SLS. The FD-2SLS-KR estimator yields the lowest coefficient estimate of lagged consumption (0.40). The own price elasticity is −0.34 and significant. The income effect is small 0.21 but significant and the bootlegging effect is small 0.083 and significant. The Arellano and Bond (1991) GMM two-step estimator yields a lagged consumption coefficient estimate of 0.70 and an own price elasticity of −0.40, both highly significant. Table 8.2 gives the Stata output replicating the two-step Arellano and Bond estimator using the (*xtabond2*, twostep). This gives the robust standard errors proposed by Windmeijer (2005). This could have been obtained with *xtabond*, robust. Note that the two-step Sargan test for over-identification does not reject the null, but this could be due to the bad power of this test for $N = 46$ and $T = 28$. Not all the moment conditions are used and in fact the collapse option was invoked to reduce these moment conditions. The test for first-order serial correlation rejects the null of no first-order serial correlation, but it does not reject the null that there is no second-order serial correlation. This is what one expects in a first differenced equation with the original untransformed disturbances assumed to be not serially correlated. Table 8.3 gives the Stata output for the system GMM Blundell and Bond (1998) estimator using the *xtabond2* com-

Table 8.2 Robust Arellano and Bond GMM estimates of cigarette demand

```
. xtabond2 lnc L.(lnc) lnrp lnrpn lnrdi dum3 dum8 dum10-dum29, gmm(L.(lnc ),
collapse) iv(lnrp lnrpn lnrdi dum3 dum8 dum10-dum29)  noleveleq robust nomata
twostep
Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------
Group variable: state                           Number of obs      =      1288
Time variable : yr                              Number of groups   =        46
Number of instruments = 53                      Obs per group: min =        28
Wald chi2(26) =    5492.89                                     avg =     28.00
Prob > chi2   =      0.000                                     max =        28
------------------------------------------------------------------------------
             |              Corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnc |
         L1. |   .6993342   .0688738    10.15   0.000     .5643441    .8343242
        lnrp |  -.3956542   .0659587    -6.00   0.000    -.5249309   -.2663775
       lnrpn |  -.1054687   .0846056    -1.25   0.213    -.2712926    .0603552
       lnrdi |   .1258248   .0359738     3.50   0.000     .0553173    .1963322

------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -4.74  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   1.82  Pr > z =  0.069

Hansen test of overid. restrictions: chi2(27)    =  32.25  Prob > chi2 =  0.223
Warning: Sargan/Hansen tests are weak when instruments are many.
------------------------------------------------------------------------------
Time dummies are not shown here to save space.
```

mand developed by Roodman (2009). The moment conditions on equations in levels in addition to the moment conditions on the first differenced equation are used as described in Sect. 8.5. Once again the collapse option was used to reduce the number of moment conditions. Sargan test for over-identification does not reject the null, and the tests for first-order and second-order serial correlation yield the expected diagnostics. System GMM yields a lagged consumption coefficient estimate of 0.70 and an own price elasticity of -0.42, both highly significant, but with higher standard errors than the corresponding Arellano and Bond estimators reported in Table 8.2.

Table 8.4 applies the Keane and Runkle (1992) estimator to the cigarette demand example assuming the regressors are strictly exogenous using the *xtkr* command in Stata. This is the 2sls Keane–Runkle estimator for cigarette demand. The lagged dependent variable is 0.71 and significant. The short-run price elasticity is -0.31 and significant. Table 8.5 applies the Keane and Runkle (1992) estimator to the differenced cigarette demand regression assuming the regressors are strictly exogenous. The lagged dependent variable is 0.40 and significant. The short-run price elasticity is -0.34 and significant.

8.8.2 Example 2: Democracy and Education

Acemoglu et al. (2005) revisited the argument that education promotes democracy both because it enables a 'culture of democracy' to develop and because it leads to

Table 8.3 System GMM estimates of cigarette demand

```
. xtabond2 lnc L.(lnc) lnrp lnrpn lnrdi dum3 dum8 dum10-dum29,
gmm(L.(lnc),collapse) iv(lnrp lnrpn lnrdi dum3 dum8 dum10-dum29) robust nomata
twostep
Dynamic panel-data estimation, two-step system GMM
------------------------------------------------------------------------------
Group variable: state                           Number of obs      =      1334
Time variable : yr                              Number of groups   =        46
Number of instruments = 55                      Obs per group: min =        29
Wald chi2(26) =   6937.61                                      avg =     29.00
Prob > chi2   =     0.000                                      max =        29
------------------------------------------------------------------------------
             |              Corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnc |
         L1. |   .6961172    .078724     8.84   0.000     .5418211    .8504133
        lnrp |  -.4152584   .0962569    -4.31   0.000    -.6039185   -.2265982
       lnrpn |  -.0034555   .0500122    -0.07   0.945    -.1014777    .0945667
       lnrdi |   .0925415   .0276142     3.35   0.001     .0384187    .1466644

------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -4.63  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   1.92  Pr > z =  0.055

Hansen test of overid. restrictions: chi2(28)   =  26.79  Prob > chi2 =  0.530
Warning: Sargan/Hansen tests are weak when instruments are many.
------------------------------------------------------------------------------
Time dummies are not shown here to save space.
```

Table 8.4 Keane and Runkle estimates of cigarette demand in levels

```
. xtkr lnc lnrp lnrpn lnrdi  (l.lnc = l.lnrp l.lnrpn l.lnrdi)


Keane-Runkle (1992) Regression
Number of Obs: 1334                Number of Panel Units: 46
------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lnrp |  -.3107293   .0223926   -13.88   0.000    -.3546181   -.2668406
       lnrpn |    .070636   .0192248     3.67   0.000     .0329561     .108316
       lnrdi |  -.0146637   .0097829    -1.50   0.134    -.0338379    .0045105
             |
         lnc |
         L1. |   .7081463    .031187    22.71   0.000      .647021    .7692717
             |
    constant |   2.706642   .2853597     9.49   0.000     2.147347    3.265937
------------------------------------------------------------------------------
Instruments: lnrp lnrpn lnrdi constant L.lnrp L.lnrpn L.lnrdi
```

Table 8.5 Keane and Runkle estimates of cigarette demand in differences

```
. xtkr d.lnc d.lnrp d.lnrpn d.lnrdi (d.l.lnc = l(1/2).lnrp l(1/2).lnrpn l(1/2).lnrdi)


Keane-Runkle (1992) Regression
Number of Obs: 1288              Number of Panel Units: 46
------------------------------------------------------------------------------
      D.lnc |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnrp |
        D1. | -.3411479   .0184998   -18.44   0.000    -.3774069    -.304889
            |
      lnrpn |
        D1. |  .0827428   .0195492     4.23   0.000     .044427    .1210586
            |
      lnrdi |
        D1. |  .2085542   .0207969    10.03   0.000     .167793    .2493153
            |
        lnc |
        LD. |  .3959478   .0298621    13.26   0.000    .3374192    .4544764
            |
   constant | -.0077665   .0006011   -12.92   0.000   -.0089447   -.0065884
------------------------------------------------------------------------------
Instruments: D.lnrp D.lnrpn D.lnrdi constant L.lnrp L2.lnrp L.lnrpn L2.lnrpn L.lnrdi L2.lnrdi
```

greater prosperity, which is also thought to cause political development. There is a rich literature that argues that education broadens men's outlooks, increases their norms of tolerance, restrains them from adhering to extremist and monistic doctrines, and increases their capacity to make rational electoral choices. In fact, some argue that if we cannot say that a high level of education is a sufficient condition for democracy, the available evidence does suggest that it comes close to being a necessary condition. Acemoglu et al. (2005) show that the cross-sectional relationship between schooling and democracy disappears when country fixed effects are included in the regression. Democracy is measured using the Freedom House Political Rights Index (from 1 to 7). These are transformed to lie between 0 and 1, with 1 corresponding to the most democratic set of institutions. The data is a 5 yearly panel (1960–2000). Education is measured by the average years of schooling in the total population of age 25 and above. This varied between 0.04 and 12.18 with an average of 4.44. Acemoglu et al. (2005) also estimated their dynamic democracy equation using the two-step robust Arellano and Bond (1991) estimator, see their Table 8.1, column (iv).[6] Table 8.6 replicates their results using Stata with the (*xtabond2*, twostep) command. Note that lagged education is negative and insignificant using the robust standard errors. The diagnostics are fine, with the Sargan test not rejecting the over-identification conditions and the tests for serial correlation finding first order but not second-order serial correlation. However, Table 8.7 shows that applying system GMM to the same equation yields proper diagnostics and over turns the result on education. In fact, the

Table 8.6 Robust Arellano and Bond estimates of the democracy equation

```
. xtabond2 dem L.(dem educ) yr* if year>=1960&year<=2000&year>=(indyear+5),
gmm(L.(dem)) iv(L.educ yr*) noleveleq robust nomata
Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: code_numeric                    Number of obs      =       667
Time variable : year_numeric                    Number of groups   =       104
Number of instruments = 50                      Obs per group: min =         4
Wald chi2(8)   =    116.61                                     avg =      6.41
Prob > chi2    =     0.000                                     max =         7
------------------------------------------------------------------------------
             |              Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         dem |
         L1. |    .507231   .0958393     5.29   0.000     .3193895    .6950726
        educ |
         L1. |  -.0165729   .0216582    -0.77   0.444    -.0590222    .0258764
         yr4 |  -.1068231   .0623675    -1.71   0.087    -.2290612    .0154149
         yr5 |  -.1954687   .0620037    -3.15   0.002    -.3169936   -.0739437
         yr6 |  -.1564205   .0577853    -2.71   0.007    -.2696776   -.0431633
         yr7 |  -.0735737   .0451653    -1.63   0.103    -.1620961    .0149488
         yr8 |   -.056575   .0377228    -1.50   0.134    -.1305103    .0173602
         yr9 |   -.049678   .0284248    -1.75   0.081    -.1053895    .0060334
        yr10 |  -.0162374   .0231068    -0.70   0.482    -.0615258     .029051
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -5.82  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   0.25  Pr > z =  0.806

Hansen test of overid. restrictions: chi2(41)    =  45.05  Prob > chi2 =  0.306
Warning: Sargan/Hansen tests are weak when instruments are many.
------------------------------------------------------------------------------
```

lagged education coefficient estimate is now positive and significant. This result was reported by Bobba and Coviello (2007).

8.9 Selected Applications

There are hundreds of applications of the dynamic error component model. Here are a few of them:

(1) Holtz-Eakin, Newey and Rosen (1988) formulate a coherent set of procedures for estimating and testing vector autoregressions (VAR) with panel data. The model builds upon Chamberlain (1984) study and allows for nonstationary individual effects. It is applied to the study of dynamic relationships between wages and hours worked in two samples of American males. The data are based on a sample of 898 males from the PSID covering the period 1968–81. Two variables are considered for each individual, log of annual average hourly earnings, and log of annual hours of work. Some of the results are checked using data from the National Longitudinal Survey of Men 45–59. Tests for parameter stationarity, minimum lag length, and causality are performed. Holtz-Eakin, Newey and Rosen (1988) emphasize the importance of testing for the appropriate lag length before testing for causality, especially in short panels. Otherwise, misleading results on causality can be obtained. They suggest a simple method of estimating VAR equations with panel data that

Table 8.7 System GMM estimates of the democracy equation

```
. xtabond2 dem L.(dem educ) yr* if year>=1960&year<=2000&year>=(indyear+5),
gmm(L.(dem)) iv(L.educ yr*) robust nomata
Dynamic panel-data estimation, one-step system GMM
------------------------------------------------------------------------------
Group variable: code_numeric                    Number of obs      =       765
Time variable : year_numeric                    Number of groups   =       108
Number of instruments = 60                      Obs per group: min =         1
Wald chi2(9)   =   1706.36                                     avg =      7.08
Prob > chi2    =     0.000                                     max =         8
------------------------------------------------------------------------------
             |              Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dem   |
         L1. |   .5802369   .0609314     9.52   0.000     .4608136    .6996603
       educ  |
         L1. |   .0360906   .0059827     6.03   0.000     .0243646    .0478165
         yr4 |   .0926967   .0274313     3.38   0.001     .0389323    .1464611
         yr6 |   .0355514   .0337632     1.05   0.292    -.0306232     .101726
         yr7 |   .1027438   .0374142     2.75   0.006     .0294132    .1760743
         yr8 |   .0864848   .0318197     2.72   0.007     .0241193    .1488503
         yr9 |   .0708039   .0333395     2.12   0.034     .0054597    .1361482
        yr10 |   .0769046   .0375558     2.05   0.041     .0032967    .1505125
        yr11 |   .0664129    .032606     2.04   0.042     .0025063    .1303195
       _cons |   .0154328   .0350348     0.44   0.660    -.0532341    .0840996
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -6.07  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   0.44  Pr > z =  0.660

Hansen test of overid. restrictions: chi2(50)   =  49.68  Prob > chi2 =  0.486
Warning: Sargan/Hansen tests are weak when instruments are many.
------------------------------------------------------------------------------
```

has a straightforward GLS interpretation. This is based on applying instrumental variables to the quasi-differenced autoregressive equations. They demonstrate how inappropriate methods that deal with individual effects in a VAR context can yield misleading results. Another application of these VAR methods with panel data is Holtz-Eakin, Newey and Rosen (1989) who study the dynamic relationships between local government revenues and expenditures. The data are based on 171 municipal governments over the period 1972–80. It is drawn from the Annual Survey of Governments between 1973 and 1980 and the Census of Governments conducted in 1972 and 1977. The main findings include the following:

(1) Lags of one or two years are sufficient to summarize the dynamic interrelationships in local public finance.
(2) There are important intertemporal linkages among expenditures, taxes, and grants.
(3) Results of the stationarity test cast doubt over the stability of parameters over time.
(4) Contrary to previous studies, this study finds that past revenues help predict current expenditures, but past expenditures do not alter the future path of revenues.

(2) Blundell et al. (1992) apply the Arellano and Bond estimator to a panel of 532 UK manufacturing companies over the period 1975–86. They study the importance of Tobin's Q in the determination of investment decisions. Tobin's Q is allowed to be endogenous and possibly correlated with the firm-specific effects. Utilizing past variables as instruments, Tobin's Q effect is found to be small but significant. These results are sensitive to the choice of dynamic specification, exogeneity assumptions, and measurement error in Q.

(3) Becker, Grossman and Murphy (1994) estimate a rational addiction model for cigarettes using a panel of 50 states (and the District of Columbia) over the period 1955–85. They apply fixed effects 2SLS to estimate a second-order difference equation in consumption of cigarettes, finding support for forward looking consumers and rejecting myopic behavior. Their long-run price elasticity estimate is -0.78 as compared to -0.44 for the short-run. Baltagi and Griffin (2001) apply the FD-2SLS, FE-2SLS, and GMM dynamic panel estimation methods studied in this chapter to the Becker, Grossman and Murphy rational addiction model for cigarettes. Although the results are in general supportive of rational addiction, the estimates of the implied discount rate are not precise. Baltagi and Griffin (1995) estimate a dynamic demand for liquor across 43 states over the period 1960–82. Fixed effects 2SLS as well as FD-2SLS-KR are performed. A short-run price elasticity of -0.20 and a long-run price elasticity of -0.69 are reported. Their findings support strong habit persistence, a small positive income elasticity and weak evidence of bootlegging from adjoining states.

(4) Bond et al. (2003) estimate dynamic investment equations using company panel data for manufacturing firms in Belgium, France, Germany, and the United Kingdom, covering the period 1978–1989. Using GMM first difference estimation methods, they find that cash flow and profits appear to be both statistically and quantitatively more significant in the United Kingdom than in the three continental European countries. This is consistent with the suggestion that financial constraints on investment may be relatively severe in the more market-oriented U.K. financial system.

(5) Acemoglu et al. (2008) challenge the literature that finds income per capita is strongly correlated with the level of democracy across countries. Using an unbalanced panel for 150 countries over the period 1960–2000 at five-year intervals, and after controlling for country and time specific effects, Acemoglu et al. (2008) find that this positive association vanishes. Acemoglu et al. (2008) also find an insignificant effect of income on democracy by estimating a dynamic democracy equation using the two-step robust Arellano and Bond (1991) estimator; see their Table 8.2, column (2). However, Heid, Langer and Larch (2012) over turn this result by finding a significant positive relation between income and democracy using system GMM.

(6) Baltagi, Demetriades and Law (2009) use panel data of 42 developing countries over the period 1980–2003 to address the empirical question of whether trade and financial openness can help explain the pace in financial development, as well as its variation across countries. Using Arellano and Bond (1991) GMM dynamic panel estimation, they show that both types of openness are statistically significant determinants of banking sector development. They also show that the marginal effects

of trade (financial) openness are negatively related to the degree of financial (trade) openness, indicating that relatively closed economies stand to benefit most from opening up their trade and/or capital accounts. Although these economies may be able to accomplish more by taking steps to open both their trade and capital accounts, opening up one without the other could still generate gains in terms of banking sector development.

8.10 Further Reading

The literature on dynamic panel data models continues to exhibit phenomenal growth. This is understandable given that most of our economic models are implicitly or explicitly dynamic in nature. This section suggests some additional readings not covered in this chapter.

Hsiao (2003) shows that for the random effects dynamic model, the consistency property of MLE and GLS depends upon various assumptions on the initial observations and on the way in which and T tend to infinity. Read also the Arellano and Honoré (2001) chapter in the *Handbook of Econometrics*. Arellano's (2003) book is dedicated to the study of dynamic panel data models.

Chamberlain (1984) considers the panel data model as a multivariate regression of T equations subject to restrictions and derives an efficient minimum distance estimator that is robust to residual autocorrelation of arbitrary form. Chamberlain (1984) also first differences these equations to get rid of the individual effects and derives an asymptotically equivalent estimator to his efficient minimum distance estimator based on 3SLS of the $(T-2)$ differenced equations. Building on Chamberlain's work, Arellano (1990) develops minimum chi-square tests for various covariance restrictions. These tests are based on 3SLS residuals of the dynamic error component model and can be calculated from a generalized linear regression involving the sample autocovariance and dummy variables. The asymptotic distribution of the unrestricted autocovariance estimates is derived without imposing the normality assumption. In particular, Arellano (1990) considers testing covariance restrictions for error components or first-difference structures with white noise, moving average or autoregressive schemes. If these covariance restrictions are true, 3SLS is inefficient and Arellano (1990) proposes a GLS estimator which achieves asymptotic efficiency in the sense that it has the same limiting distribution as the optimal minimum distance estimator.

Kruiniger (2007) considered the GMM estimation of a simple dynamic panel data model with no exogenous regressors. He suggests a two-step optimal linear GMM estimator that is asymptotically equivalent to the optimal nonlinear GMM estimator of Ahn and Schmidt (1997) when the data are covariance stationary. When the model has a unit root, see Chap. 12, it is shown that in most cases this optimal linear GMM estimator is superconsistent, under a variety of assumptions about the initial observations and the initial estimator.

Wansbeek and Bekker (1996) considered a simple dynamic panel data model with no exogenous regressors and disturbances u_{it} and random effects μ_i that are independent and normally distributed. They derived an expression for the optimal instrumental variable estimator, i.e., one with minimal asymptotic variance. A striking result is the difference in efficiency between the IV and ML estimators. They find that for regions of the autoregressive parameter δ which are likely in practice, ML is superior. The gap between IV (or GMM) and ML can be narrowed down by adding moment restrictions of the type considered by Ahn and Schmidt (1995). Hence, Wansbeek and Bekker (1996) find support for adding these nonlinear moment restrictions and warn against the loss in efficiency as compared with MLE by ignoring them.

Bun and Kiviet (2006) analyze the finite sample behavior of the FE, GLS, and a range of GMM estimators in dynamic panel data models with individual effects and an additional regressor. The additional regressor may be correlated with the individual effects and is predetermined. Asymptotic expansions indicate how the order of magnitude of bias of these estimators depend on N and T. For example, they show that FE and GLS are biased of $O(1/T)$ irrespective of the value of N, while the GMM estimators are biased of the order $O(1/N)$, assuming T fixed. They also reveal how the bias of the GMM estimators tends to increase with the number of moment conditions exploited. They study both GMM based on the levels equation and those based on the forward orthogonalization procedure. They provide analytic evidence on how the bias of the various estimators depends on the feedbacks and on other model characteristics such as prominence of individual effects and correlation between observed and unobserved heterogeneity. Simulation results show that none of the techniques examined dominates regarding bias and mean squared error over all parametrizations examined. For N and T of moderate size, all estimators show substantial bias and poor RMSE performance leading the authors to conclude that "..standard first-order asymptotic theory is of little use indeed to establish and rank the qualities of the estimators".

Andrews and Lu (2001) develop consistent model and moment *selection criteria* and downward testing procedures for GMM estimation that are able to select the correct model and moments with probability that goes to one as the sample size goes to infinity. This is applied to dynamic panel data models with unobserved individual effects. The selection criteria can be used to select the lag length for the lagged dependent variables, to determine the exogeneity of the regressors, and/or to determine the existence of correlation between some regressors and the individual effects. Monte Carlo experiments are performed to study the small sample performance of the selection criteria and the testing procedures and their impact on parameter estimation.

Hahn and Kuersteiner (2002) consider the simple autoregressive model given in (8.32) with $\nu_{it} \sim N(0, \Omega)$ *iid* across i, $0 < \lim(N/T) = c < \infty$, $|\delta| < 1$ and $\sum_{i=1}^{N} y_{i0}^2/N = O(1)$ and $\sum_{i=1}^{N} \mu_i^2/N = O(1)$. The MLE of δ is the FE estimator which is inconsistent for fixed T and N $\rightarrow \infty$. For large T, large N, as in cross-country studies, such that lim(N/T) = c is finite, Hahn and Kuersteiner derive a bias corrected estimator which reduces to

$$\widehat{\delta}_c = (\frac{T+1}{T})\widetilde{\delta}_{FE} + \frac{1}{T} \tag{8.45}$$

with $\sqrt{NT}(\widehat{\delta}_c - \delta) \rightarrow N(0, 1-\delta^2)$. Under the assumption of normality of the disturbances, $\widehat{\delta}_c$ is assymptotically efficient as N, T$\rightarrow \infty$ at the same rate. Monte Carlo results for T = 5, 10, 20 and N = 100, 200 show that this bias-corrected MLE has comparable bias properties to the Arellano and Bover (1995) GMM estimator and often dominates in terms of RMSE for T = 10, 20 and N = 100, 200. Kiviet (1995) showed that a bias-corrected MLE (knowing δ) has much more desirable finite sample properties than various instrumental variable estimators. However, in order to make this estimator feasible, an initial instrumental variable for δ is used and its asymptotic properties are not derived. In contrast, Hahn and Kuersteiner (2002) correction does not require a preliminary estimate of δ and its asymptotic properties are well derived. They also showed that this bias corrected MLE is not expected to be asymptotically unbiased under a unit root ($\delta = 1$). Hahn and Moon (2006) showed that this result can be extended to dynamic linear panel data models with both individual and time effects. They find that the asymptotic bias of the fixed effects estimator is the same in the two way as in the one-way dynamic linear panel model without the time effect. Hence, the same higher order bias correction approach as in Hahn and Kuersteiner (2002) can be adopted even when time effects are present. They stress that such robustness is limited only to linear models.

Hahn, Hausman and Kuersteiner (2007) consider the simple autoregressive panel data model in (8.32) with the following strong assumptions: (i) $\nu_{it} \sim IIN(0, \sigma_\nu^2)$ over i and t, (ii) stationarity conditions $(y_{i0}/\mu_i){\sim}N(\frac{\mu_i}{1-\delta}, \frac{\sigma_\nu^2}{1-\delta^2})$ and $\mu_i \sim N(0, \sigma_\mu^2)$. They show that the Arellano and Bover (1995) GMM estimator, based on the forward demeaning transformation described in Problem 8.4, can be represented as a linear combination of 2SLS estimators and therefore may be subject to a substantial finite sample bias. Based on 5000 Monte Carlo replication, they show that this indeed the case for $T = 5, 10$; $N = 100, 500$ and $\delta = 0.1, 0.3, 0.5, 0.8$ and 0.9. For example, for $T = 5$, $N = 100$ and $\delta = 0.1$, the %bias of the GMM estimator is -16%. For $\delta = 0.8$,this %bias is -28%, and for $\delta = 0.9$,this %bias is -51%. Hahn, Hausman and Kuersteiner attempt to eliminate this bias using two different approaches. The first is a second-order Taylor series type approximation and the second is a long difference estimator. The Monte Carlo results show that the second-order Taylor series type approximation does a reasonably good job except when δ is close to 1 and N is small. Based on this, the bias corrected (second-order theory) should be relatively free of bias. Monte Carlo results show that this is the case unless δ is close 1. For $T = 5$, $N = 100$ and $\delta = 0.1, 0.8, 0.9$, the %bias for this bias-corrected estimator is 0.25%, -11% and -42%, respectively.

The second-order asymptotics fails to be a good approximation around $\delta = 1$. This is due to the *weak instrument* problem; see Blundell and Bond (1998) in Sect. 8.5. In fact, the latter paper argued that the weak IV problem can be alleviated by assuming stationarity on the initial observation y_{i0}. The stationarity condition turns out to be a predominant source of information around $\delta = 1$. The stationarity condition may or may not be appropriate for particular applications, and substantial finite sample biases due to inconsistency will result under violation of stationarity. Hahn, Hausman and Kuersteiner turn to the long difference estimator to deal with weak IV around the unit circle avoiding the stationarity assumption:

$$y_{it} - y_{i1} = \delta(y_{it} - y_{i0}) + \nu_{it} - \nu_{i1}$$

Here, y_{i0} is a valid instrument. The residuals $(y_{i,T-1} - \delta y_{i,T-2})$, …, $(y_{i,2} - \delta y_{i,1})$ are also valid instruments. To make it operational, they suggest using the Arellano and Bover estimator for the first step and iterating using the long difference estimator. The bias of the 2SLS (GMM) estimator depends on four factors, the sample size, the number of instruments, the covariance between the stochastic disturbance of the structural equation and the reduced form equation, and the explained variance of the first stage reduced form. The long difference estimator increases the R^2, but decreases the covariance between the stochastic disturbance of the structural equation and the reduced form equation. This alleviates the weak instruments problem. Further, the number of instruments is smaller for the long difference specification than for the first difference GMM and therefore one should expect smaller bias. The actual properties of the long difference estimator turn out to be much better than those predicted by higher order theory especially around the unit circle. Monte Carlo results show that the long difference estimator does better than the other estimators for large δ and not significantly different for moderate δ.

Hahn, Hausman and Kuersteiner analyze the class of GMM estimators that exploit the Ahn and Schmidt (1997) complete set of moment conditions and show that a strict subset of the full set of moment restrictions should be used in estimation in order to minimize bias. They show that the long difference estimator is a good approximation to the bias minimal procedure. They report the numerical values of the biases of the Arellano and Bond, Arellano and Bover and Ahn and Schmidt estimators under near unit root asymptotics and compare them with biases for the long difference estimator as well as the bias minimal estimator. Despite the fact that the long difference estimator does not achieve small bias reduction as the fully optimal estimator it has significantly less bias than the more commonly used implementations of the GMM estimator.

Evereart (2013) suggests transforming the dynamic panel data model into orthogonal deviations from its individual backward mean. In this case, the transformed lagged dependent variable is contemporaneously uncorrelated with the idiosyncratic error term. The estimators based on this orthogonal to backward mean transformation are referred to as WGob estimators. First, in a model with no additional exogenous regressors, the WGob estimator is obtained as the LS estimator after transforming the model in orthogonal deviations from the backward mean of the lagged dependent variable. Equivalently, it is obtained by (i) adding the backward mean of the lagged dependent variable as a regressor to the model, which then serves as a proxy for the individual effects, or (ii) instrumenting the lagged dependent variable by the orthogonal deviations from its backward mean, which is similar to the Hausman and Taylor (1981) representation of the WG estimator. Second, this Hausman–Taylor representation of the WGob estimator makes it very easy to add exogenous explanatory variables to the model which (a) serve as their own instruments when they are not correlated with the individual effect or (b) can be instrumented by the deviations from their sample mean when they are correlated with the individual effect. The WGob estimator is shown to be consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and T fixed. Monte Carlo experiments further show that overall, the small sample properties of the WGob estimator are superior to those of conventional dynamic panel data estimators, i.e., it considerably outperforms conventional estimators in terms of bias, dispersion, and inference in the cases where these estimators fail while not performing much worse in all other cases.

Read Chap. 2 on dynamic panel data models in the Oxford Handbook of Panel Data by Bun and Sarafidis (2015), especially for the sensitivity of GMM estimators to the assumptions on the initial condition. Also, Chap. 4 on the incidental parameter problem in dynamic panel models by Moon, Perron and Phillips (2015) and its effect on the bias and inconsistency of dynamic panel estimators as well as model selection criteria.

8.11 Notes

1. Other methods of estimating dynamic panel data models include quasi–maximum likelihood (QML) methods conditioning on the initial observation; see Bhargava and Sargan (1983) and Hsiao, Pesaran and Tahmiscioglu (2002).
2. Arellano and Bond (1991) warn about circumstances where their proposed serial correlation test is not defined, but where Sargan's over-identification test can still be computed. This is evident for $T = 4$ where no differenced residuals two periods apart are available to compute the serial correlation test. However, for the simple autoregressive model given in (8.3), Sargan's statistic tests two linear combinations of the three moment restrictions available, i.e., $E[(\nu_{i3} - \nu_{i2})y_{i1}] = E[(\nu_{i4} - \nu_{i3})y_{i1}] = E[(\nu_{i4} - \nu_{i3})y_{i2}] = 0$.
3. Arellano and Bover (1995) also discuss a forward orthogonal deviations operator as another example of C which is useful in the context of models with predetermined variables. This transformation essentially subtracts the mean of future observations available in the sample; see Problem 8.4.
4. It may be worth emphasizing that if $T > N$, this procedure will fail since Σ_{TS} will be singular with rank N. Also, the estimation of an unrestricted P_{TS} matrix will be difficult with missing data.
5. An alternative one-step method that achieves the same asymptotic efficiency as robust GMM or LIML estimators is the maximum empirical likelihood estimation method; see Imbens (1997). This maximizes a multinomial pseudo-likelihood function subject to the orthogonality restrictions. These are invariant to normalization because they are maximum likelihood estimators. See also Newey and

Smith (2004) who give general analytical bias-corrected versions of GMM and generalized empirical likelihood estimators.

6. The data set and Stata programs were kindly provided by Pierre Yared and Daron Acemoglu.

8.12 Problems

8.1 *Arellano and Bond estimator*. For the simple autoregressive model with no regressors given in (8.3)

(a) Write the first-differenced form of this equation for $t = 5$ and $t = 6$ and list the set of valid instruments for these two periods.
(b) Show that variance–covariance matrix of the first difference disturbances is given by (8.5).
(c) Verify that (8.8) is the GLS estimator of (8.7).

8.2 Consider the Monte Carlo setup given in Arellano and Bond (1991) (p. 283) for a simple autoregressive equation with one regressor with $N = 100$ and $T = 7$.

(a) Compute the bias and mean squared error based on 100 replications of the following estimators: OLS, Within, one-step and two-step Arellano and Bond GMM estimators, two Anderson and Hsiao type estimators that use $\Delta y_{i,t-2}$ and $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$, respectively. Compare with Table 8.1, p. 284 of Arellano and Bond (1991).
(b) Compute Sargan's test of over-identifying restrictions given below (8.11) and count the number of rejections out of 100 replications. Compare with Table 8.2 of Arellano and Bond (1991).

8.3 *Sargan's test of over-identifying restrictions*. For $T = 5$, list the moment restrictions available for the simple autoregressive model given in (8.3). What over-identifying restrictions are being tested by Sargan's statistic given below (8.11)?

8.4 *Alternative transformations that wipe out the individual effects*. Consider three $(T-1) \times T$ matrices defined in (8.15) as follows: $C_1 =$ the first $(T-1)$ rows of $(I_T - \bar{J}_T)$, $C_2 =$ the first-difference operator, $C_3 =$ the forward orthogonal deviations operator which subtracts the mean of future observations from the first $(T-1)$ observations. This last matrix is given by Arellano and Bover

(1995) as

$$C_3 = \text{diag}\left[\frac{T-1}{T}, \ldots, \frac{1}{2}\right]^{1/2}$$

$$\times \begin{bmatrix} 1 & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & \cdots & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} & -\frac{1}{(T-1)} \\ 0 & 1 & -\frac{1}{(T-2)} & \cdots & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} & -\frac{1}{(T-2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \ldots & 0 & 1 & -1 \end{bmatrix}$$

Verify that each one of these C matrices satisfies

(a) $C_j \iota_T = 0$ for $j = 1, 2, 3$.
(b) $C_j'(C_j C_j')^{-1} C_j = I_T - \bar{J}_T$, the Within transformation, for $j = 1, 2, 3$.
(c) For C_3, show that $C_3 C_3' = I_{T-1}$ and $C_3' C_3 = I_T - \bar{J}_T$. Hence $C_3 = (C'C)^{-1/2} C$ for any upper triangular C such that $C \iota_T = 0$.

8.5 *Arellano and Bover estimator.*

(a) Verify that GLS on (8.19) yields (8.20).
(b) For the error component model with $\widetilde{\Omega} = \widetilde{\sigma}_\nu^2 I_T + \widetilde{\sigma}_\mu^2 J_T$ and $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_\mu^2$ denoting consistent estimates of σ_v^2 and σ_μ^2, respectively, show that $\widehat{\eta}$ in (8.20) can be written as

$$\widehat{\eta} = \left[\sum_{i=1}^{N} W_i'(I_T - \bar{J}_T)W_i + \widetilde{\theta}^2 T \sum_{i=1}^{N} \bar{w}_i m_i' \left(\sum_{i=1}^{N} m_i m_i'\right)^{-1} \sum_{i=1}^{N} m_i \bar{w}_i'\right]^{-1}$$
$$\times \left[\sum_{i=1}^{N} W_i'(I_T - \bar{J}_T)y_i + \widetilde{\theta}^2 T \sum_{i=1}^{N} \bar{w}_i m_i' \left(\sum_{i=1}^{N} m_i m_i'\right)^{-1} \sum_{i=1}^{N} m_i \bar{y}_i\right]$$

where $\bar{w}_i = W_i' \iota_T / T$ and $\widetilde{\theta}^2 = \widetilde{\sigma}_\nu^2 / (T\widetilde{\sigma}_\mu^2 + \widetilde{\sigma}_\nu^2)$. These are the familiar expressions for the HT, AM, and BMS estimators for the corresponding choices of m_i. (Hint: see the proof in the Appendix of Arellano and Bover (1995)).

8.6 For $T = 4$ and the simple autoregressive model considered in (8.3)

(a) What are the moment restrictions given by (8.25)? Compare with Problem 8.3.
(b) What are the additional moment restrictions given by (8.26)?
(c) Write down the system of equations to be estimated by 3SLS using these additional restrictions and list the matrix of instruments for each equation.

8.7 *Dynamic demand for cigarettes*. Consider the Baltagi and Levin (1986) cigarette demand example for 46 states described in Sect. 8.9. This data, updated from 1963–92, is available on the Springer web site as cigar.txt.

(a) Estimate equation (8.43) using 2SLS, FD-2SLS and their Keane and Runkle (1992) version. (Assume only $\ln C_{i,t-1}$ is endogenous).
(b) Estimate question (8.43) using the Within and FE-2SLS and perform the Hausman-type test based on FE-2SLS versus FD-2SLS.
(c) Perform the Hausman-type test based on 2SLS versus FD-2SLS.
(d) Perform the Anderson and Hsiao (1982) estimator for equation (8.43).
(e) Replicate the two-step robust Arellano and Bond (1991) GMM estimator for equation (8.43) given in Table 8.2.
(f) Replicate the System GMM estimator for equation (8.43) given in Table 8.3.

Hint: Some of the results are available in Table 1 of Baltagi, Griffin and Xiong (2000).

8.8 Consider the Arellano and Bond (1991) *dynamic employment* equation for 140 UK companies over the period 1979–1984. Stata has this data set as *abdata1*. Replicate Table 4 of Arellano and Bond (1991) (p. 290) columns (a) and (b) using *xtabond*. Perform the serial correlation test using *estat abond* and the Sargan test for overidentification using *estat sargan*.

8.9 *Democracy and Education*. Consider the Acemoglu et al. (2005) dynamic democracy equation described in Sect. 8.8. Replicate all the estimation results in Table 1 of Acemoglu et al. (2005, p. 46). Check the sensitivity of these results to running system GMM rather than Arellano and Bond GMM using *xtabond2*, see Bobba and Coviello (2007)? Does the use of robust standard errors using the Windmeijer (2005) small sample correction change the significance of these coefficients?

8.10 *Homicide rates*. In Chap. 3, Problem 3.18, we replicated Neumayer (2003)'s two-way fixed effects estimates using a panel of homicide data from up to 117 countries over the period 1980–97. Neumayer (2003) also ran dynamic panel data estimation using Arellano and Bond GMM; see column 3 of his Table III on p. 632. He found insignificant lagged effects of homicide. However, zero first-order serial correlation was not rejected. This renders the use of lagged homicide rates as instruments invalid. The data set and Stata code are provided on the author's university web page. (http://www2.lse.ac.uk/geographyAndEnvironment/whosWho/profiles/neumayer/replicationdatasets2.aspx).

(a) Replicate column 3 of Table III of Neumayer (2003) (p. 629) which reports Arellano and Bond GMM. Perform the Sargan over-identification test and zero first and second-order tests for serial correlation. What do you conclude?
(b) Perform system GMM rather than Arellano and Bond GMM using *xtdpdsys* and the corresponding diagnostics. What do you conclude?

8.11 *Dynamic Investment and Tobin's q*. For the investment equation based on Tobin's q described in Problem 4.20, perform the dynamic regressions given in Table IV of Schaller (1990) (p. 317), only do that using the Arellano and Bond (1991) estimation procedure. Report the diagnostics on serial correlation and Sargan over-identification test.

8.12 *Democracy Does Cause Growth.* Acemoglu et al. (2019) provide evidence that democracy has a significant and robust positive effect on log GDP per capita. They estimate a dynamic panel data model of log GDP per capita for 175 countries over the period 1960–2010 using 4 lags on the dependent variable. They introduce a new dichotomous measure of democracy that consolidates the information from several sources. Their results show that democratizations increase GDP per capita by about 20% in the long run. The data set and Stata programs are available on Acemoglu's web site https://economics.mit.edu/faculty/acemoglu/data.

(a) Replicate the first 4 columns of Table 2 of Acemoglu et al. (2019) which run the FE dynamic panel model with one, two, four, and eight lags of log (GDP per capita) and democracy. Both country and time dummies are included. In particular, verify that their preferred specification of 4 lags, reported in column 3, yields a positive and significant effect of democracy on log (GDP per capita) with a long-run effect of 21.24 and standard error of (7.22).

(b) Replicate the next 4 columns of Table 2 of Acemoglu et al. (2019) which run the Arellano and Bond (1991) one-step dynamic panel model with one, two, four, and eight lags of log (GDP per capita) and democracy. Time dummies are included. In particular, verify that their preferred specification of 4 lags, reported in column 7, yields a long-run effect of democracy on log (GDP per capita) of 16.45 with a standard error of (8.43).

8.13 *Quasi–Maximum Likelihood (QML) estimator.* Using Kripfganz (2016) *xtdpdqml* command in Stata, replicate his illustration using the Arellano and Bond (1991) dynamic employment empirical application in Problem 8.8. Unlike the original study, the dynamics is restricted to one lag on the dependent variable while Arellano and Bond (1991) used two lags, and there are no distributed lag on the exogenous variables when Arellano and Bond (1991) used up to two lags. The illustration also omits the industry output variable arguing that Arellano and Bond (1991) found it insignificant. This applies QML of Hsiao, Pesaran and Tahmiscioglu (2002) as an alternative to the GMM estimator of Arellano and Bond (1991).

8.14 *Dynamic Hausman–Taylor Two-Stage Estimator.* Using the gravity model for foreign direct investment (FDI) of Egger and Pfaffermayr (2004) considered in Problem 7.17 where they applied a static Hausman and Taylor (1981) model, you are asked to replicate the results reported in Table 4 of Kripfganz and Schwarz (2018) (p. 538) now for a dynamic version of this gravity model using the *xtseqreg* command in Stata. Are the results sensitive to introducing dynamics?

8.15 *Dynamic Earnings Equation.* In Sect. 7.5, we considered a Hausman and Taylor (1981) estimator for a static earnings equation using the study of Cornwell and Rupert (1988); see Table 7.5. This used data on 595 individuals drawn from the PSID for the period 1976–1982. Using this data estimate a dynamic version of this earnings equation using the Kripfganz and Schwarz (2018) two-step estimator. (Hint: use the *xtseqreg* command). Are the results sensitive to introducing dynamics? Is the lagged dependent variable significant? What is the returns to schooling? Do females make less than males and is that significant?

References

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2005. From education to democracy? *American Economic Review* 95: 44–49.

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2008. Income and democracy. *American Economic Review* 98: 808–842.

Acemoglu, D., S. Naidu, P. Restrepo, and J.A. Robinson. 2019. Democracy Does Cause Growth. *Journal of Political Economy* 127 (1): 47–100.

Ahn, S.C., and P. Schmidt. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 5–27.

Ahn, S.C., and P. Schmidt. 1997. Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics* 76: 309–321.

Alonso-Borrego, C., and M. Arellano. 1999. Symmetrically normalized instrumental variable estimation using panel data. *Journal of Business and Economic Statistics* 17: 36–49.

Alvarez, J., and M. Arellano. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71: 1121–1159.

Amemiya, T., and T.E. MaCurdy. 1986. Instrumental-variable estimation of an error components model. *Econometrica* 54: 869–881.

Anderson, T.W., and C. Hsiao. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47–82.

Andrews, D.W.K., and B. Lu. 2001. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics* 101: 123–164.

Arellano, M. 1989. A note on the Anderson-Hsiao estimator for panel data. *Economics Letters* 31: 337–341.

Arellano, M. 1990. Some testing for autocorrelation in dynamic random effects models. *Review of Economic Studies* 57: 127–134.

Arellano, M. 2003. *Panel data econometrics*. Oxford: Oxford University Press.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Arellano, M., and O. Bover. 1995. Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics* 68: 29–51.

Arellano, M., and B. Honoré. 2001. Panel data models: Some recent developments. Chapter 53. In *Handbook of econometrics*, (eds.) J. Heckman and E. Leamer, 3229–3296. North-Holland, Amsterdam.

Baltagi, B.H., P.O. Demetriades, and S.H. Law. 2009. Financial development and openness: Evidence from panel data. *Journal of Development Economics* 89: 285–296.

Baltagi, B.H., and J.M. Griffin. 1995. A dynamic demand model for liquor: The case for pooling. *Review of Economics and Statistics* 77: 545–553.

Baltagi, B.H., and J.M. Griffin. 2001. The econometrics of rational addiction: The case of cigarettes. *Journal of Business and Economic Statistics* 19: 449–454.

Baltagi, B.H., J.M. Griffin, and W. Xiong. 2000. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82: 117–126.

Baltagi, B.H., and D. Levin. 1986. Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation, and advertising reconsidered. *Review of Economics and Statistics* 68: 148–155.

Becker, G.S., M. Grossman, and K.M. Murphy. 1994. An empirical analysis of cigarette addiction. *American Economic Review* 84: 396–418.

Bhargava, A., and J.D. Sargan. 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51: 1635–1659.

Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.

Blundell, R., and S. Bond. 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19: 321–340.

Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli. 1992. Investment and Tobin's q: Evidence from company panel data. *Journal of Econometrics* 51: 233–257.

Bobba, M., and D. Coviello. 2007. Weak instruments and weak identification in estimating the effects of education on democracy. *Economics Letters* 96: 301–306.

Bond, S., C. Bowsher, and F. Windmeijer. 2001. Criterion-based inference for GMM in autoregressive panel data models. *Economics Letters* 73: 379–388.

Bond, S., J.A. Elston, J. Mairesse, and B. Mulkay. 2003. Financial factors and investment in Belgium, France, Germany, and the United Kingdom: A Comparison using company panel data. *The Review of Economics and Statistics* 85: 153–165.

Bowsher, C.G. 2002. On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* 77: 211–220.

Breusch, T.S., G.E. Mizon, and P. Schmidt. 1989. Efficient estimation using panel data. *Econometrica* 57: 695–700.

Bun, M.J.G., and J.F. Kiviet. 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132: 409–444.

Bun, M.J.G., and V. Sarafidis. 2015. Dynamic panel data models. Chapter 3. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 76–110. Oxford: Oxford University Press.

Bun, M.J.G., and F. Windmeijer. 2010. The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13: 95–126.

Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.

Chamberlain, G. 1984. Panel data. Chapter 22. In *Handbook of econometrics*, (eds.) Z. Griliches, and M. Intrilligator, 1247–1318. North-Holland, Amsterdam.

Cornwell, C., and P. Rupert. 1988. Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 3: 149–155.

Dessi, R., and D. Robertson. 2003. Debt, incentives and performance: Evidence from UK panel data. *The Economic Journal* 113: 903–919.

Egger, P., and M. Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19: 227–246.

Evereart, G. 2013. Orthogonal to backward mean transformation for dynamic panel data models. *Econometrics Journal* 16: 179–221.

Griliches, Z., and J.A. Hausman. 1986. Errors in variables in panel data. *Journal of Econometrics* 31: 93–118.

Griliches, Z., and J. Mairesse. 1998. Production functions: The search for identification. In *Essays in honour of Ragnar Frisch, Econometric Society Monograph Series*, (ed.) S. Strom. Cambridge: Cambridge University Press.

Hahn, J., J. Hausman, and G. Kuersteiner. 2007. Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics* 140: 574–617.

Hahn, J., and G. Kuersteiner. 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70: 1639–1657.

Hahn, J., and H.R. Moon. 2006. Reducing bias of MLE in a dynamic panel model. *Econometric Theory* 22: 499–512.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.

Hayakawa, K. 2009. On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics* 153: 133–135.

Heid, B., J. Langer, and M. Larch. 2012. Income and democracy: Evidence from system GMM estimates. *Economics Letters* 116: 166–169.

Holtz-Eakin, D., W. Newey, and H.S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–1395.

Holtz-Eakin, D., W. Newey, and H.S. Rosen. 1989. The revenues–expenditures nexus: Evidence from local government data. *International Economic Review* 30: 415–429.

Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu. 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109: 107–150.

Hsiao, C. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.

Imbens, G. 1997. One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies* 64: 359–383.

Judson, R.A., and A.L. Owen. 1999. Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters* 65: 9–15.

Keane, M., and T. Neal. 2016. The Keane and Runkle estimator for panel-data models with serial correlation and instruments that are not strictly exogenous. *Stata Journal* 16: 523–549.

Keane, M.P., and D.E. Runkle. 1992. On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *Journal of Business and Economic Statistics* 10: 1–9.

Kiviet, J.F. 1995. On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68: 53–78.

Kripfganz, S. 2016. Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal* 16: 1013–1038.

Kripfganz, S., and C. Schwarz. 2018. Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics* 34: 526–546.

Kruiniger, H. 2007. An efficient linear GMM estimator for the covariance stationary AR(1)/unit root model for panel data. *Econometric Theory* 23: 519–535.

Moon, H.R., B. Perron, and P.C.B. Phillips. 2015. Incidental parameters and dynamic panel modeling. Chapter 4. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 111–148. Oxford: Oxford University Press.

Neumayer, E. 2003. Good policy can lower violent crime: Evidence from a cross-national panel of homicide rates, 1980–97. *Journal of Peace Research* 40: 619–640.

Newey, W.K., and R.J. Smith. 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72: 214–253.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

Roodman, D.M. 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71: 135–158.

Schaller, H. 1990. A re-examination of the q theory of investment using U.S. firm data. *Journal of Applied Economics* 5: 209–325.

Schmidt, P., S.C. Ahn, and D. Wyhowski. 1992. Comment. *Journal of Business and Economic Statistics* 10: 10–14.

Tauchen, G. 1986. Statistical properties of generalized method of moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics* 4: 397–416.

Wansbeek, T.J., and P. Bekker. 1996. On IV, GMM and ML in a dynamic panel data model. *Economics Letters* 51: 145–152.

Wansbeek, T.J., and T. Knapp. 1999. Estimating a dynamic panel data model with heterogeneous trends. *Annales D'Économie et de Statistique* 55–56: 331–349.

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51.

Ziliak, J.P. 1997. Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business and Economic Statistics* 15: 419–431.

Unbalanced Panel Data Models

9

9.1 Introduction

So far we have dealt only with "complete panels" or "balanced panels", i.e., cases where the individuals are observed over the entire sample period. Incomplete panels are more likely to be the norm in typical economic empirical settings. For example, in collecting data on US airlines over time, a researcher may find that some firms have dropped out of the market while new entrants emerged over the sample period observed. Similarly, while using labor or consumer panels on households, one may find that some households moved and can no longer be included in the panel. Additionally, if one is collecting data on a set of countries over time, a researcher may find that some countries can be traced back longer than others. These typical scenarios lead to "unbalanced" or "incomplete" panels. This chapter deals with the econometric problems associated with these incomplete panels and how they differ from the complete panel data case. Throughout this chapter, the panel data are assumed to be incomplete due to randomly missing observations. Nonrandomly missing data and rotating panels will be considered in Chap. 10.[1] Section 9.2 starts with the simple one-way error component model case with unbalanced data and surveys the estimation methods proposed in the literature. Section 9.4 treats the more complicated two-way error component model with unbalanced data. Section 9.5 looks at how some of the tests introduced earlier in the book are affected by the unbalanced panel, while Sect. 9.6 gives some extensions of these unbalanced panel data methods to the nested error component model.

9.2 The Unbalanced One-Way Error Component Model

To simplify the presentation, we analyze the case of two cross-sections with unequal number of time-series observations. Then, we generalize the analysis to the case of N cross-sections. Let n_1 be the shorter time series observed for the first cross-section

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_9

$(i = 1)$, and n_2 be the extra time-series observations available for the second cross-section $(i = 2)$.[2] Stacking the n_1 observations for the first individual on top of the $(n_1 + n_2)$ observations on the second individual, we get

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{9.1}$$

where y_1 and y_2 are vectors of dimensions n_1 and $n_1 + n_2$, respectively. X_1 and X_2 are matrices of dimensions $n_1 \times K$ and $(n_1 + n_2) \times K$, respectively. In this case, $u'_1 = (u_{11}, \ldots, u_{1,n_1})$, $u'_2 = (u_{21}, \ldots, u_{2,n_1}, \ldots, u_{2,n_1+n_2})$ and the variance–covariance matrix is given by

$$\Omega = \begin{bmatrix} \sigma_\nu^2 I_{n_1} + \sigma_\mu^2 J_{n_1 n_1} & 0 & 0 \\ 0 & \sigma_\nu^2 I_{n_1} + \sigma_\mu^2 J_{n_1 n_1} & \sigma_\mu^2 J_{n_1 n_2} \\ 0 & \sigma_\mu^2 J_{n_2 n_1} & \sigma_\nu^2 I_{n_2} + \sigma_\mu^2 J_{n_2 n_2} \end{bmatrix} \tag{9.2}$$

where $u' = (u'_1, u'_2)$, I_{n_i} denotes an identity matrix of order n_i, and $J_{n_i n_j}$ denotes a matrix of ones of dimension $n_i \times n_j$. Note that all the nonzero off-diagonal elements of Ω are equal to σ_μ^2. Therefore, if we let $T_j = \sum_{i=1}^{j} n_i$ for $j = 1, 2$, then Ω is clearly block-diagonal, with the jth block

$$\Omega_j = (T_j \sigma_\mu^2 + \sigma_\nu^2) \bar{J}_{T_j} + \sigma_\nu^2 E_{T_j} \tag{9.3}$$

where $\bar{J}_{T_j} = J_{T_j}/T_j$, $E_{T_j} = I_{T_j} - \bar{J}_{T_j}$ and there is no need for the double subscript anymore. Using the Wansbeek and Kapteyn (1982) trick extended to the unbalanced case, Baltagi (1985) derived

$$\Omega_j^r = (T_j \sigma_\mu^2 + \sigma_\nu^2)^r \bar{J}_{T_j} + (\sigma_\nu^2)^r E_{T_j} \tag{9.4}$$

where r is any scalar. Let $w_j^2 = T_j \sigma_\mu^2 + \sigma_\nu^2$, then the Fuller and Battese (1974) transformation for the unbalanced case is the following:

$$\sigma_\nu \Omega_j^{-1/2} = (\sigma_\nu / w_j) \bar{J}_{T_j} + E_{T_j} = I_{T_j} - \theta_j \bar{J}_{T_j} \tag{9.5}$$

where $\theta_j = 1 - \sigma_\nu / w_j$, and $\sigma_\nu \Omega_j^{-1/2} y_j$ has a typical element $(y_{jt} - \theta_j \bar{y}_{j.})$ with $\bar{y}_{j.} = \sum_{t=1}^{T_j} y_{jt}/T_j$. Note that θ_j varies for each cross-sectional unit j depending on T_j. Hence, GLS can be obtained as a simple weighted least squares (WLS) as in the complete panel data case. The basic difference, however, is that in the incomplete panel data case, the weights are crucially dependent on the lengths of the time series available for each cross-section.

The above results generalize in two directions: (i) the same analysis applies no matter how the observations for the two firms overlap; (ii) the results extend from the two cross-sections to the N cross-sections case. The proof is simple. Since the off-diagonal elements of the covariance matrix are zero for observations belonging to different firms, Ω remains block-diagonal as long as the observations are ordered by firms. Also, the nonzero off-diagonal elements are all equal to σ_μ^2. Hence, $\Omega_j^{-1/2}$ can be derived along the same lines shown above.

In general, the regression model with unbalanced one-way error component disturbances is given by

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T_i \tag{9.6}$$
$$u_{it} = \mu_i + \nu_{it}$$

where X_{it} is a $(K-1) \times 1$ vector of regressors, $\mu_i \sim \text{IIN}(0, \sigma^2_\mu)$ and independent of $\nu_{it} \sim \text{IIN}(0, \sigma^2_\nu)$. This model is unbalanced in the sense that there are N individuals observed over varying time-period length (T_i for $i = 1, \ldots, N$). Writing this equation in vector form, we have

$$y = \alpha\iota_n + X\beta + u = Z\delta + u \tag{9.7}$$
$$u = Z_\mu\mu + \nu$$

where y and Z are of dimensions $n \times 1$ and $n \times K$, respectively, $Z = (\iota_n, X)$, $\delta' = (\alpha', \beta')$, $n = \sum T_i$, $Z_\mu = \text{diag}(\iota_{T_i})$ and ι_{T_i} is a vector of ones of dimension T_i. $\mu = (\mu_1, \mu_2, \ldots, \mu_N)'$ and $\nu = (\nu_{11}, \ldots, \nu_{1T_1}, \ldots, \nu_{N1}, \ldots, \nu_{NT_N})'$.

The ordinary least squares (OLS) on the unbalanced data is given by

$$\widehat{\delta}_{OLS} = (Z'Z)^{-1}Z'y \tag{9.8}$$

OLS is the best linear unbiased estimator when the variance component σ^2_μ is equal to zero. Even when σ^2_μ is positive, OLS is still unbiased and consistent, but its standard errors are biased (see Moulton 1986). The OLS residuals are denoted by $\widehat{u}_{OLS} = y - Z\widehat{\delta}_{OLS}$.

The Within estimator can be obtained by first transforming the dependent variable y and X, the exogenous regressors excluding the intercept, using the matrix $Q = \text{diag}(E_{T_i})$, and then applying OLS to the transformed data:

$$\widetilde{\beta} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y} \tag{9.9}$$

where $\widetilde{X} = QX$, $\widetilde{y} = Qy$. The estimate of the intercept can be retrieved as follows: $\widetilde{\alpha} = (\bar{y}_{..} - \bar{X}_{..}\widetilde{\beta})$ where the dot indicates summation and the bar indicates averaging, for example, $\bar{y}_{..} = \sum\sum y_{it}/n$. Least squares dummy variables (LSDV) estimator was shown to be equivalent to the Within estimator for balanced panels in Chap. 2. This result remains true for unbalanced panels since Q is still the orthogonal projection on the matrix of individual dummy variables. See also Wooldridge (2010) and Abrevaya (2013). Following Amemiya (1971), the Within residuals $\widetilde{u}$ for the unbalanced panel are given by

$$\widetilde{u} = y - \widetilde{\alpha}\iota_n - X\widetilde{\beta} \tag{9.10}$$

The Between estimator $\widehat{\delta}_{Between}$ is obtained as follows:

$$\widehat{\delta}_{Between} = (Z'PZ)^{-1}Z'Py \tag{9.11}$$

where $P = \text{diag}[\bar{J}_{T_i}]$, and the Between residuals are denoted by $\hat{u}^b = y - Z\widehat{\delta}_{Between}$.

GLS using the true variance components is obtained as follows:

$$\widehat{\delta}_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y \tag{9.12}$$

where $\Omega = \sigma_\nu^2 \Sigma = E(uu')$ with

$$\Sigma = I_n + \rho Z_\mu Z_\mu' = \text{diag}(E_{T_i}) + \text{diag}[(1+\rho T_i)\bar{J}_{T_i}] \tag{9.13}$$

and $\rho = \sigma_\mu^2/\sigma_\nu^2$. Note that $(1+\rho T_i) = (w_i^2/\sigma_\nu^2)$ where $w_i^2 = (T_i\sigma_\mu^2 + \sigma_\nu^2)$ was defined in (9.4). Therefore, GLS can be obtained by applying OLS on the transformed variables y^* and Z^*, i.e.,

$$\widehat{\delta} = (Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^*$$

where $Z^* = \sigma_\nu \Omega^{-1/2} Z$, $y^* = \sigma_\nu \Omega^{-1/2} y$ and

$$\sigma_\nu \Omega^{-1/2} = \text{diag}(E_{T_i}) + \text{diag}[(\sigma_\nu/w_i)\bar{J}_{T_i}] \tag{9.14}$$

as described in (9.5).

We now focus on methods of estimating the variance components, which are described in more detail in Baltagi and Chang (1994).

9.2.1 ANOVA Methods

The ANOVA method is one of the most popular methods in the estimation of variance components. The ANOVA estimators are method of moments type estimators, which equate quadratic sums of squares to their expectations and solve the resulting linear system of equations. For the balanced model, ANOVA estimators are best quadratic unbiased (BQU) estimators of the variance components (see Searle 1971). Under normality of the disturbances, these ANOVA estimators are minimum variance unbiased. For the unbalanced one-way model, BQU estimators of the variance components are a function of the variance components themselves. Still, unbalanced ANOVA methods are available (see Searle 1987), but optimal properties beyond unbiasedness are lost. In what follows, we generalize some of the ANOVA methods described in Chap. 2 to the unbalanced case. In particular, we consider the two quadratic forms defining the Within and Between sums of squares:

$$q_1 = u'Qu \quad \text{and} \quad q_2 = u'Pu \tag{9.15}$$

where $Q = \text{diag}[E_{T_i}]$ and $P = \text{diag}[\bar{J}_{T_i}]$. Since the true disturbances are not known, we follow the Wallace and Hussain (1969) suggestion by substituting OLS residuals $\widehat{u}_{OLS}$ for u in (9.15). Upon taking expectations, we get

$$\begin{aligned} E(\widehat{q}_1) &= E(\hat{u}_{OLS}'Q\hat{u}_{OLS}) = \delta_{11}\sigma_\mu^2 + \delta_{12}\sigma_\nu^2 \\ E(\widehat{q}_2) &= E(\hat{u}_{OLS}'P\hat{u}_{OLS}) = \delta_{21}\sigma_\mu^2 + \delta_{22}\sigma_\nu^2 \end{aligned} \tag{9.16}$$

where $\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}$ are given by

$$\begin{aligned} \delta_{11} &= \text{tr}((Z'Z)^{-1}Z'Z_\mu Z_\mu' Z) - \text{tr}((Z'Z)^{-1}Z'PZ(Z'Z)^{-1}Z'Z_\mu Z_\mu' Z) \\ \delta_{12} &= n - N - K + \text{tr}((Z'Z)^{-1}Z'PZ) \\ \delta_{21} &= n - 2\text{tr}((Z'Z)^{-1}Z'Z_\mu Z_\mu' Z) + \text{tr}((Z'Z)^{-1}Z'PZ(Z'Z)^{-1}Z'Z_\mu Z_\mu' Z) \\ \delta_{22} &= N - \text{tr}((Z'Z)^{-1}Z'PZ) \end{aligned}$$

Equating $\widehat{q}_i$ to its expected value $E(\widehat{q}_i)$ in (9.16) and solving the system of equations, we get the Wallace and Hussain (WH) type estimators of the variance components.

Alternatively, we can substitute Within residuals in the quadratic forms given in (9.15) to get $\widetilde{q}_1 = \widetilde{u}'Q\widetilde{u}$ and $\widetilde{q}_2 = \widetilde{u}'P\widetilde{u}$ as suggested by Amemiya (1971) for the balanced case. The expected values of $\widetilde{q}_1$ and $\widetilde{q}_2$ are given by

$$E(\widetilde{q}_1) = (n - N - K + 1)\sigma_v^2$$

$$E(\widetilde{q}_2) = (N - 1 + \text{tr}[(X'QX)^{-1}X'PX] - \text{tr}[(X'QX)^{-1}X'\bar{J}_nX])\sigma_\nu^2 + \left[n - \left(\sum_{i=1}^{N} T_i^2/n\right)\right]\sigma_\mu^2 \quad (9.17)$$

Equating $\widetilde{q}_i$ to its expected value $E(\widetilde{q}_i)$ in (9.17), we get the Amemiya-type estimators of the variance components

$$\widehat{\sigma}_\nu^2 = \widetilde{u}'Q\widetilde{u}/(n - N - K + 1) \quad (9.18)$$

$$\widehat{\sigma}_\mu^2 = \frac{\widetilde{u}'P\widetilde{u} - \{N - 1 + \text{tr}[(X'QX)^{-1}X'PX] - \text{tr}[(X'QX)^{-1}X'\bar{J}_nX]\}\widehat{\sigma}_\nu^2}{n - \sum_{i=1}^{N} T_i^2/n}$$

Next, we follow the Swamy and Arora (1972) suggestion of using the Between and Within regression mean square errors to estimate the variance components. In fact, their method amounts to substituting Within residuals in q_1 and Between residuals in q_2, to get $\widetilde{q}_1 = \widetilde{u}'Q\widetilde{u}$ and $\hat{q}_2^b = \hat{u}^{b\prime}P\hat{u}^b$. Since $\widetilde{q}_1$ is exactly the same as that for the Amemiya method, the Swamy and Arora (SA) type estimator of $\widehat{\sigma}_\nu^2$ is the same as that given in Eq. (9.18). The expected value of $\hat{q}_2^b$ is given by

$$E(\hat{q}_2^b) = [n - \text{tr}((Z'PZ)^{-1}Z'Z_\mu Z_\mu' Z)]\sigma_\mu^2 + (N - K)\sigma_\nu^2 \quad (9.19)$$

Equating $E(\hat{q}_2^b)$ to $\hat{q}_2^b$ one gets the following estimator of σ_μ^2:

$$\widehat{\sigma}_\mu^2 = \frac{\hat{u}^{b\prime}P\hat{u}^b - (N - K)\widehat{\sigma}_\nu^2}{n - \text{tr}((Z'PZ)^{-1}Z'Z_\mu Z_\mu' Z)} \quad (9.20)$$

Note that $\hat{u}^{b'}P\hat{u}^b$ can be obtained as the OLS residual sum of squares from the regression involving $\sqrt{T_i}\bar{y}_{i.}$ on $\sqrt{T_i}\bar{Z}_{i.}$.

Finally, we consider Henderson's method III (see Fuller and Battese 1974) which will be denoted by HFB. This method utilizes the fitting constants method described in Searle (1971, p. 489). Let

$$R(\mu) = y'Z_\mu(Z_\mu' Z_\mu)^{-1}Z_\mu' y = \sum_{i=1}^{N}(y_{i.}^2/T_i);\ R(\delta \mid \mu) = \tilde{y}'\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\tilde{y}$$

$$R(\delta) = y'Z(Z'Z)^{-1}Z'y \quad \text{and} \quad R(\mu \mid \delta) = R(\delta \mid \mu) + R(\mu) - R(\delta)$$

Then Henderson's (1953) method III estimators are given by

$$\widehat{\sigma}_\nu^2 = \frac{y'y - R(\delta \mid \mu) - R(\mu)}{n - K - N + 1}$$

$$\widehat{\sigma}_\mu^2 = \frac{R(\mu \mid \delta) - (N - 1)\widehat{\sigma}_\nu^2}{n - \text{tr}(Z_\mu' Z(Z'Z)^{-1}Z'Z_\mu)} \quad (9.21)$$

9.3 Maximum Likelihood Estimators

Maximum likelihood (ML) estimates of the variance components and regression coefficients are obtained by maximizing the following log-likelihood function:

$$\log L = -(n/2)\log(2\pi) - (n/2)\log\sigma_v^2 - \tfrac{1}{2}\log|\Sigma| \\ -(y - Z\delta)'\Sigma^{-1}(y - Z\delta)/2\sigma_\nu^2 \qquad (9.22)$$

where ρ and Σ are given in (9.13). The first-order conditions give closed-form solutions for $\widehat{\delta}$ and $\widehat{\sigma}_\nu^2$ conditional on $\widehat{\rho}$:

$$\widehat{\delta} = (Z'\widehat{\Sigma}^{-1}Z)^{-1}Z'\widehat{\Sigma}^{-1}y \\ \widehat{\sigma}_\nu^2 = (y - Z\delta)'\widehat{\Sigma}^{-1}(y - Z\delta)/n \qquad (9.23)$$

However, the first-order condition based on ρ is nonlinear in ρ even for known values of δ and σ_ν^2.

$$0 = \frac{\partial \log L}{\partial \rho} = \frac{1}{2}\text{tr}(Z'\Sigma^{-1}Z) + \frac{1}{2\sigma_\nu^2}(y - Z\delta)'\Sigma^{-1}Z_\mu Z_\mu'\Sigma^{-1}(y - Z\delta) \qquad (9.24)$$

A numerical solution by means of iteration is necessary for $\widehat{\rho}$. The second derivative of $\log L$ with respect to ρ is given by

$$\frac{\partial^2 \log L}{\partial\rho\partial\rho} = \frac{1}{2}\text{tr}\{(Z_\mu'\Sigma^{-1}Z_\mu)(Z_\mu'\Sigma^{-1}Z_\mu)\} \\ -\frac{1}{\sigma_\nu^2}\{(y - Z\delta)'\Sigma^{-1}Z_\mu(Z_\mu'\Sigma^{-1}Z_\mu)Z_\mu'\Sigma^{-1}(y - Z\delta)\} \qquad (9.25)$$

Starting with an initial value of ρ_0, one obtains $\widehat{\Sigma}_0$ from (9.13) and $\widehat{\delta}_0$ and $\widehat{\sigma}_{\nu 0}^2$ from (9.23). The updated value ρ_1 is given from the following formula:

$$\rho_1 = \rho_0 - s\left[\frac{\partial^2 \log L}{\partial\rho\partial\rho}\right]_0^{-1}\left[\frac{\partial \log L}{\partial\rho}\right]_0 \qquad (9.26)$$

where the subscript $_0$ means evaluated at $\widehat{\Sigma}_0$, $\widehat{\delta}_0$, and $\widehat{\sigma}_{\nu 0}^2$, and s is a step size which is adjusted by step halving.[3] For the computational advantage of this algorithm as well as other algorithms like the Fisher scoring algorithm, see Jennrich and Sampson (1976) and Harville (1977). Maximum likelihood estimators are functions of sufficient statistics and are consistent and asymptotically efficient; see Harville (1977) for a review of the properties, advantages and disadvantages of ML estimators.

The ML approach has been criticized on grounds that it does not take into account the loss of degrees of freedom due to the regression coefficients in estimating the variance components. Patterson and Thompson (1971) remedy this by partitioning the likelihood function into two parts, one part depending only on the variance components and is free of the regression coefficients. Maximizing this part yields the restricted maximum likelihood estimator (REML). REML estimators of the variance components are asymptotically equivalent to the ML estimators, however, little is known about their finite sample properties, and they reduce to the ANOVA estimators under several balanced data cases.

9.3.1 Minimum Norm and Minimum Variance Quadratic Unbiased Estimators (MINQUE and MIVQUE)

Under normality of the disturbances, Rao's (1971a) MINQUE and MIVQUE procedures for estimating the variance components are identical. Since we assume normality, we will focus on MIVQUE. Basically, the MIVQUE of a linear combination of the variance components, $p_\mu\sigma_\mu^2 + p_\nu\sigma_\nu^2$, is obtained by finding a symmetric matrix G such that $\text{var}(y'Gy) = 2\ \text{tr}\{\sigma_\mu^2(GZ_\mu Z_\mu') + \sigma_\nu^2 G\}^2$ is minimized subject to the conditions that $y'Gy$ is an unbiased estimator of $(p_\mu\sigma_\mu^2 + p_\nu\sigma_\nu^2)$ and is invariant to any translation of the δ parameter. These yield the following constraints:

(1) GZ = 0;
(2) $\text{tr}(GZ_\mu Z_\mu') = p_\mu$ and $\text{tr}(G) = p_\nu$.

Rao (1971b) showed that the MIVQUE estimates of the variance components are given by

$$\begin{bmatrix} \hat{\sigma}_\mu^2 \\ \hat{\sigma}_\nu^2 \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

where $\gamma_{11} = \text{tr}(Z_\mu Z_\mu' R Z_\mu Z_\mu' R)$, $\gamma_{12} = \text{tr}(Z_\mu Z_\mu' RR)$, $\gamma_{22} = \text{tr}(RR)$, $\delta_1 = y'RZ_\mu Z_\mu' Ry$, $\delta_2 = y'RRy$, and $R = (\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1})/\sigma_\nu^2$. It is clear that MIVQUE requires a priori values of the variance components, and the resulting estimators possess the minimum variance property only if these a priori values coincide with the true values. Therefore, MIVQUE are only "locally best" and "locally minimum variance". If one iterates on the initial values of the variance components, the iterative estimators (IMIVQUE) become biased after the first iteration and MINQUE properties are not preserved. Two priors for the MINQUE estimator used in practice are (i) the identity matrix, denoted by (MQ0) and (ii) the ANOVA estimator of Swamy and Arora denoted by (MQA). Under normality, if one iterates until convergence, IMINQUE, IMIVQUE, and REML will be identical.[4]

9.3.2 Monte Carlo Results

Baltagi and Chang (1994) performed an extensive Monte Carlo study using a simple as well as a multiple regression with unbalanced one-way error component disturbances. The degree of unbalance in the sample as well as the variance component ratio ρ were varied across the experiments. The total number of observations as well as the total variance were fixed across the experiments to allow comparison of MSE for the various estimators considered. Some of the basic results of the Monte Carlo study suggest the following:

(1) As far as the estimation of regression coefficients is concerned, the simple ANOVA type feasible GLS estimators compare well with the more complicated estimators such as ML, REML, and MQA and are never more than 4%

above the MSE of true GLS. However, MQ0 is not recommended for large ρ and unbalanced designs.

(2) For the estimation of the remainder variance component σ_ν^2 the ANOVA, MIVQUE(A), ML, and REML estimators show little difference in relative MSE performance. However, for the individual specific variance component estimation, σ_μ^2, the ANOVA type estimators perform poorly relative to ML, REML, and MQA when the variance component ratio $\rho > 1$ and the pattern is severely unbalanced. MQ0 gives an extremely poor performance for severely unbalanced patterns and large ρ and is not recommended for these cases.

(3) Better estimates of the variance components, in the MSE sense, do not necessarily imply better estimates of the regression coefficients. This echoes similar findings for the balanced panel data case.

(4) Negative estimates of the variance components occurred when the true value of σ_μ^2 was zero or close to zero. In these cases, replacing these negative estimates by zero did not lead to much loss in efficiency.

(5) Extracting a balanced panel out of an unbalanced panel by either maximizing the number of households observed or the total number of observations in the balanced panel leads in both cases to an enormous loss in efficiency and is not recommended.[5]

9.4 Empirical Example: Hedonic Housing

Baltagi and Chang (1994) apply the various unbalanced variance components methods to the data set collected by Harrison and Rubinfeld (1978) for a study of hedonic housing prices and the willingness to pay for clean air. This data is available on the Springer website as Hedonic.xls. The total number of observations is 506 census tracts in the Boston area in 1970, and the number of variables is 14. Belsley, Kuh and Welsch (1980) identify 92 towns, consisting of 15 within Boston and 77 in its surrounding area. Thus, it is possible to group these data and analyze them as an unbalanced one-way model with random group effects. The group sizes range from one to 30 observations. The dependent variable is the median value (MV) of owner-occupied homes. The regressors include two structural variables, RM the average number of rooms, and AGE representing the proportion of owner units built prior to 1940. In addition, there are eight neighborhood variables: B, the proportion of blacks in the population; LSTAT, the proportion of population that is lower status; CRIM, the crime rate; ZN, the proportion of 25000 square feet residential lots; INDUS, the proportion of nonretail business acres; TAX, the full value property tax rate (\$/\$10000); PTRATIO, the pupil-teacher ratio; and CHAS represents the dummy variable for Charles River: = 1 if a tract bounds the Charles. There are also two accessibility variables, DIS the weighted distances to five employment centers in the Boston region, and RAD the index of accessibility to radial highways. One more regressor is an air pollution variable NOX, the annual average nitrogen oxide concentration in parts per hundred million.[6] Moulton (1987) performed the Breusch and

Table 9.1 One-way unbalanced variance components estimates for the Harrison–Rubinfeld hedonic housing equation. Dependent variable: MV

	OLS	Within	SA	WH	HFB	ML	REML	MQ0	MQA
Intercept	9.76	–	9.68	9.68	9.67	9.68	9.67	9.68	9.67
	(0.15)		(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
CRIM	−0.12	−0.63	−0.72	−0.74	−0.72	−0.72	−0.71	−0.73	−0.71
	(0.12)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)
ZN	0.08	–	0.04	0.07	0.02	0.03	0.01	0.06	0.01
	(0.51)		(0.69)	(0.68)	(0.70)	(0.69)	(0.71)	(0.69)	(0.71)
INDUS	0.02	–	0.21	0.16	0.24	0.22	0.24	0.18	0.24
	(0.24)		(0.43)	(0.43)	(0.45)	(0.44)	(0.46)	(0.43)	(0.45)
CHAS	0.91	−0.45	−0.01	−0.06	−0.13	−0.12	−0.14	−0.08	−0.14
	(0.33)	(0.30)	(0.29)	(0.30)	(0.29)	(0.29)	(0.29)	(0.30)	(0.29)
NOX	−0.64	−0.56	−0.59	−0.58	−0.59	−0.59	−0.59	−0.59	−0.59
	(0.11)	(0.14)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)	(0.12)
RM	0.63	0.93	0.92	0.91	0.92	0.92	0.92	0.91	0.92
	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
AGE	0.09	−1.41	−0.93	−0.87	−0.96	−0.94	−0.97	−0.90	−0.96
	(0.53)	(0.49)	(0.46)	(0.49)	(0.46)	(0.46)	(0.46)	(0.48)	(0.46)
DIS	−1.91	0.80	−1.33	−1.42	−1.27	−1.30	−1.25	−1.38	−1.26
	(0.33)	(0.71)	(0.46)	(0.46)	(0.46)	(0.47)	(0.47)	(0.46)	(0.47)
RAD	0.96	–	0.97	0.96	0.97	0.97	0.98	0.96	0.97
	(0.19)		(0.28)	(0.28)	(0.29)	(0.28)	(0.30)	(0.28)	(0.29)
TAX	−0.42	–	−0.37	−0.38	−0.37	−0.37	−0.37	−0.38	−0.37
	(0.12)		(0.19)	(0.19)	(0.19)	(0.19)	(0.20)	(0.19)	(0.20)
PTRATIO	−3.11	–	−2.97	−2.95	−2.99	−2.98	−2.99	−2.96	−2.99
	(0.50)		(0.98)	(0.96)	(1.01)	(0.98)	(1.02)	(0.97)	(1.02)
B	0.36	0.66	0.58	0.57	0.58	0.58	0.58	0.57	0.58
	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
LSTAT	−3.71	−2.45	−2.85	−2.90	−2.82	−2.84	−2.82	−2.88	−2.82
	(0.25)	(0.26)	(0.24)	(0.25)	(0.24)	(0.24)	(0.24)	(0.25)	(0.24)
$\hat{\sigma}^2_\nu$	–	–	0.017	0.020	0.017	0.017	0.017	0.019	0.017
$\hat{\sigma}^2_\mu$	–	–	0.017	0.016	0.019	0.018	0.020	0.017	0.020

*Approximate standard errors are given in parentheses. $n = 506$ observations for $N = 92$ towns.
Source Baltagi and Chang (1994), Reproduced by permission of Elsevier Science Publishers B.V. (North Holland)

Pagan (1980) Lagrange multiplier (LM) test on this data set and found compelling evidence against the exclusion of random group effects.[7]

Table 9.1 shows the OLS, Within, ANOVA, ML, REML, and MIVQUE type estimates using the entire data set of 506 observations for 92 towns. Unlike the drastic difference between OLS and the Within estimators which were analyzed in Moulton (1987), the various ANOVA, MLE, and MIVQUE type estimators, reported in Table 9.1, give similar estimates. Exceptions are ZN, INDUS, and CHAS estimates which vary across methods, but are all statistically insignificant. For the statistically significant variables, AGE varies from −0.87 for WH to −0.97 for REML, and DIS varies from −1.25 for REML to −1.42 for WH.[8] These results were verified using Stata and TSP. The higher the crime rate, air pollution, tax rate, pupil-teacher ratio, proportion of the population in lower status, the older the home and the greater

Table 9.2 Hedonic housing equation: Maximum Likelihood Estimator

```
Random-effects ML regression                    Number of obs      =       506
Group variable (i) : town                       Number of groups   =        92

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       5.5
                                                               max =        30

                                                LR chi2(13)        =    604.11
Log likelihood  =  236.26918                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        crim |  -.0071948   .0010277    -7.00   0.000     -.009209   -.0051806
          zn |   .0000286   .0006894     0.04   0.967    -.0013226    .0013799
       indus |   .0022167   .0043906     0.50   0.614    -.0063887    .0108221
        chas |  -.0119739    .028971    -0.41   0.679    -.0687561    .0448083
         nox |  -.0058672   .0012282    -4.78   0.000    -.0082744     -.00346
          rm |   .0092024   .0011643     7.90   0.000     .0069204    .0114843
         age |   -.000943   .0004614    -2.04   0.041    -.0018473   -.0000387
         dis |  -.1298569   .0469261    -2.77   0.006    -.2218304   -.0378834
         rad |   .0971024   .0284233     3.42   0.001     .0413938     .152811
         tax |  -.0003741   .0001895    -1.97   0.048    -.0007456   -2.59e-06
          pt |  -.0297989   .0097987    -3.04   0.002    -.0490041   -.0105938
           b |   .5778527   .0999609     5.78   0.000      .381933    .7737724
         lst |  -.2837924     .02405   -11.80   0.000    -.3309295   -.2366552
       _cons |   9.675679   .2069417    46.76   0.000     9.270081    10.08128
-------------+----------------------------------------------------------------
    /sigma_u |   .1337509   .0132895    10.06   0.000      .107704    .1597979
    /sigma_e |   .1304801   .0045557    28.64   0.000     .1215512    .1394091
-------------+----------------------------------------------------------------
         rho |   .5123767   .0546929                      .4060176    .6178576
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0: chibar2(01)=  172.71 Prob>=chibar2 = 0.000
```

the distance from employment centers in Boston, the lower is the median value of the house. Similarly, the more rooms a house has and the more accessible to radial highways the higher is the median value of that home. Table 9.2 produces the maximum likelihood estimates using Stata. The likelihood ratio for H_0; $\sigma_\mu^2 = 0$ is 172.7. This is asymptotically distributed as χ_1^2 and is significant. The Breusch Pagan LM test for H_0 is 240.8. This is asymptotically distributed as χ_1^2 and is also significant. The Hausman specification test based on the contrast between the fixed effects and random effects estimators in Stata yields a χ_8^2 statistic of 66.1 which is statistically significant. Table 9.3 reproduces the Swamy and Arora estimator using Stata.

In conclusion, for the regression coefficients, both the Monte Carlo and the empirical illustration indicate that the computationally simple ANOVA estimates compare favorably with the computationally demanding ML, REML, and MQA type estimators. For the variance components, the ANOVA methods are recommended except when ρ is large and the unbalancedness of the data is severe. For these cases, ML, REML, or MQA are recommended. As a check for misspecification, one should perform at least one of the ANOVA methods and one of the ML methods to see if the estimates differ widely. This is the Maddala and Mount (1973) suggestion for the balanced data case and applies as well for the unbalanced data case.

Table 9.3 Hedonic housing equation: Swamy and Arora estimator

```
Random-effects GLS regression                   Number of obs      =       506
Group variable (i) : town                       Number of groups   =        92

R-sq:  within  = 0.6682                         Obs per group: min =         1
       between = 0.8088                                        avg =       5.5
       overall = 0.7875                                        max =        30

Random effects u_i ~ Gaussian                   Wald chi2(13)      =   1169.62
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------- theta --------------------
  min      5%       median        95%      max
0.2915   0.2915     0.5514      0.7697   0.8197

------------------------------------------------------------------------------
          mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        crim |  -.0072338   .0010346    -6.99   0.000    -.0092616   -.0052061
          zn |   .0000396   .0006878     0.06   0.954    -.0013084    .0013876
       indus |   .0020794   .0043403     0.48   0.632    -.0064273    .0105861
        chas |  -.0105913   .0289598    -0.37   0.715    -.0673515     .046169
         nox |   -.005863   .0012455    -4.71   0.000    -.0083041   -.0034219
          rm |   .0091774   .0011792     7.78   0.000     .0068662    .0114885
         age |  -.0009272   .0004647    -2.00   0.046    -.0018379   -.0000164
         dis |  -.1328825   .0456826    -2.91   0.004    -.2224186   -.0433463
         rad |   .0968634   .0283495     3.42   0.001     .0412994    .1524274
         tax |  -.0003747    .000189    -1.98   0.047    -.0007452   -4.25e-06
          pt |   -.029723   .0097538    -3.05   0.002    -.0488402   -.0106059
           b |   .5750649    .101031     5.69   0.000     .3770479     .773082
         lst |  -.2851401   .0238546   -11.95   0.000    -.3318942   -.2383859
       _cons |   9.677802   .2071417    46.72   0.000     9.271811    10.08379
-------------+----------------------------------------------------------------
     sigma_u |  .12973801
     sigma_e |  .13024876
         rho |  .49803548   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In another application studying the damage associated with proximity to a hazardous waste site, Mendelsohn et al. (1992) use panel data on repeated single-family home sales in the harbor area surrounding New Bedford, Massachusetts over the period 1969-88. Note that one observes the dependent variable, in this case the value of the house, only when an actual sale occurs. Therefore, these data are “unbalanced” with different time-period intervals between sales, and different numbers of repeated sales for each single-family house over the period observed. These comprised 780 properties and 1916 sales. Mendelsohn et al. (1992) used first-differenced and fixed effects estimation methods to control for specific individual housing characteristics. Information on the latter variables are rarely available or complete. They find a significant reduction in housing values, between 7000 and 10000 (1989 dollars), as a result of these houses’ proximity to hazardous waste sites.

Baltagi, Egger and Pfaffermayr (2003) consider an unbalanced panel of bilateral export flows from the EU15 countries, USA, and Japan to their 57 most important trading partners over the period 1986–1998. They estimate a three-way gravity equation with importer, exporter, and time fixed effects as well as pairwise interaction effects. These effects include time-invariant factors like distance, common borders, island nation, land locked, common language, colonies, etc. These fixed effects as well as the interaction terms are found to be statistically significant. Omission of these effects can result in biased and misleading inference.

9.5 The Unbalanced Two-Way Error Component Model

Wansbeek and Kapteyn (1989), henceforth WK, consider the regression model with unbalanced two-way error component disturbances

$$\begin{aligned} y_{it} &= X'_{it}\beta + u_{it} \quad i = 1, \ldots, N_t \quad t = 1, \ldots, T \\ u_{it} &= \mu_i + \lambda_t + \nu_{it} \end{aligned} \tag{9.27}$$

where $N_t (N_t \leqslant N)$ denotes the number of individuals observed in year t, with $n = \sum_t N_t$. Let D_t be the $(N_t \times N)$ matrix obtained from I_N by omitting the rows corresponding to individuals not observed in year t. Define

$$\Delta = (\Delta_1, \Delta_2) \equiv \begin{bmatrix} D_1 & D_1 \iota_N & \\ \vdots & & \ddots \\ D_T & & & D_T \iota_N \end{bmatrix} \tag{9.28}$$

where $\Delta_1 = (D'_1, \ldots, D'_T)'$ is $n \times N$ and $\Delta_2 = \text{diag}[D_t \iota_N] = \text{diag}[\iota_{N_t}]$ is $n \times T$. The matrix Δ gives the dummy-variable structure for the incomplete data model. Note that WK order the data on the N individuals in T consecutive sets, so that t runs slowly and i runs fast. This is exactly the opposite ordering that has been used so far in the text. For complete panels, $\Delta_1 = (\iota_T \otimes I_N)$ and $\Delta_2 = I_T \otimes \iota_N$.

9.5.1 The Fixed Effects Model

If the μ_i and λ_t are fixed, one has to run the regression given in (9.27) with the matrix of dummies given in (9.28). Most likely, this will be infeasible for large panels with many households or individuals, and we need the familiar Within transformation. This was easy for the balanced case and extended easily to the unbalanced one-way case. However, for the unbalanced two-way case, WK showed that this transformation is a little complicated but nevertheless manageable. To see this, we need some more matrix results.

Note that $\Delta_N \equiv \Delta'_1\Delta_1 = \text{diag}[T_i]$ where T_i is the number of years individual i is observed in the panel. Also, $\Delta_T \equiv \Delta'_2\Delta_2 = \text{diag}[N_t]$ and $\Delta_{TN} \equiv \Delta'_2\Delta_1$ is the $(T \times N)$ matrix of zeros and ones indicating the absence or presence of a household in a certain year. For complete panels, $\Delta_N = TI_N$, $\Delta_T = NI_T$ and $\Delta_{TN} = \iota_T \iota_N' = J_{TN}$. Define $P_{[\Delta]} = \Delta(\Delta'\Delta)^- \Delta'$, then the Within transformation is $Q_{[\Delta]} = I_n - P_{[\Delta]}$. For the two-way unbalanced model with $\Delta = (\Delta_1, \Delta_2)$ given by (9.28), WK show that

$$P_{[\Delta]} = P_{\Delta_1} + P_{[Q_{[\Delta_1]}\Delta_2]} \tag{9.29}$$

The proof is sketched out in problem 9.6. Therefore,

$$Q_{[\Delta]} = Q_{[\Delta_1]} - Q_{[\Delta_1]}\Delta_2(\Delta_2' Q_{[\Delta_1]}\Delta_2)^- \Delta_2' Q_{[\Delta_1]} \tag{9.30}$$

Davis (2002) generalizes the WK Within transformation to the three-way, four-way, and higher order error component models. Davis shows that the Within transformation can be applied in stages to the variables in the regression, just like in (9.30). This reduces the computational burden considerably. For example, consider a three-way error component model, representing products sold at certain locations and observed over some time period. These fixed effects are captured by three dummy variables matrices $\Delta = [\Delta_1, \Delta_2, \Delta_3]$. In order to get the Within transformation, Davis (2002) applies (9.29) twice and obtains $Q_{[\Delta]} = Q_{[A]} - P_{[B]} - P_{[C]}$ where $A = \Delta_1$, $B = Q_{[A]}\Delta_2$, and $C = Q_{[B]}Q_{[A]}\Delta_3$; see problem 9.7. This idea generalizes readily to higher order fixed effects error components models.

9.5.2 The Random Effects Model

In vector form, the incomplete two-way random effects model can be written as

$$u = \Delta_1 \mu + \Delta_2 \lambda + \nu \tag{9.31}$$

where $\mu' = (\mu_1, \ldots, \mu_N)$, $\lambda' = (\lambda_1, \ldots, \lambda_T)$ and ν are random variables described exactly as in the two-way error component model considered in Chap. 3. μ, λ, and ν are independent of each other and among themselves with zero means and variances σ_μ^2, σ_λ^2 and σ_ν^2, respectively. In this case,

$$\begin{aligned}\Omega = E(uu') &= \sigma_\nu^2 I_n + \sigma_\mu^2 \Delta_1 \Delta_1' + \sigma_\lambda^2 \Delta_2 \Delta_2' \\ &= \sigma_\nu^2 (I_n + \phi_1 \Delta_1 \Delta_1' + \phi_2 \Delta_2 \Delta_2') = \sigma_\nu^2 \Sigma\end{aligned} \tag{9.32}$$

with $\phi_1 = \sigma_\mu^2/\sigma_\nu^2$ and $\phi_2 = \sigma_\lambda^2/\sigma_\nu^2$. Using the general expression for the inverse of $(I + XX')$, see problem 9.8, Wansbeek and Kapteyn (1989) obtain the inverse of Σ as

$$\Sigma^{-1} = V - V\Delta_2 \widetilde{P}^{-1} \Delta_2' V \tag{9.33}$$

where

$$\begin{aligned} V &= I_n - \Delta_1 \widetilde{\Delta}_N^{-1} \Delta_1' && (n \times n) \\ \widetilde{P} &= \widetilde{\Delta}_T - \Delta_{TN} \widetilde{\Delta}_N^{-1} \Delta_{TN}' && (T \times T) \\ \widetilde{\Delta}_N &= \Delta_N + (\sigma_\nu^2/\sigma_\mu^2) I_N && (N \times N) \\ \widetilde{\Delta}_T &= \Delta_T + (\sigma_\nu^2/\sigma_\lambda^2) I_T && (T \times T) \end{aligned}$$

Note that we can no longer obtain the simple Fuller and Battese (1973) transformation for the unbalanced two-way model. The expression for Σ^{-1} is messy and asymmetric in individuals and time, but it reduces computational time considerably relative to inverting Σ numerically. Davis (2002) shows that the WK results can be generalized to an arbitrary number of random error components. In fact, for a three-way random error component, like the one considered in problem 9.8, the added random error component η adds an extra $\sigma_\eta^2 \Delta_3 \Delta_3'$ term to the variance–covariance given in (9.32). Therefore, Σ remains of the $(I + XX')$ form and its inverse can be obtained by repeated application of this inversion formula. This idea generalizes readily to higher order unbalanced random error component models. WK suggest an ANOVA type quadratic unbiased estimator (QUE) of the variance components based on the Within residuals. In fact, the MSE of the Within regression is unbiased for σ_ν^2 even under the random effects specification. Let $e = y - X\widetilde{\beta}$ where $\widetilde{\beta}$ denote the Within estimates and define

$$q_W = e'Q_{[\Delta]}e \tag{9.34}$$

$$q_N = e'\Delta_2\Delta_T^{-1}\Delta_2'e = e'P_{[\Delta_2]}e \tag{9.35}$$

$$q_T = e'\Delta_1\Delta_N^{-1}\Delta_1'e = e'P_{[\Delta_1]}e \tag{9.36}$$

By equating q_W, q_N, and q_T to their expected values and solving these three equations one gets QUE of σ_ν^2, σ_μ^2 and σ_λ^2. WK also derive the ML iterative first-order conditions as well as the information matrix under normality of the disturbances. These will not be reproduced here, and the reader is referred to the WK article for details. A limited Monte Carlo experiment was performed using 50 replications and three kinds of data designs: complete panel data, 20% random attrition, and a rotating panel. This was done using a simple regression with a Nerlove type X for fixed $\sigma_\mu^2 = 400$, $\sigma_\lambda^2 = 25$ and $\sigma_\nu^2 = 25$. The regression coefficients were fixed at $\alpha = 25$ and $\beta = 2$, and the number of individuals and time periods were $N = 100$ and $T = 5$, respectively. The results imply that the QUE of the variance components are in all cases at least as close to the true value as the MLE so that iteration on these values does not seem to pay off. Also, GLS gives nearly identical results to MLE as far as the regression coefficient estimates are concerned. Therefore, WK recommend GLS over MLE in view of the large difference in computational cost.

Baltagi, Song and Jung (2002a) reconsider the unbalanced two-way error component given in (9.27) and (9.28) and provide alternative analysis of variance (ANOVA), minimum norm quadratic unbiased (MINQUE), and restricted maximum likelihood (REML) estimation procedures. These are similar to the methods studied in Sect. 9.2 for the unbalanced one-way error component model. The mean squared error performance of these estimators are compared using Monte Carlo experiments. Focusing on the estimates of the variance components, the computationally more demanding MLE, REML, MIVQUE estimators are recommended especially if the unbalanced pattern is severe. However, focusing on the regression coefficients estimates, the simple ANOVA methods perform just as well as the computationally demanding MLE, REML, and MIVQUE methods and are recommended.

9.6 Testing for Individual and Time Effects Using Unbalanced Panel Data

In Chap. 4, we derived the Breusch and Pagan (1980) LM test for H_0; $\sigma_\mu^2 = \sigma_\lambda^2 = 0$ in a complete panel data model with two-way error component disturbances. Baltagi and Li (1990b) derived the corresponding LM test for the unbalanced two-way error component model. This model is given by (9.27), and the variance–covariance matrix of the disturbances is given by (9.32). Following the same derivations as given in Sect. 4.2 (see problem 9.8), one can show that under normality of the disturbances

$$\partial\Omega/\partial\sigma_\mu^2 = \Delta_1\Delta_1'; \partial\Omega/\partial\sigma_\lambda^2 = \Delta_2\Delta_2' \quad \text{and} \quad \partial\Omega/\partial\sigma_\nu^2 = I_n \tag{9.37}$$

with

$$\text{tr}(\Delta_2\Delta_2') = \text{tr}(\Delta_2'\Delta_2) = \text{tr}(\text{diag}[N_t]) = \sum_{t=1}^{T} N_t = n \tag{9.38}$$

and

$$\text{tr}(\Delta_1'\Delta_1) = \text{tr}(\text{diag}[T_i]) = \sum_{i=1}^{N} T_i = n \tag{9.39}$$

Substituting these results in (4.17) and noting that under H_0, $\widetilde{\Omega}^{-1} = (1/\widetilde{\sigma}_\nu^2)I_n$, where $\widehat{\sigma}_\nu^2 = \widetilde{u}'\widetilde{u}/NT$ and $\widetilde{u}$ denote the OLS residuals, one gets

$$\widetilde{D} = (\partial L/\partial\theta)\,|_{\theta=\widetilde{\theta}} = (n/2\widetilde{\sigma}_\nu^2)\begin{bmatrix} A_1 \\ A_2 \\ 0 \end{bmatrix} \tag{9.40}$$

where $\theta' = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$ and $\widetilde{\theta}$ denotes the restricted MLE of θ under H_0. Also, $A_r = [(\widetilde{u}'\Delta_r\Delta_r'\widetilde{u}/\widetilde{u}'\widetilde{u}) - 1]$ for $r = 1, 2$. Similarly, one can use (4.19) to obtain the information matrix

$$\widetilde{J} = (n/2\widetilde{\sigma}_\nu^4)\begin{bmatrix} M_{11}/n & 1 & 1 \\ 1 & M_{22}/n & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{9.41}$$

where $M_{11} = \sum_{i=1}^{N} T_i^2$ and $M_{22} = \sum_{t=1}^{T} N_t^2$. This makes use of the fact that

$$\text{tr}(\Delta_2\Delta_2')^2 = \sum_{t=1}^{T} N_t^2; \ \text{tr}(\Delta_1\Delta_1')^2 = \sum_{i=1}^{N} T_i^2 \tag{9.42}$$

and

$$\text{tr}[(\Delta_1\Delta_1')(\Delta_2\Delta_2')] = \sum_{t=1}^{T} \text{tr}[(D_tD_t')J_{N_t}] = \sum_{t=1}^{T} \text{tr}(J_{N_t}) = \sum_{t=1}^{T} N_t = n$$

Using (9.40) and (9.41) one gets the LM statistic

$$LM = \widetilde{D}'\widetilde{J}^{-1}\widetilde{D} = (\tfrac{1}{2})n^2[A_1^2/(M_{11} - n) + A_2^2/(M_{22} - n)] \tag{9.43}$$

which is asymptotically distributed as χ^2_2 under the null hypothesis. For computational purposes, one need not form the Δ_r matrices to compute $A_r (r = 1, 2)$. In fact,

$$\widetilde{u}\Delta_1\Delta_1'\widetilde{u} = \sum_{i=1}^{N} \widetilde{u}_{i.}^2 \quad \text{where} \quad \widetilde{u}_{i.} = \sum_{t=1}^{T_i} \widetilde{u}_{it} \tag{9.44}$$

and

$$\widetilde{u}'\Delta_2\Delta_2'\widetilde{u} = \sum_{t=1}^{T} \widetilde{u}_{.t}^2 \quad \text{where} \quad \widetilde{u}_{.t} = \sum_{i=1}^{N_t} \widetilde{u}_{it} \tag{9.45}$$

Equation (9.45) is obvious, since $\Delta_2 = \text{diag}[\iota_{N_t}]$, and (9.44) can be similarly obtained, by restacking the residuals such that the faster index is t. The LM statistic given in (9.43) is easily computed using least squares residuals, and retains a similar form to that of the complete panel data case. In fact, when $N_t = N$, (9.43) reverts back to the LM statistic derived in Breusch and Pagan (1980). Also, (9.43) retains the additive property exhibited in the complete panel data case, i.e., if H_0; $\sigma^2_\mu = 0$, the LM test reduces to the first term of (9.43), whereas if H_0; $\sigma^2_\lambda = 0$, the LM test reduces to the second term of (9.43). Both test statistics are asymptotically distributed as χ^2_1 under the respective null hypotheses.

These variance components cannot be negative and therefore H_0; $\sigma^2_\mu = 0$ has to be against a one-sided alternative H_1; $\sigma^2_\mu > 0$. Moulton and Randolph (1989) derived the one-sided LM_1 statistic

$$LM_1 = n[2(M_{11} - n)]^{-1/2} A_1 \tag{9.46}$$

which is the square root of the first term in (9.43). Under weak conditions as $n \to \infty$ and $N \to \infty$ the LM_1 statistic has an asymptotic standard normal distribution under H_0. However, Moulton and Randolph (1989) showed that this asymptotic $N(0, 1)$ approximation can be poor even in large samples. This occurs when the number of regressors is large or the intra-class correlation of some of the regressors is high. They suggest an alternative standardized Lagrange multiplier SLM given by

$$SLM = \frac{LM_1 - E(LM_1)}{\sqrt{\text{var}(LM_1)}} = \frac{d - E(d)}{\sqrt{\text{var}(d)}} \tag{9.47}$$

where $d = (\tilde{u}'D\tilde{u})/\tilde{u}'\tilde{u}$ and $D = \Delta_1\Delta_1'$. Using the results on moments of quadratic forms in regression residuals (see, for example, Evans and King 1985), we get

$$E(d) = \text{tr}(D\bar{P}_Z)/p$$

where $p = [n - (K + 1)]$ and

$$\text{var}(d) = 2\{p\ \text{tr}(D\bar{P}_Z)^2 - [\text{tr}(D\bar{P}_Z)]^2\}/p^2(p + 2)$$

Under H_0, this SLM has the same asymptotic $N(0, 1)$ distribution as the LM_1 statistic. However, the asymptotic critical values for the SLM are generally closer to the exact critical values than those of the LM_1 statistic. Similarly, for H_0; $\sigma^2_\lambda = 0$, the one-sided LM test statistic is the square root of the second term in (9.43), i.e.,

$$LM_2 = n[2(M_{22} - n)]^{-1/2} A_2 \tag{9.48}$$

Honda's (1985) "handy" one-sided test for the two-way model with unbalanced data is simply

$$HO = (LM_1 + LM_2)/\sqrt{2}$$

It is also easy to show, see Baltagi, Chang and Li (1998), that the locally mean most powerful (LMMP) one-sided test suggested by King and Wu (1997) for the unbalanced two-way error component model is given by

$$KW = \frac{\sqrt{M_{11} - n}}{\sqrt{M_{11} + M_{22} - 2n}} LM_1 + \frac{\sqrt{M_{22} - n}}{\sqrt{M_{11} + M_{22} - 2n}} LM_2 \tag{9.49}$$

where LM_1 and LM_2 are given by (9.46) and (9.48), respectively. Both HO and KW are asymptotically distributed as $N(0, 1)$ under H_0. These test statistics can be standardized and the resulting SLM given by $\{d - E(d)\}/\sqrt{\text{var}(d)}$ where $d = \tilde{u}' D\tilde{u}/\tilde{u}'\tilde{u}$ with

$$D = \frac{1}{2}\frac{n}{\sqrt{M_{11} - n}}(\Delta_1\Delta_1') + \frac{1}{2}\frac{n}{\sqrt{M_{22} - n}}(\Delta_2\Delta_2') \tag{9.50}$$

for Honda's (1985) version, and

$$D = \frac{n}{\sqrt{2}\sqrt{M_{11} + M_{22} - 2n}}[(\Delta_1\Delta_1') + (\Delta_2\Delta_2')] \tag{9.51}$$

for the King and Wu (1997) version of this test. $E(d)$ and $\text{var}(d)$ are obtained from the same formulas shown (9.47) using the appropriate D matrices.

Since LM_1 and LM_2 can be negative for a specific application, especially when one or both variance components are small and close to zero, one can use the Gourieroux, Holly and Monfort (1982) (GHM) test which is given by

$$\chi_m^2 = \begin{cases} LM_1^2 + LM_2^2 & \text{if } LM_1 > 0, LM_2 > 0 \\ LM_1^2 & \text{if } LM_1 > 0, LM_2 \leqslant 0 \\ LM_2^2 & \text{if } LM_1 \leqslant 0, LM_2 > 0 \\ 0 & \text{if } LM_1 \leqslant 0, LM_2 \leqslant 0 \end{cases} \tag{9.52}$$

χ_m^2 denotes the mixed χ^2 distribution. Under the null hypothesis,

$$\chi_m^2 \sim \left(\tfrac{1}{4}\right)\chi^2(0) + \left(\tfrac{1}{2}\right)\chi^2(1) + \left(\tfrac{1}{4}\right)\chi^2(2)$$

where $\chi^2(0)$ equals zero with probability one.[10] The weights $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ follow from the fact that LM_1 and LM_2 are asymptotically independent of each other and the results in Gourieroux, Holly and Monfort (1982). This proposed test has the advantage over the Honda and KW tests in that it is immune to the possible negative values of LM_1 and LM_2.

Baltagi, Chang and Li (1998) compare the performance of these tests using Monte Carlo experiments for an unbalanced two-way error component model. The results of the Monte Carlo experiments show that the nominal size of the Honda and King-Wu tests based on asymptotic critical values are inaccurate for all unbalanced patterns considered. However, the nominal size of the standardized version of these tests is closer to the true significance value and is recommended. This confirms similar results for the unbalanced one-way error component model by Moulton and Randolph

(1989). In cases where at least one of the variance components is close to zero, the Gourieroux, Holly and Monfort (1982) test is found to perform well in Monte Carlo experiments and is recommended. All the tests considered have larger power as the number of individuals N in the panel and/or the variance components increase. In fact, for typical labor or consumer panels with large N, the Monte Carlo results show that the power of these tests is one except for cases where the variance components comprise less than 10% of the total variance.

9.7 The Unbalanced Nested Error Component Model

Baltagi, Song and Jung (2001) extend the ANOVA, MINQUE, and MLE estimation procedures described in Sect. 9.2 to the unbalanced nested error component regression model. For this model, the incomplete panel data exhibits a natural nested grouping. For example, data on firms may be grouped by industry, data on states by region, data on individuals by profession and data on students by schools.[11] The unbalanced panel data regression model is given by

$$y_{ijt} = x'_{ijt}\beta + u_{ijt}, \ i = 1, \ldots, M, \ j = 1, \ldots, N_i \quad \text{and} \quad t = 1, \ldots, T_i, \tag{9.53}$$

where y_{ijt} could denote the output of the jth firm in the ith industry for the tth time period. x_{ijt} denotes a vector of k nonstochastic inputs. The disturbances are given by

$$u_{ijt} = \mu_i + \nu_{ij} + \varepsilon_{ijt}, \ i = 1, \ldots, M, \ j = 1, \ldots, N_i \quad \text{and} \quad t = 1, \ldots, T_i, \tag{9.54}$$

where μ_i denotes the ith unobservable industry specific effect which is assumed to be IID$(0, \sigma^2_\mu)$, ν_{ij} denotes the nested effect of the jth firm within the ith industry which is assumed to be IID$(0, \sigma^2_\nu)$, and ε_{ijt} denotes the remainder disturbance which is also assumed to be IID$(0, \sigma^2_\varepsilon)$. The μ_i's, ν_{ij}'s, and ε_{ijt}'s are independent of each other and among themselves. This is a nested classification in that each successive component of the error term is imbedded or "nested" within the preceding component; see Graybill (1961, p. 350). This model allows for unequal number of firms in each industry as well as different number of observed time periods across industries. Detailed derivation of the variance–covariance matrix Ω, the Fuller and Battese (1973) transformation, as well as ANOVA, MINQUE, and MLE methods are given in Baltagi, Song and Jung (2001) and will not be reproduced here. Baltagi, Song and Jung (2001) compared the performance of these estimators using Monte Carlo experiments. While the MLE and MIVQUE methods perform the best in estimating the variance components and the standard errors of the regression coefficients, the simple ANOVA methods perform just as well in estimating the regression coefficients. These estimation methods are also used to investigate the productivity of public capital in private production. In a companion paper, Baltagi, Song and Jung (2002b) extend the Lagrange Multiplier tests described in Sect. 9.5 to the unbalanced nested error component model. Later, Baltagi, Song and Jung (2002c) derived the the Lagrange Multiplier tests for the unbalalnced nested error component model with serially correlated disturbances.

9.7.1 Empirical Example: Nested States Public Capital Productivity

In Chap. 2, example 3, we estimated a Cobb–Douglas production function investigating the productivity of public capital in each state's private output. This was based on a panel of 48 states over the period 1970-86. The data was provided by Munnell (1990). Here, we group these states into nine geographical regions with the Middle Atlantic region, for example, containing three states: New York, New Jersey and Pennsylvania and the Mountain region containing eight states: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, and Nevada. In this case, the primary group would be the regions, the nested group would be the states, and these are observed over 17 years. The dependent variable y is the gross state product and the regressors include the private capital stock (K) computed by apportioning the Bureau of Economic Analysis (BEA) national estimates. The public capital stock is measured by its components: highways and streets (KH), water and sewer facilities (KW), and other public buildings and structures (KO), all based on the BEA national series. Labor (L) is measured by the employment in nonagricultural payrolls. The state unemployment rate is included to capture the business cycle in a given state. All variables except the unemployment rate are expressed in natural logarithm

$$y_{ijt} = \alpha + \beta_1 K_{ijt} + \beta_2 KH_{ijt} + \beta_3 KW_{ijt} + \beta_4 KO_{ijt} + \beta_5 L_{ijt} + \beta_6 \, Unemp_{ijt} + u_{it} \tag{9.55}$$

where $i = 1, 2, \ldots, 9$ regions, $j = 1, \ldots, N_i$ with N_i equaling three for the Middle Atlantic region and eight for the Mountain region and $t = 1, 2, \ldots, 17$. The data is unbalanced only in the differing number of states in each region. The disturbances follow the nested error component specification given by (9.54).

Table 9.4 gives the OLS, Within, ANOVA, MLE, REML, and MIVQUE type estimates using this unbalanced nested error component model. The OLS estimates show that the highways and streets and water and sewer components of public capital have a positive and significant effect upon private output, whereas that of other public buildings and structures is not significant. Because OLS ignores the state and region effects, the corresponding standard errors and t-statistics are biased; see Moulton (1986). The Within estimator shows that the effect of KH and KW is insignificant, whereas that of KO is negative and significant. The primary region and nested state effects are significant using several LM tests developed in Baltagi, Song and Jung (2002b). This justifies the application of the feasible GLS, MLE, and MIVQUE methods. For the variance components estimates, there are no differences in the estimate of σ^2_ε. But estimates of σ^2_μ and σ^2_ν vary. $\widehat{\sigma}^2_\mu$ is as low as 0.0015 for SA and MLE and as high as 0.0029 for HFB. Similarly, $\widehat{\sigma}^2_\nu$ is as low as 0.0043 for SA and as high as 0.0069 for WK. This variation had little effect on estimates of the regression coefficients or their standard errors. For all estimators of the random effects model, the highways and streets and water and sewer components of public capital had a positive and significant effect, while the other public buildings and structures had a negative and significant effect upon private output. These results were verified using TSP and Stata. In fact Table 9.5 replicates the MLE column of Table 9.4, while Table 9.6 replicates the REML column of Table 9.4, using the *xtmixed* command in Stata.

Table 9.4 Cobb–Douglas production function estimates with unbalanced nested error components 1970–1986, nine regions, 48 States

Variable	OLS	Within	WH	WK	SA	HFB	MLE	REML	MV1	MV2	MV3
Intercept	1.926	–	2.082	2.131	2.089	2.084	2.129	2.127	2.083	2.114	2.127
	(0.053)		(0.152)	(0.160)	(0.144)	(0.150)	(0.154)	(0.157)	(0.152)	(0.154)	(0.156)
K	0.312	0.235	0.273	0.264	0.274	0.272	0.267	0.266	0.272	0.269	0.267
	(0.011)	(0.026)	(0.021)	(0.022)	(0.020)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.021)
L	0.550	0.801	0.742	0.758	0.740	0.743	0.754	0.756	0.742	0.750	0.755
	(0.016)	(0.030)	(0.026)	(0.027)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
KH	0.059	0.077	0.075	0.072	0.073	0.075	0.071	0.072	0.075	0.072	0.072
	(0.015)	(0.031)	(0.023)	(0.024)	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
KW	0.119	0.079	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
	(0.012)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
KO	0.009	−0.115	−0.095	−0.102	−0.094	−0.096	−0.100	−0.101	−0.095	−0.098	−0.100
	(0.012)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Unemp	−0.007	−0.005	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
σ_ε^2	0.0073	0.0013	0.0014	0.0014	0.0014	0.0014	0.0013	0.0014	0.0014	0.0014	0.0014
σ_μ^2	–	–	0.0027	0.0022	0.0015	0.0029	0.0015	0.0019	0.0027	0.0017	0.0017
σ_ν^2	–	–	0.0045	0.0069	0.0043	0.0044	0.0063	0.0064	0.0046	0.0056	0.0063

The dependent variable is log of gross state product. Standard errors are given in parentheses.
Source Baltagi, Song and Jung (2001), reproduced by permission of Elsevier Science Publishers B.V. (North-Holland)

Other empirical applications of the nested error component model include Montmarquette and Mahseredjian (1989) who study whether schooling matters in educational achievements in Montreal's Francophone public elementary schools. Also, Antweiler (2001) who derives the maximum likelihood estimator for an unbalanced nested three-way error component model. This is applied to the problem of explaining the determinants of pollution concentration (measured by the log of atmospheric sulfuric dioxide) at 293 observation stations located in 44 countries over the time period 1971–96. This data is highly unbalanced in that out of a total of 2621 observations, about a third of these are from stations in one country, the United States. Also, the time period of observation is not necessarily continuous. Comparing the results of maximum likelihood for a nested versus a simple (non-nested) unbalanced error component model, Antweiler (2001) finds that the scale elasticity coefficient estimate for the nested model is less than half that for the non-nested model. Scale elasticity is the coefficient of log of economic intensity as measured by GDP per square kilometer. This is also true for the estimate of the income effect which is negative and much lower in absolute value for the nested model than the non-nested model. Finally, the estimate of the composition effect which is the coefficient of the log of the country's capital abundance is higher for the nested model than for the non-nested model.

Davis (2002) applies OLS, Within, MIVQUE, and MLE procedures to a three-way unbalanced error component model using data on film revenues for six movie theaters

Table 9.5 Nested Public Capital Equation: MLE Estimator

```
. xtmixed y k l kh kw ko unemp || region: || state:, mle

Mixed-effects ML regression                     Number of obs      =       816

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         region |        9         51       90.7        136
          state |       48         17       17.0         17
-----------------------------------------------------------

                                                Wald chi2(6)       =  18829.06
Log likelihood =  1430.5016                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           k |   .2671485   .0212591    12.57   0.000     .2254814    .3088155
           l |   .7540721   .0261868    28.80   0.000     .7027468    .8053973
          kh |   .0709766    .023041     3.08   0.002     .0258171    .1161362
          kw |   .0761188   .0139248     5.47   0.000     .0488266     .103411
          ko |  -.0999956   .0169366    -5.90   0.000    -.1331908   -.0668005
       unemp |  -.0058983   .0009031    -6.53   0.000    -.0076684   -.0041282
       _cons |   2.128824   .1543854    13.79   0.000     1.826234    2.431414
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
region: Identity             |
                   sd(_cons) |    .038087   .0170591      .0158316     .091628
-----------------------------+------------------------------------------------
state: Identity              |
                   sd(_cons) |   .0792193   .0093861      .0628027    .0999273
-----------------------------+------------------------------------------------
                sd(Residual) |   .0366893    .000939      .0348944    .0385766
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  1154.73   Prob > chi2 = 0.0000
```

near New Haven, CT, observed over a six week period in 1998. Some of the reasons for unbalancedness in the data occurs because (i) not all films are shown at all locations, (ii) films start and stop being shown at theaters during the observation period, and (iii) data on revenues are missing due to nonresponse. The estimates obtained reveal a complex set of asymmetric cross-theater price elasticities of demand. These estimates are useful for the analysis of the impact of mergers on pricing, and for determining the appropriate extent of geographic market definition in these markets.

In Chap. 7, we showed that the Within estimator is equivalent to OLS on the augmented regression which adds the covariate averages $\overline{X}_{i.}$ to the original model; see Mundlak (1978) and problem 7.8. Wooldridge (2010) shows that this Mundlak result still holds for unbalanced panels. In Chap. 4, we also showed that the Within estimator can be obtained using the Chamberlain (1982) approach of projecting the individual effects upon covariates from all time periods. Abrevaya (2013) shows that the Chamberlain (1982) approach is not immediately applicable for unbalanced

Table 9.6 Nested Public Capital Equation: REML Estimator

```
. xtmixed y k l kh kw ko unemp || region: || state:, reml

Mixed-effects REML regression                   Number of obs      =       816

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         region |        9         51       90.7        136
          state |       48         17       17.0         17
-----------------------------------------------------------

                                                Wald chi2(6)       =  18382.38
Log restricted-likelihood =  1404.7101          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           k |   .2660309   .0215471    12.35   0.000     .2237993    .3082624
           l |    .755506   .0264556    28.56   0.000     .7036539     .807358
          kh |   .0718856   .0233478     3.08   0.002     .0261248    .1176464
          kw |   .0761553   .0139952     5.44   0.000     .0487251    .1035854
          ko |  -.1005397   .0170173    -5.91   0.000    -.1338931   -.0671864
       unemp |  -.0058815   .0009093    -6.47   0.000    -.0076636   -.0040994
       _cons |   2.126996   .1574865    13.51   0.000     1.818328    2.435664
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
region: Identity             |
                   sd(_cons) |   .0435474   .0186293      .0188289    .1007163
-----------------------------+------------------------------------------------
state: Identity              |
                   sd(_cons) |   .0802738   .0095512      .0635763    .1013567
-----------------------------+------------------------------------------------
                sd(Residual) |   .0368008   .0009442       .034996    .0386987
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  1162.40   Prob > chi2 = 0.0000
```

panels. Instead, Abrevaya introduces a modified Chamberlain approach in which the projection depends upon the form of missingness for a given individual. This leads to orthogonality conditions that depend upon the form of exogeneity assumption maintained. These orthogonality conditions are used in a GMM framework to develop estimators of the model (and projection) parameters as well as tests of strict exogeneity and random effects.

Baltagi and Liu (2020) derive the best linear unbiased prediction (BLUP) for an unbalanced panel data model. Starting with a simple one-way error component regression model with unbalanced panel data and random effects considered in Sect. 9.2, the paper generalizes the BLUP derived by Taub (1979) from the balanced panel data case; see Sect. 2.5 to the unbalanced panel data case. In fact, the BLUP for individual i who is observed for T_i periods corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual. This correction applies no matter what future S period we are forecasting this individual's behavior. Next, the paper derives the BLUP for an unequally spaced panel data model with serial correlation of the AR(1) type in the remainder disturbances considered by Baltagi and Wu (1999) in Sect. 5.2.5. This in turn extends the BLUP for a panel data model with AR(1) type remainder disturbances derived by Baltagi and Li (1991) in Sect. 5.2.6 from the balanced to the unequally spaced panel data case. They illustrate these forecasts with an earnings equation using the NLS young women data over the period 1968–1988 employed by Drukker (2003). The data is unbalanced and the forecast of the logarithm of wage for the last available year for that individual is computed using OLS, FE, RE, FE with AR(1), RE with AR(1). They find that the random effects model with an AR(1) term predicts significantly better than all other models. Problem 9.15 asks the reader to replicate the results in Baltagi and Liu (2020).

Read Chap. 5 of the Oxford Handbook of Panel Data by Bai, Liao and Liang (2015) which deals with unbalanced panel models with interactive effects. They propose new algorithms that allow for various types of unbalanced panels and show their performance using Monte Carlo experiments.

9.8 Notes

1. Other methods of dealing with missing data include (i) imputing the missing values and analyzing the filled-in data by complete panel data methods; (ii) discarding the nonrespondents and weighting the respondents to compensate for the loss of cases; see (Little, 1988) and the section on nonresponse adjustments in Kasprzyk et al. (1989).
2. This analysis assumes that the observations of the individual with the shortest time series are nested in a specific manner within the observations of the other individual. However, the same derivations apply for different types of overlapping observations.
3. Note that if the updated value is negative, it is replaced by zero and the iteration continues until the convergence criterion is satisfied.
4. It is important to note that ML and restricted ML estimates of the variance components are by definition nonnegative. However, ANOVA and MINQUE methods can produce negative estimates of the variance component σ_μ^2. In these cases, the negative variance estimates are replaced by zero. This means that the resulting variance component estimator is $\widetilde{\sigma}_\mu^2 = \max(\widehat{\sigma}_\mu^2, 0)$ which is no longer unbiased.
5. Problem 90.2.3 in *Econometric Theory* by Baltagi and Li (1990a) demonstrated analytically that for a random error component model, one can construct a simple unbiased estimator of the variance components using the entire unbalanced panel that is more efficient than the BQU estimator using only the subbalanced pattern (see problem 9.5). Also, Chowdhury (1991) showed that for the fixed effects error

component model, the Within estimator based on the entire unbalanced panel is efficient relative to any Within estimator based on a subbalanced pattern.

6. The variable descriptions are from Table IV of Harrison and Rubinfeld (1978). See Belsley, Kuh and Welsch (1980) for a listing of the data and further diagnostic analysis of these data. Moulton (1986) used these data to show the inappropriate use of OLS in the presence of random group effects and Moulton (1987) applied a battery of diagnostic tools to this data set.
7. Later, Moulton and Randolph (1989) found that asymptotic critical values of the one-sided LM test can be very poor, and suggested a standardized LM test whose asymptotic critical-value approximations are likely to be much better than those of the LM statistic. They applied it to this data set and rejected the null hypothesis of no random group effect using an exact critical value.
8. Note that the Amemiya-type estimator is not calculated for this data set since there are some regressors without Within variation.
9. If the data were arranged differently, one would get the generalized inverse of an $(N \times N)$ matrix rather than that of $(T \times T)$ as in P. Since $N > T$ in most cases, this choice is most favorable from the point of view of computations.
10. Critical values for the mixed χ^2_m are 7.289, 4.231, and 2.952 for $\alpha = 0.01, 0.05$, and 0.1, respectively.
11. See problem 3.14 for an introduction to the balanced nested error component model.

9.9 Problems

9.1 *Variance–covariance matrix of unbalanced panels*. (a) Show that the variance–covariance matrix of the disturbances in (9.1) is given by (9.2).

(b) Show that the two nonzero block matrices in (9.2) can be written as in (9.3).

(c) Show that $\sigma_\nu \Omega_j^{-1/2} y_j$ has a typical element $(y_{jt} - \theta_j \bar{y}_{j.})$, where $\theta_j = 1 - \sigma_\nu/\omega_j$ and $\omega_j^2 = T_j \sigma_\mu^2 + \sigma_\nu^2$.

9.2 *Wallace and Hussain type estimators for the variance components of a one-way unbalanced panel data model.*

(a) Verify the $E(\widehat{q}_1)$ and $E(\widehat{q}_2)$ equations given in (9.16).
(b) Verify $E(\widetilde{q}_1)$ and $E(\widetilde{q}_2)$ given in (9.17).
(c) Verify $E(\hat{q}_2^b)$ given in (9.19).

9.3 Using the Monte Carlo setup for the unbalanced one-way error component model considered by Baltagi and Chang (1994), compare the various estimators of the variance components and the regression coefficients considered in Sect. 9.3.2.

9.4 *Hedonic housing*. Using the Harrison and Rubinfeld (1978) data published in Belsley, Kuh and Welsch (1980) and provided on the Springer website as Hedonic.xls, reproduce Table 9.1. Perform the Hausman test based on the fixed effects and the random effects contrast. Perform the LM test for H_0; $\sigma_\mu^2 = 0$.

9.5 *Comparison of variance component estimators using balanced vs. unbalanced data*. This exercise is based on problem 90.2.3 in *Econometric Theory* by Baltagi and Li (1990a). Consider the following unbalanced one-way analysis of variance model

$$y_{it} = \mu_i + \nu_{it} \quad i = 1, \ldots, N \quad t = 1, 2, \ldots, T_i$$

where for simplicity's sake no explanatory variables are included. y_{it} could be the output of firm i at time period t and μ_i could be the managerial ability of firm i, whereas ν_{it} is a remainder disturbance term. Assume that $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ independent of each other. Let T be the maximum overlapping period over which a complete panel could be established ($T \leqslant T_i$ for all i). In this case, the corresponding vector of balanced observations on y_{it} is denoted by y_b and is of dimension NT. One could estimate the variance components using this complete panel as follows:

$$\widehat{\sigma}_\nu^2 = y_b'(I_N \otimes E_T)y_b/N(T-1)$$

and

$$\sigma_\mu^2 = [y_b'(I_N \otimes \bar{J}_T)y_b/NT] - (\widehat{\sigma}_\nu^2/T)$$

where $E_T = I_T - \bar{J}_T$, $\bar{J}_T = J_T/T$ and J_T is a matrix of ones of dimension T. $\widehat{\sigma}_\nu^2$ and $\widehat{\sigma}_\mu^2$ are the best quadratic unbiased estimators (BQUE) of the variance components based on the complete panel. Alternatively, one could estimate the variance components from the entire unbalanced panel as follows:

$$\widetilde{\sigma}_\nu^2 = y'\text{diag}(E_{T_i})y/(n-N)$$

where $n = \sum_{i=1}^N T_i$ and $E_{T_i} = I_{T_i} - \bar{J}_{T_i}$. Also, $\sigma_i^2 = (T_i\sigma_\mu^2 + \sigma_\nu^2)$ can be estimated by $\widetilde{\sigma}_i^2 = y_i'\bar{J}_{T_i}y_i$, where y_i denotes the vector of T_i observations on the ith individual. Therefore, there are N estimators of σ_μ^2 obtained from $(\widetilde{\sigma}_i^2 - \widetilde{\sigma}_\nu^2)/T_i$ for $i = 1, \ldots, N$. One simple way of combining them is to take the average

$$\widetilde{\sigma}_\mu^2 = \sum_{i=1}^N [(\widetilde{\sigma}_i^2 - \widetilde{\sigma}_\nu^2)/T_i]/N = \left\{ y'\text{diag}[\bar{J}_{T_i}/T_i]y - \sum_{i=1}^N \widetilde{\sigma}_\nu^2/T_i \right\} /N$$

(a) Show that $\widetilde{\sigma}_\nu^2$ and $\widetilde{\sigma}_\mu^2$ are unbiased estimators σ_ν^2 and σ_μ^2.
(b) Show that $\text{var}(\widetilde{\sigma}_\nu^2) \leqslant \text{var}(\widehat{\sigma}_\nu^2)$ and $\text{var}(\widetilde{\sigma}_\mu^2) \leqslant \text{var}(\widehat{\sigma}_\mu^2)$. (Hint: See solution 90.2.3 in *Econometric Theory* by Koning (1991).)

9.6 *Fixed effects for the two-way unbalanced panel data model*. For $X = (X_1, X_2)$, the generalized inverse of $(X'X)$ is given by

$$(X'X)^- = \begin{bmatrix} (X_1'X_1)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X_1'X_1)^- X_1'X_2 \\ I \end{bmatrix} (X_2' Q_{[X_1]} X_2)^- [-X_2'X_1(X_1'X_1)^- \; I]$$

see Davis (2002), Appendix A. Use this result to show that $P_{[X]} = P_{[X_1]} + P_{[Q_{[X_1]}X_2]}$. (Hint: premultiply this expression by X, and post-multiply by X'.) This verifies (9.29).

9.7 *Fixed effects for the three-way unbalanced panel data model.* Consider the three-way error component model described in problem 3.15. The panel data can be unbalanced and the matrices of dummy variables are $\Delta = [\Delta_1, \Delta_2, \Delta_3]$ with

$$u = \Delta_1 \mu + \Delta_2 \lambda + \Delta_3 \eta + \nu$$

where μ, λ, and ν are random variables defined (9.31) and the added random error η has mean zero and variance σ_η^2. All random errors are independent among themselves and with each other. Show that $P_{[\Delta]} = P_{[A]} + P_{[B]} + P_{[C]}$ where $A = \Delta_1$, $B = Q_{[A]}\Delta_2$, and $C = Q_{[B]}Q_{[A]}\Delta_3$. This is Corollary 1 of Davis (2002). (Hint: apply (9.29) twice. Let $X_1 = \Delta_1$ and $X_2 = (\Delta_2, \Delta_3)$. Using problem 9.6, we get $P_{[X]} = P_{[\Delta_1]} + P_{[Q_{[\Delta_1]}X_2]}$. Now, $Q_{[\Delta_1]}X_2 = Q_{[\Delta_1]}(\Delta_2, \Delta_3) = [B, Q_{[A]}\Delta_3]$. Applying (9.29) again we get $P_{[B, Q_{[A]}\Delta_3]} = P_{[B]} + P_{[Q_{[B]}Q_{[A]}\Delta_3]}$.)

9.8 *Random effects for the unbalanced three-way panel data model.* (a) For Δ_1 and Δ_2 defined in (9.28), verify that $\Delta_N \equiv \Delta_1'\Delta_1 = \text{diag}[T_i]$ and $\Delta_T \equiv \Delta_2'\Delta_2 = \text{diag}[N_t]$. Show that for the complete panel data case $\Delta_1 = \iota_T \otimes I_N$, $\Delta_2 = I_T \otimes \iota_N$, $\Delta_N = TI_N$, and $\Delta_T = NI_T$.

(b) Under the complete panel data case, verify that $\Delta_{TN} \equiv \Delta_2'\Delta_1$ is J_{TN} and $Q = E_T \otimes E_N$, see Chap. 3, Eq. (3.3) and problem 3.1.

(c) Let $X = (X_1, X_2)$ with $|I + XX'| \neq 0$. Using the result that $[I_n + XX']^{-1} = I_n - X(I + X'X)^{-1}X'$, apply the partitioned inverse formula for matrices to show that $(I + XX')^{-1} = \widetilde{Q}_{[X_2]} - \widetilde{Q}_{[X_2]}X_1S^{-1}X_1'\widetilde{Q}_{[X_2]}$ where $\widetilde{Q}_{[X_2]} = I - X_2(I + X_2'X_2)^{-1}X_2' = (I + X_2X_2')^{-1}$ and $S = I + X_1'\widetilde{Q}_{[X_2]}X_1$. This is lemma 2 of Davis (2002).

(d) Apply the results in part (c) using $X = (\frac{\sigma_\mu}{\sigma_\nu}\Delta_1, \frac{\sigma_\lambda}{\sigma_\nu}\Delta_2)$ to verify Σ^{-1} given in (9.33).

(e) Derive $E(q_W)$, $E(q_N)$, and $E(q_T)$ given in (9.34), (9.35), and (9.36).

9.9 Using the Monte Carlo setup for the unbalanced two-way error component model considered by Wansbeek and Kapteyn (1989), compare the MSE performance of the variance components and the regression coefficients estimates.

9.10 *Breusch and Pagan LM test for unbalanced panel data*. Assuming normality on the disturbances, verify (9.37), (9.40) and (9.41).

9.11 *Locally mean most powerful one-sided test for unbalanced panel data*. Verify that the King and Wu (1997) test for the unbalanced two-way error component model is given by (9.49).

9.12 *Standardized Honda and King and Wu tests for unbalanced panel data*. Verify that the SLM version of the KW and HO tests are given by (9.47) with D defined in (9.50) and (9.51).

9.13 *Nested MLE and REML*. Using the Munnell (1990) data set considered in the empirical example in Sect. 9.7.1, estimate the Cobb–Douglas production function investigating the productivity of public capital in each state's private output using nested MLE and REML as shown in the Stata output in Tables 9.5 and 9.6.

9.14 Consider Mundlak's (1978) augmented regression in (7.35) except now allow for unbalanced panel data. Show that OLS on this augmented regression yields the *unbalanced* Within estimator for β given by (9.9).

9.15 *Forecasting with Unbalanced Panels*. This is based on Baltagi and Liu (2020). (a) Derive the BLUP for an unbalanced one-way error component S periods ahead. Show that this predictor corrects the GLS prediction by a fraction of the mean of the GLS residuals corresponding to that ith individual with differing number of observations for each individual over time. (b) Using the unbalanced NLS young women data over the period 1968–1988 employed by Drukker (2003) and available in Stata, estimate the earnings equation using OLS, FE, RE, FE with AR(1), RE with AR(1) and replicate Table 6 of Baltagi and Liu (2020). (c) Forecast the logarithm of wage for the last year available for that individual (not using this last observation in estimation). Using the MSE criteria, show that the RE with AR(1) has the best performance.

References

Abrevaya, J. 2013. The projection approach for unbalanced panel data. *Econometrics Journal* 16: 161–178.

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13.

Antweiler, W. 2001. Nested random effects estimation in unbalanced panel data. *Journal of Econometrics* 101: 295–313.

Bai, J., Y. Liao and J. Liang. 2015. Unbalanced panel data models with interactive effects, Chapter 5. In *Oxford handbook of panel data*, (ed.) Baltagi, B.H., Oxford: Oxford University Press, 149–170.

Baltagi, B.H. 1985. Pooling cross-sections with unequal time-series lengths. *Economics Letters* 18: 133–136.

Baltagi, B.H., and Y.J. Chang. 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics* 62: 67–89.

Baltagi, B.H., Y.J. Chang, and Q. Li. 1998. Testing for random individual and time effects using unbalanced panel data. *Advances in Econometrics* 13: 1–20.

Baltagi, B.H., P. Egger, and M. Pfaffermayr. 2003. A generalized design for bilateral trade flow models. *Economics Letters* 80: 391–397.

Baltagi, B.H., and Q. Li. 1990a. A comparison of variance components estimators using balanced versus unbalanced data, Problem 90.2.3. *Econometric Theory* 6: 283–285.

Baltagi, B.H., and Q. Li. 1990b. A lagrange multiplier test for the error components model with incomplete panels. *Econometric Reviews* 9: 103–107.

Baltagi, B.H., and Q. Li. 1991. A transformation that will circumvent the problem of autocorrelation in an error component model. *Journal of Econometrics* 48: 385–393.

Baltagi, B.H., and L. Liu. 2020. Forecasting with unbalanced panel data. *Journal of Forecasting* 39(5): 709–724.

Baltagi, B.H., S.H. Song, and B.C. Jung. 2001. The unbalanced nested error component regression model. *Journal of Econometrics* 101: 357–381.

Baltagi, B.H., S.H. Song, and B.C. Jung. 2002a. A comparative study of alternative estimators for the unbalanced two-way error component regression model. *Econometrics Journal* 5: 480–493.

Baltagi, B.H., S.H. Song, and B.C. Jung. 2002b. Simple LM tests for the unbalanced nested error component regression model. *Econometric Reviews* 21: 167–187.

Baltagi, B.H., S.H. Song, and B.C. Jung. 2002c. LM tests for the unbalanced nested panel data regression model with serially correlated errors. *Annales D'Économie et de Statistique* 65: 219–268.

Baltagi, B.H., and P.X. Wu. 1999. Unequally spaced panel data regressions with AR (1) disturbances. *Econometric Theory* 15: 814–823.

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.

Breusch, T.S., and A.R. Pagan. 1980. The lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.

Chowdhury, G. 1991. A comparison of covariance estimators for complete and incomplete panel data models. *Oxford Bulletin of Economics and Statistics* 53: 88–93.

Davis, P. 2002. Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics* 106: 67–95.

Drukker, D.M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3 (2): 168–177.

Evans, M.A., and M.L. King. 1985. Critical value approximations for tests of linear regression disturbances. *Australian Journal of Statistics* 27: 68–83.

Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.

Fuller, W.A., and G.E. Battese. 1974. Estimation of linear models with cross-error structure. *Journal of Econometrics* 2: 67–78.

Gourieroux, C., A. Holly, and A. Monfort. 1982. Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50: 63–80.

Graybill, F.A. 1961. *An introduction to linear statistical models*. New York: McGraw-Hill.

Harrison, D., and D.L. Rubinfeld. 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5: 81–102.

Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320–340.

Henderson, C.R. 1953. Estimation of variance components. *Biometrics* 9: 226–252.

Honda, Y. 1985. Testing the error components model with non-normal disturbances. *Review of Economic Studies* 52: 681–690.

Jennrich, R.I., and P.F. Sampson. 1976. Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* 18: 11–17.

Kasprzyk, D., G.J. Duncan, G. Kalton, and M.P. Singh. 1989. *Panel surveys*. New York: Wiley.

King, M.L., and P.X. Wu. 1997. Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* 16: 131–156.

Koning, R.H. 1991. A comparison of variance components estimators using balanced versus unbalanced data, Solution 90.2.3. *Econometric Theory* 6: 425–427.

Little, R.J.A. 1988. Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics* 6: 287–297.

Maddala, G.S., and T.D. Mount. 1973. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association* 68: 324–328.

Mendelsohn, R., D. Hellerstein, M. Huguenin, R. Unsworth, and R. Brazee. 1992. Measuring hazardous waste damages with panel models. *Journal of Environmental Economics and Management* 22: 259–271.

Montmarquette, C., and S. Mahseredjian. 1989. Does school matter for educational achievement? *A two-way nested-error components analysis, Journal of Applied Econometrics* 4: 181–193.

Moulton, B.R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32: 385–397.

Moulton, B.R. 1987. Diagnostics for group effects in regression analysis. *Journal of Business and Economic Statistics* 5: 275–282.

Moulton, B.R., and W.C. Randolph. 1989. Alternative tests of the error components model. *Econometrica* 57: 685–693.

Mundlak, Y. 1978. On the pooling of time series and cross-section data. *Econometrica* 46: 69–85.

Munnell, A. 1990. How does public infrastructure affect regional economic performance? *New England Economic Review* 3–22.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545–554.

Rao, C.R. 1971a. Estimation of variance and covariance components -MINQUE theory. *Journal of Multivariate Analysis* 1: 257–275.

Rao, C.R. 1971b. Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1: 445–456.

Searle, S.R. 1971. *Linear Models*. New York: John Wiley.

Searle, S.R. 1987. *Linear Models for Unbalanced Data*. New York: Wiley.

Swamy, P.A.V.B., and S.S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40: 261–275.

Taub, A.J. 1979. Prediction in the context of the variance-components model. *Journal of Econometrics* 10: 103–108.

Wallace, T.D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37: 55–72.

Wansbeek, T.J., and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.

Wansbeek, T.J., and A. Kapteyn. 1989. Estimation of the error components model with incomplete panels. *Journal of Econometrics* 41: 341–361.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross-Section and Panel Data*. Massachusetts: MIT Press.

10 Special Topics

10.1 Measurement Error and Panel Data

Micro-panel data on households, individuals, and firms are highly likely to exhibit measurement error. In Chap. 1, we cited Duncan and Hill (1985) who found serious measurement error in average hourly earnings in the panel study of income dynamics (PSID). This got worse for a two-year recall as compared to a one-year recall. Bound et al. (1990) use two validation data sets to study the extent of measurement error in labor market variables. The first data set is the panel study of income dynamics validation study (PSIDVS) which uses a two-wave panel survey taken in 1983 and 1987 from a single large manufacturing company. The second data set matches panel data on earnings from the 1977 and 1978 waves of the U.S. Current Population Survey (CPS) to Social Security earnings records for those same individuals. They find that biases from measurement errors could be very serious for hourly wages and unemployment spells, but not severe for annual earnings. In analyzing data from household budget surveys, total expenditure and income are known to contain measurement error.

Econometric textbooks emphasize that measurement error in the explanatory variables result in bias and inconsistency of the OLS estimates, and the solution typically involves the existence of *extraneous* instrumental variables or additional assumptions to identify the model parameters. Using panel data, Griliches and Hausman (1986) showed that one can identify and estimate a variety of errors in variables models *without* the use of external instruments. Let us illustrate their approach with a simple regression with random individual effects:

$$y_{it} = \alpha + \beta x_{it}^* + u_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \tag{10.1}$$

where the error follows a one-way error component model

$$u_{it} = \mu_i + \nu_{it} \tag{10.2}$$

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_10

and the x_{it}^* are observed only with error

$$x_{it} = x_{it}^* + \eta_{it} \tag{10.3}$$

In this case, $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$, $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ and $\eta_{it} \sim \text{IID}(0, \sigma_\eta^2)$ are all independent of each other. Additionally, x_{it}^* is independent of u_{it} and η_{it}. In terms of observable variables, the model becomes

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it} \tag{10.4}$$

where

$$\epsilon_{it} = \mu_i + \nu_{it} - \beta\eta_{it} \tag{10.5}$$

It is clear that OLS on (10.4) is inconsistent, since x_{it} is correlated with η_{it} and therefore ϵ_{it}. We follow Wansbeek and Koning (1991) by assuming that the variance–covariance matrix of x denoted by $\Sigma_x(T \times T)$ is the same across individuals, but otherwise of general form over time. In vector form, the model becomes

$$y = \alpha\iota_{NT} + x\beta + \epsilon \tag{10.6}$$

with

$$\epsilon = (\iota_T \otimes \mu) + \nu - \beta\eta; \quad \mu' = (\mu_1, \ldots, \mu_N)$$
$$\nu = (\nu_{11}, \ldots, \nu_{N1}, \ldots, \nu_{1T}, \ldots, \nu_{NT})$$

and

$$\eta' = (\eta_{11}, \ldots, \eta_{N1}, \ldots, \eta_{1T}, \ldots, \eta_{NT})$$

Note that the data are ordered such that the faster index is over individuals. Now consider any matrix P that wipes out the individual effects. P must satisfy $P\iota_T = 0$ and let $Q = P'P$ For example, $P = I_T - (\iota_T\iota_T'/T)$ is one such matrix, and the resulting estimator is the Within estimator. In general, for any Q, the estimator of β is given by

$$\begin{aligned}\widehat{\beta} &= x'(Q \otimes I_N)y/x'(Q \otimes I_N)x \\ &= \beta + x'(Q \otimes I_N)(\nu - \beta\eta)/x'(Q \otimes I_N)x\end{aligned} \tag{10.7}$$

For a fixed T, taking probability limits as the limit of expectations of the numerator and denominator as $N \to \infty$, we get

$$\frac{1}{N}E[x'(Q \otimes I_N)(\nu - \beta\eta)] = -\frac{1}{N}\beta\,\text{tr}[(Q \otimes I_N)E(\eta\eta')] = -\beta\sigma_\eta^2\,\text{tr}Q$$
$$\frac{1}{N}E[x'(Q \otimes I_N)x] = \frac{1}{N}\,\text{tr}[(Q \otimes I_N)(\Sigma_x \otimes I_N)] = \text{tr}Q\Sigma_x$$

and

$$\begin{aligned}\text{plim}\widehat{\beta} &= \beta - \beta\sigma_\eta^2(\text{tr}Q/\text{tr}Q\Sigma_x) \\ &= \beta(1 - \sigma_\eta^2\phi)\end{aligned} \tag{10.8}$$

where $\phi \equiv (\text{tr}Q/\text{tr}Q\Sigma_x) > 0$. Griliches and Hausman (1986) used various Q transformations like the Within estimator and difference estimators to show that although these transformations wipe out the individual effect, they may aggravate the measurement-error bias. Also, consistent estimators of β and σ_η^2 can be obtained by combining these inconsistent estimators. There are actually $\frac{1}{2}T(T-1)-1$ linearly independent Q transformations. Let Q_1 and Q_2 be two choices for Q and $\phi_i = \text{tr}(Q_i)/\text{tr}(Q_i\Sigma_x)$ be the corresponding choices for ϕ, for $i = 1, 2$. Then $\text{plim}\widehat{\beta}_i = \beta(1-\sigma_\eta^2\phi_i)$, and by replacing $\text{plim}\widehat{\beta}_i$ by $\widehat{\beta}_i$ itself, one can solve these two equations in two unknowns to get

$$\widehat{\beta} = \frac{\phi_1\widehat{\beta}_2 - \phi_2\widehat{\beta}_1}{\phi_1 - \phi_2} \tag{10.9}$$

and

$$\widehat{\sigma}_\eta^2 = \frac{\widehat{\beta}_2 - \widehat{\beta}_1}{\phi_1\widehat{\beta}_2 - \phi_2\widehat{\beta}_1} \tag{10.10}$$

In order to make these estimators operational, ϕ_i is replaced by $\widehat{\phi}_i$, where $\widehat{\phi}_i = \text{tr}(Q_i)/\text{tr}(Q_i\widehat{\Sigma}_x)$. Note that $P = I_T - (\iota_T\iota_T')/T$ yields the Within estimator, while $P = L'$, where L' is the $(T-1)\times T$ matrix defined in Chap. 8, yields the first-difference estimator. Other P matrices suggested by Griliches and Hausman (1986) are based on differencing the data j periods apart, $(y_{it} - y_{i,t-j})$, thus generating "different lengths" difference estimators. The remaining question is how to combine these consistent estimators of β into an efficient estimator of β. The generalized method of moments (GMM) approach can be used and this is based upon fourth-order moments of the data. Alternatively, under normality one can derive the asymptotic covariance matrix of the $\widehat{\beta}_i$ which can be consistently estimated by second-order moments of the data. Using the latter approach, Wansbeek and Koning (1991) showed that for m different consistent estimators of β given by $b = (\widehat{\beta}_1, \ldots, \widehat{\beta}_m)'$ based on m different Q_i

$$\sqrt{N}[b - \beta(\iota_m - \sigma_\eta^2\phi)] \sim N(0, V)$$

where

$$\begin{aligned} \phi &= (\phi_1, \ldots, \phi_m)' \\ V &= F'\{\sigma_v^2\Sigma_x \otimes I_T + \beta^2\sigma_\eta^2(\Sigma_x + \sigma_\eta^2 I_N)\otimes I_T\}F \end{aligned} \tag{10.11}$$

and F is the $(T^2\times m)$ matrix with ith column $f_i = \text{vec}\, Q_i/(\text{tr}Q_i\Sigma_x)$. By minimizing $[b - \beta(\iota_m - \sigma_\eta^2\phi)]'V^{-1}[b - \beta(\iota_m - \sigma_\eta^2\phi)]$ one gets the asymptotically efficient estimators (as far as they are based on b) of β and σ_v^2 given by

$$\widehat{\beta} = \left\{\frac{\phi'\widehat{V}^{-1}b}{\phi'\widehat{V}^{-1}\phi} - \frac{\iota'\widehat{V}^{-1}b}{\iota'\widehat{V}^{-1}\phi}\right\} \Big/ \left\{\frac{\phi'\widehat{V}^{-1}\iota}{\phi'\widehat{V}^{-1}\phi} - \frac{\iota'\widehat{V}^{-1}\iota}{\iota'\widehat{V}^{-1}\phi}\right\} \tag{10.12}$$

and

$$\widehat{\sigma}_\nu^2 = \left\{\frac{\phi'\widehat{V}^{-1}\iota}{\phi'\widehat{V}^{-1}b} - \frac{\iota'\widehat{V}^{-1}\iota}{\iota'\widehat{V}^{-1}b}\right\} \Big/ \left\{\frac{\phi'\widehat{V}^{-1}\phi}{\phi'\widehat{V}^{-1}b} - \frac{\iota'\widehat{V}^{-1}\phi}{\iota'\widehat{V}^{-1}b}\right\} \tag{10.13}$$

with $\sqrt{N}(\widehat{\beta} - \beta, \widehat{\sigma}_\nu^2 - \sigma_\nu^2)$ asymptotically distributed as $N(0, W)$ and

$$W = \frac{1}{\Delta}\begin{bmatrix} \beta^2\phi'V^{-1}\phi & \beta(\iota_m - \sigma_\eta^2\phi)'V^{-1}\phi \\ & (\iota_m - \sigma_\eta^2\phi)'V^{-1}(\iota_m - \sigma_\eta^2\phi) \end{bmatrix} \tag{10.14}$$

where

$$\Delta = \beta^2(\iota_m - \sigma_\eta^2\phi)'V^{-1}(\iota_m - \sigma_\eta^2\phi)(\phi' V^{-1}\phi) - \beta^2[\phi' V^{-1}(\iota_m - \sigma_\eta^2\phi)]^2 \quad (10.15)$$

Griliches and Hausman (1986) classic paper shows that measurement-error exacerbates the bias in the fixed effects estimator and it can be reduced by differencing estimators that decrease the bias the further the differencing periods are apart. Griliches and Hausman (1986) argue that their results can be extended to the case of several independent variables provided that the measurement errors in the explanatory variables are mutually uncorrelated, or correlated with a known correlation structure. Under some stringent assumptions these results can be extended to the case of serially correlated η_{it}. They illustrate their approach by estimating a labor demand relationship using data on $N = 1242$ U.S. manufacturing firms over six years (1972–77) drawn from the National Bureau of Economic Research R&D panel.

Biorn (1996) also gives an extensive treatment for the case where the model disturbances u_{it} in Eq. (10.1) are white noise, i.e., without any error component, and the case where η_{it}, the measurement error, is autocorrelated over time. For all cases considered, Biorn derives the asymptotic bias of the Within, Between, various difference estimators and the GLS estimator as either N or T tend to ∞. Biorn shows how the different panel data transformations implied by these estimators affect measurement error differently. Biorn (2000) proposes GMM estimators that use either (A) equations in differences with level values as instruments, or (B) equations in levels with differenced values as instruments. The conditions needed for the consistency of the (B) procedures under individual heterogeneity are stronger than for the (A) procedures. These procedures are illustrated for a simple regression of log of gross production on log of material input for the manufacture of textiles. The data uses $N = 215$ firms observed over $T = 8$ years 1983–90 and obtained from the annual Norwegian manufacturing census. For this empirical illustration, Biorn shows that adding the essential two-period difference orthogonality conditions to the one-period conditions in the GMM algorithm may significantly increase estimation efficiency. However, redundant orthogonality conditions are of little practical use. Overall, the GMM estimates based on the level equations are more precise than those based on differenced equations.

Read Chap. 11 of the Oxford Handbook of Panel data entitled measurement error in panel data by Meijer, Spierdijk and Wansbeek (2015). This chapter takes the reader through a simple panel data model with measurement error and explains how panels can decrease the inconsistency of OLS. Then it shows how the fixed effects transformation and differencing that wipe out the individual effects are affected by measurement error. In fact, they derive an ordering of the reduction in bias due to these various estimators. This reproduces the Hausman and Griliches (1986) result that the inconsistency due to measurement error decreases when differences are taken further apart in time. Much more on how measurement error affects the random effects estimator, dynamic estimators, identification issues. This is essential reading that supplements this section.

10.2 Rotating Panels

Biorn (1981) considers the case of rotating panels, where in order to keep the *same* number of households in the survey, the fraction of households that drops from the sample in the second period is replaced by an equal number of new households that are freshly surveyed. This is a necessity in survey panels where the same household may not want to be interviewed again and again. In the study by Biorn and Jansen (1983) based on data from the Norwegian household budget surveys, half the sample is rotated in each period. In other words, half the households surveyed drop from the sample each period and are replaced by new households. To illustrate the basics of rotating panels, let us assume that $T = 2$ and that half the sample is being rotated each period. In this case, without loss of generality, households $1, 2, \ldots, (N/2)$ are replaced by households $N + 1, N + 2, \ldots, N + (N/2)$ in period 2. It is clear that only households $(N/2) + 1, (N/2) + 2, \ldots, N$ are observed over two periods.[1] In this case there are $3N/2$ distinct households, only $N/2$ households of which are observed for two periods. In our case, the first and last $N/2$ households surveyed are only observed for one period. Now consider the usual one-way error component model

$$u_{it} = \mu_i + \nu_{it}$$

with $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Order the observations such that the faster index is that of households and the slower index is that of time. This is different from the ordering we used in Chap. 2. In this case, $u' = (u_{11}, u_{21}, \ldots, u_{N1}, u_{N/2+1,2}, \ldots, u_{3N/2,2})$ and

$$E(uu') = \Omega = \begin{bmatrix} \sigma^2 I_{N/2} & 0 & 0 & 0 \\ 0 & \sigma^2 I_{N/2} & \sigma_\mu^2 I_{N/2} & 0 \\ 0 & \sigma_\mu^2 I_{N/2} & \sigma^2 I_{N/2} & 0 \\ 0 & 0 & 0 & \sigma^2 I_{N/2} \end{bmatrix} \tag{10.16}$$

where $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$. It is easy to see that Ω is block-diagonal and that the middle block has the usual error component model form $\sigma_\mu^2(J_2 \otimes I_{N/2}) + \sigma_\nu^2(I_2 \otimes I_{N/2})$. Therefore

$$\Omega^{-1/2} = \begin{bmatrix} \frac{1}{\sigma} I_{N/2} & 0 & 0 \\ 0 & \left(\frac{1}{\sigma_1^*}\bar{J}_2 + \frac{1}{\sigma_\nu}E_2\right) \otimes I_{N/2} & 0 \\ 0 & 0 & \frac{1}{\sigma} I_{N/2} \end{bmatrix} \tag{10.17}$$

where $E_2 = I_2 - \bar{J}_2$, $\bar{J}_2 = J_2/2$ and $\sigma_1^{*2} = 2\sigma_\mu^2 + \sigma_\nu^2$. By premultiplying the regression model by $\Omega^{-1/2}$ and performing OLS one gets the GLS estimator of the rotating panel. In this case, one divides the first and last $N/2$ observations by σ. For the middle N observations, with $i = (N/2) + 1, \ldots, N$ and $t = 1, 2$, quasi-demeaning similar to the usual error component transformation is performed, i.e., $(y_{it} - \theta^* \bar{y}_{i.})/\sigma_\nu$ with $\theta^* = 1 - (\sigma_\nu/\sigma_1^*)$ and $\bar{y}_{i.} = (y_{i1} + y_{i2})/2$. A similar transformation is also performed on the regressors. In order to make this GLS estimator feasible, we need estimates of σ_μ^2 and σ_ν^2. One consistent estimator of σ_ν^2 can be obtained from the middle N observations or simply the households that are observed over two periods. For these observations, σ_ν^2 is estimated consistently from the Within residuals

$$\widetilde{\sigma}_\nu^2 = \sum_{t=1}^{2} \sum_{i=N/2+1}^{N} [(y_{it} - \bar{y}_{i.}) - (x_{it} - \bar{x}_{i.})'\widetilde{\beta}_{Within}]^2/N \tag{10.18}$$

whereas the total variance can be estimated consistently from the least squares mean square error over the entire sample

$$\widetilde{\sigma}^2 = \widetilde{\sigma}_\nu^2 + \widetilde{\sigma}_\mu^2 = \sum_{t=1}^{2} \sum_{i=1}^{3N/2} (y_{it} - x_{it}'\widehat{\beta}_{OLS})^2/(3N/2) \tag{10.19}$$

Note that we could have reordered the data such that households observed over one period are stacked on top of households observed over two time periods. This way the rotating panel problem becomes an unbalanced panel problem with N households observed over one period and $N/2$ households observed for two periods. In fact, except for this different way of ordering the observations, one can handle the estimation as in Chap. 9.

This feasible GLS estimation can be easily derived for other rotating schemes. In fact, the reader is asked to do that for $T = 3$ with $N/2$ households rotated every period, and $T = 3$ with $N/3$ households rotated every period (see problem 10.2). For the estimation of more general rotation schemes as well as maximum likelihood estimation under normality, see Biorn (1981). The analysis of rotating panels can also be easily extended to a set of seemingly unrelated regressions, simultaneous equations, or a dynamic model. Biorn and Jansen (1983) consider a rotating panel of 418 Norwegian households, one half of which are observed in 1975 and 1976 and the other half in 1976 and 1977. They estimate a complete system of consumer demand functions using maximum likelihood procedures.

Rotating panels allow the researcher to test for the existence of "time-in-sample" bias effects mentioned in Chap. 1. These correspond to a significant change in response between the initial interview and a subsequent interview when one would expect the same response.[2] With rotating panels, the fresh group of individuals that are added to the panel with each wave provide a means of testing for time-in-sample bias effects. Provided that all other survey conditions remain constant for all rotation groups at a particular wave, one can compare these various rotation groups (for that wave) to measure the extent of rotation group bias. This has been done for various labor force characteristics in the Current Population Survey. While the findings indicate a pervasive effect of rotation group bias in panel surveys, the survey conditions do not remain the same in practice and hence it is hard to disentangle the effects of time-in-sample bias from other effects.

10.3 Pseudo-Panels

For some countries, panel data may not exist. Instead the researcher may find annual household surveys based on a large random sample of the population. Examples of some of these cross-sectional consumer expenditure surveys include: the British Family Expenditure Survey, also a number of household surveys from less developed countries sponsored by the World Bank. Examples of repeated cross-section surveys in the USA include the Current Population Survey, the National Health Interview Survey, the Consumer Expenditure Survey, the National Crime Survey, and the survey of small business finances from the Federal Reserve Board, to mention a few.

For these repeated cross-section surveys, it may be impossible to track the same household over time as required in a genuine panel. Instead, Deaton (1985) suggests tracking cohorts and estimating economic relationships based on cohort means rather than individual observations. One cohort could be the set of all males born between 1945 and 1950. This *birth cohort* is well defined and can be easily identified from the data. Deaton (1985) argued that these pseudo-panels do not suffer the attrition problem that plagues genuine panels and may be available over longer time periods compared to genuine panels.[3] In order to illustrate the basic ideas involved in constructing a pseudo-panel, we start with the set of T independent cross-sections given by

$$y_{it} = x'_{it}\beta + \mu_i + \nu_{it} \quad t = 1, \ldots, T \tag{10.20}$$

Note that the individual subscript i corresponds to a new and most likely different set of individuals in each period. This is why it is denoted by i(t) to denote that each period different individuals are sampled, making these individuals time dependent. For ease of exposition, we continue the use of the subscript i and assume that the same number of households N is randomly surveyed each period. Define a set of C cohorts, each with a fixed membership that remains the same throughout the entire period of observation. Each individual observed in the survey belongs to exactly one cohort. Averaging the observations over individuals in each cohort, one gets

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_{ct} + \bar{\nu}_{ct} \quad c = 1, \ldots, C \quad t = 1, \ldots, T \tag{10.21}$$

where $\bar{y}_{ct}$ is the average of y_{it} over all individuals belonging to cohort c at time t. Since the economic relationship for the individual includes an individual fixed effect, the corresponding relationship for the cohort will also include a fixed cohort effect. However, $\bar{\mu}_{ct}$ now varies with t, because it is averaged over a different number of individuals belonging to cohort c at time t. These $\bar{\mu}_{ct}$ are most likely correlated with the x_{it} and a random effect specification will lead to inconsistent estimates. On the other hand, treating the $\bar{\mu}_{ct}$ as fixed effects leads to an identification problem, unless $\bar{\mu}_{ct} = \bar{\mu}_c$ and is invariant over time. The latter assumption is plausible if the number of observations in each cohort is very large. In this case,

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_c + \bar{\nu}_{ct} \quad c = 1, \ldots, C \quad t = 1, \ldots, T \tag{10.22}$$

For this pseudo-panel with T observations on C cohorts, the fixed effects estimator $\widetilde{\beta}_W$, based on the Within cohort transformation $\widetilde{y}_{ct} = \bar{y}_{ct} - \bar{y}_c$, is a natural candidate for estimating β. Note that the cohort *population* means are genuine panels in that, at the population level, the groups contain the same individuals over time. However, as Deaton (1985) argued, the sample-based averages of the cohort means, $\bar{y}_{ct}$, can only estimate the unobserved population cohort means with measurement error. Therefore, one has to correct the Within estimator for measurement error using estimates of the errors in measurement variance–covariance matrix obtained from the individual survey data. Details are given in Deaton (1985) and Verbeek and Nijman (1993). Deaton (1985) shows that his proposed measurement-error corrected Within-groups estimator for the static model with individual effects is *consistent* for a fixed number of observations per cohort. Verbeek and Nijman (1993) modify Deaton's estimator to achieve consistency for a fixed number of time periods and a fixed number of individuals per cohort. If the number of individuals in each cohort is large, so that the average cohort size $n_c = N/C$ tends to infinity, then the measurement errors as well as their estimates tend to zero and the Within cohort estimator of β is asymptotically identical to Deaton's (1985) estimator of β, denoted by $\widetilde{\beta}_D$. In fact, when n_c is large, most applied researchers ignore the measurement error problem and compute the Within cohort estimator of β (see Browning, Deaton and Irish, 1985). The last study involved sixteen cohorts, seven-time periods with an average cohort size of 190.

There is an obvious trade-off in the construction of a pseudo-panel. The larger the number of cohorts, the smaller is the number of individuals per cohort. In this case, C is large and the pseudo-panel is based on a large number of observations. However, the fact that n_c is not large implies that the sample cohort averages are not precise estimates of the population cohort means. In this case, we have a large number C of imprecise observations. In contrast, a pseudo-panel constructed with a smaller number of cohorts (C) and therefore more individuals per cohort (n_c) is trading a large pseudo-panel with imprecise observations for a smaller pseudo-panel with more precise observations. It is important to note that $n_c \to \infty$ is a crucial condition for the consistency of the Within cohort estimator. However, the bias of the Within cohort estimator may be substantial even for large n_c, see Verbeek (2008). On the other hand, Deaton's estimator is consistent for β, for finite n_c when either C or T tend to infinity. How to choose the cohorts under study is very important. For example, in order to minimize the measurement error variance, the individuals in each cohort should be as homogeneous as possible. However, to maximize the variation in the pseudo-panel, and get precise estimates, the different cohorts should be as heterogeneous as possible. Verbeek (2008) empasizes the fact that a necessary condition for consistent estimation is that all exogenous variables exhibit time-varying cohort-specific variation. This is not necessarily satisfied in empirical applications.

Moffitt (1993) extends Deaton's (1985) analysis to the estimation of dynamic models with repeated cross-sections. By imposing certain restrictions, Moffitt shows that linear and nonlinear models, with and without fixed effects, can be identified and consistently estimated with pseudo-panels. Moffitt (1993) gives an instrumental variable interpretation for the Within estimator based on the pseudo-panel using cohort dummies, and a set of time dummies interacted with the cohort dummies. Because n_c is assumed to tend to ∞, the measurement error problem is ignored. Since different individuals are sampled in each period, the lagged dependent variable is not observed. Moffitt suggests replacing the unknown $y_{i,t-1}$ by a fitted value obtained from observed data at time $t-1$. Moffitt (1993) illustrates his estimation method for the linear fixed effects life-cycle model of labor supply using repeated cross-sections from the U.S. Current Population Survey (CPS). The sample included white males, ages 20–59, drawn from 21 waves over the period 1968-88. In order to keep the estimation problem manageable, the data were randomly sub-sampled to include a total of 15500 observations. Moffitt concludes that there is a considerable amount of parsimony achieved in the specification of age and cohort effects. Also, individual characteristics are considerably more important than either age, cohort, or year effects. Blundell, Meghir and Neves (1993) use the annual UK Family Expenditure Survey covering the period 1970–84 to study the intertemporal labor supply and consumption of married women. The total number of households considered was 43671. These were allocated to ten different cohorts depending on the year of birth. The average number of observations per cohort was 364. Their findings indicate reasonably sized intertemporal labor supply elasticities.

Collado (1997) proposes measurement-error corrected estimators for dynamic models with individual effects using time series of independent cross-sections. A GMM estimator corrected for measurement error is proposed that is consistent as the number of cohorts, tends to infinity for a fixed T and a fixed number of individuals per cohort. In addition, a measurement-error corrected Within-groups estimator is proposed which is consistent as T tends to infinity. Monte Carlo simulations are performed to study the small sample properties of the estimators proposed. Some of the main results indicate that the measurement-error correction is important, and that corrected estimators reduce the bias obtained. Also, for small T, GMM estimators are better than Within-groups estimators.

Verbeek and Vella (2005) review the identification conditions for consistent estimation of a linear dynamic model from repeated cross-sections. They show that Moffitt's (1993) estimator is inconsistent, unless the exogenous variables are either time-invariant or exhibit no autocorrelation. They propose an alternative instrumental variable estimator, corresponding to the Within estimator applied to the pseudo-panel of cohort averages. This estimator is consistent under the same conditions as those suggested by Collado (1997). However, Verbeek and Vella argue that those conditions are not trivially satisfied in applied work.

Girma (2000) suggests an alternative GMM method of estimating linear dynamic models from a time series of independent cross-sections. Unlike the Deaton (1985) approach of averaging across individuals in a cohort, Girma suggests a quasi-differencing transformation across pairs of individuals that belong to the same group. The asymptotic properties of the proposed GMM estimators are based upon having a large number of individuals per group-time cell. This is in contrast to the Deaton-type estimator which requires the number of group/time periods to grow without limit. Some of the other advantages of this method include the fact that no aggregation is involved, the dynamic response parameters can freely vary across groups, and the presence of unobserved individual specific heterogeneity is explicitly allowed for.

McKenzie (2001) considers the problem of estimating dynamic models with unequally spaced pseudo-panel data. Surveys in developing countries are often taken at unequally spaced intervals and this unequal spacing, in turn, imposes nonlinear restrictions on the parameters.[4] nonlinear least squares, minimum distance, and one-step estimators are suggested that are consistent and asymptotically normal for finite T as the number of individuals per cohort is allowed to pass to infinity. In another paper, McKenzie (2004) allows for parameter heterogeneity among cohorts and argues that in many practical applications, it is important to investigate whether there are systematic differences between cohorts. McKenzie (2004) develops an asymptotic theory for pseudo-panels using sequential and diagonal path limit techniques following the work of Phillips and Moon (1999) for nonstationary panels. McKenzie uses 20 years of household survey data (1976–1996) from the Taiwanese personal income distribution survey, to quantify the degree of inter-cohort parameter heterogeneity. He finds that younger cohorts experienced faster consumption growth over the sample period than older cohorts.

Inoue (2008) considers the efficient estimation of pseudo-panels when the number of individuals per cohort n_c is large relative to the number of cohorts C, and the number of time periods T. Inuoe shows that the OLS estimator, ignoring the time-invariant cohort fixed effects, converges to a random variable rather than a constant. Also, the fixed effects estimator employing cohort effects is consistent but the associated t-statistics are not asymptotically Normal. Inoue proposes efficient GMM estimators using the orthogonality conditions implied by the grouping into cohorts and provides t-tests that are valid even in the presence of time invariant cohort effects. Inoue suggests using the GMM over-identification test as a test for the validity of cohort selection.

10.4 Short-Run Versus Long-Run Estimates in Pooled Models

Applied studies using panel data find that the *Between* estimator (which is based on the cross-sectional component of the data) tends to give long-run estimates while the *Within* estimator (which is based on the time-series component of the data) tends to give short-run estimates. This agrees with the folk wisdom that cross-sectional studies tend to yield long-run responses while time-series studies tend to yield short-run responses. Both are consistent estimates of the same regression coefficients as long as the disturbances are uncorrelated with the explanatory variables. In fact, Hausman's specification test is based on the difference between these estimators (see Chap. 4). Rejection of the null implies that the random individual effects are correlated with the explanatory variables. This means that the *Between* estimator is inconsistent while the *Within* estimator is consistent since it sweeps away the individual effects. In these cases, the applied researcher settles on the *Within* estimator rather than the *Between* or *GLS* estimators. (See Mundlak, 1978 for additional support of the *Within* estimator). Baltagi and Griffin (1984) argue that in panel data models, the difference between the *Within* and *Between* estimators is due to dynamic misspecification. The basic idea is that even with a rich panel data set, long-lived lag effects coupled with the shortness of the time series is a recipe for dynamic underspecification. This is illustrated using Monte Carlo experiments. Egger and Pfaffermayr (2004) show that the asymptotic bias of the *Within* and *Between* estimators as estimates of short-run and long-run effects depend upon the memory of the data generating process, the length of the time series and the importance of the cross-sectional variation in the explanatory variables. Griliches and Hausman (1986) attribute the difference between the *Within* and *Between* estimators to measurement error in panel data (see Sect. 10.1). Mairesse (1990) tries to explain why these two estimators differ in economic applications using three samples of large manufacturing firms in France, Japan, and the USA over the period 1967–79, and a Cobb–Douglas production function. Mairesse (1990) compares *OLS*, *Between* and the *Within* estimators using levels and first-differenced regressions with and without constant returns to scale. Assuming constant returns to scale, he finds that the *Between* estimates of the elasticity of capital are of the order of 0.31 for France, 0.47 for Japan and 0.22 for the USA, whereas the *Within* estimates are lower, varying from 0.20 for France to 0.28 for Japan and 0.21 for the USA. Mairesse (1990) argues that if the remainder error ν_{it} is correlated with the explanatory variables, then the *Within* estimator will be inconsistent, while the *Between* estimator is much less affected by these correlations because the ν_{it} are averaged and practically wiped out for large enough T. This is also the case when measurement error in the explanatory variables is present. In fact, if these measurement errors are not serially correlated from one year to the next, the *Between* estimator tends to minimize their importance by averaging. In contrast, the *Within* estimator magnifies the variability of these measurement errors and increases the resulting bias.

10.5 Heterogeneous Panels

For panel data studies with large N and small T, it is usual to pool the observations, assuming homogeneity of the slope coefficients. The latter is a testable assumption which is quite often rejected, see Chap. 4. Moreover, with the increasing time dimension of panel data sets, some researchers including Robertson and Symons (1992) and Pesaran and Smith (1995) have questioned the poolability of the data across heterogeneous units. Instead, they argue in favor of heterogeneous estimates that can be combined to obtain homogeneous estimates if the need arises. To make this point, Robertson and Symons (1992) studied the properties of some panel data estimators when the regression coefficients vary across individuals, i.e., they are *heterogeneous* but are assumed *homogeneous* in estimation. This is done for both stationary and nonstationary regressors. The basic conclusion is that severe biases can occur in dynamic estimation even for relatively small parameter variation. They consider the case of say two countries ($N = 2$), where the asymptotics depend on $T \to \infty$. Their true model is a simple *heterogeneous static* regression model with one regressor

$$y_{it} = \beta_i x_{it} + \nu_{it} \quad i = 1, 2 \quad t = 1, \ldots, T \tag{10.23}$$

where ν_{it} is independent for $i = 1, 2$, and β_i varies across $i = 1, 2$. However, their estimated model is *dynamic and homogeneous* with $\beta_1 = \beta_2 = \beta$ and assumes an identity covariance matrix for the disturbances:

$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + w_{it} \quad i = 1, 2 \tag{10.24}$$

The regressors are assumed to follow a stationary process $x_{it} = \rho x_{i,t-1} + \epsilon_{it}$ with $|\rho| < 1$ but different variances σ_i^2 for $i = 1, 2$. Seemingly unrelated regression estimation with the equality restriction imposed and an identity covariance matrix reduces to OLS on this system of two equations. Robertson and Symons (1992) obtain the probability limits of the resulting $\widehat{\lambda}$ and $\widehat{\beta}$ as $T \to \infty$. They find that the coefficient λ of $y_{i,t-1}$ is overstated, while the mean effect of the regressors (the x_{it}) is underestimated. In case the regressors are random walks ($\rho = 1$), then plim $\widehat{\lambda} = 1$ and plim$\widehat{\beta} = 0$. Therefore, false imposition of parameter homogeneity, and dynamic estimation of a static model when the regressors follow a random walk lead to perverse results. Using Monte Carlo experiments they show that the dynamics become misleading even for T as small as 40, which corresponds to the annual postwar data period. Even though these results are derived for $N = 2$, one regressor and no lagged dependent variable in the true model, Robertson and Symons (1992) show that the same phenomenon occurs for an empirical example of a real wage equation for a panel of 13 OECD countries observed over the period 1958-86. Parameter homogeneity across countries is rejected and the true relationship appears dynamic. Imposing false equality restriction biases the coefficient of the lagged wage upwards and the coefficient of the capital-labor ratio downwards.

For typical labor or consumer panels where N is large but T is fixed, Robertson and Symons (1992) assume that the true model is given by (10.23) with $\beta_i \sim \text{IID}(\beta, \sigma_\beta^2)$ for $i = 1, \ldots, N$, and $\nu_{it} \sim \text{IID}(0, 1)$. In addition, x_{it} is AR(1) with innovations $\epsilon_{it} \sim \text{IID}(0, 1)$ and $x_{i0} = \nu_{i0} = 0$. The estimated model is dynamic as given by (10.24), with known variance–covariance matrix I, and $\beta_i = \beta$ imposed for $i = 1, \ldots, N$. For fixed T, and random walk regressors, plim $\widehat{\lambda} > 0$ and plim $\widehat{\beta} < \beta$ as $N \to \infty$, so that the coefficient of $y_{i,t-1}$ is over-estimated and the mean effect of the β_i is underestimated. As $T \to \infty$, one gets the same result obtained previously, plim $\widehat{\lambda} = 1$ and plim $\widehat{\beta} = 0$. If the regressor x_{it} is white noise, no biases arise. These results are confirmed with Monte Carlo experiments for $T = 5$ and $N = 50$, 100 and 200. The dynamics are overstated even for $N = 50$ and $T = 5$, but they disappear as the regressor approaches white noise, and remain important for autoregressive regressors with $\rho = 0.5$. Finally, Robertson and Symons (1992) reconsider the Anderson and Hsiao (1982) estimator of a dynamic panel data model that gets rid of the individual effects by first-differencing and uses lagged first-differences of the regressors as instruments. Imposing false equality restrictions renders these instruments invalid unless x_{it} is white noise or follows a random walk. Only the second case is potentially important because many economic variables are well approximated by random walks. However, Robertson and Symons (1992) show that if x_{it} is a random walk, the instrument is orthogonal to the instrumented variable and the resulting estimator has infinite asymptotic variance, a result obtained in the stationary case by Arellano (1989). Using levels ($y_{i,t-2}$) as instruments as suggested by Arellano (1989) will not help when x_{it} is a random walk, since the correlation between the stationary variable ($y_{i,t-1} - y_{i,t-2}$) and the $I(1)$ variable $y_{i,t}$ will be asymptotically zero.

Using Monte Carlo experiments, with $T = 5$ and $N = 50$, Robertson and Symons (1992) conclude that the Anderson and Hsiao (1982) estimator is useful only when x_{it} is white noise or a random walk. Otherwise, severe biases occur when x_{it} is stationary and autocorrelated.

Pesaran and Smith (1995) consider the problem of estimating a dynamic panel data model when the parameters are individually heterogeneous and illustrate their results by estimating industry-specific UK labor demand functions. In this case the model is given by

$$y_{it} = \lambda_i y_{i,t-1} + \beta_i x_{it} + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{10.25}$$

where λ_i is IID$(\lambda, \sigma^2_\lambda)$ and β_i is IID(β, σ^2_β). Further λ_i and β_i are independent of y_{is}, x_{is} and u_{is} for all s. The objective in this case is to obtain consistent estimates of the mean values of λ_i and β_i. Pesaran and Smith (1995) present four different estimation procedures:

(1) aggregate time-series regressions of group averages;
(2) cross-section regressions of averages over time;
(3) pooled regressions allowing for fixed or random intercepts, or
(4) separate regressions for each group, where coefficients estimate are averaged over these groups.

They show that when T is small (even if N is large), all the procedures yield inconsistent estimators. The difficulty in obtaining consistent estimates for λ and β can be explained by rewriting (10.25) as

$$y_{it} = \lambda y_{i,t-1} + \beta x_{it} + \nu_{it} \tag{10.26}$$

where $\nu_{it} = u_{it} + (\lambda_i - \lambda)y_{i,t-1} + (\beta_i - \beta)x_{it}$. By continuous substitution of $y_{i,t-s}$ it is easy to see that ν_{it} is correlated with all present and past values of $y_{i,t-1-s}$ and x_{it-s} for $s \geqslant 0$. The fact that ν_{it} is correlated with the regressors of (10.26) renders the OLS estimator inconsistent, and the fact that ν_{it} is correlated with $(y_{i,t-1-s}, x_{i,t-s})$ for $s > 0$ rules out the possibility of choosing any lagged value of y_{it} and x_{it} as legitimate instruments. When both N and T are large, Pesaran and Smith (1995) show that the cross-section regression procedure will yield consistent estimates of the mean values of λ and β. Intuitively, when T is large, the individual parameters λ_i and β_i can be consistently estimated using T observations of each individual i, say $\widehat{\lambda}_i$ and $\widehat{\beta}_i$, then averaging these individual estimators, $\sum_{i=1}^N \widehat{\lambda}_i/N$ and $\sum_{i=1}^N \widehat{\beta}_i/N$, will lead to consistent estimators of the mean values of λ and β. This is known as the mean group estimator and is programmed in Stata with the *xtmg* command, see Eberhardt (2012).

Maddala et al. (1997) on the other hand argued that the heterogeneous time series estimates yield inaccurate estimates and even wrong signs for the coefficients, while the panel data estimates are not valid when one rejects the hypothesis of homogeneity of the coefficients. They argued that shrinkage estimators are superior to either heterogeneous or homogeneous parameter estimates especially for prediction purposes. In fact, Maddala et al. (1997) considered the problem of estimating short-run and long-run elasticities of residential demand for electricity and natural gas for each of 49 states over the period 1970-90. They conclude that individual heterogeneous state estimates were hard to interpret and had the wrong signs. Pooled data regressions were not valid because the hypothesis of homogeneity of the coefficients was rejected. They recommend shrinkage estimators if one is interested in obtaining elasticity estimates for each state since these give more reliable results.

In the context of dynamic demand for gasoline across 18 OECD countries over the period 1960–1990, Baltagi and Griffin (1997) argued for pooling the data as the best approach for obtaining reliable price and income elasticities. They also pointed out that pure cross-section studies cannot control for unobservable country effects, whereas pure time-series studies cannot control for unobservable oil shocks or behavioral changes occurring over time. Baltagi and Griffin (1997) compared the homogeneous and heterogeneous estimates in the context of gasoline demand based on the plausibility of the price and income elasticities as well as the speed of adjustment path to the long-run equilibrium. They found considerable variability in the parameter estimates among the heterogeneous estimators some giving implausible estimates, while the homogeneous estimators gave similar plausible short-run estimates that differed only in estimating the long-run effects. Baltagi and Griffin (1997) also compared the forecast performance of these homogeneous and heterogeneous estimators over one, five, and ten years horizon. Their findings show that the homogeneous estimators outperformed their heterogeneous counterparts based on mean squared forecast error. This result was replicated using a panel data set of 21 French regions over the period 1973–1998 by Baltagi et al. (2003). Unlike the international OECD gasoline data set, the focus on the inter-regional differences in gasoline prices and income within France posed a different type of data set for the heterogeneity versus homogeneity debate. The variation in these prices and income were much smaller than international price and income differentials. This in turn reduces the efficiency gains from pooling and favors the heterogeneous estimators, especially given the differences between the Paris region and the rural areas of France. Baltagi et al. (2003) showed that the time series estimates for each region are highly variable, unstable and offer the worst out-of-sample forecasts. Despite the fact that the shrinkage estimators proposed by Maddala et al. (1997) outperformed these individual heterogeneous estimates, they still had a wide range and were outperformed by the homogeneous estimators in out-of-sample forecasts. Baltagi, Griffin and Xiong (2000) carried out this comparison for a dynamic demand for cigarettes across 46 U.S. states over 30 years (1963–1992). Once again the results showed that the homogeneous panel data estimators beat the heterogeneous and shrinkage type estimators in RMSE performance for out-of-sample forecasts. In another application, Driver et al. (2004) utilize the Confederation of British Industry's (CBI) survey data to measure the impact of uncertainty on UK investment authorizations. The panel consists of 48 industries observed over 85 quarters 1978(Q1) to 1999(Q1). The uncertainty measure is based on the dispersion of beliefs across survey respondents about the general business situation in their industry. The heterogeneous estimators considered are OLS and 2SLS at the industry level, as well as the unrestricted SUR estimation method. Fixed effects, random effects, pooled 2SLS and restricted SUR are the homogeneous estimators considered. The panel estimates find that uncertainty has a negative, non-negligible effect on investment, while the heterogeneous estimates vary considerably across industries. Forecast performance for 12 out-of-sample quarters 1996(Q2) to 1999(Q1) are compared. The pooled homogeneous estimators outperform their heterogeneous counterparts in terms of RMSE.

Baltagi, Bresson and Pirotte (2002) reconsidered the two U.S. panel data sets on residential electricity and natural gas demand used by Maddala et al. (1997) and compared the out-of-sample forecast performance of the homogeneous, heterogeneous, and shrinkage estimators. Once again the results show that when the data is used to estimate heterogeneous models across states, individual estimates offer the worst out-of-sample forecasts. Despite the fact that shrinkage estimators outperform these individual estimates, they are outperformed by simple homogeneous panel data estimates in out-of-sample forecasts. Admittedly, these are additional case studies, but they do add to the evidence that simplicity and parsimony in model estimation offered by the homogeneous estimators yield better forecasts than the more parameter consuming heterogeneous estimators.

Hsiao and Tahmiscioglu (1997) use a panel of 561 U.S. firms over the period 1971–92 to study the influence of financial constraints on company investment. They find substantial differences across firms in terms of their investment behavior. When a homogeneous pooled model is assumed, the impact of liquidity on firm investment is seriously underestimated. The authors recommend a mixed fixed and random coefficients framework based on the recursive predictive density criteria.

Pesaran, Smith and Im (1996) investigated the small sample properties of various estimators of the long-run coefficients for a dynamic heterogeneous panel data model using Monte Carlo experiments. Their findings indicate that the mean group estimator performs reasonably well for large T. However, when T is small, the mean group estimator could be seriously biased, particularly when N is large relative to T. Hsiao, Pesaran and Tahmiscioglu (1999) suggest a Bayesian approach for estimating the mean parameters of a dynamic heterogeneous panel data model. The coefficients are assumed to be normally distributed across cross-sectional units and the Bayes estimator is implemented using Markov Chain Monte Carlo methods. Hsiao, Pesaran and Tahmiscioglu (1999) argue that Bayesian methods can be a viable alternative in the estimation of mean coefficients in dynamic panel data models even when the initial observations are treated as fixed constants. They establish the asymptotic equivalence of this Bayes estimator and the mean group estimator proposed by Pesaran and Smith (1995). The asymptotics are carried out for both N and $T \to \infty$ with $\sqrt{N}/T \to 0$. Monte Carlo experiments show that this Bayes estimator has better sampling properties than other estimators for both small and moderate size T. Hsiao, Pesaran and Tahmiscioglu also caution against the use of the mean group estimator unless T is sufficiently large relative to N. The bias in the mean coefficient of the lagged dependent variable appears to be serious when T is small and the true value of this coefficient is larger than 0.6. Hsiao, Pesaran and Tahmiscioglu apply their methods to estimate the q-investment model using a panel of 273 U.S. firms over the period 1972–93. Baltagi, Bresson and Pirotte (2004) reconsider the Tobin q investment model studied by Hsiao, Pesaran and Tahmiscioglu (1999) using a slightly different panel of 337 U.S. firms over the period 1982–1998. They contrast the out-of-sample forecast performance of 9 homogeneous panel data estimators and 11 heterogeneous and shrinkage Bayes estimators over a 5-year horizon. Results show that the average heterogeneous estimators perform the worst in terms of mean squared error, while the hierarchical Bayes estimator suggested by Hsiao, Pesaran and Tahmiscioglu (1999) performs the best. Homogeneous panel estimators and iterative Bayes estimators are a close second. In conclusion, while the performance of various estimators and their corresponding forecasts may vary in ranking from one empirical example to another, the consistent finding in all these studies is that homogeneous panel data estimators perform well in forecast performance mostly due to their simplicity, their parsimonious representation and the stability of the parameter estimates. Average heterogeneous estimators perform badly due to parameter estimate instability caused by the estimation of several parameters with short time series. Shrinkage estimators did well for some applications, especially iterative Bayes and iterative empirical Bayes.

Using data on migration to Germany from 18 source countries over the period 1967–2001, Brucker and Siliverstovs (2006) compare the performance of homogeneous and heterogeneous estimators using out-of-sample forecasts. They find that the mean group estimator performs the worst, while a fixed effects estimator performs the best in RMSE for 5 years and 10 years ahead forecasts. In general, the heterogeneous estimators performed poorly. They attribute this to the unstable regression parameters across the 18 countries, such that the gains from pooling more than offset the biases from the inter-country heterogeneity.

Rapach and Wohar (2004) show that the monetary model of exchange rate determination performs poorly on a country-by-country basis for U.S. dollar exchange rates over the post-Bretton Woods period for 18 industrialized countries for quarterly data over the period 1973:1–1997:1. However, they find considerable support for the monetary model using panel procedures. They reject tests for the homogeneity assumptions inherent in panel procedures. Hence, they are torn between obtaining panel cointegrating coefficient estimates that are much more plausible in economic terms than country-by-country estimates. Yet these estimates might be spurious since they are rejected by formal statistical test for pooling. Rapach and Wohar (2004) perform an out-of-sample forecasting exercise using the panel and country-by-country estimates employing the RMSE criteria for a 1-, 4-, 8-, 12-, and 16-step ahead quarters. For the 1-step and 4-step ahead, the RMSEs of the homogeneous and heterogeneous estimates are similar. At the 8-step ahead horizon, homogeneous estimates generate better forecasts in comparison to five of the six heterogeneous estimates. At the 16-step horizon, the homogeneous estimates have RMSE that is smaller than each of the heterogeneous estimates. In most cases the RMSE is reduced by 20%. They conclude that while there are good reasons to favor the panel estimates over the country-by country estimates of the monetary model, there are also good reasons to be suspicious of these panel estimates since the homogeneity assumption is rejected. Despite this fact, they argue that panel data estimates should not be dismissed based on tests for homogeneity alone, because they may eliminate certain biases that plague country-by-country estimates. In fact, panel estimates of the monetary model were more reliable and generated superior forecasts to those of country-by-country estimates. Rapach and Wohar (2004) suspicion of panel data estimates come from Monte Carlo evidence that show that "it is not improbable to find evidence in support of the monetary model by relying on panel estimates, even when the true data generating process is characterized by a heterogeneous structure that is not consistent with the monetary model". Another paper in this vein is Groen (2005) which utilizes a panel of vector error-correction models based on a common long-run relationship to test whether the Euro exchange rates of Canada, Japan, and the United States have a long-run link with monetary fundamentals. Out of sample forecasts show that this common long-run exchange model is superior to both the naive random walk based forecasts and the standard cointegrated VAR model-based forecasts, especially for horizons of 2–4 years.

Hoogstrate, Palm and Pfann (2000) investigate the improvement of forecasting performance using pooling techniques instead of single country forecasts for N fixed and T large. They use a set of dynamic regression equations with contemporaneously correlated disturbances. When the parameters of the models are different but exhibit some similarity, pooling may lead to a reduction in the mean squared error of the estimates and the forecasts. They show that the superiority of the pooled forecasts in small samples can deteriorate as T grows. They apply these results to growth rates of 18 OECD countries over the period 1950–1991 using an AR(3) model and an AR(3) model with leading indicators. They find that the median MSFE of OLS-based pooled forecasts is smaller than that of OLS-based individual forecasts and that a fairly large T is needed for the latter to outperform the former. They argue that this is due to the reduction in MSE due to imposing a false restriction (pooling). However, for a large enough T, the bias of the pooled estimates increase with out bound and the resulting forecasts based on unrestricted estimates will outperform the forecasts based on the pooled restricted estimates.

Gavin and Theodorou (2005) use forecasting criteria to examine the macrodynamic behavior of 15 OECD countries observed quarterly over the period 1980 to 1996. They utilize a small set of familiar, widely used core economic variables, (output, price level, interest rates, and exchange rates), omitting country-specific shocks. They find that this small set of variables and a simple VAR common model strongly support the hypothesis that many industrialized nations have similar macroeconomic dynamics. In sample, they often reject the hypothesis that coefficient vectors estimated separately for each country are the same. They argue that these rejections may be of little importance if due to idiosyncratic events since macro-time series are typically too short for standard methods to eliminate the effects of idiosyncratic factors. Panel data can be used to exploit the heterogeneous information in cross-country data, hence increasing the data and eliminating the idiosyncratic effects. They compare the forecast accuracy of the individual country models with the common models in a simulated out of sample experiment. They calculate four forecasts with increasing horizons at each point in time-one quarter ahead and four quarters ahead. For the four equations, at every horizon, the panel forecasts are significantly more accurate more often than are the individual country model forecasts. The biggest difference is for the exchange rate and the interest rate. They conclude that the superior out of sample forecasting performance of the common model supports their hypothesis that market economies tend to have a common macrodynamic patterns related to a small number of variables.

10.6 Count Panel Data

Examples of panel data where the dependent variable is a count, include the number of bids on an offshore oil lease by a firm, or the number of visits to a doctor by an individual, or the number of cigarettes smoked per day, or the number of patents filed by an R&D firm, all of which are observed over time. Though one can still treat this count data with a panel regression, the occurrence of zeroes and the discrete non-negative nature of the dependent variable suggest that perhaps a Poisson panel regression model should be used, see Cameron and Trivedi (2015) for a handbook chapter on this subject and Hausman, Hall and Griliches (1984) for popularizing this approach using panel data. The Poisson panel regression is given by

$$\Pr(Y_{it} = y_{it}/x_{it}) = \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{y_{it}!} \text{ where } y_{it} = 0, 1, 2, \ldots;\; i = 1, \ldots, N;\; t = 1, \ldots, T \tag{10.27}$$

with i denoting households, individuals, firms, countries, etc., and t denoting time. The most common specification is the loglinear model $\ln\lambda_{it} = \mu_i + x'_{it}\beta$, where μ_i denotes the *unobservable* individual specific effect. For the *fixed effects* specification for the $\mu'_i s$, the $E(y_{it}/x_{it}) = var(y_{it}/x_{it}) = \lambda_{it} = e^{\mu_i + x'_{it}\beta}$. The marginal effect of a continuous variable x_k is given by $\partial E(y_{it}/x_{it})/\partial x_k = \lambda_{it}\beta_k$. In this case, one can write the likelihood function as the product of the marginals

$$L(\beta, \lambda_{it}) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[-\lambda_{it} + y_{it}(\mu_i + x'_{it}\beta) - \ln y_{it}!\right] \tag{10.28}$$

The first-order conditions of this log likelihood are given by

$$\partial L(\beta, \lambda_{it})/\partial\beta = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \lambda_{it})x_{it} = 0$$

and

$$\partial L(\beta, \lambda_{it})/\partial\mu_i = \sum_{t=1}^{T}(y_{it} - \lambda_{it}) = \sum_{t=1}^{T}(y_{it} - e^{\mu_i} e^{x'_{it}\beta}) = 0$$

for $i = 1, \ldots, N$. Solving for μ_i in terms of β in this model gives

$$\widehat{\mu_i} = \ln\left[\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} e^{x'_{it}\widehat{\beta}}}\right]$$

The conditional Poisson likelihood, which maximizes the joint probability of $(y_{i1}, y_{i2}, \ldots, y_{iT})$ conditioning on their sum, yields a likelihood that is free of μ_i. This is similar to the conditional logit approach in panel data, see Chamberlain (1984) and Chap. 11. In fact, the sum of T independent Poissons each with parameter λ_{it} is a Poisson with parameter $\sum_{t=1}^{T} \lambda_{it}$, i.e.,

$$Pr\left[\sum_{t=1}^{T} y_{it}\right] = \frac{\exp\left(-\sum_{t=1}^{T} \lambda_{it}\right)\left(\sum_{t=1}^{T} \lambda_{it}\right)^{\left(\sum_{t=1}^{T} y_{it}\right)}}{\left(\sum_{t=1}^{T} y_{it}\right)!} \tag{10.29}$$

Since μ_i is a fixed parameter, the joint probability is the product of the marginals

$$Pr\left[y_{i1}, y_{i2}, \ldots, y_{iT}\right] = \frac{\exp\left(-\sum_{t=1}^{T} \lambda_{it}\right)\prod_{t=1}^{T} \lambda_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!} \tag{10.30}$$

Hence, the conditional likelihood is

$$\begin{aligned} Pr\left[y_{i1}, y_{i2}, \ldots, y_{iT} / \sum_{t=1}^{T} y_{it}\right] &= \frac{\exp\left(-\sum_{t=1}^{T} \lambda_{it}\right)\prod_{t=1}^{T} \lambda_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!} \\ &\quad \div \frac{\exp\left(-\sum_{t=1}^{T} \lambda_{it}\right)\left(\sum_{t=1}^{T} \lambda_{it}\right)^{\left(\sum_{t=1}^{T} y_{it}\right)}}{\left(\sum_{t=1}^{T} y_{it}\right)!} \\ &= \frac{\left(\sum_{t=1}^{T} y_{it}\right)!}{\prod_{t=1}^{T} y_{it}!}\prod_{t=1}^{T} p_{it}^{y_{it}} \end{aligned} \tag{10.31}$$

where $p_{it} = \frac{\lambda_{it}}{\sum_{t=1}^{T} \lambda_{it}} = \frac{e^{\mu_i + x'_{it}\beta}}{\sum_{t=1}^{T}(e^{\mu_i + x'_{it}\beta})} = \frac{e^{x'_{it}\beta}}{\sum_{t=1}^{T} e^{x'_{it}\beta}}$. As clear from this conditional likelihood, it is free of μ_i. Also, zero counts in every period do not contribute to this conditional log likelihood.

In contrast, the *random effects* Poisson panel data model has μ_i correlated across periods for the same individual. We will encounter this phenomena when studying random effects probits in Chap. 11. The same idea applies here for the estimation of the maximum likelihood. First, one condition on the random effects and write the joint probability $Pr(y_{i1}, y_{i2}, \ldots, y_{iT}/\mu_i) = \prod_{t=1}^{T} Pr(y_{it}/\mu_i)$, then one integrates out the effect of μ_i, i.e.,

$$Pr(y_{i1}, y_{i2}, \ldots, y_{iT}) = \int Pr(y_{i1}, y_{i2}, \ldots, y_{iT}, \mu_i)d\mu_i = \int Pr(y_{i1}, y_{i2}, \ldots, y_{iT}/\mu_i)g(\mu_i)d\mu_i \tag{10.32}$$

Assuming $Pr(y_{it}/\mu_i)$ is distributed as *Poisson*$(\lambda_{it} = e^{\mu_i + x'_{it}\beta})$ and e^{μ_i} is distributed as *Gamma* with mean 1 and variance θ, the resulting distribution is a *negative binomial* for $\sum_{t=1}^{T} y_{it}$. In fact,

$$Pr(y_{i1}, y_{i2}, \ldots, y_{iT}/\mu_i) = \frac{\exp\left(-\sum_{t=1}^{T}\lambda_{it}\right)\prod_{t=1}^{T}\lambda_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!} \tag{10.33}$$

$$= \frac{\exp\left(-e^{\mu_i}\sum_{t=1}^{T}\gamma_{it}\right)\left(\prod_{t=1}^{T}\gamma_{it}^{y_{it}}\right)(e^{\mu_i})^{\sum_{t=1}^{T} y_{it}}}{\prod_{t=1}^{T} y_{it}!}$$

where $\gamma_{it} = e^{x'_{it}\beta}$. Let $\varepsilon_i = e^{\mu_i}$ be distributed as *Gamma* with mean 1 and variance θ, then

$$g(\varepsilon_i) = \frac{\theta^\theta}{\Gamma(\theta)}\varepsilon_i^{\theta-1}\exp(-\theta\varepsilon_i) \tag{10.34}$$

for $\varepsilon_i > 0$. So that

$$Pr(y_{i1}, y_{i2}, \ldots, y_{iT}) = \frac{\theta^\theta}{\Gamma(\theta)}\frac{\prod_{t=1}^{T}\gamma_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!}\int_0^\infty \exp\left(-\varepsilon_i(\theta + \sum_{t=1}^{T}\gamma_{it})\right)\varepsilon_i^{(\theta+\sum_{t=1}^{T} y_{it})-1} d\varepsilon_i$$

$$= \frac{\theta^\theta}{\Gamma(\theta)}\frac{\prod_{t=1}^{T}\gamma_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!}\frac{\Gamma(\theta + \sum_{t=1}^{T} y_{it})}{(\theta + \sum_{t=1}^{T}\gamma_{it})^{(\theta+\sum_{t=1}^{T} y_{it})}} \tag{10.35}$$

Let $q_i = \theta/(\theta + \sum_{t=1}^{T}\gamma_{it})$, then

$$Pr(y_{i1}, y_{i2}, \ldots, y_{iT}) = \frac{\prod_{t=1}^{T}\gamma_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!}\frac{\Gamma(\theta + \sum_{t=1}^{T} y_{it})}{\Gamma(\theta)(\sum_{t=1}^{T}\gamma_{it})^{\sum_{t=1}^{T} y_{it}}} q_i^\theta(1-q_i)^{\sum_{t=1}^{T} y_{it}} \tag{10.36}$$

which is a *negative binomial*. Stata computes the fixed and random Poisson panel procedures using *xtpoisson*, fe and re. In addition, it computes alternative estimates of the Poisson panel model allowing for a Normal distribution on the random effects using Gauss–Hermite quadrature which will also be used in the estimation of random effects probit, see Chap. 11. A quadcheck should be used to check whether the integrand is well approximated by a polynomial.

The Poisson specification has been criticized in the count data literature, despite its popularity, because it has the property that its mean is equal to its variance. This is known as the *equi-dispersion* property. Empirical work most often reject this specification in favor of *over-dispersion*. To model this over-dispersion, the negative binomial specification is usually employed and this is studied for panel data by Hausman, Hall and Griliches (1984). This can also be implemented using Stata *xtnbreg*, re, and fe. The negative binomial (NB) distribution has mean λ and variance $\lambda + \alpha\lambda^{2-k}$. When $\alpha = 0$, this reverts back to the Poisson model of equi-dispersion. Stata uses $k = 0$ as the *default option*, this is the NB2. In this case the variance is quadratic in the mean, i.e., $\lambda + \alpha\lambda^2$. Another popular specification is $k = 1$, the NB1, which has a variance that is proportional to the mean, i.e., $(1 + \alpha)\lambda$. This can be implemented with Stata with the option *dispersion (constant)*.

Empirical Example: Hausman, Hall and Griliches (1984) studied the relationship between patents and R&D expenditures using panel data. We use their data comprising of 346 U.S. firms observed over the period 1975–1979. The count dependent variable is the number of patents applied for by the firm during the year (that were eventually granted). Some of the explanatory variables included are (i) real R&D spending (in 1972 dollars) and its lagged values; (ii) logarithm of the book value of capital in 1972 as a measure of size of the firm (LOGK); (iii) a dummy for whether the firm is in the scientific sector (SCISECT); (iv) a two-digit code for the applied R&D industrial classification; (v) the sum of patents applied for between 1972–1979. Tables 10.1 and 10.2 show the Stata output for the fixed effects specification using the Poisson and Negative Binomial distributions, while Tables 10.3 and 10.4 give the corresponding random effects specification including variables that are time invariant like LOGK and SCISECT. Both the Poisson and Negative Binomial fixed effects estimates show that only the current R&D spending is significant. For the random effects specification, LOGK is significant, while SCISECT is not. The current R&D spending has a larger effect on patents for the Poisson random effects than for the Poisson fixed effects (0.40 as compared to 0.32). For the Negative Binomial distribution, the difference is much smaller (0.35 as compared to 0.32).

Another problem with the Poisson specification is that it cannot explain the occurrence of *excess zeroes*. For example, health data may contain a large number of individuals that do *not* visit a doctor or smoke zero cigarettes. The *Zero-Inflated Poisson* (ZIP) in its simplest form gives a constant zero-inflation probability q to non-users and (1-q) to users. This is estimated using maximum likelihood methods and is implemented in Stata by the *ZIP* command with the option *inflate (_cons)*. Another option that can be used is Vuong, which gives a non-nested test of *ZIP* versus the Poisson model. This can be generalized to allow the zero-inflated probability q to depend on

Table 10.1 Poisson fixed effects for the R & D data

```
. xtpois PAT LOGR LOGR1 LOGR2 LOGR3 LOGR4 LOGR5 dyear2 dyear3 dyear4 dyear5,fe
note: 22 groups (110 obs) dropped due to all zero outcomes


Conditional fixed-effects Poisson regression    Number of obs      =      1620
Group variable (i): id                          Number of groups   =       324

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                Wald chi2(10)      =    245.39
Log likelihood  = -3536.3086                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         PAT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LOGR |   .3222105   .0459412     7.01   0.000     .2321674    .4122535
       LOGR1 |  -.0871295   .0486887    -1.79   0.074    -.1825576    .0082986
       LOGR2 |   .0785816    .044784     1.75   0.079    -.0091934    .1663567
       LOGR3 |     .00106   .0414151     0.03   0.980    -.0801122    .0822322
       LOGR4 |  -.0046414   .0378489    -0.12   0.902    -.0788238    .0695411
       LOGR5 |   .0026068   .0322596     0.08   0.936    -.0606209    .0658346
      dyear2 |  -.0426076    .013132    -3.24   0.001    -.0683458   -.0168695
      dyear3 |  -.0400462   .0134677    -2.97   0.003    -.0664423     -.01365
      dyear4 |  -.1571185   .0142281   -11.04   0.000    -.1850051   -.1292319
      dyear5 |  -.1980306   .0152946   -12.95   0.000    -.2280074   -.1680538
------------------------------------------------------------------------------
```

Table 10.2 Negative binomial fixed effects for the R & D data

```
. xtnbreg PAT LOGR LOGR1 LOGR2 LOGR3 LOGR4 LOGR5 dyear2 dyear3 dyear4 dyear5,fe
note: 22 groups (110 obs) dropped due to all zero outcomes

Conditional FE negative binomial regression     Number of obs      =      1620
Group variable (i): id                          Number of groups   =       324

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                Wald chi2(10)      =    117.12
Log likelihood  =  -3206.867                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         PAT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LOGR |   .3188563   .0673654     4.73   0.000     .1868225    .4508902
       LOGR1 |   -.080442   .0773311    -1.04   0.298    -.2320082    .0711241
       LOGR2 |   .0559045   .0710929     0.79   0.432    -.0834351    .1952441
       LOGR3 |  -.0128025   .0659707    -0.19   0.846    -.1421028    .1164978
       LOGR4 |   .0355272   .0620031     0.57   0.567    -.0859966    .1570511
       LOGR5 |   .0094533   .0516237     0.18   0.855    -.0917273    .1106338
      dyear2 |  -.0422643   .0249051    -1.70   0.090    -.0910773    .0065488
      dyear3 |  -.0488698   .0253965    -1.92   0.054     -.098646    .0009063
      dyear4 |  -.1606011   .0262724    -6.11   0.000    -.2120941   -.1091081
      dyear5 |  -.2154138   .0265014    -8.13   0.000    -.2673556    -.163472
       _cons |   2.423638   .1749545    13.85   0.000     2.080734    2.766543
------------------------------------------------------------------------------
```

Table 10.3 Poisson random effects for the R & D data

```
. xtpois PAT LOGR LOGR1 LOGR2 LOGR3 LOGR4 LOGR5 dyear2 dyear3 dyear4 dyear5
 LOGK  SCISECT, re

Fitting Poisson model:

Random-effects Poisson regression               Number of obs      =      1730
Group variable (i): id                          Number of groups   =       346

Random effects u_i ~ Gamma                      Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                Wald chi2(12)      =   1272.14
Log likelihood  = -5234.9265                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         PAT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        LOGR |   .4034537   .0435022     9.27   0.000      .318191    .4887165
       LOGR1 |  -.0461765   .0482224    -0.96   0.338    -.1406906    .0483376
       LOGR2 |   .1079235   .0447115     2.41   0.016     .0202905    .1955565
       LOGR3 |   .0297733   .0413235     0.72   0.471    -.0512193     .110766
       LOGR4 |   .0106957   .0377074     0.28   0.777    -.0632094    .0846008
       LOGR5 |   .0406111   .0315738     1.29   0.198    -.0212724    .1024946
      dyear2 |  -.0449624   .0131291    -3.42   0.001     -.070695   -.0192298
      dyear3 |  -.0483864   .0134018    -3.61   0.000    -.0746534   -.0221193
      dyear4 |  -.1741619   .0139702   -12.47   0.000     -.201543   -.1467809
      dyear5 |  -.2258977   .0146645   -15.40   0.000    -.2546396   -.1971557
        LOGK |   .2916932   .0393368     7.42   0.000     .2145945     .368792
     SCISECT |   .2570001   .1122716     2.29   0.022     .0369517    .4770484
       _cons |   .4107881   .1467443     2.80   0.005     .1231746    .6984016
-------------+----------------------------------------------------------------
    /lnalpha |   -.156739   .0809735                     -.3154441    .0019661
-------------+----------------------------------------------------------------
       alpha |   .8549271   .0692264                      .7294648    1.001968
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0: chibar2(01) =  2.5e+04 Prob>=chibar2 = 0.000
```

some explanatory variables. It can also be extended to allow for zero-inflated NB rather than the zero-inflated Poisson model.

Alternative explanations of excess zeroes are that they are generated by a different process altogether. For example, the decision to visit a doctor may depend on the individual, while the frequency of visits depends on the doctor, once initial contact is made. Let P_1 define the participation decision, which is usually a binary decision, modeled with a logit or probit specification, while P_2 defines the process generating the positive counts, a truncated at zero count model usually a Poisson or NB. This double hurdle model for count data can be estimated separately using maximum likelihood methods and can be easily implemented in Stata using logit/probit and *ztnb*.

Table 10.4 Negative binomial random effects for the R & D data

```
. xtnbreg PAT LOGR LOGR1 LOGR2 LOGR3 LOGR4 LOGR5 dyear2 dyear3 dyear4 dyear5
 LOGK  SCISECT, re

Fitting negative binomial (constant dispersion) model:

Random-effects negative binomial regression     Number of obs      =      1730
Group variable (i): id                          Number of groups   =       346

Random effects u_i ~ Beta                       Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                Wald chi2(12)      =    944.21
Log likelihood  = -4948.4944                    Prob > chi2        =    0.0000
```

PAT	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LOGR	.3503119	.0652818	5.37	0.000	.2223619	.4782619
LOGR1	-.0030317	.0750916	-0.04	0.968	-.1502085	.1441452
LOGR2	.1049876	.0688488	1.52	0.127	-.0299537	.2399289
LOGR3	.0163523	.0636376	0.26	0.797	-.1083752	.1410797
LOGR4	.0359425	.0587161	0.61	0.540	-.0791389	.1510239
LOGR5	.0718323	.0482887	1.49	0.137	-.0228119	.1664764
dyear2	-.0436736	.0213435	-2.05	0.041	-.085506	-.0018411
dyear3	-.0556597	.0218572	-2.55	0.011	-.098499	-.0128203
dyear4	-.1831055	.0227183	-8.06	0.000	-.2276326	-.1385784
dyear5	-.2300438	.0231525	-9.94	0.000	-.2754219	-.1846658
LOGK	.161937	.0417874	3.88	0.000	.0800351	.2438388
SCISECT	.1176419	.1066164	1.10	0.270	-.0913224	.3266063
_cons	.8995618	.1681113	5.35	0.000	.5700698	1.229054
/ln_r	.9877591	.0961426			.7993231	1.176195
/ln_s	.7009608	.1079684			.4893467	.9125748
r	2.68521	.2581631			2.224035	3.242015
s	2.015688	.2176306			1.63125	2.490728

For count panel data estimation of dynamic models that include correlated individual effects and predetermined variables, see Blundell, Griffith and Windmeijer (2002).

10.7 Notes

1. In general, for any T, as long as the fraction of the sample being rotated is greater or equal to half, then no individual will be observed more than twice.
2. The terms "panel conditioning," "reinterview effect" and "rotation group bias" are also used in the literature synonymously with "time-in-sample bias" effects.
3. Moffitt (1993) explains that many researchers prefer to use pseudo-panels like the CPS because it has larger, more representative samples and the questions asked are more consistently defined over time than the available U.S. panels.

4. Table 10.1 of McKenzie (2001) provides examples of unequally spaced surveys and their sources. See also Table 10.1 of Millimet and McDonough (2017).

10.8 Problems

10.1 *Measurement error and panel data*. This problem is based upon Griliches and Hausman (1986). Using the simple regression given in (10.1)–(10.3):

(a) Show that for the first-difference (FD) estimator of β, the expression in (10.8) reduces to

$$\text{plim}\widehat{\beta}_{FD} = \beta\left(1 - \frac{2\sigma_\eta^2}{\text{var}(\Delta x)}\right)$$

where $\Delta x_{it} = x_{it} - x_{i,t-1}$.

(b) Also show that (10.8) reduces to

$$\text{plim}\widetilde{\beta}_W = \beta\left(1 - \frac{T-1}{T}\frac{\sigma_\eta^2}{\text{var}(\widetilde{x})}\right)$$

where $\widetilde{\beta}_W$ denotes the Within estimator and $\widetilde{x}_{it} = x_{it} - \bar{x}_i$.

(c) For most economic series, the x_{it}^* are positively serially correlated exhibiting a declining correlogram, with

$$\text{var}(\Delta x) < \frac{2T}{T-1}\text{var}(\widetilde{x}) \quad \text{for } T > 2$$

Using this result, conclude that

$$|\ \text{bias}\widehat{\beta}_{FD}\ | > |\ \text{bias}\widetilde{\beta}_W\ |$$

(d) Solve the expressions in parts (a) and (b) for β and σ_η^2 and verify that the expressions in (10.9) and (10.10) reduce to

$$\widehat{\beta} = \frac{[2\widetilde{\beta}_W/\text{var}(\Delta x) - (T-1)\widehat{\beta}_{FD}/T\ \text{var}(\widetilde{x})]}{[2/\text{var}(\Delta x) - (T-1)/T\ \text{var}(\widetilde{x})]}$$

$$\sigma_\eta^2 = (\widehat{\beta} - \widehat{\beta}_{FD})\ \text{var}(\Delta x)/2\widehat{\beta}$$

(e) For $T = 2$, the Within estimator is the same as the first-difference estimator since $\frac{1}{2}\Delta x_{it} = \widetilde{x}_{it}$. Verify that the expressions in part (a) and (b) are also the same.

10.2 *Rotating panel with three waves*. For the rotating panel considered in Sect. 10.2, assume that $T = 3$ and that the number of households being replaced each period is equal to $N/2$.

(a) Derive the variance–covariance of the disturbances Ω.

(b) Derive $\Omega^{-1/2}$ and describe the transformation needed to make GLS a weighted least squares regression.
(c) How would you consistently estimate the variance components σ_μ^2 and σ_ν^2?
(d) Repeat this exercise for the case where the number of households being replaced each period is $N/3$. How about $2N/3$?

10.3 *Residential natural gas and electricity*. Download the Maddala et al. (1997) data set on residential natural gas and electricity consumption for 49 states over 21 years (1970-90) from the *Journal of Business and Economic Statistics* web site www.amstat.org/publications/jbes/ftp.html.

(a) Using this data set, replicate the individual OLS state regressions for electricity, given in Table 6, and natural gas, given in Table 8 of Maddala et al. (1997).
(b) Replicate the shrinkage estimates for electricity and natural gas given in Tables 7 and 9 of Maddala et al. (1997).
(c) Replicate the fixed effects estimator given in column 1 of Table 10.2 of Maddala et al. (1997), and the pooled OLS model given in column 2 of that table.
(d) Replicate the average OLS, the average shrinkage estimator and the average Stein-Rule estimator in Table 10.2 of Maddala et al. (1997).
(e) Redo parts (c) and (d) for the natural gas equation as given in Table 10.4 of Maddala et al. (1997).

10.4 *Patents and R&D expenditures*. Download the Hausman, Hall and Griliches (1984) panel data on patents and R&D expenditures using 346 U.S. firms observed over the period 1975–1979. See the empirical example in Sect. 10.6.

(a) Replicate Tables 10.1 and 10.2 for the fixed effects specification using the Poisson and Negative Binomial distributions.
(b) Replicate Tables 10.3 and 10.4 for the random effects specification (including variables that are time invariant like LOGK and SCISECT) using the Poisson and Negative Binomial distributions.

10.5 *Doctor's visits.* Winkelmann (2004) fits a Poisson model to explain the number of doctor's visits using panel data drawn from the GSOEP from 1995–1999. The explanatory variables include, age, age-squared, dummy for male, years of education, dummy for married, household size, active sports, good health, bad health, whether on social assistance, log(income), yearly dummies, whether self-employed, full-time, part-time, or unemployed, and quarterly dummies. The data set can be downloaded from the *Journal of Applied Econometrics* archive web site: (http://qed.econ.queensu.ca/jae/).

(a) Run the Poisson random and fixed effects regressions given in Table II on page 466 of Winkelmann (2004). How does being married or being on welfare affect the number of doctor's visits?
(b) Run the Negative Binomial random and fixed effects regressions. How does log(income) affect the number of doctor's visits?

10.6 *Hospital visits.* Geil et al. (1997) fit a negative binomial random effects model to explain the number of hospital visits using panel data on 5180 individuals drawn from 8 waves of the GSOEP from 1984–1994. The 1990, 1991, and 1993 waves were excluded because they did not provide information on hospitalization. The individuals were between the ages of 25 to 64 and excluded children, students, and retired people. This is an unbalanced panel of 30,590 observations. The explanatory variables are described in Table I of Geil et al. (1997, p. 300). These include, age, age-squared, age-cubed, dummy for male, dummy for private insurance, dummy for private insurance with copayment obligation, dummy for public insurance, dummy for voluntary public insurance, dummy for family public insurance, dummy for public insurance company legally obliged to accept all risks, dummy for public insurance with voluntary additional coverage through a private scheme, dummy for chronic conditions, dummy for handicapped, monthly net income, dummy for living outside city center, dummy for married, dummy for at least secondary education, dummy for university or technical college, dummy for passing vocational training, dummy for working in a health-related field, dummy for being in the labor force, dummy for blue collar, dummy for white collar, dummy for civil servant, dummy for self-employed, dummy for part-time, dummy for a non-German from Western countries, dummy for other non-German nationals, dummy for children below the age of 16 in the household. The data set can be downloaded from the *Journal of Applied Econometrics* archive web site: (http://qed.econ.queensu.ca/jae/).

(a) Replicate the descriptive statistics given in Table II of Geil et al. (1997, p. 301)?
(b) Run the Negative Binomial random effects regressions for males and females given in Table III of Geil et al. (1997, p. 305). Compare the estimates and their significance for males versus females with regard to their hospital visits in Germany?

10.7 *Matched panels: smoking and birthweight*. Abrevaya (2006) estimates the effect of smoking on birth outcomes from panel data (i.e., data on mothers with multiple births). Panel data allows the identification of the smoking effect from women who change their smoking behavior from one pregnancy to another. The data set contains 296,218 birth observations with 141,929 distinct mothers (identified by momid3 and idx, an index number of a mother's birth). The data set can be downloaded from the *Journal of Applied Econometrics* archive web site:

(http://qed.econ.queensu.ca/jae/2006-v21.4/abrevaya/). Birthweight (in grams) is regressed on (i)whether the mother smokes; (ii) the number of cigarettes smoked per day; (iii) the baby's gender; (iv) mother's age and age-squared; (v) whether she is a high-school graduate, had some-college, or is a college-graduate; (vi) her race and marital status; (vii) adeqcode2 and adeqcode3, which are indicators that the Kessner index = 2 or 3. This measures the adequacy of prenatal care, 2 being intermediate and 3 being inadequate; (viii) a dummy variable for no prenatal visits; (ix) petri2 and petri3, which are indicators that the first prenatal visit occurred in the 2nd or 3rd trimester. (a) Run the OLS estimates to replicate column 5 of Table IV of Abrevaya (2006, p.502). This only includes the dummy variable for smoking but not the number of cigarettes smoked. Be sure to include the dummies for the number of live births, mother's state of residence, and mother's year of birth. (b) Run the corresponding FE estimates (including the women's fixed effect) and thus wiping out the time-invariant variables. This should replicate the FE estimates in column 6 of Table IV of Abrevaya (2006, p.502). (c) Add the number of cigarettes smoked and replicate columns 11 and 12 of Table IV of Abrevaya (2006, p.502). (d) comparing the FE and OLS estimates, what do you conclude? (e) To gauge the degree of incorrect matching and its effect on the estimates reported above, Abrevaya utilized a proxy for correct matches. This dummy variable proxy takes the value 1 if the observed interval since last birth agrees with the record. A value of zero for proxy is extremely strong evidence of an incorrect match since only miscoding of the interval record or birth month could result in proxy = 0 for a correct match. Show that the fixed effects estimates for this more reliable sample yield a reduction in birth weight of 67 g for smokers which is much smaller than the overall fixed effects estimates based upon the full samples (144 g).

10.8 *European Patents*. Cincera (1997) performed count panel data regressions of patent activity using 181 international manufacturing firms investing substantial amounts in R&D over the period 1983 to 1991. The dependent variable was the number of European patent applications filed by the firm in that year. This was regressed on current and 4 year lag of log of R&D expenditures as well as log of technological spillovers. Other control variables included the firm's geographical area as well as the firm's main industry sector. The data set can be downloaded from the *Journal of Applied Econometrics* archive web site.

(a) Replicate the descriptive statistics and correlation matrix given in Table I of Cincera (1997, p. 273).
(b) Run the poisson and negative binomial regressions using robust variance–covariance matrices as described in equation (1) of Cincera (1997, p. 267).

(c) Run *xtpoisson* and *xtnbreg* using both the fe and re options on this data. This should replicate the estimates (but not the standard errors), reported in column (3) of Table II of Cincera (1997, p. 275) which is entitled conditional poisson.

References

Abrevaya, J. 2006. Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics* 21: 489–519.

Anderson, T.W., and C. Hsiao. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47–82.

Arellano, M. 1989. A note on the Anderson-Hsiao estimator for panel data. *Economics Letters* 31: 337–341.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2002. Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption. *Economics Letters* 76: 375–382.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2004. Tobin q: Forecast performance for hierarchical Bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics* 29: 107–113.

Baltagi, B.H., G. Bresson, J.M. Griffin, and A. Pirotte. 2003. Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption. *Empirical Economics* 28: 795–811.

Baltagi, B.H., and J.M. Griffin. 1984. Short and long run effects in pooled models. *International Economic Review* 25: 631–645.

Baltagi, B.H., and J.M. Griffin. 1997. Pooled estimators versus their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 77: 303–327.

Baltagi, B.H., J.M. Griffin, and W. Xiong. 2000. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82: 117–126.

Biorn, E. 1981. Estimating economic relations from incomplete cross-section/time-series data. *Journal of Econometrics* 16: 221–236.

Biorn, E. 1996. Panel data with measurement errors, Chapter 10. In *The econometrics of panel data: A handbook of the theory with applications*, (eds.) L. Mátyás, and P. Sevestre, 236–279. Dordrecht: Kluwer Academic Publishers.

Biorn, E. 2000. Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews* 19: 391–424.

Biorn, E., and E.S. Jansen. 1983. Individual effects in a system of demand functions. *Scandinavian Journal of Economics* 85: 461–483.

Blundell, R., C. Meghir, and P. Neves. 1993. Labour supply and intertemporal substitutions. *Journal of Econometrics* 59: 137–160.

Blundell, R., R. Griffith, and F. Windmeijer. 2002. Individual effects and dynamics in count data models. *Journal of Econometrics* 108: 113–131.

Bound, L., C. Brown, G.J. Duncan, and W.L. Rodgers. 1990. Measurement error in cross-sectional and longitudinal labor market surveys: Validation study evidence. In *Panel Data and labor market studies*, (eds.) J. Hartog, G. Ridder, and T. Theeuwes, 1–19. Amsterdam: North-Holland.

Browning, M., A. Deaton, and M. Irish. 1985. A profitable approach to labor supply and commodity demands over the life cycle. *Econometrica* 53: 503–543.

Brucker, H., and B. Siliverstovs. 2006. On the estimation and forecasting of international migration: How relevant is heterogeneity across countries. *Empirical Economics* 31: 735–754.

Cameron, C., and P. Trivedi. 2015. Count panel data, Chapter 8. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 233–256. Oxford: Oxford University Press.

Chamberlain, G. 1984. Panel data, Chapter 22. In *Handbook of econometrics*, (eds.) Z. Griliches, and M. Intrilligator, 1247–1318. Amsterdam: North-Holland.

Cincera, M. 1997. Patents, R&D and technological spillovers at the firm level: Some evidence from econometric count models for panel data. *Journal of Applied Econometrics* 12: 265–280.

Collado, M.D. 1997. Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics* 82: 37–62.

Deaton, A. 1985. Panel data from time series of cross-sections. *Journal of Econometrics* 30: 109–126.

Driver, C., K. Imai, P. Temple, and A. Urga. 2004. The effect of uncertainty on UK investment authorisation: Homogeneous vs. heterogeneous estimators. *Empirical Economics* 29: 115–128.

Duncan, G.J., and D.H. Hill. 1985. An investigation of the extent and consequences of measurement error in labor economic survey data. *Journal of Labor Economics* 3: 508–532.

Eberhardt, M. 2012. Estimating panel time series models with heterogeneous slopes. *Stata Journal* 12: 61–71.

Egger, P., and M. Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19: 227–246.

Gavin, W.T., and A.T. Theodorou. 2005. A common model approach to macroeconomics: Using panel data to reduce sampling error. *Journal of Forecasting* 24: 203–219.

Geil, P., A. Million, R. Rotte, and K.F. Zimmermann. 1997. Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics* 12: 295–311.

Girma, S. 2000. A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections. *Journal of Econometrics* 98: 365–383.

Griliches, Z., and J.A. Hausman. 1986. Errors in variables in panel data. *Journal of Econometrics* 31: 93–118.

Groen, J.J.J. 2005. Exchange rate predictability and monetary fundamentals in a small multi-country panel. *Journal of Money, Credit, and Banking* 37: 495–516.

Hausman, J.A., B.H. Hall, and Z. Griliches. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52: 909–938.

Hoogstrate, A.J., F.C. Palm, and G.A. Pfann. 2000. Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business and Economic Statistics* 18: 274–283.

Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu. 1999. Bayes estimation of short run coefficients in dynamic panel data models, Chapter 11. In *Analysis of panels and limited dependent variable models*, (eds.) C. Hsiao, K. Lahiri, L.F. Lee, and M.H. Pesaran, 268–296. Cambridge: Cambridge University Press.

Hsiao, C., and A.K. Tahmiscioglu. 1997. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92: 455–465.

Inoue, A. 2008. Efficient estimation and inference in linear pseudo-panel data models. *Journal of Econometrics* 142: 449–466.

Maddala, G.S., R.P. Trost, H. Li, and F. Joutz. 1997. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics* 15: 90–100.

Mairesse, J. 1990. Time-series and cross-sectional estimates on panel data: Why are they different and why should they be equal? In *Panel data and labor market studies*, (eds.) J. Hartog, G. Ridder, and J. Theeuwes, 81–95. Amsterdam: North-Holland.

McKenzie, D.J. 2001. Estimation of AR(1) models with unequally spaced pseudo-panels. *Econometrics Journal* 4: 89–108.

McKenzie, D.J. 2004. Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics* 120: 235–262.

Meijer, E., L. Spierijk, and T. Wansbeek. 2015. Measurement error in panel data, Chapter 11. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 325–362. Oxford: Oxford University Press.

Millimet, D.L., and I.K. McDonough. 2017. Dynamic panel data models with irregular spacing: With an application to early childhood development. *Journal of Applied Econometrics* 32: 725–743.

Moffitt, R. 1993. Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics* 59: 99–123.

Mundlak, Y. 1978. On the pooling of time series and cross-section data. *Econometrica* 46: 69–85.

Pesaran, M.H., and R. Smith. 1995. Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics* 68: 79–113.

Pesaran, M.H., R. Smith, and K.S. Im. 1996. Dynamic linear models for heterogenous panels, Chapter 8. In *The Econometrics of panel data: A handbook of the theory with applications*, (eds.) L. Mátyás, and P. Sevestre, 145–195. Dordrecht: Kluwer Academic Publishers.

Phillips, P.C.B., and H. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67: 1057–1111.

Rapach, D.E., and M.E. Wohar. 2004. Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23: 867–895.

Robertson, D., and J. Symons. 1992. Some strange properties of panel data estimators. *Journal of Applied Econometrics* 7: 175–189.

Verbeek, M. 2008. Pseudo-panels and repeated cross-sections, Chapter 11. In *The econometrics of panel data: fundamentals and recent developments in theory and practice*, (eds.) L. Mátyás, and P. Sevestre, 369–383. Berlin: Springer.

Verbeek, M., and T.E. Nijman. 1993. Minimum MSE estimation of a regression model with fixed effects and a series of cross-sections. *Journal of Econometrics* 59: 125–136.

Verbeek, M., and F. Vella. 2005. Estimating dynamic models from repeated cross-sections. *Journal of Econometrics* 127: 83–102.

Wansbeek, T.J., and R.H. Koning. 1991. Measurement error and panel data. *Statistica Neerlandica* 45: 85–92.

Winkelmann, R. 2004. Health care reform and the number of doctor visits-an econometric analysis. *Journal of Applied Econometrics* 19: 455–472.

11 Limited Dependent Variables and Panel Data

In many economic studies, the dependent variable is discrete, indicating, for example, that a household purchased a car or that an individual is unemployed or that he or she joined a labor union or defaulted on a loan or was denied credit. This dependent variable is usually represented by a binary choice variable $y_{it} = 1$ if the event happens and 0 if it does not for individual i at time t. In fact, if p_{it} is the probability that individual i participated in the labor force at time t, then $E(y_{it}) = 1.\ p_{it} + 0.\ (1 - p_{it}) = p_{it}$, and this is usually modeled as a function of some explanatory variables

$$p_{it} = Pr[y_{it} = 1] = E(y_{it}/x_{it}) = F(x_{it}'\beta) \tag{11.1}$$

For the linear probability model, $F(x_{it}'\beta) = x_{it}'\beta$ and the usual panel data methods apply except that $\widehat{y}_{it}$ is not guaranteed to lie in the unit interval. The standard solution has been to use the logistic or normal cumulative distribution functions that constrain $F(x_{it}'\beta)$ to be between zero and one. These probability functions are known in the literature as *logit* and *probit*, corresponding to the logistic and normal distributions, respectively.[1] For example, a worker participates in the labor force if his offered wage exceeds his unobserved reservation wage. This threshold can be described as

$$\begin{aligned} y_{it} &= 1 \ \text{ if } y_{it}^* > 0 \\ &= 0 \ \text{ if } y_{it}^* \leqslant 0 \end{aligned} \tag{11.2}$$

where $y_{it}^* = x_{it}'\beta + u_{it}$. So that

$$\Pr[y_{it} = 1] = \Pr[y_{it}^* > 0] = \Pr[u_{it} > -x_{it}'\beta] = F(x_{it}'\beta) \tag{11.3}$$

where the last equality holds as long as the density function describing F is symmetric around zero. This is true for the logistic and normal density functions.

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_11

11.1 Fixed and Random Logit and Probit Models

For panel data, the presence of individual effects complicates matters significantly. To see this, consider the fixed effects panel data model, $y_{it}^* = x_{it}'\beta + \mu_i + \nu_{it}$ with

$$\Pr[y_{it} = 1] = \Pr[y_{it}^* > 0] = \Pr[\nu_{it} > -x_{it}'\beta - \mu_i] = F(x_{it}'\beta + \mu_i) \qquad (11.4)$$

where the last equality holds as long as the density function describing F is symmetric around zero. In this case, μ_i and β are unknown parameters and as $N \to \infty$, for a fixed T, the number of parameters μ_i increases with N. This means that μ_i cannot be consistently estimated for a fixed T. This is known as the *incidental parameters problem* in statistics. For the linear panel data regression model, when T is fixed, only β was estimated consistently by first getting rid of the μ_i using the Within transformation.[2] This was possible for the linear case because the MLE of β and μ_i are asymptotically independent (see Hsiao, 2003). This is no longer the case for a qualitative limited dependent variable model with fixed T as demonstrated by Chamberlain (1980). For a simple illustration of how the inconsistency of the MLE of μ_i is transmitted into inconsistency for $\widehat{\beta}_{mle}$, see Hsiao (2003). This is done in the context of a logit model with one regressor x_{it} that is observed over two periods, with $x_{i1} = 0$ and $x_{i2} = 1$. Hsiao shows that as $N \to \infty$ with $T = 2$, plim$\widehat{\beta}_{mle} = 2\beta$, see also problem 11.4. Greene (2004a) shows that despite the large number of incidental parameters, one can still perform maximum likelihood for the fixed effects model by brute force, i.e., including a large number of dummy variables. Using Monte Carlo experiments, he shows that the fixed effects MLE is biased even when T is large. For $N = 1000$, $T = 2$ and 200 replications, this bias is 100% confirming the results derived by Hsiao (2003). However, this bias improves as T increases. For example, when $N = 1000$ and $T = 10$ this bias is 16% and when $N = 1000$ and $T = 20$ this bias is 6.9%.

The usual solution around this incidental parameters problem is to find a minimal sufficient statistic for μ_i. For the logit model, Chamberlain (1980) finds that $\sum_{t=1}^T y_{it}$ is a minimum sufficient statistic for μ_i. Therefore, Chamberlain suggests maximizing the *conditional* likelihood function

$$L_c = \prod_{i=1}^N \Pr\left(y_{i1}, \ldots, y_{iT} / \sum_{t=1}^T y_{it}\right) \qquad (11.5)$$

to obtain the *conditional logit* estimates for β. By definition of a sufficient statistic, the distribution of the data given this sufficient statistic will not depend on μ_i. For the fixed effects logit model, this approach results in a computationally convenient estimator and the basic idea can be illustrated for $T = 2$. The observations over the two periods and for all individuals are independent and the *unconditional* likelihood is given by

$$L = \prod_{i=1}^N \Pr(y_{i1})\Pr(y_{i2}) \qquad (11.6)$$

The sum $(y_{i1} + y_{i2})$ can be 0, 1 or 2. If it is 0, both y_{i1} and y_{i2} are 0 and

$$\Pr[y_{i1} = 0, y_{i2} = 0/y_{i1} + y_{i2} = 0] = 1 \tag{11.7}$$

Similarly, if the sum is 2 both y_{i1} and y_{i2} are 1 and

$$\Pr[y_{i1} = 1, y_{i2} = 1/y_{i1} + y_{i2} = 2] = 1 \tag{11.8}$$

These terms add nothing to the conditional log likelihood since log1= 0. Only the observations for which $y_{i1} + y_{i2} = 1$ matter in $\log L_c$ and these are given by

$$\Pr[y_{i1} = 0, y_{i2} = 1/y_{i1} + y_{i2} = 1] \quad \text{and} \Pr[y_{i1} = 1, y_{i2} = 0/y_{i1} + y_{i2} = 1]$$

The latter can be calculated as $\Pr[y_{i1} = 1, y_{i2} = 0]/\Pr[y_{i1} + y_{i2} = 1]$ with

$$\Pr[y_{i1} + y_{i2} = 1] = \Pr[y_{i1} = 0, y_{i2} = 1] + \Pr[y_{i1} = 1, y_{i2} = 0]$$

since the latter two events are mutually exclusive. From (11.4), the logit model yields

$$\Pr[y_{it} = 1] = \frac{e^{\mu_i + x'_{it}\beta}}{1 + e^{\mu_i + x'_{it}\beta}} \tag{11.9}$$

and

$$\Pr[y_{it} = 0] = 1 - \frac{e^{\mu_i + x'_{it}\beta}}{1 + e^{\mu_i + x'_{it}\beta}} = \frac{1}{1 + e^{\mu_i + x'_{it}\beta}}$$

Therefore

$$\Pr[y_{i1} = 1, y_{i2} = 0] = \frac{e^{\mu_i + x'_{i1}\beta}}{1 + e^{\mu_i + x'_{i1}\beta}} \frac{1}{1 + e^{\mu_i + x'_{i2}\beta}}$$

and

$$\Pr[y_{i1} = 0, y_{i2} = 1] = \frac{1}{1 + e^{\mu_i + x'_{i1}\beta}} \frac{e^{\mu_i + x'_{i2}\beta}}{1 + e^{\mu_i + x'_{i2}\beta}}$$

with

$$\Pr[y_{i1} + y_{i2} = 1] = \Pr[y_{i1} = 1, y_{i2} = 0] + \Pr[y_{i1} = 0, y_{i2} = 1]$$
$$= \frac{e^{\mu_i + x'_{i1}\beta} + e^{\mu_i + x'_{i2}\beta}}{(1 + e^{\mu_i + x'_{i1}\beta})(1 + e^{\mu_i + x'_{i2}\beta})}$$

Therefore,

$$\Pr[y_{i1} = 1, y_{i2} = 0/y_{i1} + y_{i2} = 1] = \frac{\Pr[y_{i1} = 1, y_{i2} = 0]}{\Pr[y_{i1} + y_{i2} = 1]} \tag{11.10}$$

$$= \frac{e^{\mu_i + x'_{i1}\beta}}{e^{\mu_i + x'_{i1}\beta} + e^{\mu_i + x'_{i2}\beta}} = \frac{e^{x'_{i1}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} = \frac{1}{1 + e^{(x_{i2} - x_{i1})'\beta}}$$

Similarly

$$\Pr[y_{i1} = 0, y_{i2} = 1/y_{i1} + y_{i2} = 1] = \frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} = \frac{e^{(x_{i2}-x_{i1})'\beta}}{1 + e^{(x_{i2}-x_{i1})'\beta}} \quad (11.11)$$

and neither probability involves the μ_i. Therefore, by conditioning on $y_{i1} + y_{i2}$, we swept away the μ_i. The product of terms such as these with $y_{i1} + y_{i2} = 1$ give the conditional likelihood function which can be maximized with respect to β using conventional maximum likelihood logit programs. In this case, only the observations for individuals who *switched* status are used in the estimation. A standard logit package can be used with $x'_{i2} - x'_{i1}$ as explanatory variables and the dependent variable taking the value one if y_{it} switches from 0 to 1, and zero if y_{it} switches from 1 to 0. This procedure can be easily generalized for $T > 2$ (see problem 11.2).[3]

In order to test for fixed individual effects one can perform a Hausman-type test based on the difference between Chamberlain's conditional MLE and the usual logit MLE ignoring the individual effects. The latter estimator is consistent and efficient only under the null of no individual effects and inconsistent under the alternative. Chamberlain's estimator is consistent whether H_0 is true or not, but it is inefficient under H_0 because it may not use all the data. Both estimators can be easily obtained from the usual logit ML routines. The constant is dropped and estimates of the asymptotic variances are used to form Hausman's χ^2 statistic. This will be distributed as χ^2_K under H_0. For an application of Chamberlain's conditional MLE see Winkelmann and Winkelmann (1998) who applied the conditional logit approach to study the effect of unemployment on the level of satisfaction. Using data from the first six waves of the GSOEP over the period 1984-89, the authors showed that unemployment had a large detrimental effect on satisfaction. This effect became even larger after controlling for individual specific effects. The dependent variable was based on the response to the question "How satisfied are you at present with your life as a whole?" An ordinal scale from 0 to 10 is recorded, where 0 meant "completely dissatisfied" and 10 meant "completely satisfied". Winkelmann and Winkelmann constructed a binary variable taking the value 1 if this score was above 7 and 0 otherwise. They justified this on the basis that average satisfaction was between 7 and 8 and this was equivalent to classifying individuals into those who reported above and those who reported below average satisfaction. The explanatory variables included a set of dummy variables indicating current labor market status (unemployed out of the labor force) with employed as the reference category. A good health variable defined as the absence of any chronic condition or handicap. Age, age-squared, marital status, and the duration of unemployment and its square. Since unemployment reduces income which in turn may reduce satisfaction, household income was included as a control variable to measure the non-pecuniary effect of unemployment holding income constant. Of particular concern with the measurement of life satisfaction is that individuals "anchor" their scale at different levels, rendering interpersonal comparisons of responses meaningless. This problem bears a close resemblance to the issue of cardinal versus ordinal utility. Any statistic that is calculated from a cross-section of individuals, for instance, an average satisfaction, requires cardinality of the measurement scale. This problem is closely related to the unobserved individual specific effects. Hence anchoring causes the estimator to be biased as long as it is not random but correlated with the explanatory variables. Panel data help if the metric used by individuals is time invariant. Fixed effects make inference based on intra- rather than interpersonal comparisons of satisfaction. This avoids not only the potential bias caused by anchoring, but also bias caused by other unobserved individual specific factors. Hausman's test based on the difference between a standard logit and a fixed effects logit yielded a significant χ^2 variable. After controlling for individual specific effects, this study found that unemployment had a significant and substantial negative impact on satisfaction. The non-pecuniary costs of unemployment by far exceeded the pecuniary costs associated with loss of income while unemployed.

In contrast to the fixed effects logit model, the conditional likelihood approach does not yield computational simplifications for the fixed effects *probit* model. This is why there is no *xtprobit* with a fe option in Stata. But the probit specification has been popular for the *random effects* model. In this case, $u_{it} = \mu_i + \nu_{it}$ where $\mu_i \sim \text{IIN}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Since $E(u_{it}u_{is}) = \sigma_\mu^2$ for $t \neq s$, the joint likelihood of $(y_{1t}, \ldots, y_{Nt})$ can no longer be written as the product of the marginal likelihoods of the y_{it}. This complicates the derivation of maximum likelihood which will now involve T-dimensional integrals.[4] The likelihood function is a multiple integral

$$L_i = Pr[y_{i1}, y_{i2}, .., y_{iT}/X] = \int .. \int f(u_{i1}, u_{i2}, .., u_{iT})du_{i1}du_{i2}..du_{iT} \tag{11.12}$$

which is maximized w.r.t. β and σ_μ. This gets to be infeasible if T is big. The trick is to write the joint density function as a product of the conditional density and the marginal density of μ_i. In fact,

$$f(u_{i1}, u_{i2}, .., u_{iT}, \mu_i) = f_1(u_{i1}, u_{i2}, .., u_{iT}/\mu_i)f_2(\mu_i)$$

so that

$$f(u_{i1}, u_{i2}, .., u_{iT}) = \int f_1(u_{i1}, u_{i2}, .., u_{iT}/\mu_i)f_2(\mu_i)d\mu_i$$

By conditioning on the individual effects, this T-dimensional integral problem reduces to a single integral. To see this, note that the $u_i's$ conditional on μ_i are independent, so

$$f(u_{i1}, u_{i2}, .., u_{iT}) = \int \prod_{t=1}^{T} f_1(u_{it}/\mu_i)f_2(\mu_i)d\mu_i$$

Inserting this in the likelihood, one gets

$$L_i = Pr[y_{i1}, y_{i2}, .., y_{iT}/X] = \int .. \int \int \prod_{t=1}^{T} f_1(u_{it}/\mu_i) f_2(\mu_i) d\mu_i du_{i1} du_{i2}..du_{iT}$$

The ranges of integration are independent, so we interchange the order of integration

$$L_i = Pr[y_{i1}, y_{i2}, .., y_{iT}/X] = \int \left[\int .. \int \prod_{t=1}^{T} f_1(u_{it}/\mu_i) du_{i1} du_{i2}..du_{iT} \right] f_2(\mu_i) d\mu_i$$

The terms in square brackets are the product of individual probabilities

$$L_i = Pr[y_{i1}, y_{i2}, .., y_{iT}/X] = \int \left[\prod_{t=1}^{T} (\int f_1(u_{it}/\mu_i) du_{it} \right] f_2(\mu_i) d\mu_i \quad (11.13)$$

For the probit, the individual probabilities inside the product are given by $\Phi(q_{it}(x'_{it}\beta + \mu_i))$, where $q_{it} = 2y_{it} - 1$. The payoff is that this likelihood involves only one integral. The inner integrals are standard normal CDF. This can be evaluated using the Gaussian–Hermite quadrature procedure suggested by Butler and Moffitt (1982):

$$lnL_h = \sum_{i=1}^{N} [ln\{\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} \prod_{t=1}^{T} w_h \Phi(q_{it}(x'_{it}\beta + \theta z_h))\}]$$

where H is the number of points for the quadrature, and w_h and z_h are the weights and nodes of the quadrature. Here $\theta = \sigma_\mu \sqrt{2}$, so an estimate of σ_μ can be obtained by dividing the estimate of θ by $\sqrt{2}$. This approach has the advantage of being computationally feasible even for fairly large T. The accuracy of this quadrature procedure increases with the number of evaluation points. For an application of the *random effects probit* model, see Sickles and Taubman (1986) who estimated a two-equation structural model of the health and retirement decisions of the elderly using five biennial panels of males drawn from the Retirement History Survey. Both the health and retirement variables were limited dependent variables and MLE using the Butler and Moffitt (1982) Gaussian quadrature procedure was implemented. Sickles and Taubman found that retirement decisions were strongly affected by health status, and that workers not yet eligible for social security were less likely to retire. LIMDEP and Stata provide basic routines for the random and fixed effects logit and probit models. In Stata these are the (*xtprobit* and *xtlogit*) commands with the (*fe* and *re)* options available for *xtlogit* and only *re* available for the *xtprobit*. These will be illustrated in the empirical example later on in this chapter.

Heckman (1981b) performed some limited Monte Carlo experiments on a probit model with a single regressor and a Nerlove (1971) type x_{it}. For $N = 100$, $T = 8$, $\sigma^2_\nu = 1$ and $\sigma^2_\mu = 0.5$, 1 and 3, Heckman computed the bias of the fixed effects MLE of β using 25 replications. He found at most 10% bias for $\beta = 1$ which was always *toward* zero. Replicating Heckman's design and using 100 replications, Greene (2004a) finds that the bias of the fixed effects MLE of β is of the order of 10 to 24% always *away* from zero.

Fernandez-Val (2009) characterizes the leading term of a large T expansion of the bias of the fixed effects probit MLE of β which allows him to obtain a lower bound for the first order bias that depends uniquely on T. This bias turns out to be at least 40% for $T = 2$; 20% for $T = 4$, and 10% for $T = 8$. Monte Carlo simulations confirm that these biases are between 33 and 48% for $T = 4$, and between 15 and 21% for $T = 8$. When there is a single regressor, the probit fixed effects estimates are biased away from zero, providing support for the previous numerical evidence by Greene (2004a). Interestingly, the FE estimates of the *marginal effects* exhibit no bias in the absence of heterogeneity and negligible bias for a wide variety of distributions of regressors and individual effects in the presence of heterogeneity.

Hahn and Newey (2004) consider two approaches to reducing the bias from fixed effects estimators in nonlinear models as T gets large. The first is a panel jacknife that uses the variation in the fixed effects estimators as each time period is dropped, one at a time, to form a bias corrected estimator. The second is an analytic bias correction using the bias formula obtained from an asymptotic expansion as T grows. They show that if T grows at the same rate as N, the fixed effects estimator is asymptotically biased, so that the asymptotic confidence intervals are incorrect. However, these are correct for the panel jacknife. If T grows faster than $N^{1/3}$, the analytical bias correction yields an estimator that is asymptotically normal and centered at the truth.

Cruz-Gonzalez, Fernandez-Val and Weidner (2017) wrote Stata commands *probitfe* and *logitfe* that implement the analytical and jackknife bias corrections of Fernandez-Val and Weidner (2016) to probit and logit panel data models with both individual and time effects. The methods are combinations of the leave-one-observation-out panel jackknife (PJ) of Hahn and Newey (2004) and the split-panel jackknife (SPJ) of Dhaene and Jochmans (2015) applied to the two dimensions of the panel.

Example 1 *Female Labor force participation*. This is the empirical example used in Fernandez-Val (2009). The sample is selected from waves 13 to 22 of the Panel Study of Income Dynamics (PSID) and covers ten calendar years 1979–1988. Only women aged 18–60 in 1985 who were continuously married with husbands in the labor force in each of the sample periods are included in the sample. The sample considered consists of 1461 women, 664 of whom changed labor force participation status during the sample period. The first year of the sample is excluded for use as the initial condition in the dynamic model. Table 11.1 gives the *xtlogit, fe* results for the female labor force participation as a function of the number of kids between 0-2 years of age, number of kids between 3–5 years of age, number of kids between 6–12 years of age, log of her husband's income, her age and age^2. The results show that having kids no matter what age reduces the probability of participating in the labor force. The higher the husband's income and the higher her age, the higher is her probability of participating in the labor force. This probability however decreases with age^2. All effects are significant. Unfortunately, these are not the marginal effects. Average partial effects are available using *logitfe.*

Table 11.1 Conditional Logit: Female Labor Force Participation

```
. xtlogit lfp kids0_2 kids3_5 kids6_17 loghusbandincome age age2, fe

note: multiple positive outcomes within groups encountered.

note: 797 groups (7,173 obs) dropped because of all positive or all negative outcomes.

Conditional fixed-effects logistic regression   Number of obs      =      5,97 6

Group variable: ID1979                          Number of groups   =        664

                                                Obs per group:   min =   9

                                                          avg =        9.0

                                                          max = 9

                                                LR chi2(6)         =     272.67

Log likelihood  = -2267.8037                    Prob > chi2        =     0.000 0
-----------------------------------------------------------------------------

         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval ]

-------------+---------------------------------------------------------------

     kids0_2 |  -1.086185   .0912304   -11.91   0.000    -1.264993   -.9073763

     kids3_5 |  -.6265956   .0835397    -7.50   0.000    -.7903304   -.4628607

    kids6_17 |   -.206979   .0672433    -3.08   0.002    -.3387734   -.0751847

loghusband~e |  -.3662394   .0880333    -4.16   0.000    -.5387814   -.1936974

         age |   3.641422   .6080303     5.99   0.000     2.449705     4.83314

        age2 |  -.4520102   .0807705    -5.60   0.000    -.6103174   -.293702 9
```

Table 11.2 does *xtprobit, re* for the female labor force participation empirical example 1 taken from Fernandez-Val (2009). You are asked to replicate Tables 11.1 and 11.2 in problem 11.12. One can also show that *xtprobit, re* and *xtlogit, re* are robust to *quadcheck* for this example. The quadcheck option in Stata checks the sensitivity of the estimates to the number of nodes used in computing the integral.

Problem 11.12 illustrates the *logitfe* and *probitfe* commands in Stata for the female labor force participation considered in example 1. Problem 11.13 illustrates *logitfe* and *probitfe* for the trade empirical example of Helpman, Melitz and Rubenstein (2008). These commands are useful for empirical research as they check the sensitivity of the estimates to various jackknife procedures as well as the extent of the bias in the uncorrected fixed effects estimator for the logit and probit model. Note that *xtprobit* fixed effects is not available in Stata because Chamberlain's conditional fixed effects applies specifically to the logit specification and is not applicable to the probit specification. It is also worth mentioning that the average partial effects are available using *probitfe* and *logitfe* but not available when using *xtlogit* fixed effects.

Example 2 *Beer Taxes and motor vehicle fatality rates.* Ruhm (1996) uses grouped logit analysis with fixed time and state effects to study the impact of beer taxes and a variety of alcohol-control policies on motor vehicle fatality rates. Ruhm uses panel data of 48 states (excluding Alaska, Hawaii and the District of Columbia) over the period 1982-1988. The dependent variable is $\log[p/(1-p)]$ where p is the total vehicle fatality rate per capita for state i at time t. The explanatory variables included

Table 11.2 Probit Random Effects: Female Labor Force Participation

```
. xtprobit lfp kids0_2 kids3_5 kids6_17 loghusbandincome age age2, re

Fitting comparison model:

Random-effects probit regression                Number of obs     =     13,149

Group variable: ID1979                          Number of groups  =      1,461

Random effects u_i ~ Gaussian                   Obs per group:

                                                              min =          9

                                                              avg =        9.0

                                                              max =          9

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(6)      =     330.65

Log likelihood  = -4928.3851                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------------

            lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

----------------+-------------------------------------------------------------------

        kids0_2 |  -.6865494   .0489026   -14.04   0.000    -.7823968   -.5907021

        kids3_5 |  -.4050329   .0441319    -9.18   0.000    -.4915299    -.318536

       kids6_17 |  -.1290886   .0320563    -4.03   0.000    -.1919179   -.0662594

loghusbandincome |  -.2538797   .0440499    -5.76   0.000     -.340216   -.1675435

            age |   2.119458   .2901562     7.30   0.000     1.550763    2.688154

           age2 |  -.2832263    .037195    -7.61   0.000    -.3561272   -.2103254

          _cons |   .6753673   .6574111     1.03   0.304    -.6131348    1.963869

----------------+-------------------------------------------------------------------

       /lnsig2u |   1.296086   .0686458                      1.161543     1.43063

----------------+-------------------------------------------------------------------

        sigma_u |   1.911796   .0656184                      1.787417     2.04483

            rho |   .7851756   .0115788                       .761613    .8069994

------------------------------------------------------------------------------------

LR test of rho=0: chibar2(01) = 5088.69               Prob >= chibar2 = 0.000
```

the real beer tax rate on 24 (12 oz.) containers of beer (BEERTAX), the minimum legal drinking age (MLDA) in years, the percentage of the population living in dry counties (DRY), the average number of vehicle miles per person aged 16 and over (VMILES), and the percentage of young drivers (15–24 years old) (YNGDRV). Also some dummy variables indicating the presence of alcohol regulations. These include BREATH test laws which is a dummy variable that takes the value 1 if the state authorized the police to administer pre-arrest breath test to establish probable cause for driving under the influence (DUI). JAILD which takes the value of 1 if the state passed legislation mandating jail or community service (COMSERD) for the first DUI conviction. Other variables included are the unemployment rate, real per capita income and state and time dummy variables. Details on these variables are given in Table 11.1 of Ruhm (1996). Results showed that most of the regulations had little or no impact on traffic mortality. By contrast, higher beer taxes were associated with reductions in crash deaths. Problem 11.8 asks the reader to replicate some of the results in this paper.

Grossman (2001) also ran an inverse logit model investigating the effect of multiple liability of bank share holders on bank failure rates in U.S. states before the Great Depression. Grossman found that double liability did reduce bank failures in periods where bank failures were not abnormally high. However, it did not guarantee bank stability in times of widespread financial distress. Problem 11.11 asks the reader to replicate some of the results in this paper.

The random effects probit model assume that μ_i and x_{it} are uncorrelated. Chamberlain (1980, 1984) relaxes this assumption as follows:

$$\mu_i = x_i'a + \epsilon_i \tag{11.14}$$

where $a' = (a_1', \ldots, a_T')$, $x_i' = (x_{i1}', \ldots, x_{iT}')$ and $\epsilon_i \sim \text{IID}(0, \sigma_\epsilon^2)$ independent of ν_{it}. In this case,

$$y_{it} = 1 \quad \text{if } (x_{it}'\beta + x_i'a + \epsilon_i + \nu_{it}) > 0$$

and the distribution of y_{it} conditional on x_{it} but marginal on μ_i has the probit form

$$\Pr[y_{it} = 1] = \Phi[(1 + \sigma_\epsilon^2)^{-1/2}(x_{it}'\beta + x_i'a)]$$

where Φ denotes the cumulative normal distribution function. Once again, MLE involves numerical integration, but a computationally simpler approach suggested by Chamberlain is to run simple probit on this equation to get $\widehat{\Pi}$. In this case, Π satisfies the restriction

$$\Pi = (1 + \sigma_\epsilon^2)^{-1/2}(I_T \otimes \beta' + \iota_T a')$$

Therefore, Chamberlain suggests a minimum distance estimator based on $(\widehat{\pi} - \pi)$, where $\pi = \text{vec}(\Pi')$, that imposes this restriction. Chamberlain (1984) applies both his *fixed effects logit* estimator and his *minimum distance random effects probit* estimator to a study of labor force participation of 924 married women drawn from the PSID. These estimation methods give different results especially with regard to the effect of the presence of young children on labor force participation. These different results could be attributed to the misspecification of the relationship between μ_i and the x_{it} in the random effects specification or a misspecification of the fixed effects logit model in its omission of leads and lags of the x_{it} from the structural equation.

Bover and Arellano (1997) provide extensions of the random effects probit model of Chamberlain (1984) which has applications in the analysis of binary choice, linear regression subject to censoring, and other models with endogenous selectivity. They propose a simple two-step Within estimator for limited dependent variable models, which may include lags of the dependent variable, other exogenous variables and unobservable individual effects. This estimator is based on reduced form predictions of the latent endogenous variables. It can be regarded as a member of Chamberlain's class of random effects minimum distance estimators, and as such it is consistent and asymptotically normal for fixed T. However, this Within estimator is not asymptotically efficient within the minimum distance class, since it uses a non-optimal weighting matrix. Therefore, Bover and Arellano (1997) show how one can obtain in one more step a chi-squared test statistic for over-identifying restrictions and linear GMM estimators that are asymptotically efficient. The drawbacks of this approach are the same as those for the Chamberlain probit model. Both require the availability of strictly exogenous variables and the specification of the conditional distribution of the effects. Labeaga (1999) applies the Bover and Arellano (1997) method to estimate a double-hurdle rational addiction model for tobacco consumption using an unbalanced panel of households drawn from the Spanish Permanent Survey of Consumption (SPSC). This is a panel collected by the Spanish Statistical Office for approximately 2000 households between 1977 and 1983.

11.2 Simulation Estimation of Limited Dependent Variable Models with Panel Data

Keane (1994) derived a computationally practical simulation estimator for the panel data probit model. The basic idea of simulation estimation methods is to replace intractable integrals by unbiased Monte Carlo probability simulators. This is ideal for limited dependent variable models where for a multinomial probit model, the choice probabilities involve multivariate integrals.[5] In fact, for cross-section data, the method of simulated moments (MSM) involves an $M - 1$ integration problem, where M is the number of possible choices facing the individual. For panel data, things get more complicated, because there are M choices facing any individual at each period. This means that there are M^T possible choice sequences facing each individual over the panel. Hence the MSM estimator becomes infeasible as T gets large. Keane (1994) sidesteps this problem of having to simulate M^T possible choice sequences by factorizing the method of simulated moments first-order conditions into transition probabilities. The latter are simulated using highly accurate importance sampling techniques. This method of simulating probabilities is referred to as the Geweke, Hajivassiliou, and Keane (GHK) simulator because it was independently developed by these authors. Keane performs Monte Carlo experiments and finds that even for large T and small simulation sizes, the bias in the MSM estimator is negligible. When maximum likelihood methods are feasible, Keane finds that the MSM estimator performs well relative to quadrature-based maximum likelihood methods even where the latter are based on a large number of quadrature points. When maximum likelihood is not feasible, the MSM estimator outperforms the simulated MLE even when the highly accurate GHK probability simulator is used.

The Method of Simulated Likelihood looks at the transformed likelihood in (11.13) as an expectation:

$$L_i = \int \left[\prod_{t=1}^{T} (\int f_1(u_{it}/\mu_i) du_{it} \right] f_2(\mu_i) d\mu_i = E_{\mu_i} \left[\prod_{t=1}^{T} (\int f_1(u_{it}/\mu_i) du_{it} \right] = E_{\mu_i}(h(\mu_i)) \tag{11.15}$$

This function is smooth, continuous, and continuously differentiable. If this expectation is finite, then the conditions of the law of large numbers should apply. This means that for a sample of observations μ_{i1}, ..., μ_{iR}

$$plim \frac{1}{R} \sum_{r=1}^{R} h(\mu_{ir}) = E_{\mu_i}(h(\mu_i))$$

A sample of person specific draws from the population μ_i can be generated from a random number generator. For the probit function in the Buttler and Moffitt (1982) model

$$lnL_{simulated} = \sum_{i=1}^{N} [ln\{\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \Phi(q_{it}(x'_{it}\beta + \sigma_\mu \mu_{ir}))\}]$$

Keane (1994) argues that MSM has three advantages over other practical non-maximum likelihood estimators. First, MSM is asymptotically as efficient as maximum likelihood (in simulation size) while the other estimators are not. Second, MSM can be easily extended to handle multinomial probit situations whereas the extension of the other estimators is computationally burdensome. Third, MSM can be extended to handle nonlinear systems of equations which are intractable with maximum likelihood. Keane (1994) also finds that MSM can estimate random effects models with autoregressive moving average error in about the same time necessary for estimating a simple random effects model using maximum likelihood quadrature. The extension of limited dependent variable models to allow for a general pattern of serial correlation is now possible using MSM and could prove useful for out-of-sample predictions. An example of the MSM estimator is given by Keane (1993) who estimates probit employment equations using data from the National Longitudinal Survey of Young Men (NLS). This is a sample of 5225 males aged 14–24 and interviewed 12 times over the period 1966-81. For this example, Keane concludes that relaxing the equicorrelation assumption by including an MA(1) or AR(1) component to the error term had little effect on the parameter estimates. Keane (1993) discusses simulation estimation of models more complex than probit models. He argues that it is difficult to put panel data selection models and Tobit models in an MSM framework and that the method of simulated scores (MSS) may be a preferable way to go. For another application, see Hajivassiliou (1994) who reconsiders the problem of external debt crisis of 93 developing countries observed over the period 1970-88. Using several simulation estimation methods, Hajivassiliou concludes that allowing for flexible correlation patterns changes the estimates substantially and raises doubts over previous studies that assumed restrictive correlation structures.

Zhang and Lee (2004) argue that the statistical performance of the GHK simulator may be adequate for panels with small T, but this performance deteriorates when T is larger than 50 (for a moderate amount of simulation draws). In fact, the bias of the SML estimator may become larger than its standard deviation. Zhang and Lee suggest applying the accelerated importance sampling (AIS) procedure to SML estimation of dynamic discrete choice models with long panels. Using Monte Carlo experiments, they show that this can improve upon the GHK sampler when T is large and they illustrate their method using an application on firm's dividend decisions. They collect data on quarterly dividends and earnings per share from COMPUSTAT tapes. The sample period is 54 quarters (1987:1–2002:2). Two quarters were used for getting the initial value for each firm, so T = 52. The final sample used included N =150 large U.S. industrial firms and the total number of observations NT = 7800. The results confirm that the AIS improves the performance of the GHK sampler.

11.3 Dynamic Panel Data Limited Dependent Variable Models

So far the model is static implying that, for example, the probability of buying a car is independent of the individual's past history of car purchases. If the probability of buying a car is more likely if the individual has bought a car in the past than if he or she has not, then a dynamic model that takes into account the individual's past experience is more appropriate. Heckman (1981a, b, c) gives an extensive treatment of these dynamic models and the consequences of various assumptions on the initial values on the resulting estimators. Heckman (1981c) also emphasizes the importance of distinguishing between *true state dependence* and *spurious state dependence*. In the "true" case, once an individual experiences an event like unemployment, his preferences change and he or she will behave differently in the future as compared with an identical individual that has not experienced this event in the past. In fact, it is observed that individuals with a long history of unemployment are less likely to leave unemployment. They may be less attractive for employers to hire or may become discouraged in looking for a job. In the "spurious" case, past experience has no effect on the probability of experiencing the event in the future. It is the individual's characteristics that makes him or her less likely to leave unemployment. However, one cannot properly control for all the variables that distinguish one individual's decision from another's. In this case, past experience which is a good proxy for these omitted variables shows up as a significant determinant of the future probability of occurrence of this event. Testing for true versus spurious state dependence is therefore important in these studies, but it is complicated by the presence of the individual effects or heterogeneity. In fact, even if there is no state dependence, $\Pr[y_{it}/x_{it}, y_{i,t-l}] \neq \Pr[y_{it}/x_{it}]$ as long as there are random individual effects present in the model. If in addition to the absence of the state dependence, there is also no heterogeneity, then $\Pr[y_{it}/x_{it}, y_{i,t-l}] = \Pr[y_{it}/x_{it}]$. A test for this equality can be based on a test for $\gamma = 0$ in the model

$$\Pr[y_{it} = 1/x_{it}, y_{it-1}] = F(x_{it}'\beta + \gamma y_{i,t-1})$$

using standard maximum likelihood techniques. If $\gamma = 0$ is not rejected, we ignore the heterogeneity issue and proceed as in conventional limited dependent variable models not worrying about the panel nature of the data. However, rejecting the null does not necessarily imply that there is heterogeneity since γ can be different from zero due to serial correlation in the remainder error or due to state dependence. In order to test for time dependence one has to condition on the individual effects, i.e., test $\Pr[y_{it}/y_{i,t-l}, x_{it}, \mu_i] = \Pr[y_{it}/x_{it}, \mu_i]$. In fact, if $\gamma = 0$ is rejected, Hsiao (2003) suggests testing for time dependence against heterogeneity. If heterogeneity is rejected, the model is misspecified. If heterogeneity is not rejected then one estimates the model correcting for heterogeneity. See Heckman (1981c) for an application to married women's employment decisions based on a three-year sample from the PSID. One of the main findings of this study is that neglecting heterogeneity in dynamic models overstates the effect of past experience on labor market participation. Das and van Soest (1999) use the October waves of 1984 till 1989 from the Dutch Socio-Economic Panel to study household subjective expectations about future income changes. Ignoring attrition and sample selection problems which could be serious, the authors estimate a static random effects probit model and a fixed effects conditional logit model and extend them to the case of ordered response. Using Heckman's (1981b) procedure, they also estimate a dynamic random effects model which includes a measure of permanent and transitory income. They find that income change expectations strongly depend on realized income changes in the past. In particular, those whose income fell were more pessimistic than others, while those whose income rose were more optimistic. The paper rejects rational expectations finding that households whose income has decreased in the past underestimate their future income growth. In marketing research, one can attribute consumers repeated purchases of the same brands to either state dependence or heterogeneity. For an application using household-level scanner panel data on six frequently purchased packaged products: ketchup, peanut, butter, liquid detergent, tissue, tuna, and sugar, see Erdem and Sun (2001). The authors find evidence of state dependence for all product categories except sugar. Read also Chap. 18 of the Handbook of Panel Data by Keane (2015) entitled Discrete choice models of consumer demand. Keane gives an extensive review of the use of the method of simulated methods in marketing where scanner panel data is available on consumers buying goods over several weeks. Keane sheds more light on distinguishing between state dependence and consumer heterogeneity controlling for serial correlation in tastes. One of his conclusions is that there is consensus in the marketing literature of the existence of state dependence. This has important implications for marketing actions, where price discounts not only affect current but also future demand.

Chamberlains' fixed effects conditional logit approach can be generalized to include lags of the dependent variable, provided there are *no* explanatory variables and $T \geqslant 4$, see Chamberlain (1985). Assuming the initial period y_{i0} is observed but its probability is unspecified, the model is given by

$$\Pr[y_{i0} = 1/\mu_i] = p_0(\mu_i)$$
$$\Pr[y_{it} = 1/\mu_i, y_{i0}, y_{i1}, \ldots, y_{i,t-1}] = \frac{e^{\gamma y_{i,t-1}+\mu_i}}{1+e^{\gamma y_{i,t-1}+\mu_i}} \quad t = 1, \ldots, T \tag{11.16}$$

where $p_0(\mu_i)$ is *unknown* but the logit specification is imposed from period one to T. Consider the two events

$$A = \{y_{i0} = d_0, y_{i1} = 0, y_{i2} = 1, y_{i3} = d_3\} \tag{11.17}$$

$$B = \{y_{i0} = d_0, y_{i1} = 1, y_{i2} = 0, y_{i3} = d_3\} \tag{11.18}$$

where d_0 and d_3 are either 0 or 1. If $T = 3$, inference on γ is based upon the fact that $\Pr[A/y_{i1} + y_{i2} = 1, \mu_i]$ and $\Pr[B/y_{i1} + y_{i2} = 1, \mu_i]$ do not depend upon μ_i, see problem 11.2. Honoré and Kyriazidou (2000b) consider the identification and estimation of panel data discrete choice models with lags of the dependent variable and strictly exogenous variables that allow for unobservable heterogeneity. In particular, they extend Chamberlain's (1985) fixed effects logit model in (11.16) to include *strictly exogenous* variables $x_i' = (x_{i1}, \ldots, x_{iT})$, i.e.,

$$\Pr[y_{i0} = 1/x_i', \mu_i] = p_0(x_i', \mu_i) \tag{11.19}$$

$$\Pr[y_{it} = 1/x_i', \mu_i, y_{i0}, \ldots, y_{i,t-1}] = \frac{e^{x_{it}'\beta + \gamma y_{i,t-1} + \mu_i}}{1 + e^{x_{it}'\beta + \gamma y_{i,t-1} + \mu_i}} \quad t = 1, \ldots, T$$

The crucial assumption is that the errors in the threshold-crossing model leading to (11.19) are IID over time with logistic distributions and independent of (x_i', μ_i, y_{i0}) at all time periods. Honoré and Kyriazidou (2000b) show that $\Pr(A/x_i', \mu_i, A \cup B)$ and $\Pr(B/x_i', \mu_i, A \cup B)$ will still depend upon μ_i. This means that a conditional likelihood approach will not eliminate the fixed effects. However, if $x_{i2}' = x_{i3}'$, then the conditional probabilities

$$\Pr(A/x_i', \mu_i, A \cup B, x_{i2}' = x_{i3}') = \frac{1}{1 + e^{(x_{i1} - x_{i2})'\beta + \gamma(d_0 - d_3)}} \tag{11.20}$$

$$\Pr(B/x_i', \mu_i, A \cup B, x_{i2}' = x_{i3}') = \frac{e^{(x_{i1} - x_{i2})'\beta + \gamma(d_0 - d_3)}}{1 + e^{(x_{i1} - x_{i2})'\beta + \gamma(d_0 - d_3)}}$$

do *not* depend on μ_i, see problem 11.3. If all the explanatory variables are discrete and $\Pr[x_{i2}' = x_{i3}'] > 0$, Honoré and Kyriazidou (2000b) suggest maximizing a weighted likelihood function based upon (11.20) for observations that satisfy $x_{i2}' = x_{i3}'$ and $y_{i1} + y_{i2} = 1$. The weakness of this approach is its reliance on observations for which $x_{i2}' = x_{i3}'$ which may not be useful for many economic applications. However, Honoré and Kyriazidou suggest weighing the likelihood function with weights that depend inversely on $x_{i2}' - x_{i3}'$, giving more weight to observations for which x_{i2}' is close to x_{i3}'. This is done using a kernel density $K\left(\frac{x_{i2}' - x_{i3}'}{h_N}\right)$ where h_N is a bandwidth that shrinks as N increases. The resulting estimators are consistent and asymptotically Normal under standard assumptions. However, their rate of convergence will be slower than $\sqrt{N}$ and will depend upon the number of continuous covariates in x_{it}'. The results of a small Monte Carlo study suggest that this estimator performs well and that the asymptotics provide a reasonable approximation to the finite sample behavior of the estimator.[6]

Chintagunta, Kyriazidou and Perktold (2001) apply the Honoré and Kyriazidou (2000b) method to study yogurt brand loyalty in South Dakota. They use household panel data with at least two purchases of Yoplait and Nordica yougurt brands over approximately a two-year period. They control for household effects, difference in price and whether the brand was featured in an advertisement that week or displayed in the store. They find that a previous purchase of a brand increases the probability of purchasing that brand in the next period. They also find that if one ignores household heterogeneity, this previous purchase effect is overstated.

Contoyannis, Jones and Rice (2004) utilize seven waves (1991–1997) of the british household panel survey (BHPS) to analyze the dynamics of individual health and to decompose the persistence in health outcomes in the BHPS data into components due to state dependence, serial correlation, and unobserved heterogeneity. The indicator of health is defined by a binary response to the question: " Does your health in any way limit your daily activities compared to most people of your age?" A sample of 6106 individuals resulting in 42,742 panel observations are used to estimate static and dynamic panel probit models by maximum simulated likelihood using the GHK simulator with antithetic acceleration. The dynamic models show strong positive state dependence.

Arellano and Carrasco (2003) consider a binary choice panel data model with predetermined variables. A semiparametric random effects specification is suggested as a compromise to the fixed effects specification that leaves the distribution of the individual effects unrestricted. Dependence is allowed through a nonparametric specification of the conditional expectation of the effects given the predetermined variables. The paper proposes a GMM estimator which is shown to be consistent and asymptotically normal for fixed T and large N. This method is used to estimate a labor force participation equation for women with children, using PSID data.

Carro (2007) considers the problem of estimating dynamic binary choice panel data models with fixed effects. He suggests a modified maximum likelihood estimator (MMLE) which modifies the concentrated log-likelihood to correct the first term on the asymptotic bias that comes from the estimation of fixed effects. This reduces the order of bias for the MLE from $O(1/T)$ to $O(1/T^2)$, without increasing the asymptotic variance. Even though this estimator is consistent only for $T \to \infty$, it is shown via Monte Carlo experiments to have negligible finite sample bias for logit and probit dynamic panel data models with $T = 8$. Unlike the Honoré and Kyriazidou (2000b) estimator, the MMLE does not require a logistic distribution and can allow for time dummy variables. It can also allow for more lags on the endogenous variable and could be generally applied to multinomial choice and nonlinear models.

Wooldridge (2005) suggested a simple approach for handling the *initial conditions problem* in dynamic nonlinear unobserved effects models. Three popular applications of this approach include the probit, Tobit, and Poisson panel data models. This involves parametrizing the distribution of the unobserved effects conditional on the initial value and any exogenous explanatory variables, see Chamberlain (1980, 1984). This has the advantages of being flexible and easily estimated with standard software which in turn allows the identification of partial effects on mean responses averaged over the distribution of unobservables. Its disadvantage is that misspecifying this distribution leads to inconsistent estimates. The semiparametric approach of Honoré and Kyriazidou (2000b) does not specify the distribution of the unobserved effects but requires strong assumptions on the strictly exogenous variables mentioned above. Also, it reduces the panel to individuals with no change in any discrete variables over the last two time periods. It also cannot yield partial effects on the response probabilities. For the dynamic probit model, Wooldridge (2005) assumes that for each random individual $i = 1, 2.., N$:

$$\Pr[y_{it} = 1/x_i', \mu_i, y_{i0}, \ldots, y_{i,t-1}] = \Phi(x_{it}'\beta + \lambda y_{i,t-1} + \mu_i) \text{ for } t = 1, 2, .., T.$$

where $x_i' = (x_{i1}', .., x_{iT}')$. Also, that

$$(\mu_i/x_i', y_{i0}) \sim N((x_i'\delta + \gamma y_{i,0} + \gamma_0), \sigma_\varepsilon^2)$$

and estimation can be carried out with a standard random effects probit procedure. In fact, if we write

$$\mu_i = x_i'\delta + \gamma y_{i,0} + \gamma_0 + \varepsilon_i$$

with $(\varepsilon_i/x_i', y_{i0})\tilde{\ } N(0, \sigma_\varepsilon^2)$, then y_{it} given $(x_i', \varepsilon_i, y_{i0}, \ldots, y_{i,t-1})$ follows a probit model with response probability:

$$\Phi(x_{it}'\beta + \lambda y_{i,t-1} + x_i'\delta + \gamma y_{i,0} + \gamma_0 + \varepsilon_i)$$

This means that we can estimate this model with *xtprobit, re* in Stata using as regressors $(1, x_{it}', y_{i,t-1}, x_i', y_{i,0})$. In essence, one is adding x_i', $y_{i,0}$ as extra regressors. A consistent estimator of the average partial effects can be obtained from evaluating changes or derivatives of the following expression with respect to x_{it}' or $y_{i,t-1}$ at their MLEs:

$$\frac{1}{N}\sum_{i=1}^{N} \Phi[(x_{it}'\widehat{\beta} + \widehat{\lambda} y_{i,t-1} + x_i'\widehat{\delta} + \widehat{\gamma} y_{i,0} + \widehat{\gamma_0})/\sqrt{1 + \widehat{\sigma}_\varepsilon^2}]$$

Wooldridge (2005) applies this method to the Vella and Verbeek (1998) panel data set which estimated the union wage differential for working men using the PSID, see Sect. 11.6 below. In particular, Wooldridge examines the persistence of union membership using the dynamic probit equation described above with y_{it} denoting union membership and x_i' including marital status and time dummies. Wooldridge finds that the lagged union effect is statistically significant and so is the initial union membership effect. Table 11.3 replicates the results of column 1 of Table I in Wooldridge (2005, p. 52). In the *xtprobit* command: union_1 is union lagged $(y_{i,t-1})$ and union80 is the initial value ($y_{i,0}$). See also the companion book, exercise 11.8 in Baltagi (2009). Additionally, two time-invariant variables: years of education and a dummy variable for whether the individual is black were added to the specification, finding only the latter significant. Using the average partial effects computations described above, Wooldridge computes the estimated probability of being in a union in 1987 and 1986 for married and non-married men and finds that the estimate of state dependence for union membership is 0.182 for married men and 0.173 for non-married men over these two years. Problem 11.7 asks the reader to replicate these results.

Table 11.3 Union Membership: Random-effects probit

```
. xtprobit union married union_1 union80 marr81 marr82 marr83 marr84 marr85 marr86
marr87 d81 d82 d83 d84 d85 d86 d87, re

Random-effects probit regression                Number of obs      =      3815

Group variable (i): nr                          Number of groups   =       545

Random effects u_i ~ Gaussian                   Obs per group: min =         7

                                                               avg =       7.0

                                                               max =         7

                                                Wald chi2(16)      =    403.30

Log likelihood  = -1291.2555                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------

       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

     married |   .1645174   .1086856     1.51   0.130    -.0485024    .3775373

     union_1 |   .9778509   .0845496    11.57   0.000     .8121366    1.143565

     union80 |   1.342638   .1370911     9.79   0.000     1.073944    1.611331

      marr81 |   .0557457   .1927517     0.29   0.772    -.3220408    .4335321

      marr82 |  -.1061208   .2278688    -0.47   0.641    -.5527355    .3404939

      marr83 |  -.0740984   .2307409    -0.32   0.748    -.5263423    .3781455

      marr84 |   .0019333    .248823     0.01   0.994    -.4857507    .4896174

      marr85 |   .3572686   .2346558     1.52   0.128    -.1026483    .8171854

      marr86 |   .1000752    .235037     0.43   0.670    -.3605888    .5607392

      marr87 |  -.3870529   .1835973    -2.11   0.035    -.7468971   -.0272087

         d81 |  -.0765441   .1166593    -0.66   0.512    -.3051922    .1521039

         d82 |  -.0489836   .1149461    -0.43   0.670    -.2742738    .1763065

         d83 |  -.1634238   .1154857    -1.42   0.157    -.3897717    .0629241

         d84 |  -.1246027   .1147137    -1.09   0.277    -.3494374    .1002319

         d85 |  -.3350622   .1172977    -2.86   0.004    -.5649614    -.105163

         d86 |  -.3777447   .1173934    -3.22   0.001    -.6078316   -.1476579

       _cons |  -1.644473   .1275296   -12.89   0.000    -1.894426   -1.394519
```

(continued)

Table 11.3 (continued)

```
-------------+----------------------------------------------------------------

    /lnsig2u |  -.0889939   .1235904                     -.3312267    .1532389

-------------+----------------------------------------------------------------

     sigma_u |   .9564785   .0591058                      .8473738    1.079631

         rho |   .4777662   .0308365                      .4179422    .5382349

------------------------------------------------------------------------------
```

Likelihood-ratio test of rho=0: chibar2(01) = 149.59 Prob >= chibar2 = 0.000

Using Monte Carlo experiments, Akay (2012) shows that for *short panels*, misspecification of the conditional distribution of the initial values leads to serious bias in the estimated parameters. For example, exogenous initial values assumption leads to overestimation of the true state dependence and underestimation of the variance of the unobserved individual effects. However, this bias is not a problem for panels of long duration. One of the main findings is that Wooldridge's method works very well for panels with $T \geq 5$ periods, but its performance for $T < 5$ may be highly sensitive to the specification of the auxiliary distribution of the unobserved individual effects and the explanatory variables entering the specification.

11.4 Selection Bias in Panel Data

In Chap. 9, we studied incomplete panels that had randomly missing data. In section 10.2 we studied rotating panels where, by the design of the survey, households that drop from the sample in one period are intentionally replaced in the next period. However, in many surveys, nonrandomly missing data may occur due to a variety of self-selection rules. One such self-selection rule is the problem of nonresponse of the economic agent. Nonresponse occurs, for example, when the individual refuses to participate in the survey, or refuses to answer particular questions. This problem occurs in cross-section studies, but it becomes aggravated in panel surveys. After all, panel surveys are repeated cross-sectional interviews. So, in addition to the above kinds of nonresponse, one may encounter individuals that refuse to participate in subsequent interviews or simply move or die. Individuals leaving the survey cause *attrition* in the panel. This distorts the random design of the survey and questions the representativeness of the observed sample in drawing inference about the population we are studying. Inference based on the balanced subpanel is inefficient even in randomly missing data since it is throwing away data. In nonrandomly missing data, this inference is misleading because it is no longer representative of the population. Verbeek and Nijman (1996) survey the reasons for nonresponse and distinguish between *ignorable* and *nonignorable* selection rules. This is important because, if the selection rule is ignorable for the parameters of interest, one can use the standard panel data methods for consistent estimation. If the selection rule is nonignorable, then one has to take into account the mechanism that causes the missing observations in order to obtain consistent estimates of the parameters of interest. In order to reduce the effects of attrition, *refreshment samples* are used which replace individuals who dropped from the panel by new individuals randomly sampled from the population. With these refreshment samples, it may be possible to test whether the missing data is *ignorable* or *nonignorable,* see Hirano et al. (2001).

For the one-way error component regression model

$$y_{it} = x'_{it}\beta + \mu_i + v_{it} \tag{11.21}$$

where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ independent of each other and the x_{it}. Observations on y_{it} (and possibly x_{it}) are missing if a selection variable $r_{it} = 0$ and not missing if $r_{it} = 1$. The missing data mechanism is ignorable of order one for β if $E(\mu + \nu_i/r_i) = 0$ for $i = 1, \ldots, N$, where $\mu' = (\mu_1, \ldots, \mu_N)$, $\nu'_i = (\nu_{i1}, \ldots, \nu_{iT})$ and $r'_i = (r_{i1}, \ldots, r_{iT})$. In this case, both GLS on the unbalanced panel and the balanced subpanel are consistent if $N \to \infty$. The Within estimator is consistent for both the unbalanced and balanced subpanel as $N \to \infty$ if $E(\widetilde{\nu}_i/r_i) = 0$ where $\widetilde{\nu}'_i = (\widetilde{\nu}_{i1}, \ldots, \widetilde{\nu}_{iT})$ and $\widetilde{\nu}_{it} = \nu_{it} - \overline{\nu}_i$.[7]

We now consider a simple model of nonresponse in panel data. Following Verbeek and Nijman (1996), we assume that y_{it} is observed, i.e., $r_{it} = 1$, if a latent variable $r^*_{it} \geqslant 0$. This latent variable is given by

$$r^*_{it} = z'_{it}\gamma + \epsilon_i + \eta_{it} \tag{11.22}$$

where z_{it} is a set of explanatory variables possibly including some of the x_{it}.[8] The one-way error component structure allows for heterogeneity in the selection process. The errors are assumed to be normally distributed $\epsilon_i \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $\eta_{it} \sim \text{IIN}(0, \sigma_\eta^2)$ with the only nonzero covariances being $\text{cov}(\epsilon_i, \mu_i) = \sigma_{\mu\epsilon}$ and $\text{cov}(\eta_{it}, \nu_{it}) = \sigma_{\eta\nu}$. In order to get a consistent estimator for β, a generalization of Heckman's (1979) selectivity bias correction procedure from the cross-section to the panel data case can be employed. The conditional expectation of u_{it} given selection now involves two terms. Therefore, instead of one selectivity bias correction term, there are now two terms corresponding to the two covariances $\sigma_{\mu\epsilon}$ and $\sigma_{\eta\nu}$. However, unlike the cross-sectional case, these correction terms cannot be computed from simple probit regressions and require numerical integration. Fortunately, this is only a one-dimensional integration problem because of the error component structure. Once the correction terms are estimated, they are included in the regression equation as in the cross-sectional case and OLS or GLS can be run on the resulting augmented model. For details, see Verbeek and Nijman (1996) who also warn about heteroskedasticity and serial correlation in the second step regression if the selection rule is nonignorable. Verbeek and Nijman (1996) also discuss MLE for this random effect probit model with selection bias. The computations require two-dimensional numerical integration for all individuals with $r_{it} = 0$ for at least one t.

Before one embarks on these complicated estimation procedures one should first test whether the selection rule is ignorable. Verbeek and Nijman (1992) consider a Lagrange multiplier (LM) test for H_0; $\sigma_{\nu\eta} = \sigma_{\mu\epsilon} = 0$. The null hypothesis is a sufficient condition for the selection rule to be ignorable for the random effects model. Unfortunately, this also requires numerical integration over a maximum of two dimensions and is cumbersome to use in applied work. In addition, the LM test is highly dependent on the specification of the selectivity equation and the distributional assumptions. Alternatively, Verbeek and Nijman (1992) suggest some simple Hausman-type tests based on GLS and Within estimators for the unbalanced panel and the balanced subpanel.[9] All four estimators are consistent under the null hypothesis that the selection rule is ignorable and all four estimators are inconsistent under the alternative. This is different from the usual Hausman-type test where one estimator is consistent under both the null and alternative hypotheses. Whereas the other estimator is efficient under the null, but inconsistent under the alternative. As a consequence, these tests may have low power especially if under the alternative these estimators have close asymptotic biases. On the other hand, the advantages of these tests are that they are computationally simple and do not require the specification of a selection rule to derive these tests. Let $\widehat{\delta} = (\widetilde{\beta}_W(B), \widetilde{\beta}_W(U), \widehat{\beta}_{GLS}(B), \widehat{\beta}_{GLS}(U))$ where $\widetilde{\beta}_W$ denotes the Within estimator and $\widehat{\beta}_{GLS}$ denotes the GLS estimator, $\widehat{\beta}(B)$ corresponds to an estimator of β from the balanced subpanel and $\widehat{\beta}(U)$ corresponds to an estimator of β from the unbalanced panel. Verbeek and Nijman (1992) show that the variance–covariance matrix of $\widehat{\delta}$ is given by

$$\mathrm{var}(\widehat{\delta}) = \begin{bmatrix} V_{11} & V_{22} & V_{33} & V_{44} \\ & V_{22} & V_{22}V_{11}^{-1}V_{13} & V_{44} \\ & & V_{33} & V_{44} \\ & & & V_{44} \end{bmatrix} \tag{11.23}$$

where $V_{11} = \mathrm{var}(\widetilde{\beta}_W(B))$, $V_{22} = \mathrm{var}(\widetilde{\beta}_W(U))$, $V_{33} = \mathrm{var}(\widehat{\beta}_{GLS}(B))$ and $V_{44} = \mathrm{var}(\widehat{\beta}_{GLS}(U))$. Therefore an estimate of $\mathrm{var}(\widehat{\delta})$ can be obtained from the estimated variance–covariance matrices of the four estimation procedures. Hausman-type tests can now be performed on say H_0; $R\delta = 0$, where R is a known matrix, as follows:

$$m = N\widehat{\delta}'R'[R\ \mathrm{var}(\widehat{\delta})R']^{-}R\widehat{\delta} \tag{11.24}$$

and this is distributed as χ^2 under the null with degrees of freedom equal to the rank of $[R\ \mathrm{var}(\widehat{\delta})R']$. Natural candidates for R are $R_1 = [I, 0, -I, 0]$, $R_2 = [0, I, 0, -I]$, $R_3 = [I, -I, 0, 0]$ and $R_4 = [0, 0, I - I]$. The first two are the standard Hausman tests based on the difference between the Within and GLS estimators for the balanced subpanel (R_1) and the unbalanced panel (R_2). The third is based on the difference between the Within estimators from the balanced and unbalanced panels (R_3), while the last is based on the difference between the GLS estimators from the balanced and unbalanced panels (R_4). For all four cases considered, the variance of the difference is the difference between the two variances and hence it is easy to compute. Verbeek and Nijman (1992) perform some Monte Carlo experiments verifying the *poor power* of these tests in some cases, but also illustrating their usefulness in other cases. In practice, they recommend performing the tests based on R_2 and R_4. It is important to note that these Hausman type tests are not verifying whether the selection rule is ignorable, instead they are checking for possible bias due to sample selection on observables.

Wooldridge (1995) derives some simple variable addition tests of selection bias as well as easy-to-apply estimation techniques that correct for selection bias in linear fixed effects panel data models. The auxiliary regressors are either Tobit residuals or inverse Mill's ratios and the disturbances are allowed to be arbitrarily serially correlated and unconditionally heteroskedastic. Wooldridge (1995) considers the fixed effects model where the μ_i's are correlated with x_{it}. However, the remainder disturbances ν_{it} are allowed to display arbitrary serial correlation and unconditional heteroskedasticity. The panel is unbalanced with the selection indicator vector for each individual i denoted by $s_i' = (s_{i1}, s_{i2}, \ldots, s_{it})$. When $s_{it} = 1$, it is assumed that (x_{it}', y_{it}) is observed. The fixed effects estimator is given by

$$\widetilde{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\widetilde{x}_{it}\widetilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\widetilde{x}_{it}\widetilde{y}_{it}\right) \tag{11.25}$$

where $\widetilde{x}_{it}' = x_{it}' - \left(\sum_{r=1}^{T} s_{ir}x_{ir}'/T_i\right)$, $\widetilde{y}_{it} = y_{it} - \left(\sum_{r=1}^{T} s_{ir}y_{ir}/T_i\right)$ and $T_i = \sum_{i=1}^{T} s_{it}$. A sufficient condition for the fixed estimator to be consistent and asymptotically Normal, as $N \to \infty$, is that $E(\nu_{it}/\mu_i, x_i', s_i') = 0$ for $t = 1, 2, \ldots, T$. Recall, that $x_i' = (x_{i1}', \ldots, x_{iT}')$. Under this assumption, the selection process is strictly exogenous conditional on μ_i and x_i'.

Wooldridge (1995) considers two cases. The first is when the latent variable determining selection is partially observed. Define a latent variable

$$h_{it}^* = \delta_{t0} + x_{i1}'\delta_{t1} + \ldots + x_{iT}'\delta_{tT} + \epsilon_{it} \tag{11.26}$$

where ϵ_{it} is independent of (μ_i, x_i'), δ_{tr} is a $K \times 1$ vector of unknown parameters for $r = 1, 2, \ldots, T$ and $\epsilon_{it} \sim N(0, \sigma_t^2)$.

The binary selection indicator is defined as $s_{it} = 1$ if $h_{it}^* > 0$. For this case, the censored variable $h_{it} = \max(0, h_{it}^*)$ is observed. For example, this could be a wage equation, and selection depends on whether or not individuals are working. If a person is working, the working hours h_{it} are recorded, and selection is determined by nonzero hours worked. This is what is meant by partial observability of the selection variable.

Because s_i is a function of (x_i', ϵ_i') where $\epsilon_i' = (\epsilon_{i1}, \ldots, \epsilon_{iT})$, a sufficient condition for the fixed effects estimator to be consistent and asymptotically Normal as $N \to \infty$ is $E(\nu_{it}/\mu_i, x_i', \epsilon_i') = 0$ for $t = 1, 2 \ldots, T$. The simplest alternative that imply selectivity bias is $E(\nu_{it}/\mu_i, x_i', \epsilon_i') = E(\nu_{it}/\epsilon_{it}) = \gamma\epsilon_{it}$ for $t = 1, 2, \ldots, T$, with γ being an unknown scalar. Therefore,

$$E(y_{it}/\mu_i, x_i', \epsilon_i', s_i') = E(y_{it}/\mu_i, x_i', \epsilon_i') = \mu_i + x_{it}'\beta + \gamma\epsilon_{it} \tag{11.27}$$

It follows that, if we could observe ϵ_{it} when $s_{it} = 1$, then we could test for selectivity bias by including the ϵ_{it} as an additional regressor in fixed effects estimation and testing $H_0; \gamma = 0$ using standard methods. While ϵ_{it} cannot be observed, it can be estimated whenever $s_{it} = 1$ because ϵ_{it} is simply the error of a Tobit model.

When h_{it} is observed, Wooldridge's (1995) test for selection bias is as follows:

Step 1: For each $t = 1, 2, \ldots, T$, estimate the equation

$$h_{it} = \max(0, x_i'\delta_t + \epsilon_{it}) \tag{11.28}$$

by standard Tobit, where $\delta_t' = (\delta_{t0}, \delta_{t1}', \ldots, \delta_{tT}')$ and x_i now has unity as its first element. For $s_{it} = 1$, let $\widehat{\epsilon}_{it} = h_{it} - x_i'\widehat{\delta}_t$ denote the Tobit residuals.

Step 2: Estimate the equation

$$\widetilde{y}_{it} = \widetilde{x}_{it}'\beta + \gamma\widetilde{\epsilon}_{it} + residuals \tag{11.29}$$

by pooled OLS using those observations for which $s_{it} = 1$. $\widetilde{x}_{it}$ and $\widetilde{y}_{it}$ were defined above, and

$$\widetilde{\epsilon}_{it} = \widehat{\epsilon}_{it} - \left(\sum_{r=1}^{T} s_{ir}\widehat{\epsilon}_{ir}/T\right). \tag{11.30}$$

Step 3: Test H_0; $\gamma = 0$ using the t-statistic for $\widehat{\gamma}$. A serial correlation and heteroskedasticity-robust standard error should be used unless $E[\nu_i\nu_i'/\mu_i, x_i', s_i] = \sigma_\nu^2 I_T$. This robust standard error is given in the Appendix to Wooldridge's (1995) paper.

The second case considered by Wooldridge is when h_{it} is not observed. In this case, one conditions on s_i rather than ϵ_i. Using iterated expectations, this gives

$$\begin{aligned} E(y_{it}/\mu_i, x_i', s_i') &= \mu_i + x_{it}'\beta + \gamma E(\epsilon_{it}/\mu_i, x_i', s_i') \\ &= \mu_i + x_{it}'\beta + \gamma E(\epsilon_{it}/x_i', s_i') \end{aligned} \tag{11.31}$$

If the ϵ_{it} were independent across t, then $E(\epsilon_{it}/x_i', s_i') = E(\epsilon_{it}/x_i', s_{it})$. The conditional expectation we need to estimate is $E[\epsilon_{it}/x_i', s_{it} = 1] = E[\epsilon_{it}/x_i', \epsilon_{it} > -x_i'\delta_t]$. Assuming that the var$(\epsilon_{it}) = 1$, we get $E[\epsilon_{it}/x_i', \epsilon_{it} > -x_i'\delta_t] = \lambda(x_i'\delta_t)$ where $\lambda(.)$ denotes the inverse Mills ratio.

When h_{it} is not observed, Wooldridge's (1995) test for selection bias is as follows:

Step 1: For each $t = 1, 2, \ldots, T$, estimate the equation

$$\Pr[s_{it} = 1/x_i'] = \Phi(x_i'\delta_t) \tag{11.32}$$

using standard probit. For $s_{it} = 1$, compute $\widehat{\lambda}_{it} = \lambda(x_i'\widehat{\delta}_t)$.

Step 2: Estimate the equation

$$\widetilde{y}_{it} = \widetilde{x}_{it}'\beta + \gamma\widetilde{\lambda}_{it} + \text{residuals} \tag{11.33}$$

by pooled OLS using those observations for which $s_{it} = 1$. $\widetilde{x}_{it}$ and $\widetilde{y}_{it}$ were defined above, and

$$\widetilde{\lambda}_{it} = \widehat{\lambda}_{it} - \left(\sum_{r=1}^{T} s_{ir}\widehat{\lambda}_{ir}/T_i\right)$$

Step 3: Test H_0; $\gamma = 0$ using the t-statistic for $\gamma = 0$. Again, a serial correlation and heteroskedasticity-robust standard error is warranted unless

$$E(\nu_i\nu_i'/\mu_i, x_i', s_i) = \sigma^2 I_T \qquad \text{under } H_0.$$

Both tests proposed by Wooldridge (1995) are computationally simple involving variable addition tests. These require either Tobit residuals or inverse Mills ratios obtained from probit estimation for each time period. This is followed by fixed effects estimation.

For the random effects model, Verbeek and Nijman (1992) suggest including three simple variables in the regression to check for the presence of selection bias. These are (i) the number of waves the ith individual participates in the panel, T_i, (ii) a binary variable taking the value 1 if and only if the ith individual is observed over the entire sample, $\prod_{r=1}^{T} s_{ir}$, and (iii) $s_{i,t-1}$ indicating whether the individual was present in the last period. Intuitively, testing the significance of these variables checks whether the pattern of missing observations affects the underlying regression. Wooldridge (1995) argues that the first two variables have no time variation and cannot be implemented in a fixed effects model. He suggested other variables to be used in place of $\widehat{\lambda}_{it}$ in a variable addition test during fixed effects estimation. These are $\sum_{r\neq t}^{T} s_{ir}$ and $\prod_{r\neq t}^{T} s_{ir}$. Such tests have the computational simplicity advantage and the need to only observe x_{it} when $s_{it} = 1$.[10]

Nicoletti (2006) is critical of this literature and argues that the *missing at random* condition can be verified only in two cases: (a) when additional information is available to recover the distribution of the variables affected by nonresponse, (b) when some *untestable* assumptions are imposed on the relationship between the missing variables and the probability of responding. Nicoletti classifies the different estimation approaches to take into account the problem of nonresponse into five categories: (i) propensity score methods, used in the evaluation of treatment effects; (ii) imputation methods, used in sample surveys to deal with the nonresponse problem; (iii) econometric sample selection correction methods, a la Heckman (1979); (iv) methods using external data sources such as population registers and refreshment samples, see Hirano et al. (2001); (v) partial identification methods a la Manski (1995). Nicoletti argues that the propensity score and imputation methods are based on the assumption that the data are *missing at random* and hence their validity can only be verified when the missing data are observed! On the other hand, the econometric sample selection correction methods, relax the *missing at random* assumption, and specify a joint model for the dependent variable and the dummy indicating selection given a set of explanatory variables. This approach has been criticized because of the restrictive assumptions on the joint distribution of the errors, which are untestable. Nicoletti (2006, p. 467) adds that the "...sample selection correction approach relaxes one untestable assumption by replacing it by another untestable assumption. The choice between either accepting the *missing at random* condition or imposing a joint distributional assumption is not easy. Any decision is to some extent arbitrary and cannot be submitted to a test procedure."

11.5 Censored and Truncated Panel Data Models

So far, we have studied economic relationships, say labor supply, based on a random sample of individuals where the dependent variable is one if the individual is employed and zero if the individual is unemployed. However, for these random samples, one may observe the number of hours worked if the individual is employed. This sample is censored in that the hours worked are reported as zero if the individual does not work and the regression model is known as the Tobit model.[11] Heckman and MaCurdy (1980) consider a fixed effects Tobit model to estimate a life-cycle model of female labor supply. They argue that the individual effects have a specific meaning in a life-cycle model and therefore cannot be assumed independent of the x_{it}. Hence, a fixed effects rather than a random effects specification is appropriate. For this fixed effects Tobit model:

$$y_{it}^* = x_{it}'\beta + \mu_i + \nu_{it} \tag{11.34}$$

with $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$ and

$$\begin{aligned} y_{it} &= y_{it}^* \text{ if } y_{it}^* > 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{11.35}$$

where y_{it} could be the expenditures on a car or a house, or the number of hours worked. This will be zero if the individual does not buy a car or a house or if the individual is unemployed.[12] As in the fixed effects probit model, the μ_i cannot be swept away and as a result β and σ_ν^2 cannot be estimated consistently for T fixed, since the inconsistency in the μ_i is transmitted to β and σ_ν^2. Heckman and MaCurdy (1980) suggest estimating the log-likelihood using iterative methods. Using Monte Carlo experiments with $N = 1000$, $T = 2, 3, 5, 8, 10$ and 20, Greene (2004a, b) finds that the MLE for the Tobit model with fixed effects exhibits almost no bias even though in each data set in the design, roughly 40 to 50% of the observations were censored. For the truncated panel data regression model, Greene finds some downward bias in the estimates toward 0. He also finds that the estimated standard deviations are biased downward in all cases. For the truncated regression model, Greene (2004b) finds that the MLE/FE estimator underestimates everything, the slopes, the standard errors and the marginal effects.

Honoré (1992) suggested trimmed least absolute deviations and trimmed least squares estimators for truncated and censored regression models with fixed effects defined in (11.34). These are semiparametric estimators with no distributional assumptions necessary on the error term. The main assumption is that the remainder error ν_{it} is independent and identically distributed conditional on the x_{it} and the μ_i for $t = 1, \ldots, T$. Honoré (1992) exploits the symmetry in the distribution of the latent variables and finds that when the true values of the parameters are known, trimming can transmit the same symmetry in distribution to the observed variables. This generates orthogonality conditions which must hold at the true value of the parameters. Therefore, the resulting GMM estimator is consistent provided the orthogonality conditions are satisfied at a unique point in the parameter space. Honoré (1992) shows that these estimators are consistent and asymptotically normal. Monte Carlo results show that as long as $N \geqslant 200$, the asymptotic distribution is a good approximation of the small sample distribution. However, if N is small, the small sample distribution of these estimators is skewed.

Honoré and Kyriazidou (2000a) review estimators for censored regression and sample selection panel data models with unobserved individual specific effects and show how they can be easily extended to other Tobit-type models. The proposed estimators are semiparametric and do not require the parametrization of the distribution of the unobservables. However, they do require that the explanatory variables be strictly exogenous. This rules out lags of the dependent variables among the regressors. The general approach exploits stationarity and exchangeability assumptions on the models' transitory error terms in order to construct moment conditions that do not depend on the individual specific effects.

Kyriazidou (1997) studies the panel data sample selection model, also known as the Type 2 Tobit model with

$$y^*_{1it} = x'_{1it}\beta_1 + \mu_{1i} + \nu_{1it} \tag{11.36}$$

$$y^*_{2it} = x'_{2it}\beta_2 + \mu_{2i} + \nu_{2it} \tag{11.37}$$

where

$$\begin{aligned} y_{1it} &= 1 \quad \text{if} \quad y^*_{1it} > 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

and

$$\begin{aligned} y_{2it} &= y^*_{2it} \quad \text{if} \quad y_{1it} = 1 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Kyriazidou suggests estimating β_1 by one of the estimation methods for discrete choice models with individual effects that were discussed in Sect. 11.1. Next, μ_{2i} is eliminated by first-differencing the data for which y^*_{2it} is observed. With this sample selection, Kyriazidou (1997) focuses on individuals for whom $x'_{1it}\beta_1 = x'_{1is}\beta_1$. For these individuals, the same first-differencing that will eliminate the fixed effects will also eliminate the sample selection. This suggests a two-step Heckman procedure where β_1 is estimated in the first step and then β_2 is estimated by applying OLS to the first differences but giving more weight to observations for which $(x_{1it} - x_{1is})'\widehat{\beta_1}$ is close to zero. This weighting can be done using a Kernel whose bandwidth h_N shrinks to zero as the sample size increases. The resulting estimator is $\sqrt{Nh_N}$ consistent and asymptotically normal. Monte Carlo results for $N = 250$, 1000 and 4000 and $T = 2$ indicate that this estimator works well for sufficiently large data sets. However, it is quite sensitive to the choice of the bandwidth parameters.

Charlier, Melenberg and van Soest (2001) apply the methods proposed by Kyriazidou (1997) to a model of expenditure on housing for owners and renters using an endogenous switching regression. The data is based on three waves of the Dutch Socio-Economic Panel from 1987–1989. The share of housing in total expenditure is modeled using a household specific effect, family characteristics, constant-quality prices, and total expenditure, where the latter is allowed to be endogenous. Estimates from a random effects model are compared to estimates from a linear panel data model in which selection only enters through the fixed effects, and a Kyriazidou type estimator allowing for fixed effects and a more general type of selectivity. Hausman-type tests reject the random effects and linear panel data models as too restrictive. However, the overidentification restrictions of the more general semiparametric fixed effects model of Kyriazidou (1997) were rejected suggesting possible misspecification.

Honoré (1993) also considers the dynamic Tobit model with fixed effects, i.e.,

$$y^*_{it} = x'_{it}\beta + \lambda y_{i,t-1} + \mu_i + \nu_{it} \tag{11.38}$$

with $y_{it} = \max\{0, y_{it}^*\}$ for $i = 1, \ldots, N; t = 1, \ldots, T$. The basic assumption is that ν_{it} is IID$(0, \sigma_\nu^2)$ for $t = 1, \ldots, T$, conditional on y_{i0}, x_{it} and μ_i. Honoré (1993) shows how to trim the observations from a dynamic Tobit model so that the symmetry conditions are preserved for the observed variables at the true values of the parameters. These symmetry restrictions are free of the individual effects and no assumption is needed on the distribution of the μ_i or their relationship with the explanatory variables. These restrictions generate orthogonality conditions which are satisfied at the true value of the parameters. The orthogonality conditions can be used in turn to construct method of moments estimators. Honoré (1993) does not prove that the true values of the parameters are the only values in the parameter space where the orthogonality conditions are satisfied. This means that the resulting GMM estimator is not necessarily consistent. Using Monte Carlo experiments, Honoré shows that MLE for a dynamic Tobit model with fixed effects performs poorly, whereas the GMM estimator performs quite well, when λ is the only parameter of interest. The assumption that the ν_{it} are IID is too restrictive, especially for a dynamic model. Honoré relaxes this assumption to the case of stationary ν_{it} for $t = 1, \ldots, T$ conditional on the x_{it} and the μ_i. Still, this assumption is likely to be violated by many interesting economic models.

Hu (2002) proposes a method for estimating a censored dynamic panel data model with individual fixed effects and lagged *latent* dependent variables. Censoring destroys a certain symmetry between the latent variables. Hu shows that one can artificially truncate the observations in such a way that the symmetry is restored. Based on the restored symmetry, orthogonality conditions are constructed and GMM estimation can be implemented. Although it is hard to prove identification for nonlinear GMM, Hu shows that based on the moment conditions, one can still construct valid asymptotic confidence intervals for the parameters of interest. This is applied to matched data from the 1973 and 1978 March CPS and social security adminstration earnings records to estimate a dynamic earnings model for a sample of men living in the South during 1957-1973, by race. The results suggest that white men's earnings process appears to be more persistent than that of black men (conditional on individual heterogeneity).

Arellano, Bover and Labeaga (1999) consider a linear autoregressive model for a latent variable which is only partly observed due to a selection mechanism:

$$y_{it}^* = \alpha y_{i,t-1}^* + \mu_i + \nu_{it} \tag{11.39}$$

with $|\alpha| < 1$ and $E(\nu_{it}/y_{i1}^*, \ldots, y_{i,t-1}^*) = 0$. The variable y_{it}^* is observed subject to endogenous selection. Arellano, Bover and Labeaga (1999) show that the intractability of this dynamic model subject to censoring using a single time series can be successfully handled using panel data by noting that individuals without censored past observations are exogenously selected. They propose an asymptotic least squares method to estimate features of the distribution of the censored endogenous variable conditional on its past. They apply these methods to a study of female labor supply and wages using two different samples from the PSID covering the periods 1970–1976 and 1978–1984.

Vella and Verbeek (1999) suggest two-step estimators for a wide range of parametric panel data models with censored endogenous variables and sample selection bias. This generalizes the treatment of sample selection models by Nijman and Verbeek (1992) to a wide range of selection rules. This also generalizes the panel data dummy endogenous regressor model in Vella and Verbeek (1998) by allowing for other forms of censored endogenous regressors. In addition, this analysis shows how Wooldridge's (1995) estimation procedures for sample selection can be applied to more general specifications. The two-step procedure derives estimates of the unobserved heterogeneity responsible for the endogeneity/selection bias in the first step. These in turn are included as additional regressors in the primary equation. This is computationally simple compared to maximum likelihood procedures, since it requires only one-dimensional numerical integration. The panel nature of the data allows adjustment, and testing, for two forms of endogeneity and/or sample selection bias. Furthermore, it allows for dynamics and state dependence in the reduced form. This procedure is applied to the problem of estimating the impact of weekly hours worked on the offered hourly wage rate:

$$\begin{aligned} w_{it} &= x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + m(\text{hours}_{it};\beta_3) + \mu_i + \eta_{it} \qquad (11.40)\\ \text{hours}^*_{it} &= x'_{3,it}\theta_1 + \text{hours}_{i,t-1}\theta_2 + \alpha_i + \nu_{it}\\ \text{hours}_{it} &= \text{hours}^*_{it} \quad \text{if hours}^*_{it} > 0\\ \text{hours}_{it} &= 0, \quad w_{it} \text{ not observed if hours}^*_{it} \leqslant 0. \end{aligned}$$

Here, w_{it} represents log of the hourly wage for individual i at time t; $x_{1,it}$ and $x_{3,it}$ are variables representing individual characteristics, $x_{2,it}$ are work place characteristics for individual i; hours^*_{it} and hours_{it} represent desired and observed number of hours worked; m denotes a polynomial of known length with unknown coefficients β_3. This is estimated using data for young females from the NLSY for the period 1980-87. This included a total of 18400 observations of which 12039 observations report positive hours of work in a given period.

Semykina and Wooldridge (2010) consider the estimation of panel data models with sample selection, as well as endogenous explanatory variables and unobserved heterogeneity. They extend the approach of Wooldridge (1995) of testing for selection bias to allow for some variables to be correlated with idiosyncratic errors. In fact, they propose simple variable addition tests to detect selection biase that is due to violations of strict exogeneity. They also propose two estimation procedures that correct for selection in the presence of endogenous regressors, assuming that appropriate instruments are available. The tests are based on the FE-2SLS estimator, thereby permitting arbitrary correlation between unobserved heterogeneity and explanatory variables. The first correction procedure is parametric relying on the assumption that the errors in the selection equation are normally distributed. The second procedure estimates the model parameters semiparametrically using series estimators.

Semykina and Wooldridge (2013) argue that, for dynamic panel data models with arbitrary selection patterns, the use of first-differencing looses much of the data. Instead, they model the conditional expectation of the unobserved effect as a linear function of the exogenous variables and the initial condition as in Wooldridge (2005). Using backward substitution for the lagged dependent variable, they obtain an equation that contains lags of the exogenous explanatory variables and the initial condition, but no lags of the dependent variable. As a result, selection correction reduces to a contemporaneous selection problem of the type studied in Wooldridge (1995) with strictly exogenous variables. Assuming normality and focusing on selection (period by period) simplifies the derivation of the correction term. Once the correction term is obtained, the augmented equation can be consistently estimated by nonlinear least squares (NLS) or GMM. Semykina and Wooldridge (2013) also propose a simple test for selection bias that is based on the addition of a selection term to the first-difference equation and testing for the significance of this term. The methods are applied to estimating dynamic earnings equations for women using the Panel Study of Income Dynamics over the period 1980–1992. Since it is necessary to observe the initial condition, only females for whom 1980 earnings are available are included in the sample. The final sample consists of 579 women or 6948 observations over the 12-year period (1981–1992).

Frederiksen, Honoré and Hu (2007) consider an alternative way of modeling dynamic discrete choice panel models by using the duration in the current state as a covariate. They propose estimators that allow for group-specific effect in parametric and semiparametric versions of the model. The proposed method is illustrated with an empirical analysis of job durations allowing for firm-level effects using Danish data on all employees of all establishments in the private sector observed over the period 1980 to 2000.

11.6 Empirical Applications

There are many empirical applications illustrating the effects of attrition bias; see Hausman and Wise (1979) for a study of the Gary Income Maintenance Experiment. For this experimental panel study of labor supply response, the treatment effect is an income guarantee/tax rate combination. People who benefit from this experiment are more likely to remain in the sample. Therefore, the selection rule is nonignorable, and attrition can overestimate the treatment effect on labor supply. For the Gary Income Maintenance Experiment, Hausman and Wise (1979) found little effect of attrition bias on the experimental labor supply response. Similar results were obtained by Robins and West (1986) for the Seattle and Denver Income Maintenance Experiments. For the latter sample, attrition was modest (11% for married men and 7% for married women and single heads during the period studied) and its effect was not serious enough to warrant extensive correction procedures.

Ridder (1992) studied the determinants of the total number of trips using the first seven waves of the dutch transportation panel (DTP). This panel was commissioned by the Department of Transportation in the Netherlands to evaluate the effect of price increases on the use of public transportation. The first wave of interviews was conducted in March 1984. There is heavy attrition in the DTP with only 38% of the original sample participating in all seven waves of the panel. Ridder (1992) found that nonrandom attrition from the DTP did not bias time-constant regression coefficients. However, it did bias the time-varying coefficients. Ridder (1992) also found that the restrictions imposed by the standard Hausman and Wise (1979) model for nonrandom attrition on the correlations between individual effects and random shocks may even prevent the detection of nonrandom attrition.

Nijman and Verbeek (1992) studied the effects of nonresponse on the estimates of a simple life-cycle consumption function using a Dutch panel of households interviewed over the period April 1984–March 1987. Several tests for attrition bias were performed, and the model was estimated using (i) one wave of the panel, (ii) the balanced subpanel, and (iii) the unbalanced panel. For this application, attrition bias was not serious. The balanced subpanel estimates had implausible signs, while the one-wave estimates and the unbalanced panel estimates gave reasonably close estimates with the latter having lower standard errors.

Ziliak and Kniesner (1998) examine the importance of sample attrition in a life-cycle labor supply using both a Wald test comparing attriters to nonattriters and variable addition tests based on formal models of attrition. Estimates using waves I-XXII of the PSID (interview years 1968-1989) show that nonrandom attrition is of little concern when estimating prime age male labor supply because the effect of attrition is absorbed into fixed effects in labor supply.

Dionne, Gagne′, and Vanasse (1998) estimate a cost model based on an incomplete panel of Ontario trucking firms. The data consists of 445 yearly observations of general freight carriers in Ontario observed over the period 1981-88. It includes 163 firms for which information is available for 2.7 years on average. The cost-input demand system is jointly estimated with a bivariate probit selection model of entry and exit from the sample. A test for selectivity bias reveals potential bias related to exit but not entry from the sample.

Vella and Verbeek (1998) estimate the union premium for young men over a period of declining unionization (1980-87). The panel data is taken from the NLSY and includes 545 full time working males who completed their schooling by 1980. The probability of union membership is estimated using a dynamic random effects probit model. The coefficient of lagged union status is estimated at 0.61 with a standard error of 0.07 indicating a positive and statistically significant estimate of state dependence. OLS estimates of the wage equation yield a union wage effect of 15% to 18% depending on whether occupational status dummies are included or not. These estimates are contaminated by endogeneity. The corresponding fixed effects estimates are much lower yielding 7.9% to 8.0%. These estimates eliminate only the endogeneity operating through the individual specific effects. Thus, any time-varying endogeneity continues to contaminate these estimates. Including correction terms based on the estimated union model yield negative significant coefficients and reveal selection bias. This indicates that workers who receive lower wages, after conditioning on their characteristics and in the absence of unions, are most likely to be in the union. This is consistent with the findings that minority groups who are lower paid for discriminatory reasons have a greater tendency to seek union employment than whites. Vella and Verbeek conclude that the union effect is approximately 21% over the period studied. However, the return to unobserved heterogeneity operating through union status is substantial, making the union premium highly variable among individuals. Moreover, this union premium is sensitive to the pattern of sorting into union employment allowed in the estimation. Problem 11.6 asks the reader to replicate some of the results in this paper.

11.7 Empirical Example: Nurses Labor Supply

Shortage of nurses is a problem in several countries. It is an unsettled question whether increasing wages constitute a viable policy for extracting more labor supply from nurses. Askildsen, Baltagi and Holmås (2003) use a unique matched panel data set of Norwegian nurses covering the period 1993-1998 to estimate wage elasticities. The data set collected from different official data registers and Statistics Norway includes detailed information on 19,638 individuals over 6 years totalling 69,122 observations. Female nurses younger than 62 years of age who were registered with a complete nursing qualification and employed by municipalities or counties were included in the sample. For the sample of female nurses considered, the average age was 37 years with 35% of the nurses being single. The majority of these nurses worked in somatic hospitals (62%) or nursing homes (20%) with the remaining nurses engaged in home nursing (10%), at psychiatric institutions (5%), in health services (1%), and others (3%). Senior nurses comprised only 2% of the sample, while 16% were ward nurses, 20% were nursing specialists and the remaining majority (62%) worked as staff nurses. The average years of experience during the sample period was 12.5 years, and the average number of children below 18 years of age was 1.2. Nurses with children below the age of 3 comprised 22% of the sample, while those with children between the ages of 3 and 7 comprised 29% of the sample.

Verbeek and Nijman (1992) proposed simple tests for sample selection in panel data models. One test is to include variables measuring whether the individual is observed in the previous period, whether the individual is observed in all periods and the total number of periods the individual is observed, see Sect. 11.4. The null hypothesis says that these variables should not be significant in our model if there are no sample selection problems. Another test, a Hausman-type test, compares the fixed effects estimator from the balanced sample as opposed to an unbalanced sample. Both tests rejected the null hypothesis of no sample selection.

Table 11.4 reproduces the conditional logit model estimates as the first step of the Kyriazidou (1997) estimator. A number of variables were used that characterized the regions and municipalities where the individuals live (centrality, female work participation rates, availability of kindergarten and whether there is a hospital in the municipality). These variables were closely connected to the participation decision, and conditional on this are assumed not to affect hours of work. Job-related variables were excluded since they were not observed for those who did not participate. The conditional logit estimates were then used to construct kernel weights with the bandwidth set to with h = 1. A Hausman test based on the weighted and unweighted estimates gave a value of the test statistic ($\chi^2_{23} = 821.27$) that clearly rejected the null hypothesis of no selection. As instruments for the wage of nurses, the authors used the financial situation of the municipality, measured by lagged net financial surplus in preceding period. Also, the lagged mean wage of auxiliary nurses work-

Table 11.4 Participation equation. Conditional logit Estimates

Educated as nursing specialist	0.6155**
	(0.0685)
Age	0.1125**
	(0.0340)
Age2	-0.0035**
	(0.0004)
Single	-0.1256*
	(0.0550)
Number of children	-0.2640**
	(0.0457)
Children < 3	-0.1424**
	(0.0424)
Children 3 – 7	0.0725
	(0.0385)
Children > 7	-0.0619
	(0.0369)
Disable	-1.2678**
	(0.2240)
Hospital in municipality	0.6463**
	(0.0617)
Availability kindergarten	0.3303
	(0.2511)
Participation rate	0.0345**
	(0.0073)
East-Norway	0.5275**
	(0.1198)
South-Norway	0.8874**
	(0.1448)
West-Norway	-0.8383**
	(0.1318)
Mid-Norway	1.4610**
	(0.1429)
Municipality size	-0.0082**
	(0.0003)
Centrality level 1	-0.0471
	(0.1615)
Centrality level 2	-0.5202**
	(0.1809)
Centrality level 3	-0.5408**
	(0.1387)
Log likelihood	-22287.461
Number of observations	61464

Standard errors in parentheses. ** and * is statistically different from zero at one and five percent significance level, respectively. *Source* Askildsen, Baltagi and Holmås 2003. Reproduced by permission of Wiley

Table 11.5 Nurses' Labor Supply

Ln wage	0.2078*
	(0.0942)
Shift work	-0.0111**
	(0.0007)
Shift work 2	-0.00000
	(0.00001)
Hour_35.5	-0.0397**
	(0.0053)
Disable	-0.2581**
	(0.0261)
Age	-0.0098*
	(0.0048)
Age2	0.0003**
	(0.00003)
Single	0.0205**
	(0.0035)
Number of children	-0.0991**
	(0.0035)
Children < 3	-0.0495**
	(0.0028)
Children 3 - 7	-0.0177**
	(0.0024)
Children > 7	-0.0307**
	(0.0021)
Psychiatric	0.0466**
	(0.0092)
Home nursing	-0.0206**
	(0.0067)
Health service	-0.0567**
	(0.0148)
Nursing home	-0.0177**
	(0.0059)
Other	0.0024
	(0.0078)
Nursing specialist	0.0144*
	(0.0067)
Ward nurse	-0.0004
	(0.0076)
Senior nurse	0.0057
	(0.0123)
East Norway	-0.0622**
	(0.0131)
South Norway	-0.0802**
	(0.0170)
West Norway	-0.1157**
	(0.0218)
Mid Norway	-0.1011**
	(0.0188)
Municipality size	0.0002*
	(0.00007)
Constant	-0.0068**
	(0.0014)
Number of observations	121622

Standard errors in parentheses. ** and * is statistically different from zero at one and five percent significance level, respectively. *Source* Askildsen, Baltagi and Holmås (2003). Reproduced by permission of Wiley ing in the same municipality as the nurse, and each nurse's work experience. These variables are assumed to affect wages of nurses but not their hours of work. The instruments pass the Hausman test of over-identifying restrictions. The results of the Kyriazidou instrumental variable estimator are given in Table 11.5. Age had a significant negative effect. Nurses worked shorter hours as they became older but to a diminishing degree. The effect of family variables was as expected. Being single had a positive and significant effect on hours of work. The presence of children in the home had a negative impact on hours of work. Nurses working in psychiatric institutions worked longer hours compared to the base category somatic hospitals, whereas shorter hours were supplied by nurses engaged in home nursing, as well as in nursing homes. Labor supply was highest in the less densely populated Northern Norway (the base category). This may reflect the fact that hours of work were not allowed to vary as much in these areas. Compared to a staff nurse, which served as the base work type category, nursing specialists and senior nurses all worked longer hours. The estimated wage elasticity after controlling for individual heterogeneity, sample selection, and instrumenting for possible endogeneity was 0.21. Individual and institutional features were statistically significant and important for working hours. Contractual arrangements as represented by shift work were also important for hours of work, and omitting information about this common phenomenon will underestimate the wage effect.

11.8 Further Reading

Das (2003) presents estimators for nonparametric panel data models with additive fixed effects. Das analyzes a model in which the entire regressor vector, consisting of time-varying as well as time-invariant regressors, is correlated with the individual effect but there are insufficient exclusion restrictions to permit direct instrumental variables estimation of the model. This is applied to the problem of estimating the returns to education in a cohort of mature men aged 25-40 from the 1976 and 1981 waves of young men's cohort in the National Longitudinal Survey. Log of hourly wage is specified as a semiparametric function of education and experience, where both variables are treated as endogenous and correlated with the unobserved individual effects. Other time-invariant effects such as race, and time-varying variables, such as union affiliation, marital status, and health enter additively and are assumed to be exogenous. The nonparametric estimates indicate that the average returns to education are 0.076 for those men with high school or less education, 0.134 for those with 13 to 16 years of education and 0.188 for those with college or more. Although the last estimate is not significant. This is compared to an OLS estimate of 0.081 and an IV estimate of 0.092. Das concludes that endogeneity and nonlinearity may be jointly important in the returns to schooling.

Wooldridge (1997) considers the estimation of multiplicative, unobserved components panel data models without imposing a strict exogeneity assumption on the conditioning variables. A robust method of moments estimator is proposed which requires only a conditional mean assumption. This applies to binary choice models with multiplicative unobserved effects, and models containing parametric nonlinear transformations of the endogenous variables. This model is particularly suited to nonnegative explained variables, including count variables. In addition, it can also be applied to certain nonlinear Euler equations. Wooldridge (1999) offers some distribution-free estimators for multiplicative unobserved components panel data models. Requiring only the correct specification of the conditional mean, the multinomial quasi-conditional MLE is shown to be consistent and asymptotically normal. This estimation method is popular for estimating fixed effects count models, see Hausman, Hall and Griliches (1984). Wooldridge's results show that it can be used to obtain consistent estimates even when the dependent variable y_{it} is not a vector of counts. In fact, y_{it} can be a binary response variable, a proportion, a nonnegative continuously distributed random variable, or it can have discrete and continuous characteristics. Neither the distribution of y_{it} nor its temporal dependence are restricted. Additional orthogonality conditions can be used in a GMM framework to improve the efficiency of the estimator. Finally, Wooldridge (2000) proposes a method of estimating very general, nonlinear, dynamic, unobserved effects panel data models with feedback. Wooldridge shows how to construct the likelihood function for the conditional maximum likelihood estimator in dynamic, unobserved effects models where not all conditioning variables are strictly exogenous. A useful innovation is the treatment of the initial conditions which offers a flexible, relatively simple alternative to existing methods.

Hansen (1999) considers the estimation of threshold panel regressions with individual specific effects. This is useful for situations where the regression function is not identical across all observations in the sample. In fact, the observations are divided into two regimes depending on whether a threshold variable q_{it} is smaller or larger than the threshold γ:

$$y_{it} = \mu_i + \beta_1' x_{it} 1(q_{it} \leqslant \gamma) + \beta_2' x_{it} 1(q_{it} > \gamma) + \nu_{it}$$

where $1(.)$ is the indicator function. The regimes are distinguished by differing slopes β_1 and β_2. Hansen (1999) proposes a least squares procedure to estimate the threshold and regression slopes using fixed effects transformations. Non-standard asymptotic theory with T fixed and $N \to \infty$ is developed to allow the construction of confidence intervals and test of hypotheses. This method is applied to a panel of 565 U.S. firms observed over the period 1973-87 to test whether financial constraints affect investment decisions. Hansen finds overwhelming evidence of a double threshold effect which separates the firms based on their debt to asset ratio. The weakness of this approach is that it does not allow for heteroskedasticity, lagged dependent variables, endogenous variables , and random effects.

Honoré, Vella and Verbeek (2008) give a comprehensive review of parametric and semiparametric estimation methods for panel models with attrition, selection bias and censoring under different distributional assumptions.

The reader is encouraged to read Chap. 6 of the Oxford Handbook of Panel Data by Greene (2015) entitled Panel data models for discrete choice models, also Chap. 7 by Lee (2015) entitled Panel conditional and multinomial logit estimators.

11.9 Notes

1. For the probit model

$$F(x_{it}'\beta) = \Phi(x_{it}'\beta) = \int_{-\infty}^{x_{it}'\beta} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

and for the logit model

$$F(x_{it}'\beta) = \frac{e^{x_{it}'\beta}}{1 + e^{x_{it}'\beta}}$$

2. Note that for this nonlinear panel data model, it is not possible to get rid of the μ_i by taking differences or performing the Within transformation as in Chap. 2. Graham, Hahn and Powell (2009) show that there is no incidental parameter problem for a panel quantile model with fixed effects and $T = 2$. In fact, the maximum likelihood estimator is numerically equivalent to the least absolute deviations estimator of the differenced model which wipes out the individual effects. This estimator is consistent with T fixed asymptotics as long as the regressors are strictly exogenous.
3. This is programmed in Stata as *xtlogit* with the fe option and will be illustrated later on in this chapter.
4. On the other hand, if there are no random individual effects, the joint likelihood will be the product of the marginals and one can proceed as in the usual cross-sectional limited dependent variable case, see Greene (2003).
5. For a good read of simulation methods for limited dependent variable panel models, see Keane (1993).
6. The main limitations of the Honoré and Kyriazidou (2000b) approach are (i) the assumption that the errors in the underlying threshold-crossing model are independent over time and (ii) the assumption that $x_{i2}' - x_{i3}'$ has support in a neighborhood of 0. The latter restriction rules out time dummies.
7. Other sufficient conditions for consistency of these estimators are given by Verbeek and Nijman (1996). These are derived for specific selection rules. One interesting and practical sufficient condition that emerges is that the Within estimator is consistent and free of selectivity bias if the probability of being in the sample is constant over time. In this case, the correction for selectivity bias is time-invariant and hence is absorbed in the individual effect term.
8. If the selection rule is unknown, identification problems arise regarding the parameters of interest (see Verbeek and Nijman, 1996).
9. Verbeek and Nijman (1992) show that under nonresponse, the conditions for consistency of the Within estimator are weaker than those for the random effects GLS estimator. This means that the Within estimator is more robust to nonresponse bias than GLS.
10. It is important to point out that both Verbeek and Nijman (1992) as well as Wooldridge (1995) assume that the unobservable effects and the idiosyncratic errors in the selection process are normally distributed. Kyriazidou's (1997) treatment of sample selection leaves the distributions of all unobservables unspecified.

11. Alternatively, one could condition on the set of continuously working individuals, i.e., use only the sample with positive hours of work. In this case the sample is considered truncated (see Greene, 2003).
12. Researchers may also be interested in panel data economic relationships where the dependent variable is a count of some individual actions or events, see Sect. 10.6.

11.10 Problems

11.1 *Fixed effects logit model*. In Sect. 11.1 we considered the fixed effects logit model with $T = 2$.

(a) In this problem, we look at $T = 3$ and we ask the reader to compute the conditional probabilities that would get rid of the individual effects by conditioning on $\sum_{t=1}^{3} y_{it}$. Note that this sum can now be 0, 1, 2, or 3. (Hint: First show that terms in the conditional likelihood function, which are conditioned upon $\sum_{t=1}^{3} y_{it} = 0$ or 3 add nothing to the likelihood. Then focus on terms that condition on $\sum_{t=1}^{3} y_{it} = 1$ or 2.)
(b) Show that for $T = 10$, one has to condition on the sum being $1, 2, \ldots, 9$. One can see that the number of probability computations are increasing. To convince yourself, write down the probabilities conditioning on $\sum_{t=1}^{10} y_{it} = 1$.

11.2 *Dynamic fixed effects logit model with no regressors*. Consider the Chamberlain (1985) fixed effects conditional logit model with a lagged dependent variable given in (11.16). Show that for $T = 3$, $\Pr[A/y_{il} + y_{i2} = 1, \mu_i]$ and therefore $\Pr[B/y_{i1} + y_{i2} = 1, \mu_i]$ do not depend on μ_i. Note that A and B are defined in (11.17) and (11.18), respectively.

11.3 *Dynamic fixed effects logit model with regressors*. Consider the Honoré and Kyriazidou (2000b) fixed effects logit model given in (11.19).

(a) Show that for $T = 3$, $\Pr[A/x_i', \mu_i, A \cup B]$ and $\Pr[B/x_i', \mu_i, A \cup B]$ both depend on μ_i. This means that the conditional likelihood approach will *not* eliminate the fixed effect μ_i.
(b) If $x_{i2}' = x_{i3}'$, show that $\Pr[A/x_i', \mu_i, A \cup B, x_{i2}' = x_{i3}']$ and $\Pr[B/x_i', \mu_i, A \cup B,$
$x_{i2}' = x_{i3}']$ do *not* depend on μ_i.

11.4 *Equivalence of two estimators of the fixed effects logit model*. This is based on Abrevaya (1997). Consider the fixed effects logit model given in (11.4) with $T = 2$. In (11.10) and (11.11) we showed the conditional maximum likelihood of β, call it $\widehat{\beta}_{CML}$, can be obtained by running a logit estimator of the dependent variable $1(\Delta y = 1)$ on the independent variables Δx for the subsample of observations satisfying $y_{i1} + y_{i2} = 1$. Here $1(\Delta y = 1)$ is an indicator function taking the value one if $\Delta y = 1$. Therefore, $\widehat{\beta}_{CML}$ maximizes the log-likelihood

$$\ln L_c(\beta) = \sum_{i \epsilon \vartheta} [1(\Delta y = 1) \ln F(\Delta x \beta) + 1(\Delta y = -1) \ln(1 - F(\Delta x \beta))]$$

where $\vartheta = \{i : y_{i1} + y_{i2} = 1\}$.

(a) Maximize the unconditional log-likelihood for (11.4) given by

$$\ln L(\beta, \mu_i) = \sum_{i=1}^{N} \sum_{t=1}^{2} [y_{it} \ln F(x_{it}'\beta + \mu_i) + (1 - y_{it}) \ln(1 - F(x_{it}'\beta + \mu_i))]$$

with respect to μ_i and show that i

$$\widehat{\mu}_i = \begin{cases} -\infty & \text{if } y_{i1} + y_{i2} = 0 \\ -(x_{i1} + x_{i2})'\beta/2 & \text{if } y_{i1} + y_{i2} = 1 \\ +\infty & \text{if } y_{i1} + y_{i2} = 2 \end{cases}$$

(b) Concentrate the likelihood by plugging $\widehat{\mu}_i$ in the unconditional likelihood and show that

$$\ln L(\beta, \widehat{\mu}_i) = \sum_{i \epsilon \vartheta} 2[1(\Delta y = 1) \ln F(\Delta x'\beta/2) + 1(\Delta y = -1) \ln(1 - F(\Delta x'\beta/2))]$$

Hint: Use the symmetry of F and the fact that

$$1(\Delta y = 1) = y_{i2} = 1 - y_{i1} \quad \text{and} \quad 1(\Delta y = -1) = y_{i1} = 1 - y_{i2} \quad \text{for } i \epsilon \vartheta.$$

(c) Conclude that $\ln L(\beta, \widehat{\mu}_i) = 2 \ln L_c(\beta/2)$. This shows that a scale adjusted maximum likelihood estimator is equivalent to the conditional maximum likelihood estimator, i.e., $\widehat{\beta}_{ML} = 2\widehat{\beta}_{CML}$. Whether a similar result hold for $T > 2$ remains an open question.

11.5 *Binary Response Model Regression* (BRMR). This is based on problem 95.5.4 in *Econometric Theory* by Baltagi (1995). Davidson and MacKinnon (1993) derive an artificial regression for testing hypotheses in a binary response model. For the fixed effects model described in (11.4), the reader is asked to derive the BRMR to test $H_o; \mu_i = 0$, for $i = 1, 2, \ldots, N$. Show that if $F(.)$ is the Logistic (or Normal) cumulative distribution function, this BRMR is simply a weighted least squares regression of logit (or probit) residuals, ignoring the fixed effects, on the matrix of regressors X and the matrix of individual dummies. The test statistic in this case is the explained sum of squares from this BRMR. See solution 95.5.4 in *Econometric Theory* by Gurmu (1996).

11.6 *Union membership*. Using the Vella and Verbeek (1998) study considered in Sect. 11.6, download their data set which is posted on the *Journal of Applied Econometrics* web site and

(a) Replicate their descriptive statistics given in Table I. Confirm that the unconditional union premium is around 15%.
(b) Replicate their random effects probit estimates of union membership given in Table II.
(c) Replicate the wage regressions with union effects given in Table III.
(d) Replicate the wage regressions under unrestricted sorting given in Table V.

11.7 *Initial Condition*. Using the Wooldridge (2005) study considered in Sect. 11.3, download the data set posted on the *Journal of Applied Econometrics* web site and

(a) Replicate the results given in Table I in that article using (*xtprobit, re*) in Stata.
(b) Replicate the average partial effects for union membership for married and non-married men for 1986 and 1987. Verify that the estimates of state dependence in union membership are 0.182 for married men and 0.173 for non-married men, respectively.

11.8 *Beer taxes and motor vehicle fatality rates*. Ruhm (1996) considered the effect of beer taxes and a variety of alcohol-control policies on motor vehicle fatality rates, see Sect. 11.6. The data is for 48 states (excluding Alaska, Hawaii, and the District of Columbia) over the period 1982-1988. Some of the variables in this data set can be downloaded from the Stock and Watson (2003) web site at www.aw.com/stock_watson. Using this data set

(a) Replicate the descriptive statistics given in Table 1 of Ruhm (1996, p. 441).
(b) Replicate to the extent possible the results in Table 2 of Ruhm (1996, p. 444), and verify that beer tax is effective in reducing motor vehicle fatality rates, whereas the breath test law is not. How are the standard errors affected by using the robust option for the variance–covariance matrix?
(c) Test the significance of the state and time dummies. What do you conclude?

11.9 *Identification in a dynamic binary choice panel data model*. This is based on the Appendix of Honoré and Tamer (2006, pp. 627–628). Suppose that (y_{i1}, y_{i2}, y_{i3}) is a random vector such that
$P(y_{i1} = 1/\mu_i) = p_1(\mu_i)$
and
$P(y_{it} = 1/\mu_i, y_{i1}, .., y_{it-1}) = F(\mu_i + \gamma y_{it-1})$ for $t = 2, 3$,
where $p_1(\mu_i)$ is an unknown strictly increasing distribution function taking values between 0 and 1. Show that the sign of γ is identified. (Hint:Consider the probabilities

$P[(y_{i1}, y_{i2}, y_{i3}) = (0, 1, 0)/\mu_i]$ and $P[(y_{i1}, y_{i2}, y_{i3}) = (0, 0, 1)/\mu_i]$. Show that the sign of the difference of these two probabilities identifies the sign of γ).

11.10 *The magazine industry*. Willis (2006) re-examines the study of Cecchetti (1986) on price adjustment behavior in the magazine industry. Cecchetti assumes that a firm's pricing rules are fixed for non-overlapping three-year intervals and estimates the model using a conditional logit specification a la Chamberlain (1980). The data set consists of 38 unique magazines (mag_id) observed over the years 1953–1979. The dependent variable "pr_ch" is a dummy variable referring to whether the price changed between January of the current year and January of the following year. "y_ch" refers to the number of years since the magazine price change; "inf_ch" refers to the cumulative change in inflation since the last price change and "mags_ch" refers to the cumulative change in magazine industry sales since the previous price change. The "group_id" variable corresponds to the groups used in a Chamberlain conditional logit estimation. This can be downloaded from the *Journal of Applied Econometrics* web site. Replicate Table 1 of Willis (2006, p. 342). Column (1) runs a logit regression; while column (2) runs the same logit regression with magazine fixed effects; column (3) runs the Chamberlain conditional logit model with "group_id" allowing three-year intervals for a price change; column (5) runs a logit with magazine and time fixed effects.

11.11 *Bank Failure and Multiple Liability*. Grossman (2001) investigated the effect of multiple liabilities of bank share holders on bank failure rates in U.S. states before the Great Depression. Grossman found that double liability did reduce bank failures in periods where bank failures were not abnormally high. However, it did not guarantee bank stability in times of widespread financial distress. Table 2 of Grossman (2001, p. 151) runs an inverse logit regression of state bank failures (measured by either the failure rate or the asset failure rate) in state i at time t, on its lagged value, and the failure rate among national banks in that state at time t as well as a dummy variable for multiple liabilities which takes the value one if the state had double, triple, or unlimited liability for banks for three consecutive years, and zero otherwise.

(a) Replicate the results in Table 2 of Grossman (2001, p. 151) using the *Bank-Failure* stata data set. How are the standard errors and significance affected by robustifying these regressions?
(b) How are the results affected by including state and time dummies. What do you conclude?

11.12 *Female Labor force participation.* This is the empirical example used in Fernandez-Val (2009) and used as example 1 in this chapter. This data is available in Stata as *lfp_psid*.

(a) Replicate the conditional logit fe results in Table 11.1 in this chapter using *xtlogit, fe*.
(b) Add fixed time effects to the conditional logit fe in part (a). Test the significance of these fixed time effects.
(c) Using *logitfe* run the model in part (a) uncorrected for bias with no time effects.
(d) Using *logitfe* run the model in part (a) uncorrected for bias with individual and time effects.
(e) Add lagged labor force participation, so the model is dynamic and run *logitfe* as in part (d).
(f) Run the dynamic model in part (e) with analytical bias correction option restricting the trimming parameter to 1.
(g) Run the *logitfe* commands with the jackknife options allowing for the *six* different types of jackknife corrections described in Cruz-Gonzalez, Fernandez-Val and Weidner (2017).
(h) Redo part (a) and (b) using *xtprobit, re* rather than *xtlogit, fe.*
(i) Redo parts (c) and (h) using *probitfe* rather than *logitfe*.

11.13 *To Trade or Not to Trade*. This is based on the empirical application to bilateral trade flows between countries using data from Helpman, Melitz and Rubinstein (2008). The data set includes trade flows for 158 countries over the period from 1970 to 1997. This is used to fit probit and logit fe models by Cruz-Gonzalez, Fernandez-Val and Weidner (2017) for the probability of positive trade between country pairs in 1986. The data structure is a pseudo-panel where the two dimensions index countries, with id as importers and jd as exporters. The dependent variable trade is an indicator equal to one if country i imports from country j, and equal to zero otherwise. The specification is a gravity trade equation with explanatory variables including log(distance), border, legal origin, language, colony, currency, federal trade agreements, islands, religion, landlock. These variables are defined in Helpman, et al. (2008) and Cruz-Gonzalez, Fernandez-Val and Weidner (2017). However, the latter add lagged bilateral trade to account for possible state dependence in trade decisions. The data set is available as *trade* in Stata when you install probitfe and logitfe.

(a) Replicate Table 1 of Cruz-Gonzalez, Fernandez-Val and Weidner (2017, p. 534) using *felogit.*
(b) Replicate Table 2 of Cruz-Gonzalez, Fernandez-Val and Weidner (2017, p. 535) using *feprobit.*

References

Abrevaya, J. 1997. The equivalence of two estimators of the fixed-effects logit model. *Economics Letters* 55: 41–43.

Akay, A. 2012. Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics* 27: 1189–1204.

Arellano, M., O. Bover and J.M. Labeaga. 1999. Autoregressive models with sample selectivity for panel data, Chapter 2. In *Analysis of panels and limited dependent variable models*, (eds.) Hsiao, C., K. Lahiri, L.F. Lee and H. Pesaran, 23–48. Cambridge: Cambridge University Press.

Arellano, M., and R. Carrasco. 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics* 115: 125–157.

Askildsen, J.E., B.H. Baltagi, and T.H. Holmås. 2003. Wage policy in the health care sector: A panel data analysis of nurses' labour supply. *Health Economics* 12: 705–719.

Baltagi, B.H. 1995. Testing for fixed effects in logit and probit models using an artificial regression, Problem 95.5.4. *Econometric Theory* 11: 1179.

Baltagi, B.H. 2009. *A companion to econometric analysis of panel data*. Chichester: Wiley.

Bover, O., and M. Arellano. 1997. Estimating dynamic limited variable models from panel data. *Investigaciones Economicas* 21: 141–165.

Butler, J.S., and R. Moffitt. 1982. A computationally efficient quadrature procedure for the one factor multinominal probit model. *Econometrica* 50: 761–764.

Carro, J.M. 2007. Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* 140: 503–528.

Cecchetti, S. 1986. The frequency of price adjustment: A study of the newsstand prices of magazines. *Journal of Econometrics* 31: 255–274.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47: 225–238.

Chamberlain, G. 1984. Panel data, Chapter 22 In *Handbook of econometrics* (eds.) Griliches, Z. and M. Intrilligator, 1247–1318. North-Holland, Amsterdam.

Chamberlain, G. 1985. Heterogeneity, omitted variable bias and duration dependence, Chapter 1. In *Longitudinal analysis of labor market data*, (eds.) Heckman, J.J. and B. Singer, 3–38. Cambridge: Cambridge University Press.

Charlier, E., B. Melenberg, and A. van Soest. 2001. An analysis of housing expenditure using semiparametric models and panel data. *Journal of Econometrics* 101: 71–107.

Chintagunta, P., E. Kyriazidou, and J. Perktold. 2001. Panel data analysis of household brand choices. *Journal of Econometrics* 103: 111–153.

Contoyannis, P., A.M. Jones, and N. Rice. 2004. The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics* 19: 473–503.

Cruz-Gonzalez, M., I. Fernandez-Val, and M. Weidner. 2017. Bias corrections for probit and logit models with two-way fixed effects. *The Stata Journal* 17 (3): 517–545.

Das, M. 2003. Identification and sequential estimation of panel data models with insufficient exclusion restrictions. *Journal of Econometrics* 114: 297–328.

Das, M., and A. van Soest. 1999. A panel data model for subjective information on household income growth. *Journal of Economic Behavior & Organization* 40: 409–426.

Davidson, R., and J.G. MacKinnon. 1993. *Estimation and inference in econometrics*. New York: Oxford University Press.

Dhaene, G., and K. Jochmans. 2015. Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82 (3): 991–1030.

Dionne, G., R. Gagné, and C. Vanasse. 1998. Inferring technological parameters from incomplete panel data. *Journal of Econometrics* 87: 303–327.

Erdem, T., and B. Sun. 2001. Testing for choice dynamics in panel data. *Journal of Business and Economic Statistics* 19: 142–152.

Fernandez-Val, I. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150: 71–85.

Fernandez-Val, I., and M. Weidner. 2016. Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics* 192: 291–312.
Frederiksen, A., B.E. Honoré, and L. Hu. 2007. Discrete time duration models with group-level heterogeneity. *Journal of Econometrics* 141: 1014–1043.
Graham, B.S., J. Hahn, and J.L. Powell. 2009. The incidental parameter problem in a non-differentiable panel data model. *Economics Letters* 105: 181–182.
Greene, W.H. 2003. *Econometric analysis*. New Jersey: Prentice Hall.
Greene, W.H. 2004a. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* 7: 98–119.
Greene, W.H. 2004b. Fixed effects bias due to the incidental parameters problem in the Tobit model. *Econometric Reviews* 23: 125–147.
Greene, W.H. 2015. Panel data models for discrete choice models, Chapter 6. In *Oxford handbook of panel data*, (ed.) Baltagi, B.H., 171–201 Oxford: Oxford University Press.
Grossman, R.S. 2001. Double liability and bank risk taking. *Journal of Money, Credit and Banking* 33: 143–159.
Gurmu, S. 1996. Testing for fixed effects in logit and probit models using an artificial regression, solution 95.5.4. *Econometric Theory* 12: 872–874.
Hahn, J., and W. Newey. 2004. Jacknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72: 1295–1319.
Hajivassiliou, V.A. 1994. Estimation by simulation of the external debt repayment problems. *Journal of Applied Econometrics* 9: 109–132.
Hansen, B.E. 1999. Threshold effects in non-dynamic panels: Estimation, testing and inference. *Journal of Econometrics* 93: 345–368.
Hausman, J.A., and D. Wise. 1979. Attrition bias in experimental and panel data: The gary income maintenance experiment. *Econometrica* 47: 455–473.
Hausman, J.A., B.H. Hall, and Z. Griliches. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52: 909–938.
Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.
Heckman, J.J. 1981a. Statistical models for discrete panel data. In *Structural analysis of discrete data with econometric applications*, (eds.) C.F. Manski, and D. McFadden, 114–178. Cambridge: MIT Press.
Heckman, J.J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In *Structural analysis of discrete data with econometric applications*, (eds.) C.F. Manski, and D. McFadden, 179–195. Cambridge: MIT Press.
Heckman, J.J. 1981c. Heterogeneity and state dependence. In *Studies in labor markets*, (ed.) S. Rosen, 91–139. Chicago: Chicago University Press.
Heckman, J.J., and T.E. MaCurdy. 1980. A life-cycle model of female labor supply. *Review of Economic Studies* 47: 47–74.
Helpman, E., M. Melitz, and Y. Rubinstein. 2008. Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics* 123: 441–487.
Hirano, K., G.W. Imbens, G. Ridder, and D. Rubin. 2001. Combining panel data sets with attrition and refreshment samples. *Econometrica* 69: 1645–1660.
Honoré, B.E. 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* 60: 533–565.
Honoré, B.E. 1993. Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics* 59: 35–61.
Honoré, B.E., and E. Kyriazidou. 2000a. Estimation of Tobit-type models with individual specific effects. *Econometric Reviews* 19: 341–366.
Honoré, B.E., and E. Kyriazidou. 2000b. Panel data discrete choice models with lagged dependent variables. *Econometrica* 68: 839–874.
Honoré, B.E., and E. Tamer. 2006. Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74: 611–629.

Honoré, B.E., F. Vella, and M. Verbeek, 2008. Attrition, selection bias and censored regressions, Chapter 12 in L. Matyas and P. Sevestre (eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Springer, Berlin, 385–418.
Hsiao, C. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.
Hu, L. 2002. Estimation of a censored dynamic panel data model with an application to earnings dynamics. *Econometrica* 70: 2499–2517.
Keane, M. 1993. Simulation estimation for panel data models with limited dependent variables. In *Handbook of statistics*, (eds.) G.S. Maddala, C.R. Rao, and H.D. Vinod. Amsterdam: North-Holland.
Keane, M. 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62: 95–116.
Keane, M. 2015. Discrete choice models of consumer demand, Chapter 18. In *Oxford Handbook of Panel Data* , (ed.) Baltagi, B.H., 548–582. Oxford: Oxford University Press.
Kyriazidou, E. 1997. Estimation of a panel data sample selection model. *Econometrica* 65: 1335–1364.
Labeaga, J.M. 1999. A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics* 93: 49–72.
Lee, M.J. 2015. Panel conditional and multinomial logit estimators, Chapter 7. In *Oxford handbook of panel data*, (ed.) Baltagi, B.H., 202–232. Oxford: Oxford University Press.
Manski, C.F. 1995. *Identification problems in the social sciences*. Cambridge: Harvard University Press.
Nerlove, M. 1971. Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. *Econometrica* 39: 359–382.
Nicoletti, C. 2006. Nonresponse in dynamic panel data models. *Journal of Econometrics* 132: 461–489.
Nijman, T.E., and M. Verbeek. 1992. Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics* 7: 243–257.
Ridder, G. 1992. An empirical evaluation of some models for non-random attrition in panel data. *Structural Change and Economic Dynamics* 3: 337–355.
Robins, P.K., and R.W. West. 1986. Sample attrition and labor supply response in experimental panel data: A study of alternative correction procedures. *Journal of Business and Economic Statistics* 4: 329–338.
Ruhm, C.J. 1996. Alcohol policies and highway vehicle fatalities. *Journal of Health Economics* 15: 435–454.
Semykina, A., and J.M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157: 375–380.
Semykina, A., and J.M. Wooldridge. 2013. Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics* 28: 47–61.
Sickles, R.C., and P. Taubman. 1986. A multivariate error components analysis of the health and retirement study of the elderly. *Econometrica* 54: 1339–1356.
Stock, J.H., and M.W. Watson. 2003. *Introduction to econometrics*. Boston: Addison Wesley.
Vella, F., and M. Verbeek. 1998. Whose wages do unions raise? A dynamic model of unionism and wage determination for young men. *Journal of Applied Econometrics* 13: 163–168.
Vella, F., and M. Verbeek. 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics* 90: 239–263.
Verbeek, M., and T.E. Nijman. 1992. Testing for selectivity bias in panel data models. *International Economic Review* 33: 681–703.
Verbeek, M. and T.E. Nijman. 1996. Incomplete panels and selection bias, Chapter 18. In *The econometrics of panel data: A handbook of the theory with applications*, (eds.) Mátyás, L. and P. Sevestre, 449–490. Dordrecht: Kluwer Academic Publishers.
Willis, J.L. 2006. Magazine prices revisited. *Journal of Applied Econometrics* 21: 337–344.
Winkelmann, L., and R. Winkelmann. 1998. Why are the unemployed so unhappy? Evidence from panel data. *Economica* 65: 1–15.

Wooldridge, J.M. 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68: 115–132.

Wooldridge, J.M. 1997. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13: 667–678.

Wooldridge, J.M. 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90: 77–97.

Wooldridge, J.M. 2000. A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters* 68: 245–250.

Wooldridge, J.M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20: 39–54.

Zhang, W., and L.F. Lee. 2004. Simulation estimation of dynamic discrete choice panel models with accelerated importance samplers. *Econometrics Journal* 7: 120–142.

Ziliak, J.P., and T.J. Kniesner. 1998. The importance of sample attrition in life cycle labor supply estimation. *Journal of Human Resources* 33: 507–530.

Nonstationary Panels 12

12.1 Introduction

With the growing use of cross-country data over time to study purchasing power parity, growth convergence, and international R&D spillovers, the focus of panel data econometrics has shifted toward studying the asymptotics of macro-panels with large N (number of countries) and large T (length of the time series) rather than the usual asymptotics of micro-panels with large N and small T. The limiting distribution of double-indexed integrated processes has been extensively studied by Phillips and Moon (1999), Phillips and Moon (2000). The fact that T is allowed to increase to infinity in macro-panel data generated two strands of ideas. The first rejected the homogeneity of the regression parameters implicit in the use of a pooled regression model in favor of heterogeneous regressions, i.e., one for each country; see Pesaran and Smith (1995), Im, Pesaran and Shin (2003), Pesaran, Shin and Smith (1999), to mention a few. This literature critically relies on T being large to estimate each country's regression separately. This literature warns against the use of standard pooled estimators such as FE to estimate the dynamic panel data model arguing that they are subject to large potential bias when the parameters are heterogeneous across countries and the regressors are serially correlated. Another strand of the literature applied time-series procedures to panels, worrying about non-stationarity, spurious regressions and cointegration; see the surveys by Baltagi and Kao (2000), Choi (2015a, b), Breitung and Pesaran (2008), and Breitung (2015). Consider for example, the Penn World Table (PWT) which have been used to study growth convergence and purchasing power parity among various countries (https://www.rug.nl/ggdc/productivity/pwt/). Phillips and Moon (2000) argue that the time-series components of the variables used in the PWT, like per capita GDP growth, have strong non-stationarity, a feature which we have paid no attention to in the previous chapters. This is understandable given that micro-panels deal with large N and small T. With large N, large T macro-panels, non-stationarity deserves more attention. In particular,

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_12

time-series fully modified estimation techniques that account for endogeneity of the regressors and correlation and heteroskedasticity of the residuals can now be combined with fixed and random effects panel estimation methods. Some of the distinctive results that are obtained with nonstationary panels are that many test statistics and estimators of interest have Normal limiting distributions. This is in contrast to the nonstationary time-series literature where the limiting distributions are complicated functionals of Weiner processes. Several unit root tests applied in the time-series literature have been extended to panel data. When the panel data are both heterogeneous and nonstationary, issues of combining individual unit root tests applied on each time series are tackled by Im, Pesaran and Shin (2003), Maddala and Wu (1999), and Choi (2001). Using panel data, one can avoid the problem of spurious regression; see Kao (1999) and Phillips and Moon (1999). Unlike the single time- series spurious regression literature, the panel data spurious regression estimates give a consistent estimate of the true value of the parameter as both N and T tend to ∞. This is because, the panel estimator averages across individuals and the information in the independent cross-section data in the panel leads to a stronger overall signal than the pure time-series case. Of course letting both N and T tend to ∞ brings in a new host of issues dealing with how to do asymptotic analysis. Most papers in this literature adopt the *sequential* asymptotic theory, which assumes that $T \to \infty$ followed by N. However, from a practitioner's viewpoint, sequential asymptotic theory seems somewhat artificial because one is dealing with data where both T and N are large together. Phillips and Moon (1999) provide *joint* asymptotic analysis of pooled estimators in panel regressions with nonstationary regressors when the underlying regression disturbances follow stationary processes. Under the additional condition, $N/T \to 0$, they show that sequential asymptotic results for their pooled estimators would be equivalent to the joint ones. See also Kauppi (2000) for a joint asymptotic analysis of pooled estimators in panels containing near-integrated regressors with heterogeneous localizing parameters.

Most applications of time-series methods applied to panels include panel unit root tests, panel cointegration tests and the estimation of long-run average relations. Examples from the purchasing power parity literature and real exchange rate stationarity include O'Connell (1998), Pedroni (2001), Choi (2001), Groen and Kleibergen (2003), Smith et al. (2004), to mention a few. On interest rates, see Moon and Perron (2007); on health care expenditures, see McCoskey and Selden (1998), and Baltagi and Moscone (2010); on empirical growth, see Eberhardt and Teal (2011). On international R&D spillovers, see Kao, Chiang and Chen (1999). On savings and investment models, see Mark, Ogaki and Sul (2005).

However, the use of such panel data methods are not without their critics; see Maddala, Wu and Liu (2000) who argue that panel data unit root tests do not rescue purchasing power parity (PPP). In fact, the results on PPP with panels are mixed depending on the group of countries studied, the period of study, and the type of unit root test used. More damaging is the argument by Maddala, Wu and Liu (2000) that for PPP, panel data tests are the wrong answer to the low power of unit root tests in single time series. After all, the null hypothesis of a single unit root is different from the null hypothesis of a panel unit root for the PPP hypothesis. Using the same line of criticism, Maddala also argued that panel unit root tests did not help settle the question of growth convergence among countries. However, it was useful in spurring much needed research into dynamic panel data models. See also Banerjee, Marcellino and Osbat (2004) in the empirical section of this chapter who criticize existing panel unit root tests for assuming that cross-unit cointegrating relationships among the countries are not present. They warn that the empirical size of these tests is substantially higher than their nominal size. Panel unit root tests have been also criticized because they assume cross-sectional independence. This is restrictive as macro-time series exhibit significant cross-sectional correlation among the countries in the panel. This correlation has been modeled using a dynamic factor model by Bai and Ng (2004), Moon and Perron (2004), and Phillips and Sul (2003). Alternative panel unit root tests that account for cross-section dependence include Choi (2002), Chang (2002), and Pesaran (2007), to mention a few.[1] In their survey of cross-country growth empirics, Eberhardt and Teal (2011) emphasize the importance of heterogeneity, non-stationarity, and cross-section dependence: *heterogeneity* in the production technology and in the unobservable total factor productivity across countries, *non-stationarity* in the macro-time series like GDP and capital, and *cross-section dependence* in macro-productivity analysis due to strong inter-economy linkages.

This chapter studies the first generation panel unit root tests assuming cross-sectional independence in Sect. 12.2, while Sect. 12.3 discusses the second generation panel unit root tests allowing for cross-sectional dependence. Section 12.4 studies the spurious regression in panel models, while Sect. 12.5 considers various panel cointegration tests. Section 12.6 discusses estimation and inference in panel cointegration models, while Sect. 12.7 illustrates the panel unit root tests using three examples, the first is on purchasing power parity, the second on international R&D spillover, and the third on OECD health care expenditures. Section 12.8 gives some additional readings.

12.2 Panel Unit Roots Tests Assuming Cross-Sectional Independence

Testing for unit roots in time-series studies is now common practice among applied researchers and has become an integral part of econometric courses. Testing for unit roots in panels has been made popular by Levin, Lin and Chu (2002), Im, Pesaran and Shin (2003), Harris and Tzavalis (1999), Maddala and Wu (1999), Choi (2001), and Hadri (2000).[2] Levin, Lin and Chu (2002), hereafter LLC, generalized this model to allow for fixed effects, individual deterministic trends, and heterogeneous serially correlated errors. They assumed that both N and T tend to infinity. However, T increases at a faster rate than N with $N/T \to 0$. Even though this literature grew from time-series and panel data, the way in which N, the number of cross-section units, and T, the length of the time series, tend to infinity is crucial for determining asymptotic properties of estimators and tests proposed for nonstationary panels; see

Phillips and Moon (1999). Several approaches are possible including (i) sequential limits where one index, say N, is fixed and T is allowed to increase to infinity, giving an intermediate limit. Then by letting N tend to infinity subsequently, a sequential limit theory is obtained. Phillips and Moon (2000) argued that these sequential limits are easy to derive and are helpful in extracting quick asymptotics. However, Phillips and Moon (1999) provided a simple example that illustrates how sequential limits can sometimes give misleading asymptotic results. (ii) The second approach, used by Levin, Lin and Chu (2002), is to allow the two indexes, N and T to pass to infinity along a specific diagonal path in the two-dimensional array. This path can be determined by a monotonically increasing functional relation of the type $T = T(N)$ which applies as the index $N \to \infty$. Phillips and Moon (2000) showed that the limit theory obtained by this approach is dependent on the specific functional relation $T = T(N)$ and the assumed expansion path may not provide an appropriate approximation for a given (T, N) situation. (iii) The third approach is a joint limit theory allowing both N and T to pass to infinity simultaneously without placing specific diagonal path restrictions on the divergence. Some control over the relative rate of expansion may have to be exercised in order to get definitive results. Phillips and Moon argued that, in general, joint limit theory is more robust than either sequential limit or diagonal path limit. However, it is usually more difficult to derive and requires stronger conditions such as the existence of higher moments that will allow for uniformity in the convergence arguments. The muti-index asymptotic theory in Phillips and Moon (1999, 2000) is applied to joint limits in which $T, N \to \infty$ and $(T/N) \to \infty$, i.e., to situations where the time- series sample is large relative to the cross-section sample. However, the general approach given there is also applicable to situations in which $(T/N) \to 0$ although different limit results will generally be obtained in that case.

12.2.1 Levin, Lin and Chu Test

Suppose that we have a panel of countries over time. LLC argued that individual unit root tests for each country have limited power against alternative hypotheses with highly persistent deviations from equilibrium. This is particularly severe in small samples. LLC suggest a more powerful panel unit root test than performing individual unit root test for each country.

The maintained hypothesis is that

$$\Delta y_{it} = \rho y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it}, \qquad m = 1, 2, 3 \tag{12.1}$$

with d_{mt} indicating the vector of deterministic variables and α_{mi} the corresponding vector of coefficients for model $m = 1, 2, 3$. In particular, d_{1t} ={empty set}; $d_{2t} = \{1\}$, and $d_{3t} = \{1, t\}$. The null hypothesis is that each country time series contains a unit root (H_0: $\rho = 0$) against the alternative that each country time series is stationary (H_0: $\rho \neq 0$). Since the lag order p_i is unknown, LLC suggest a three-step procedure to implement their test.

Step 1: Perform separate augmented Dickey–Fuller (ADF) regressions for each cross-section

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \qquad m = 1, 2, 3 \tag{12.2}$$

The lag order p_i is permitted to vary across individuals.

For a given T, choose a maximum lag order $p_{\max}$ and then use the t-statistic of $\widehat{\theta}_{iL}$ to determine if a smaller lag order is preferred. (These t-statistics are distributed $N(0, 1)$ under the null hypothesis ($\theta_{iL} = 0$), both when $\rho_i = 0$ and when $\rho_i < 0$).

Once p_i is determined, two auxiliary regressions are run to get *orthogonalized residuals*:

Run Δy_{it} on $\Delta y_{i,t-L}(L = 1, \ldots, p_i)$ and d_{mt} to get residuals $\widehat{e}_{it}$

Run $y_{i,t-1}$ on $\Delta y_{i,t-L}(L = 1, \ldots, p_i)$ and d_{mt} to get residuals $\widehat{v}_{i,t-1}$

Standardize these residuals to control for different variances across i

$$\widetilde{e}_{it} = \widehat{e}_{it}/\widehat{\sigma}_{\varepsilon i} \qquad \textit{and} \qquad \widetilde{v}_{i,t-1} = \widehat{v}_{it}/\widehat{\sigma}_{\varepsilon i}$$

where $\widehat{\sigma}_{\varepsilon i}$ = standard error from each ADF regression, for $i = 1, \ldots, N$.

Step 2: Estimate the ratio of long-run to short-run standard deviations. Under the null hypothesis of a unit root, the long-run variance of (12.1) can be estimated by

$$\widehat{\sigma}^2_{yi} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y^2_{it} + 2 \sum_{L=1}^{\bar{K}} w_{\bar{K}L} \left[\frac{1}{T-1} \sum_{t=2+L}^{T} \Delta y_{it} \Delta y_{i,t-L} \right] \tag{12.3}$$

where $\bar{K}$ is a truncation lag that can be data dependent. $\bar{K}$ must be obtained in a manner that ensures the consistency of $\widehat{\sigma}^2_{yi}$. For a Bartlett kernel, $w_{\bar{K}L} = 1 - (L/(\bar{K} + 1))$. For each country i, the ratio of the long-run standard deviation to the innovation standard deviation is estimated by $\widehat{s}_i = \widehat{\sigma}_{yi}/\widehat{\sigma}_{\varepsilon i}$. The average standard deviation is estimated by $\widehat{S}_N = \frac{1}{N} \sum_{i=1}^{N} \widehat{s}_i$.

Step 3: Compute the panel test statistics. Run the pooled regression

$$\widetilde{e}_{it} = \rho \widetilde{v}_{i,t-1} + \widetilde{\varepsilon}_{it}$$

based on $N\widetilde{T}$ observations where $\widetilde{T} = T - \overline{p} - 1$, and $\overline{p} = \sum_{i=1}^{N} p_i/N$. $\overline{p}$ is the average lag order of individual ADF regressions. The conventional t-statistic for $H_0 : \rho = 0$ is $t_\rho = \frac{\widehat{\rho}}{\widehat{\sigma}(\widehat{\rho})}$ where

$$\widehat{\rho} = \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \widetilde{v}_{i,t-1} \widetilde{e}_{it} \Big/ \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \widetilde{v}^2_{i,t-1}$$

$$\widehat{\sigma}(\widehat{\rho}) = \widehat{\sigma}_{\widetilde{\varepsilon}} \Big/ \left[\sum_{i=1}^{N} \sum_{t=2+p_i}^{T} \widetilde{v}^2_{i,t-1} \right]^{1/2}$$

and

$$\widehat{\sigma}_{\widetilde{\varepsilon}}^2 = \frac{1}{N\widetilde{T}} \sum_{i=1}^{N} \sum_{t=2+p_i}^{T} (\widetilde{e}_{it} - \widehat{\rho}\widetilde{v}_{i,t-1})^2$$

is the estimated variance of $\widetilde{\varepsilon}_{it}$.

Compute the adjusted t-statistic

$$t_\rho^* = \frac{t_\rho - N\widetilde{T}\widehat{S}_N \widehat{\sigma}_{\widetilde{\varepsilon}}^{-2} \widehat{\sigma}(\widehat{\rho}) \mu_{m\widetilde{T}}^*}{\sigma_{m\widetilde{T}}^*} \tag{12.4}$$

where $\mu_{m\widetilde{T}}^*$ and $\sigma_{m\widetilde{T}}^*$ are the mean and standard deviation adjustments provided by Table 2 of LLC. This table also includes suggestions for the truncation lag parameter $\bar{K}$ for each time series $\widetilde{T}$. LLC show that t_ρ^* is asymptotically distributed as $N(0, 1)$.

The asymptotics require $\sqrt{N_T}/T \to 0$ where N_T emphasizes that the cross-sectional dimension N is an arbitrary monotonically increasing function of T. LLC argue that this is relevant for micro-panel data where T is allowed to grow slower than N_T. Other divergence speeds such as $N_T/T \to 0$ and $N_T/T \to$ constant are sufficient, but not necessary.

Computationally, the LLC method requires a specification of the *number of lags* used in each cross-section ADF regression (p_i), as well as *kernel choices* used in the computation of S_N. In addition, you must specify the exogenous variables used in the test equations. You may elect to include no exogenous regressors, or to include individual constant terms (fixed effects), or to employ constants and trends. LLC panel unit root can be performed with EViews; see Table 12.1 and empirical example 1 in Sect. 12.7. In Stata one can use the *xtunitroot* command. This is illustrated using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (a) of that problem issues the command *xtunitroot llc lnrxrate if oecd, lags(aic 10) kernel(bartlett nwest)* and finds that we can reject the null of panel unit roots for ln(exchange rates) using the LLC test.

LLC suggest using their panel unit root test for panels of moderate size with N between 10 and 250 and T between 25 and 250. They argue that the standard panel procedures may not be computationally feasible or sufficiently powerful for panels of this size. However, for very large T, they argue that individual unit root time-series tests will be sufficiently powerful to apply for each cross-section. Also, for very large N, and very small T, they recommend the usual panel data procedures. The Monte Carlo simulations performed by LLC indicate that the normal distribution provides a good approximation to the empirical distribution of the test statistic even in relatively small samples and also that the panel unit root test provides dramatic improvements in power over separate unit root tests for each cross-section.

The proposed LLC test has its limitations. The test crucially depends upon the *independence* assumption across cross-sections and is not applicable if cross-sectional correlation is present. Second, the assumption that *all* cross-sections have or do not have a unit root is restrictive.

O'Connell (1998) showed that the Levin and Lin tests suffered from significant size distortion in the presence of correlation among contemporaneous cross-sectional error terms. O'Connell highlighted the importance of controlling for cross-sectional dependence when testing for a unit root in panels of real exchange rates. He showed that, controlling for cross-sectional dependence, no evidence against the null of a random walk can be found in panels of up to 64 real exchange rates.

Harris and Tzavalis (1999) also derived unit root tests for (12.1) with $d_{mt} = \{\text{empty set}\}, \{1\}$, or $\{1, t\}$ when the time dimension of the panel T is *fixed*. This is the typical case for micro-panel studies. The main results are

$$\begin{array}{ll} d_{mt} & \widehat{\rho} \\ \{\text{empty set}\} & \sqrt{N}\,(\widehat{\rho}-1) \Rightarrow \mathrm{N}\left(0, \frac{2}{T(T-1)}\right) \\ \{1\} & \sqrt{N}\left(\widehat{\rho}-1+\frac{3}{T+1}\right) \Rightarrow \mathrm{N}\left(0, \frac{3(17T^2-20T+17)}{5(T-1)(T+1)^3}\right) \\ \{1,t\} & \sqrt{N}\left(\widehat{\rho}-1+\frac{15}{2(T+2)}\right) \Rightarrow \mathrm{N}\left(0, \frac{15(193T^2-728T+1147)}{112(T+2)^3(T-2)}\right) \end{array} \tag{12.5}$$

Harris and Tzavalis (1999) also showed that the assumption that T tends to infinity at a faster rate than N as in Levin, Lin and Chu rather than T fixed as in the case of micro-panels yields tests which are substantially undersized and have low power especially when T is small. In another paper, Harris and Tzavalis (2004) suggest a similar unit root test statistic for dynamic panel data with fixed effects. The test is based on the LM principle and is derived under the assumption that T is fixed. It is shown that the limiting distribution of the test statistic is standard normal. The similarity of the test with respect to both the initial conditions of the panel and the fixed effects is achieved by allowing for a trend in the model using a parametrization that has the same interpretation under both the null and alternative hypotheses. They re-examine the stationarity of real stock prices and dividends across 572 U.S. companies over a relatively short period of time, 1975–1994. Their results suggest that while real stock prices contain a unit root, real dividends are trend stationary. This is illustrated with Stata using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (b) of that problem issues the Stata command *xtunitroot ht lnrxrate if oecd, demean* and finds that we can reject the null of panel unit roots for ln(exchange rates) using the Harris and Tzavalis test.

De Blander and Dhaene (2012) extend the fixed T approach of Harris and Tzavalis (1999) to the case where the errors are generated by a stationary AR(1) process. T may be as small as 5, while still allowing for fixed effects, individual deterministic time trends, and a homogeneous ADF(1) term. The tests are based on a bias-corrected OLS estimator of the autoregressive parameter, obtained by numerically inverting the expression for the probability limit. This bias correction, together with an appropriately modified standard error, yields an asymptotically normal t-statistic under the null hypothesis of a unit root, as $N \to \infty$ with T fixed. They examine a weak version of the Law of One Price in the European car market since the start of stage three of the EMU in 1999. They find strong evidence for mean-reversion in the time series of cross-country car (log) price differences across EMU countries.

12.2.2 Im, Pesaran and Shin Test

The Levin, Lin and Chu test is restrictive in the sense that it requires ρ to be homogeneous across i. As Maddala pointed out, the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Im, Pesaran and Shin (2003) (IPS) allow for a *heterogeneous* coefficient of y_{it-1} and propose an alternative testing procedure based on averaging individual unit root test statistics. IPS suggest an average of the ADF tests when u_{it} is serially correlated with different serial correlation properties across cross-sectional units, i.e., the model given in (12.2). The null hypothesis is that each series in the panel contains a unit root, i.e., H_0: $\rho_i = 0$ for all i and the alternative hypothesis allows for some (but not all) of the individual series to have unit roots, i.e.,

$$H_1\colon \begin{cases} \rho_i < 0 \; for \quad i = 1, 2, \ldots, N_1 \\ \rho_i = 0 \; for \; i = N_1 + 1, \ldots, N. \end{cases} \tag{12.6}$$

Formally, *it requires the fraction of the individual time series that are stationary to be nonzero*, i.e., $\lim_{N\to\infty}(N_1/N) = \delta$ where $0 < \delta \leq 1$. This condition is necessary for the consistency of the panel unit root test. The IPS t-bar statistic is defined as the average of the individual ADF statistics as

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho_i}, \tag{12.7}$$

where t_{ρ_i} is the individual t-statistic for testing H_0: $\rho_i = 0$ for all i in (12.6). In case the lag order is always zero ($p_i = 0$ for all i), IPS provide simulated critical values for $\bar{t}$ for different number of cross-sections N, series length T, and Dickey–Fuller regressions containing intercepts only or intercepts and linear trends. In the general case where the lag order p_i may be nonzero for some cross-sections, IPS show that a properly standardized $\bar{t}$ has an asymptotic $N(0, 1)$ distribution, starting from the well-known result in time series that for a fixed N

$$t_{\rho_i} \Rightarrow \frac{\int_0^1 W_{iZ} dW_{iZ}}{\left[\int_0^1 W_{iZ}^2\right]^{1/2}} = t_{iT} \tag{12.8}$$

as $T \to \infty$, where $\int W(r)dr$ denotes a Weiner integral with the argument r suppressed in (12.8). IPS assume that t_{iT} are IID and have finite mean and variance. Then

$$\frac{\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} t_{iT} - \frac{1}{N}\sum_{i=1}^{N} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \text{var}\left[t_{iT}|\rho_i = 0\right]}} \Rightarrow \text{N}(0, 1) \tag{12.9}$$

as $N \to \infty$ by the Lindeberg–Levy central limit theorem. Hence

$$t_{IPS} = \frac{\sqrt{N}\left(\bar{t} - \frac{1}{N}\sum_{i=1}^{N} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \text{var}\left[t_{iT}|\rho_i = 0\right]}} \Rightarrow \text{N}(0, 1) \tag{12.10}$$

as $T \to \infty$ followed by $N \to \infty$ sequentially. The values of $E[t_{iT}|\rho_i = 0]$ and $\text{var}[t_{iT}|\rho_i = 0]$ have been computed by IPS via simulations for different values of T and p_i's. In Monte Carlo experiments, they show that if a large enough lag order is selected for the underlying ADF regressions, then the small sample performance of the t-bar test is reasonably satisfactory and generally better than the LLC test.

IPS can be performed with EViews; see Table 12.1 and empirical example 1 in Sect. 12.7. This is also illustrated with Stata using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (c) of that problem issues the Stata command *xtunitroot ips lnrxrate if oecd, lags(aic 10),* and finds that we can reject the null of panel unit roots for ln(exchange rates) using the IPS test.

McCoskey and Selden (1998) applied the IPS test for testing unit root for per capita national health care expenditures (HE) and gross domestic product (GDP) for a panel of 20 OECD countries. McCoskey and Selden rejected the null hypothesis that these two series contain unit roots, see problem 12.6. Gerdtham and Löthgren (2000) argued that the stationarity found by McCoskey and Selden are driven by the omission of time trends in their ADF regression in (12.6). Using the IPS test with a time trend, Gerdtham and Löthgren found that both HE and GDP are nonstationary. They concluded that HE and GDP are cointegrated around linear trends.

Westerlund and Breitung (2013) emphasize important facts about the LLC and IPS panel unit root tests, some of which are well known and others ignored by researchers. These include the following: (1) The IPS and LLC statistics are standard normally distributed as $N \to \infty$ even if T is fixed; (2) The LLC test can be more powerful than the IPS test; (3) Deterministic components need not be treated as in the Dickey and Fuller approach; (4) Incidental trends reduce the local power of the LLC test; (5) The initial condition may affect the asymptotic properties of the tests. (6) The GMM approach can also be used in the unit root case; (7) Lag augmentation does not remove the effects of serial correlation; (8) The consistency of the LLC test depends on the long-run variance estimator; (9) Cross-section dependence leads to deceptive inference; (10) The IPS and LLC tests fail under cross-unit cointegration; (11) Sequential limits need not imply joint limits. They warn the researcher not to approach the panel unit root testing problem from a too narrow and stylized perspective.

12.2.3 Breitung's Test

The LLC and IPS tests require $N \to \infty$ such that $N/T \to 0$, i.e., N should be small enough relative to T. This means that both tests may not keep nominal size well when either N is small or N is large relative to T. In fact, the simulation results of Im, Pesaran and Shin (2003) show that both IPS and LLC have size distortions as N gets large relative to T. Breitung (2000) studies the local power of LLC and IPS test statistics against a sequence of local alternatives. Breitung finds that the LLC and IPS tests suffer from a dramatic loss of power if individual-specific trends are included. This is due to the bias correction that also removes the mean under the sequence of local alternatives. Breitung suggests a test statistic that does not employ a bias adjustment whose power is substantially higher than that of LLC or the IPS tests using Monte Carlo experiments. The simulation results indicate that the power of LLC and IPS tests is very sensitive to the specification of the deterministic terms.

Breitung (2000) test statistic without bias adjustment is obtained as follows: Step 1 is the same as LLC but only $\Delta y_{i,t-L}$ is used in obtaining the residuals $\widehat{e}_{it}$ and $\widehat{v}_{i,t-1}$. The residuals are then adjusted (as in LLC) to correct for individual-specific variances. Step 2, the residuals $\widetilde{e}_{it}$ are transformed using the forward orthogonalization transformation employed by Arellano and Bover (1995):

$$e_{it}^{*} = \sqrt{\frac{T-t}{(T-t+1)}}\left(\widetilde{e}_{it} - \frac{\widetilde{e}_{i,t+1} + \cdots + \widetilde{e}_{i,T}}{T-t}\right).$$

Also,

$$\begin{aligned} v_{i,t-1}^{*} &= \widetilde{v}_{i,t-1} - \widetilde{v}_{i,1} - \frac{t-1}{T}\widetilde{v}_{iT} \text{ with intercept and trend;} \\ &= \widetilde{v}_{i,t-1} - \widetilde{v}_{i,1} \text{ with intercept, no trend;} \\ &= \widetilde{v}_{i,t-1} \text{ with no intercept or trend.} \end{aligned}$$

The last step is to run the pooled regression

$$e_{it}^{*} = \rho v_{i,t-1}^{*} + \varepsilon_{it}^{*}$$

and obtain the t-statistic for H_0; $\rho = 0$ which has in the limit a standard $N(0,1)$ distribution. Note that no kernel computations are required. Breitung's test can be performed with EViews; see Table 12.1 and empirical example 1 in Sect. 12.7. This is also illustrated with Stata using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (d) of that problem issues the Stata command *xtunitroot breitung lnrxrate if oecd, lags(3) robust trend* and finds that we cannot reject the null of panel unit roots for ln(exchange rates) when we include a trend using the Breitung test.

12.2.4 Combining *p*-Value Tests

Let G_{iT_i} be a unit root test statistic for the ith group in (12.1) and assume that as the time-series observations for the ith group $T_i \to \infty$, $G_{iT_i} \Rightarrow G_i$ where G_i is a nondegenerate random variable. Let p_i be the asymptotic p-value of a unit root test for cross-section i, i.e., $p_i = F\left(G_{iT_i}\right)$, where $F(\cdot)$ is the distribution function of the random variable G_i. Maddala and Wu (1999) and Choi (2001) proposed a Fisher type test

$$P = -2\sum_{i=1}^{N} \ln p_i \tag{12.11}$$

which combines the p-values from unit root tests for each cross-section i to test for unit root in panel data. Note that $-2\ln p_i$ has a χ^2 distribution with two degrees of freedom. This means that P is distributed as χ^2 with $2N$ degrees of freedom as $T_i \to \infty$ for finite N. Maddala and Wu (1999) argued that the IPS and Fisher tests relax the restrictive assumption of the LLC test that ρ_i is the same under the alternative. Both the IPS and Fisher tests combine information based on individual unit root tests. However, the Fisher test has the advantage over the IPS test in that it does not require a balanced panel. Also, the Fisher test can use different lag lengths in the individual ADF regressions and can be applied to any other unit root tests. The disadvantage is that the p-values have to be derived by Monte Carlo simulations. Maddala and Wu (1999) find that the Fisher test with bootstrap-based critical values performs the best and is the preferred choice for testing nonstationarity as the null and also in testing for cointegration in panels. Choi (2001) proposes two other test statistics besides Fisher's inverse chi-square test statistic P. The first is the inverse normal test $Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i)$ where Φ is the standard normal cumulative distribution function. Since $0 \leqslant p_i \leqslant 1$, $\Phi^{-1}(p_i)$ is a $N(0, 1)$ random variable and as $T_i \to \infty$ for all $i, Z \Rightarrow N(0, 1)$. The second is the logit test $L = \sum_{i=1}^{N} \ln(\frac{p_i}{1-p_i})$ where $\ln(\frac{p_i}{1-p_i})$ has the logistic distribution with mean 0 and variance $\pi^2/3$. As $T_i \to \infty$ for all i, $\sqrt{m}L \Rightarrow t_{5N+4}$ where $m = \frac{3(5N+4)}{\pi^2 N(5N+2)}$. Choi (2001) echoes similar advantages for these three combining p-values tests: (1) the cross-sectional dimension, N, can be either finite or infinite, (2) each group can have different types of non-stochastic and stochastic components, (3) the time- series dimension, T, can be different for each i, and (4) the alternative hypothesis would allow some groups to have unit roots while others may not.

When N is large, Choi (2001) proposed a modified P test,

$$P_m = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (-2 \ln p_i - 2) \tag{12.12}$$

since $E\left[-2 \ln p_i\right] = 2$ and $\text{var}\left[-2 \ln p_i\right] = 4$. Applying the Lindeberg–Lévy central limit theorem to (12.12) we get $P_m \Rightarrow N(0, 1)$ as $T_i \to \infty$ followed by $N \to \infty$. The distribution of the Z statistic is invariant to infinite N, and $Z \Rightarrow N(0, 1)$ as $T_i \to \infty$ and then $N \to \infty$. Also, the distribution of $\sqrt{m}L \approx \frac{1}{\sqrt{\pi^2 N/3}} \sum_{i=1}^{N} \ln(\frac{p_i}{1-p_i}) \Rightarrow N(0, 1)$ by the Lindeberg–Lévy central limit theorem as $T_i \to \infty$ and then $N \to \infty$. Therefore, Z and $\sqrt{m}L$ can be used without modification for infinite N. Simulation results for $N = 5, 10, 25, 50$ and 100, and $T = 50$ and 100 show that the empirical size of all the tests is reasonably close to the 0.05 nominal size when N is small. P and P_m show mild size distortions at $N = 100$, while Z and IPS show the most stable size. All tests become more powerful as N increases. The combined p-values tests have superior size-adjusted power to the IPS test. In fact, the power of the Z test is in some cases more than three times that of the IPS test. Overall, the Z test seems to outperform the other tests and is recommended.

The combining p-values Fisher's inverse chi-square test statistic P can be performed with EViews; see Table 12.1 and empirical example 1 in Sect. 12.7. This is also illustrated with Stata using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (e) of that problem issues the Stata command *xtunitroot fisher lnrxrate if oecd, dfuller lags(3) trend* and finds that we can reject the null of panel unit roots for ln(exchange rates) using the Fisher combining p-values test. Stata performs the inverse chi-square test statistic P,the inverse normal test Z, and also the inverse logit test L, while EViews only computes the inverse chi-square test statistic P.

Choi (2001) applied the combining p-values tests and the IPS test given in (12.7) to panel data of monthly U.S. real exchange rates sampled from 1973:3 to 1996:3. The combining p-values tests provided evidence in favor of the PPP hypothesis while the IPS test did not. Choi claimed that this is due to the improved finite sample power of the combination tests. Maddala and Wu (1999) and Maddala, Wu and Liu (2000) find that the Fisher test is superior to the IPS test which in turn is more powerful than the LLC test. They argue that these panel unit root tests still do not rescue the PPP hypothesis. When allowance is made for the deficiency in the panel data unit root tests and panel estimation methods, support for PPP turns out to be weak.

Choi (2002) considers four instrumental variable estimators of an error component model with stationary and nearly nonstationary regressors. The remainder disturbances follow an autoregressive process whose order as well as parameters vary across individuals. The IV estimators considered include the Within-IV, Within-IV-OLS, Within-IV-GLS, and IV-GLS estimators. Using sequential and joint limit theories, Choi shows that, under proper conditions, all the estimators have normal distributions in the limit as N and $T \to \infty$. Simulation results show that the efficiency rankings of the estimators crucially depend on the type of regressor and the number of instruments. The Wald tests for coefficient restrictions keep reasonable nominal size as $N \to \infty$ and its power depends upon the number of instruments and the degree of serial correlation and heterogeneity in the errors.

12.2.5 Residual-Based LM Test

Hadri (2000) derives a residual-based Lagrange multiplier (LM) test where the null hypothesis is that there is no unit root in any of the series in the panel against the alternative of a unit root in the panel. This is a generalization of the KPSS test from time series to panel data. It is based on OLS residuals of y_{it} on a constant, or on a constant and a trend. In particular, Hadri (2000) considers the following two models:

$$y_{it} = r_{it} + \varepsilon_{it} \quad i = 1, \cdots, N \quad t = 1, \ldots, T;$$

and

$$y_{it} = r_{it} + \beta_i t + \varepsilon_{it} \tag{12.13}$$

where $r_{it} = r_{i,t-1} + u_{it}$ is a random walk. $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $u_{it} \sim \text{IIN}(0, \sigma_\mu^2)$ are mutually independent normals that are IID across i and over t. Using back substitution, model (12.13) becomes

$$y_{it} = r_{io} + \beta_i t + \sum_{s=1}^{t} u_{is} + \varepsilon_{it} = r_{io} + \beta_i t + v_{it} \tag{12.14}$$

where $v_{it} = \sum_{s=1}^{t} u_{is} + \varepsilon_{it}$. The stationarity hypothesis is simply H_0; $\sigma_u^2 = 0$ in which case $v_{it} = \varepsilon_{it}$. The LM statistic is given by

$$LM_1 = \frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} S_{it}^2\right) / \widehat{\sigma}_\varepsilon^2$$

where $S_{it} = \sum_{s=1}^{t} \widehat{\varepsilon}_{is}$ are the partial sum of OLS residuals $\widehat{\varepsilon}_{is}$ from (12.14) and $\widehat{\sigma}_{\varepsilon}^2$ is a consistent estimate of σ_{ε}^2 under the null hypothesis H_0. A possible candidate is $\widehat{\sigma}_{\varepsilon}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^2$.

Hadri (2000) suggested an alternative LM test that allows for heteroskedasticity across i, say $\sigma_{\varepsilon i}^2$. This is in fact

$$LM_2 = \frac{1}{N}\left(\sum_{i=1}^{N}\left(\frac{1}{T^2}\sum_{t=1}^{T} S_{it}^2/\widehat{\sigma}_{\varepsilon i}^2\right)\right)$$

The test statistic is given by $Z = \sqrt{N}(LM - \xi)/\zeta$ and is asymptotically distributed as $N(0, 1)$, where $\xi = \frac{1}{6}$ and $\zeta^2 = \frac{1}{45}$ if the model only includes a constant, and $\xi = \frac{1}{15}$ and $\zeta^2 = \frac{11}{6300}$, otherwise. Hadri (2000) shows using Monte Carlo experiments that the empirical size of the test is close to its nominal 5% level for sufficiently large N and T.

Hadri's test can be performed with EViews; see Table 12.1 and empirical example 1 in Sect. 12.7. This is also illustrated with Stata using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (f) of that problem issues the Stata command *xtunitroot hadri lnrxrate if oecd, kernel(bartlett) trend* and finds that we can reject the null hypothesis of stationarity for ln(exchange rates) using the Hadri test.

Shin and Snell (2006) propose a mean group KPSS test statistic based on the mean of the KPSS stationarity test statistics from each panel unit. These are computed using parametric estimations of the long-run variance of the underlying serially correlated disturbances. Using both *sequential* and *joint* asymptotic theory, they show that the proposed statistic has a standard normal limiting distribution under the null hypothesis of stationarity. They also emphasize that the joint asymptotic approach predicts that unless N/T is small, the asymptotics will fail. This is confirmed by their Monte Carlo experiments.

Extensive simulations have been conducted to explore the finite sample performance of panel unit root tests. Choi (2001), for example, studied the small sample properties of IPS t-bar test in (12.7) and Fisher's test in (12.11). Choi's major findings were the following:

1. The empirical sizes of the IPS and the Fisher tests are reasonably close to their nominal size 0.05 when N is small. But the Fisher test shows mild size distortions at $N = 100$, which is expected from the asymptotic theory. Overall, the IPS t-bar test has the most stable size.
2. In terms of the size-adjusted power, the Fisher test seems to be superior to the IPS t-bar test.
3. When a linear time trend is included in the model, the power of all tests decrease considerably.

Karlsson and Löthgren (2000) compare the LLC and IPS tests for various size panels. They warn that for large T, panel unit root tests have high power and there is the potential risk of concluding that the *whole* panel is stationary even when there is only a *small proportion* of stationary series in the panel. For small T, panel unit root tests have low power and there is the potential risk of concluding that the whole panel is nonstationary even when there is a large proportion of stationary series in the panel. They suggest careful analysis of both the individual and panel unit root test results to fully assess the stationarity properties of the panel.

Hlouskova and Wagner (2006) perform a large-scale Monte Carlo to study the size and power of *first generation* panel unit root and stationarity tests. LLC, Harris and Tzavalis (1999), Breitung (2000), IPS, Maddala and Wu (1999), Hadri (2000), and Hadri and Larsson (2005) tests are considered. Size and power comparisons are performed for various values of (N, T) varying over the range (10, 15, 20, 25, 50, 100, 200) and various values of moving average and first-order autoregressive parameters. They find that the panel stationarity tests of Hadri (2000), Hadri and Larsson (2005) perform poorly. The null hypothesis of stationarity is rejected as soon as sizeable serial correlation of either the moving average or first-order autoregressive type are present. The picture for the panel unit root tests shows no dominant performance of one test over the others for all cases considered.

12.3 Panel Unit Roots Tests Allowing for Cross-Sectional Dependence

Pesaran (2004) suggests a simple test of error cross-section dependence (CD) that is applicable to a variety of panel models including stationary and unit root dynamic heterogeneous panels with short T and large N. The proposed test is based on an average of pair-wise correlation coefficients of OLS residuals from the individual regressions in the panel rather than their squares as in the Breusch–Pagan LM test:

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\widehat{\rho}_{ij}\right)$$

where $\widehat{\rho}_{ij} = \sum_{t=1}^{T} e_{it}e_{jt}/(\sum_{t=1}^{T} e_{it}^2)^{1/2}(\sum_{t=1}^{T} e_{jt}^2)^{1/2}$, with e_{it} denoting the OLS residuals based on T observations for each $i = 1, \ldots, N$. Under the null hypothesis of cross-section independence $CD \sim N(0, 1)$. Monte Carlo experiments show that the standard Breusch–Pagan LM test performs badly for $N > T$ panels, whereas Pesaran's CD test performs well even for small T and large N. This can be computed in Stata using the command *xtcd*. This is illustrated using the Penn World Table exchange rates for OECD countries in problem 12.6, part (g). The Stata command *xtcd lnrxrate if oecd* yields an observed CD test statistic of 56.55 which is significant.

Dynamic factor models have been used to capture cross-section correlation. In fact, factor models have been used to study world business cycles as well as common macro-shocks like international financial crises or oil price shocks. Factor models also offer a significant reduction in the number of sources of cross-sectional dependence in panel data, and they allow for heterogeneous response to common shocks through heterogeneous factor loadings. For example, Moon and Perron (2004) consider the following model:

$$y_{it} = \alpha_i + y_{it}^0$$
$$y_{it}^0 = \rho_i y_{i,t-1}^0 + \epsilon_{it}$$

where ϵ_{it} are unobservable error terms with a factor structure and α_i are fixed effects. ϵ_{it} is generated by M unobservable random factors f_t and idiosyncratic shocks e_{it} as follows:

$$\epsilon_{it} = \Lambda_i' f_t + e_{it}$$

where Λ_i are nonrandom factor loading coefficient vectors and the number of factors M is unknown. Each ϵ_{it} contains the common random factor f_t, generating the correlation among the cross-sectional units of ϵ_{it} and y_{it}. The extent of the correlation is determined by the factor loading coefficients Λ_i, i.e., $E(y_{it}y_{jt}) = \Lambda_i' E(f_t f_t')\Lambda_j$. Moon and Perron treat the factors as nuisance parameters and suggest pooling defactored data to construct a unit root test. Let Q_Λ be the matrix projecting onto the space orthogonal to the factor loadings. The defactored data is YQ_Λ and the defactored residuals eQ_Λ no longer have cross-sectional dependence, where Y is a $T \times N$ matrix whose ith column contains the observations for cross-sectional unit i.

Let $\sigma_{e,i}^2$ be the variance of e_{it}, $w_{e,i}^2$ be the long-run variance of e_{it}, and $\lambda_{e,i}$ be the one-sided long-run variance of e_{it}. Also, let σ_e^2, w_e^2, and λ_e be their cross-sectional averages, and ϕ_e^4 be the cross-sectional average of $w_{e,i}^4$. The pooled bias-correlated estimate of ρ is

$$\widehat{\rho}_{pool}^{+} = \frac{\text{tr}(Y_{-1}Q_\Lambda Y') - NT\lambda_e^N}{\text{tr}(Y_{-1}Q_\Lambda Y_{-1}')}$$

where Y_{-1} is the matrix of lagged data. Moon and Perron suggest two statistics to test H_0; $\rho_i = 1$ for all $i = 1, \ldots, M$ against the alternative hypothesis H_A; $\rho_i < 1$ for some i. These are

$$t_a = \frac{\sqrt{N}T(\widehat{\rho}_{pool}^{+} - 1)}{\sqrt{\frac{2\phi_e^4}{w_e^4}}}$$

and

$$t_b = \sqrt{N}T(\widehat{\rho}_{pool}^{+} - 1)\sqrt{\frac{1}{NT^2}\,\text{tr}\,(Y_{-1}Q_\Lambda Y_{-1}')\frac{w_e^2}{\phi_e^4}}.$$

These tests have a standard $N(0, 1)$ limiting distribution where N and T tend to infinity such that $(N/T) \to 0$. Moon and Perron also show that estimating the factors by principal components and replacing w_e^2 and ϕ_e^4 by consistent estimates lead to feasible statistics with the same limiting distribution.

Phillips and Sul (2003) consider the following common time factor model on the disturbances that can impact individual series differently

$$u_{it} = \delta_i \theta_t + \varepsilon_{it}$$

where $\theta_t \sim \text{IIN}(0, 1)$ across time. δ_i are "idiosyncratic share" parameters that measure the impact of the common time effects on series i. $\varepsilon_{it} \sim \text{IIN}(0, \sigma_i^2)$ over t, with ε_{it} independent of ε_{js} and θ_s for all $i \neq j$ and for all s, t. This model is in effect a one-factor model which is independently distributed over time. $E(u_{it}u_{js}) = \delta_i\delta_j$ and there is no cross-sectional correlation if $\delta_i = 0$ for all i, and identical cross-section correlation when $\delta_i = \delta_j = \delta_0$ for all i, j. Phillips and Sul propose an orthogonalization procedure based on iterated method of moments estimation to eliminate the common factor which is different from principal components. They suggest a series of unit root tests based on these orthogonalized data. The statistic that performs best in their simulation is a combination of p-values of individual unit root tests as in Choi (2001), i.e., $Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N-1}\Phi^{-1}(p_i)$. The sum is over $N-1$ components, since the orthogonalization they propose reduces the cross-sectional dimension by 1. The null hypothesis is rejected for large values of the Z-statistic.

Bai and Ng (2004) consider the following dynamic factor model:

$$y_{it} = \alpha_i + \Lambda_i' f_t + y_{it}^0$$
$$y_{it}^0 = \rho_i y_{i,t-1}^0 + \varepsilon_{it}$$

They test separately the stationarity of the factors and the idiosyncratic component. To do so, they obtain consistent estimates of the factors regardless of whether residuals are stationary or not. They accomplish this by estimating factors on first-differenced data and cumulating these estimated factors. Bai and Ng suggest pooling results from individual ADF tests on the estimated defactored data by combining p-values as in Maddala and Wu (1999) and Choi (2001):

$$P_{\widehat{e}}^c = \frac{-2\sum_{i=1}^N \ln p_{\widehat{e}}^c(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1)$$

where $p_{\widehat{e}}^c(i)$ is the p-value of the ADF test (without any deterministic component) on the estimated idiosyncratic shock for cross-section i. The Bai and Ng (2004) test is available in EViews and is applied to the purchasing power parity example in Sect. 12.7.

Choi (2002) uses an error component model given by

$$y_{it} = \alpha_i + f_t + y_{it}^0$$
$$y_{it}^0 = \rho_i y_{i,t-1}^0 + \varepsilon_{it}.$$

This is a restricted factor model where the cross-sections respond homogeneously to the single common factor f_t in contrast to the factor models considered above. Choi suggests demeaning the data by GLS as in Elliott, Rothenberg and Stock (1996) and taking cross-sectional means to obtain a new variable $\widetilde{y}_{it} \simeq y_{it}^0 - y_{i1}^0$ which is independent in the cross-sectional dimension as both N and T tend to infinity. Choi combines p-values from individual ADF tests as in Choi (2001). The resulting tests have a standard $N(0, 1)$ distribution. In addition, Choi suggests using an ADF test for the hypothesis that the common component f_t is nonstationary. To do so, he proposes using the cross-sectional means (at each t) of the residuals from the GLS regression used to demean the data, i.e.,

$$\widehat{f}_t = \frac{1}{N}\sum_{i=1}^N (y_{it} - \widehat{\alpha}_i).$$

Pesaran (2007) suggests a simpler way of getting rid of cross-sectional dependence than estimating the factor loading. His method is based on augmenting the usual ADF regression with the lagged cross-sectional mean and its first difference to capture the cross-sectional dependence that arises through a single factor model. This is called the cross-sectionally augmented Dickey–Fuller (CADF) test. This simple CADF regression is

$$\Delta y_{it} = \alpha_i + \rho_i^* y_{i,t-1} + d_0 \overline{y}_{t-1} + d_1 \Delta \overline{y}_t + \varepsilon_{it}$$

where $\overline{y}_t$ is the average at time t of all N observations. The presence of the lagged cross-sectional average and its first-difference accounts for the cross-sectional dependence through a factor structure. If there is serial correlation in the error term or the factor, the regression must be augmented as usual in the univariate case, but lagged first differences of both y_{it} and $\overline{y}_t$ must be added which leads to

$$\Delta y_{it} = \alpha_i + \rho_i^* y_{i,t-1} + d_0 \overline{y}_{t-1} + \sum_{j=0}^{p} d_{j+1} \Delta \overline{y}_{t-j} + \sum_{k=1}^{p} c_k \Delta y_{i,t-k} + \varepsilon_{it}$$

where the degree of augmentation can be chosen by an information criterion or sequential testing. After running this CADF regression for each unit i in the panel, Pesaran averages the t-statistics on the lagged value (called $CADF_i$) to obtain the CIPS statistics

$$CIPS = \frac{1}{N} \sum_{i=1}^{N} CADF_i.$$

The joint asymptotic limit of the CIPS statistic is nonstandard and critical values are provided for various choices of N and T. The t-tests based on this regression should be devoid of $\Lambda_i' f_t$ in the limit and therefore free of cross-sectional dependence. The limiting distribution of these tests are different from the Dickey–Fuller distribution due to the presence of the cross-sectional average of the lagged level. Pesaran uses a truncated version of the IPS test that avoids the problem of moment calculation. In addition, the t-tests are used to formulate a combination test based on the inverse normal principle. Experimental results show that these tests perform well. The CIPS test can be performed with Stata using the command *multipurt* which uses *xtfisher* and produces Maddala and Wu (1999) p-values as well as CIPS test. This is illustrated using the Penn World Table exchange rates for OECD countries in problem 12.6. Part (h) of that problem issues the Stata command *multipurt lnrxrate if oecd, lags(3)* and finds that we cannot reject the null hypothesis of panel unit root for ln(exchange rates) using the CIPS test. EViews also has the CIPS test and this is applied to the purchasing power parity example in Sect. 12.7.

Baltagi, Bresson and Pirotte (2007) study the performance of several panel unit root tests when spatial effects are present that account for cross-section correlation. See Chap. 13 for a review of spatial correlation in panel data. Unlike factor models, the structure of the spatial dependence can be related to location and distance, both in a geographic space as well as a more general economic or social network space. Using some commonly used spatial error processes, the spatial autoregressive (SAR) and the spatial moving average (SMA) error process and the spatial error components model (SEC), Baltagi, Bresson and Pirotte (2007) perform Monte Carlo simulations to compare the performance of panel unit root tests. These include first generation panel unit root tests that ignore cross-section correlation like LLC, IPS, Breitung

(2000), the Maddala and Wu test (1999), and the Choi tests (2001), also, second generation panel unit root tests that account for cross-section correlation including the Chang IV test (2002), the Choi (2002) test, the Phillips and Sul test (2003), and the Pesaran test (2007). Note that while the alternative hypothesis may be different among these panel unit root tests, the null hypothesis is the same: $H_0 : \rho_i \,(= \rho) = 1$, for all $i = 1, 2, \ldots, N$. The results show that there can be considerable size distortions in panel unit root tests when spatial dependence is present. Tests that explicitly allow for the cross-sectional dependence have better performance than other classic panel unit root tests that assume cross-sectional independence. For the SAR specification, we get the largest size distortions of the panel unit root tests. In contrast, the SMA specification of cross-sectional dependence leads to lower size distortions. For the applied econometrician, the message from these experiments is that size distortions of panel unit root tests is highly sensitive to the underlying spatial dependence specification and to the sparseness of the spatial weight matrix.

Gengenbach, Palm and Urbain (2010) compare the small sample performance of several panel unit root tests that account for cross-section dependence with up to two common factors. These include Pesaran (2007), Moon and Perron (2004), Bai and Ng (2004), among others. They also apply these tests to an empirical study of purchasing power parity. Among their findings is that the test procedures of Bai and Ng (2004), Moon and Perron (2004) exhibit size distortions if the number of common factors is misspecified.

Wang et al. (2010) generalize the nonlinear IV panel unit root test of Chang (2002) to the case where there exist some common factors in the panel. Following Bai and Ng (2004), they eliminate cross-sectional dependence using the method of principal components. Next, they apply Chang (2002) nonlinear IV test to the treated data. They show that this modified Chang test, like the original Chang test, has a standard normal limiting distribution under the null hypothesis of panel unit root. Their simulation results suggest that this modified test performs better than Chang's original test when the cross-sectional dependence is moderate to high.

12.4 Spurious Regression in Panel Data

Kao (1999) and Phillips and Moon (1999) derived the asymptotic distributions of the least squares dummy variable estimator and various conventional statistics from the spurious regression in panel data.

Suppose that y_t and X_t are unit root nonstationary time-series variables with long-run variance matrix

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix}$$

Then $\beta = \Omega_{yx}\Omega_{xx}^{-1}$ can be interpreted as a classical long-run regression coefficient relating the two nonstationary variables y_t and X_t. When Ω has deficient rank, β is a cointegrating coefficient because $y_t - \beta X_t$ is stationary. Even in the absence of time-series cointegration, β is a measure of a statistical long-run correlation between y_t and X_t. Phillips and Moon (1999) extend this concept to panel regressions with nonstationary data. In this case, heterogeneity across individuals i can be characterized by heterogeneous long-run covariance matrices Ω_i. Then Ω_i are randomly drawn from a population with mean $\Omega = E(\Omega_i)$. In this case

$$\beta = E(\Omega_{y_i x_i})E(\Omega_{x_i x_i})^{-1} = \Omega_{yx}\Omega_{xx}^{-1}$$

is the regression coefficient corresponding to the average long-run covariance matrix Ω.

Phillips and Moon (1999) study various regressions between two panel vectors that may or may not have cointegrating relations, and present a fundamental framework for studying sequential and joint limit theories in nonstationary panel data. The panel models considered allow for four cases: (i) panel spurious regression, where there is no time-series cointegration, (ii) heterogeneous panel cointegration, where each individual has its own specific cointegration relation, (iii) homogeneous panel cointegration where individuals have the same cointegration relation, and (iv) near-homogeneous panel cointegration, where individuals have slightly different cointegration relations determined by the value of a localizing parameter. Phillips and Moon (1999) investigated these four models and developed panel asymptotics for regression coefficients and tests using both sequential and joint limit arguments. In all cases considered, the pooled estimator is consistent and has a normal limiting distribution. In fact, for the spurious panel regression, Phillips and Moon (1999) showed that under quite weak regularity conditions, the pooled least squares estimator of the slope coefficient β is $\sqrt{N}$ consistent for the long-run average relation parameter β and has a limiting normal distribution. They also showed that a limiting cross-section regression with time averaged data is also $\sqrt{N}$ consistent for β and has a limiting normal distribution. This is different from the pure time-series spurious regression where the limit of the OLS estimator of β is a nondegenerate random variate that is a functional of Brownian motions and is therefore not consistent for β. The idea in Phillips and Moon (1999) is that independent cross-section data in the panel adds information and this leads to a stronger overall signal than the pure time-series case. Pesaran and Smith (1995) studied limiting cross-section regressions with time averaged data and established consistency with restrictive assumptions on the heterogeneous panel model. This differs from Phillips and Moon (1999) in that the former use an average of the cointegrating coefficients which is different from the long-run average regression coefficient. This requires the existence of cointegrating time-series relations, whereas the long-run average regression coefficient β is defined irrespective of the existence of individual cointegrating relations and relies only on the long-run average variance matrix of the panel. Phillips and Moon (1999) also showed that for the homogeneous and near-homogeneous cointegration cases, a consistent estimator of the long-run regression coefficient can be constructed which they call a pooled FM estimator. They showed that this estimator has faster convergence rate than the simple cross-section and time- series estimators. Pedroni (2000) and Kao and Chiang (2000) also investigated limit theories for various estimators of the homogeneous panel cointegration regression model. See also Phillips and Moon (2000) for a concise review. In fact, the latter paper also shows how to extend the above ideas to models with individual effects in the data generating process. For the panel spurious regression with individual-specific deterministic trends, estimates of the trend coefficients are obtained in the first step and the detrended data is pooled and used in least squares regression to estimate β in the second step. Two different detrending procedures are used based on OLS and GLS regressions. OLS detrending leads to an asymptotically more efficient estimator of the long-run average coefficient β in pooled regression than GLS detrending. Phillips and Moon (2000) explain that "the residuals after time series GLS detrending have more cross section variation than they do after OLS detrending and this produces great variation in the limit distribution of the pooled regression estimator of the long run average coefficient."

Moon and Phillips (1999) investigate the asymptotic properties of the Gaussian MLE of the localizing parameter in local to unity dynamic panel regression models with deterministic and stochastic trends. Moon and Phillips find that for the homogeneous trend model, the Gaussian MLE of the common localizing parameter is $\sqrt{N}$ consistent, while for the heterogeneous trends model, it is inconsistent. The latter inconsistency is due to the presence of an infinite number of incidental parameters (as $N \to \infty$) for the individual trends. Unlike the fixed effects dynamic panel data model where this inconsistency due to the incidental parameter problem disappears as $T \to \infty$, the inconsistency of the localizing parameter in the Moon and Phillips model persists even when both N and T go to infinity. Moon and Phillips (2000) show that the local to unity parameter in a simple panel near-integrated regression model can be consistently estimated using pooled OLS. When deterministic trends are present, pooled panel estimators of the localizing parameter are asymptotically biased. Some techniques are developed to obtain consistent estimates of this localizing parameter but only in the region where it is negative. These methods are used to show how to perform efficient trend extraction for panel data. They are also used to deliver consistent estimates of distancing parameters in nonstationary panel models where the initial conditions are in the distant past. The joint asymptotics in the paper rely on $N/T \to 0$, so that the results are most relevant in panels where T is large relative to N.

Consider the nonstationary dynamic panel data model

$$
\begin{aligned}
y_{it} &= \alpha_{i,0} + \alpha_{i,1} t + y_{it}^{0} \\
y_{it}^{0} &= \beta y_{i,t-1}^{0} + u_{i,t}
\end{aligned}
$$

with $\beta = \exp(c/T)$. Moon and Phillips (2000) focused on estimating the localizing parameter c in β which characterizes the local behavior of the unit root process. Information about c is useful for the analysis of the power properties of unit root tests, cointegration tests, the construction of confidence intervals for the long-run autoregressive coefficient, the development of efficient detrending methods, and the construction of point optimal invariant tests for a unit root and cointegrating rank. Moon and Phillips (2000) show that when $c \leq 0$, it is possible to estimate this local parameter consistently using panel data. In turn, they show how to extract the deterministic trend efficiently using this consistent estimate of c.

Sun (2004) proposed a new class of estimators of the long-run average relationship in nonstationary panel time series. The estimators are based on the long-run average variance estimate using bandwidth equal to T. It is shown that the new estimators are consistent and asymptotically normal under both the sequential limit, wherein $T \to \infty$ followed by $N \to \infty$, and the joint limit where $T, N \to \infty$ simultaneously. The rate condition for the joint limit to hold is relaxed to $N^{1/2}/T \to 0$, which is less restrictive than the rate condition $N/T \to 0$, as imposed by Moon and Phillips (1999).

12.5 Panel Cointegration Tests

Like the panel unit root tests, panel cointegration tests can be motivated by the search for more powerful tests than those obtained by applying individual time-series cointegration tests. The latter tests are known to have low power especially for short T and short span of the data which is often limited to post-war annual data. In the case of purchasing power parity and convergence in growth, economists pool data on similar countries, like G7, OECD or Euro countries, in the hopes of adding cross-sectional variation to the data that will increase the power of unit root tests or panel cointegration tests.

12.5.1 Residual-Based DF and ADF Tests (Kao Tests)

Consider the panel regression model

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it}, \tag{12.15}$$

where y_{it} and x_{it} are $I(1)$ and non-cointegrated. For $z_{it} = \{\mu_i\}$, Kao (1999) proposed DF- and ADF-type unit root tests for e_{it} as a test for the null of no cointegration. The DF-type tests can be calculated from the fixed effects residuals

$$\widehat{e}_{it} = \rho \widehat{e}_{it-1} + \nu_{it}, \tag{12.16}$$

where $\widehat{e}_{it} = \widetilde{y}_{it} - \widetilde{x}_{it}'\widehat{\beta}$ and $\widetilde{y}_{it} = y_{it} - \overline{y}_{i.}$. In order to test the null hypothesis of *no cointegration*, the null can be written as H_0: $\rho = 1$. The OLS estimate of ρ and the t -statistic are given as

$$\widehat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it}\widehat{e}_{it-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it}^2}$$

and

$$t_\rho = \frac{(\widehat{\rho} - 1)\sqrt{\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{e}_{it-1}^2}}{s_e}, \tag{12.17}$$

where $s_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{e}_{it} - \widehat{\rho}\widehat{e}_{it-1}\right)^2$. Kao proposed the following four DF type tests:

$$DF_\rho = \frac{\sqrt{N}T(\widehat{\rho}-1) + 3\sqrt{N}}{\sqrt{10.2}},$$
$$DF_t = \sqrt{1.25}t_\rho + \sqrt{1.875N},$$
$$DF_\rho^* = \frac{\sqrt{N}T(\widehat{\rho}-1) + \dfrac{3\sqrt{N}\widehat{\sigma}_\nu^2}{\widehat{\sigma}_{0\nu}^2}}{\sqrt{3 + \dfrac{36\widehat{\sigma}_\nu^4}{5\widehat{\sigma}_{0\nu}^4}}},$$

and

$$DF_t^* = \frac{t_\rho + \dfrac{\sqrt{6N}\widehat{\sigma}_\nu}{2\widehat{\sigma}_{0\nu}}}{\sqrt{\dfrac{\widehat{\sigma}_{0\nu}^2}{2\widehat{\sigma}_\nu^2} + \dfrac{3\widehat{\sigma}_\nu^2}{10\widehat{\sigma}_{0\nu}^2}}},$$

where $\widehat{\sigma}_\nu^2 = \widehat{\sum}_{yy} - \widehat{\sum}_{yx}\widehat{\sum}_{xx}^{-1}$ and $\widehat{\sigma}_{0\nu}^2 = \widehat{\Omega}_{yy} - \widehat{\Omega}_{yx}\widehat{\Omega}_{xx}^{-1}$. While DF_ρ and DF_t are based on the strong exogeneity of the regressors and errors, DF_ρ^* and DF_t^* are for the cointegration with endogenous relationship between regressors and errors. For the ADF test, we can run the following regression:

$$\widehat{e}_{it} = \rho\widehat{e}_{it-1} + \sum_{j=1}^{p}\vartheta_j\Delta\widehat{e}_{it-j} + \nu_{itp}. \qquad (12.18)$$

With the null hypothesis of no cointegration, the ADF test statistics can be constructed as

$$ADF = \frac{t_{ADF} + \dfrac{\sqrt{6N}\widehat{\sigma}_\nu}{2\widehat{\sigma}_{0\nu}}}{\sqrt{\dfrac{\widehat{\sigma}_{0\nu}^2}{2\widehat{\sigma}_\nu^2} + \dfrac{3\widehat{\sigma}_\nu^2}{10\widehat{\sigma}_{0\nu}^2}}} \qquad (12.19)$$

where t_{ADF} is the t-statistic of ρ in (12.18). The asymptotic distributions of DF_ρ, DF_t, DF_ρ^*, DF_t^*, and ADF converge to a standard normal distribution $N(0, 1)$ by sequential limit theory. Kao (1999) cointegration test can be performed with EViews; see Table 12.3 and empirical example 2 in Sect. 12.7.

12.5.2 Residual-Based LM Test

McCoskey and Kao (1998) derived a residual-based test for the *null of cointegration* rather than the null of no cointegration in panels. This test is an extension of the LM test and the locally best invariant (LBI) test for an MA unit root in the time-series literature. Under the null, the asymptotics no longer depend on the asymptotic properties of the estimating spurious regression, rather the asymptotics of the estimation of a cointegrated relationship are needed. For models which allow the cointegrating vector to change across the cross-sectional observations, the asymptotics depend merely on the time-series results as each cross-section is estimated independently. For models with common slopes, the estimation is done jointly and therefore the asymptotic theory is based on the joint estimation of a cointegrated relationship in panel data.

For the residual based test of the null of cointegration, it is necessary to use an efficient estimation technique of cointegrated variables. In the time-series literature a variety of methods have been shown to be efficient asymptotically. These include the fully modified (FM)-OLS estimator and the dynamic least squares (DOLS) estimator. For panel data, Kao and Chiang (2000) showed that both the FM and DOLS methods can produce estimators which are asymptotically normally distributed with zero means.

The model presented allows for varying slopes and intercepts:

$$y_{it} = \alpha_i + x'_{it}\beta_i + e_{it}, \tag{12.20}$$

$$x_{it} = x_{it-1} + \varepsilon_{it} \tag{12.21}$$

$$e_{it} = \gamma_{it} + u_{it}, \tag{12.22}$$

and

$$\gamma_{it} = \gamma_{it-1} + \theta u_{it},$$

where u_{it} are $\text{IID}\left(0, \sigma_u^2\right)$. The null hypothesis of cointegration is equivalent to $\theta = 0$.

The test statistic proposed by McCoskey and Kao (1998) is defined as follows:

$$LM = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}S_{it}^2}{\widehat{\sigma}_e^2}, \tag{12.23}$$

where S_{it} is partial sum process of the residuals, $S_{it} = \sum_{j=1}^{t}\widehat{e}_{ij}$, and $\widehat{\sigma}_e^2$ is defined in McCoskey and Kao. The asymptotic result for the test is

$$\sqrt{N}(LM - \mu_\nu) \Rightarrow N(0, \sigma_\nu^2). \tag{12.24}$$

The moments, μ_ν and σ_ν^2, can be found through Monte Carlo simulation. The limiting distribution of LM is then free of nuisance parameters and robust to heteroskedasticity. However, simulation studies have shown that the asymptotic theory often provides a poor approximation to the empirical test distribution; see Westerlund (2005b, 2006a). Also, this test is not equipped to deal with cross-sectional dependence. Westerlund (2006a) proposes a simple procedure to reduce the size distortions of the panel LM test for cointegration suggested by McCoskey and Kao (1998). The procedure splits the sample into even and odd numbered observations, and applies the panel LM test to each subsample. The two tests are then combined using the Bonferroni principle. The Monte Carlo evidence suggests that this procedure can lead to substantial reduction in size distortions when the equilibrium errors are autoregressive. Westerlund (2006c) also extends the McCoskey and Kao (1998) LM test by allowing for the possibility of multiple structural breaks in both the level and trend of a cointegrated panel regression. Test statistics are derived when the locations of the breaks are known a priori and when they are determined endogenously from the data. Westerlund applies these tests to re-examine the solvency of the current account, and finds evidence of cointegration between saving and investment once a level break is accommodated.

For another application, see McCoskey and Kao (1998) who revisited the relationship between urbanization levels and output. In fact, they tested the long-run stability of a production function including urbanization using nonstationary panel data techniques. They applied the IPS test described in Sect. 12.2.2, and the LM test given by (12.23). Their results show that a long-run relationship between urbanization, output per worker, and capital per worker cannot be rejected for the sample of 30 developing countries or the sample of 22 developed countries over the period 1965–89. They do find, however, that the sign and magnitude of the impact of urbanization vary considerably across the countries. These results offer new insights and potential for dynamic urban models rather than the simple cross-section approach.

12.5.3 Pedroni Tests

Pedroni (2000, 2004) also proposed several tests for the null hypothesis of *no cointegration* in a panel data model that allow for considerable heterogeneity. His tests can be classified into two categories. The first set is similar to the tests discussed above, and involve averaging test statistics for cointegration in the time series across cross-sections. For the second set, the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms.

The first set of statistics includes a form of the average of the Phillips and Ouliaris (1990) statistic:

$$\tilde{Z}_\rho = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T}(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i)}{(\sum_{t=1}^{T}\hat{e}_{it-1}^2)}, \tag{12.25}$$

where $\hat{e}_{it}$ is estimated from (12.15), and $\hat{\lambda}_i = \frac{1}{2}\left(\widehat{\sigma}_i^2 - \widehat{s}_i^2\right)$, for which $\widehat{\sigma}_i^2$ and $\widehat{s}_i^2$ are individual long-run and contemporaneous variances of the residual $\hat{e}_{it}$. For his second set of statistics, Pedroni defines four panel variance ratio statistics. Let $\hat{\Omega}_i$ be a consistent estimate of Ω_i, the long-run variance–covariance matrix. Define $\hat{L}_i$ to be the lower triangular Cholesky composition of $\hat{\Omega}_i$ such that in the scalar case $\hat{L}_{22i} = \hat{\sigma}_\varepsilon$ and $\hat{L}_{11i} = \hat{\sigma}_u^2 - \dfrac{\hat{\sigma}_{u\varepsilon}^2}{\hat{\sigma}_\varepsilon^2}$ is the long-run conditional variance. Here we consider only one of these statistics:

$$Z_{t_{\hat{\rho}NT}} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}(\hat{e}_{it-1}\Delta\hat{e}_{it} - \hat{\lambda}_i)}{\sqrt{\tilde{\sigma}_{NT}^2(\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{L}_{11i}^{-2}\hat{e}_{it-1}^2)}}, \tag{12.26}$$

where $\tilde{\sigma}_{NT} = \dfrac{1}{N}\sum_{i=1}^{N}\dfrac{\hat{\sigma}_i^2}{\hat{L}_{11i}^2}$.

It should be noted that Pedroni bases his test on the average of the numerator and denominator terms, respectively, rather than the average for the statistic as a whole. Using results on convergence of functionals of Brownian motion, Pedroni finds the following result:

$$Z_{t_{\hat{\rho}_{NT}}} + 1.73\sqrt{N} \Rightarrow N(0, 0.93).$$

Note that this distribution applies to the model including an intercept and not including a time trend. Asymptotic results for other model specifications can be found in Pedroni (2000). The intuition on these tests with varying slopes is not straightforward. The convergence in distribution is based on individual convergence of the numerator and denominator terms. What is the intuition of rejection of the null hypothesis? Using the average of the overall test statistic allows more ease in interpretation: rejection of the null hypothesis means that enough of the individual cross-sections have statistics "far away" from the means predicted by theory were they to be generated under the null.

Pedroni (1999) derived asymptotic distributions and critical values for several residual-based tests of the null of no cointegration in panels where there are multiple regressors. The model includes regressions with individual-specific fixed effects and time trends. Considerable heterogeneity is allowed across individual members of the panel with regards to the associated cointegrating vectors and the dynamics of the underlying error process. By comparing results from individual countries and the panel as a whole, Pedroni (2001) rejects the strong PPP hypothesis and finds that no degree of cross-sectional dependency would be sufficient to overturn the rejection of strong PPP. Pedroni (1999, 2004) cointegration tests can be performed with EViews; see Table 12.4 and empirical example 2 in Sect. 12.7.

12.5.4 Likelihood-Based Cointegration Test

Larsson, Lyhagen and Löthgren (2001) presented a likelihood-based (LR) panel test of cointegrating rank in heterogeneous panel models based on the average of the individual rank trace statistics developed by Johansen (1995). The proposed LR-bar statistic is very similar to the IPS t-bar statistic in (12.7)–(12.10). In Monte Carlo simulation, Larsson, Lyhagen and Löthgren investigated the small sample properties of the standardized LR-bar statistic. They found that the proposed test requires a large time-series dimension. Even if the panel has a large cross-sectional dimension, the size of the test will be severely distorted. EViews computes a Fisher-type Maddala and Wu (1999) combining p-values cointegration test based on the Johansen cointegration trace test and maximum eigenvalue test; see Table 12.5 and empirical example 2 in Sect. 12.7. Larsson and Lyhagen (2007) extend Larsson, Lyhagen and Löthgren (2001) to the case where cross-sectional correlation is allowed.

Groen and Kleibergen (2003) proposed a likelihood-based framework for cointegrating analysis in panels of a fixed number of vector error correction models. This improves on Larsson, Lyhagen and Löthgren (2001) since it allows cross-sectional correlation. Maximum likelihood estimators of the cointegrating vectors are constructed using iterated generalized method of moments (GMM) estimators. Using these estimators, Groen and Kleibergen construct likelihood ratio statistics to test for a common cointegration rank across the individual vector error correction models, both with heterogeneous and homogeneous cointegrating vectors. Groen and Kleibergen (2003) applied this likelihood ratio test to a data set of exchange rates and appropriate monetary fundamentals. They found strong evidence for the validity of the monetary exchange rate model within a panel of vector correction models for three major European countries, whereas the results based on individual vector error correction models for each of these countries separately are less supportive.

Banerjee, Marcellino and Osbat (2004) show that both univariate and multivariate panel cointegration tests can be substantially over-sized in the presence of cross-unit cointegration. They argue that the panel cointegration literature assume a unique cointegrating vector in each unit, either homogeneous (Kao 1999) or heterogeneous (Pedroni 1999) across the units of the panel. Also, the studies by Groen and Kleibergen (2003) and Larsson, Lyhagen and Löthgren (2001) that developed techniques, a la Johansen's maximum likelihood method, allow for multiple cointegrating vectors in each unit. However, these models allow for cross-unit dependence through the effects of the dynamics of short run, but no account is taken of the possibility of long-run cross-unit dependence induced by the existence of cross-unit cointegrating relationships. Banerjee, Marcellino and Osbat (2004) show through Monte Carlo simulations that the consequences of using panel cointegrated methods when the restriction of no cross-unit cointegration is violated are dramatic. They also confirm the gains in efficiency when the use of the panel approach is justified. Hence, they suggest testing for the validity of no cross-unit cointegration hypothesis prior to applying panel cointegration methods. Specifically, they recommend the extraction of the common trends from each unit using the Johansen ML method, and then testing for cointegration among these trends to rule out the existence of cross-unit cointegration. Their simulation results show that this procedure works well in practice.

Gengenbach, Palm and Urbain (2006) analyze the properties of Kao (1999) and Pedroni (1999) tests for non-cointegration under cross-sectional correlation using the unobserved common factor structure of Bai and Ng (2004). They show that under this common factor model, these test statistics are no longer asymptotically normal, and converge at the rate T rather than $\sqrt{NT}$. They suggest extracting the common factors and individual components from the observed data directly and then testing for no cointegration using residual-based panel tests applied to the defactored data.

12.5.5 Finite Sample Properties

McCoskey and Kao (1998) conducted Monte Carlo experiments to compare the size and power of different residual-based tests for cointegration in heterogeneous panel data: varying slopes and varying intercepts. Two of the tests are constructed under the null hypothesis of no cointegration. These tests are based on the average ADF test and Pedroni's pooled tests in (12.25)–(12.26). The third test is based on the null hypothesis of cointegration which is based on the McCoskey and Kao LM test in (12.23). The test of the null hypothesis was originally proposed in response to the low power of the tests of the null of no cointegration, especially in the time-series case. Further, in cases where economic theory predicted a long-run steady- state relationship, it seemed that a test of the null of cointegration rather than the null of no cointegration would be appropriate. The results from the Monte Carlo study showed that the McCoskey and Kao LM test outperforms the other two tests.

Gutierrez (2003) performed Monte Carlo experiments and compared some of the panel cointegration tests proposed by Kao (1999), Pedroni (2000), Larsson, Lyhagen and Löthgren (2001). The Kao and Pedroni tests assume that either *all* the relationships are not cointegrated or *all* the relationships are cointegrated, while the Larsson, Lyhagen and Löthgren (2001) test assumes that *all* N cross-sections have at most r cointegrating relationships against the alternative of a higher rank. Gutierrez (2003) finds that for a large T panel, when the power of these tests is high, the *whole* panel may be erroneously modeled as cointegrated when only a *small fraction* of the relationships are actually cointegrated. Also, for a small T panel, when the power of these tests is low, there is a risk of modeling the *whole* panel as not cointegrated even when a *large fraction* of the relationships are actually cointegrated. For $N = 10, 25, 100$; $T = 10, 50, 100$ and the proportion of cointegrated relationships varying between $0, 0.1, 0.2, \ldots, 1$, Gutierrez (2003) finds that for small $T = 10$, and as N increases, Kao's tests show higher power than the Pedroni tests. But, this power is still fairly low even when $N = 100$. As T gets large, the Pedroni tests have higher power than the Kao tests. Both tests performed better than the the Larsson, Lyhagen and Löthgren (2001) LR-bar test.

Wagner and Hlouskova (2010) compare the performance of both single equation and system panel cointegration tests. The tests considered include those of Pedroni (1999, 2004), Westerlund (2005a), Larsson, Lyhagen and Löthgren (2001), Breitung (2005). Among the single equation tests for the null hypothesis of no cointegration two of Pedroni's tests applying the ADF principle perform best, whereas all other tests are severely undersized and have very low power in many circumstances. This power was dismal for $T \leq 25$. Pedroni's tests are also the ones least affected by the presence of an $I(2)$ component, short-run cross-sectional correlation or cross-unit cointegration. Both of Westerlund's tests were severely undersized. The system tests show very bad performance for the small values of T. They also suffer to a certain extent from a too large cross-sectional dimension. The use of finite sample correction factors only partly overcomes these problems. Both system tests are sensitive with respect to the presence of an $I(2)$ component and are not very sensitive with respect to stable autoregressive roots approaching the unit circle. The studied forms of cross-sectional correlation and cross-unit cointegration do not lead to a sizeable deterioration of the tests' performance compared to the baseline case.

12.6 Estimation and Inference in Panel Cointegration Models

For panel cointegrated regression models, the asymptotic properties of the estimators of the regression coefficients and the associated statistical tests are different from those of the time-series cointegration regression models. Some of these differences have been emphasized by Kao and Chiang (2000), Phillips and Moon (1999), Pedroni (2000, 2004) and Mark and Sul (2003), to mention a few. The panel cointegration models are directed at studying questions that surround long-run economic relationships typically encountered in macroeconomic and financial data. Such a long-run relationship is often predicted by economic theory and it is then of central interest to estimate the regression coefficients and test whether they satisfy theoretical restrictions. Phillips and Moon (1999) and Pedroni (2000) proposed a fully modified (FM) estimator, while Kao and Chiang (2000) propose an alternative approach based on a panel dynamic least squares (DOLS) estimator.

Kao and Chiang (2000) consider the following panel regression:

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + u_{it}, \tag{12.27}$$

where $\{y_{it}\}$ are 1×1, β is a $k \times 1$ vector of the slope parameters, z_{it} is the deterministic component, and $\{u_{it}\}$ are the stationary disturbance terms. $\{x_{it}\}$ are $k \times 1$ integrated processes of order one for all i, where

$$x_{it} = x_{it-1} + \varepsilon_{it}.$$

The assumption of cross-sectional independence is maintained. Under these specifications, (12.27) describes a system of cointegrated regressions, i.e., y_{it} is cointegrated with x_{it}. The OLS estimator of β is

$$\widehat{\beta}_{OLS} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}\right]. \tag{12.28}$$

It is easy to show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}' \xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\zeta_{2i}\right], \tag{12.29}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{u}_{it} \Rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\zeta_{1i}\right] \tag{12.30}$$

using sequential limit theory, where

$$\begin{array}{ccc} z_{it} & E\left[\zeta_{1i}\right] & E\left[\zeta_{2i}\right] \\ 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 \\ \mu_i & -\frac{1}{2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui} & \frac{1}{6}\Omega_{\varepsilon i} \\ (\mu_i, t) & -\frac{1}{2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui} & \frac{1}{15}\Omega_{\varepsilon i} \end{array} \tag{12.31}$$

and

$$\Omega_i = \begin{bmatrix} \Omega_{ui} & \Omega_{u\varepsilon i} \\ \Omega_{\varepsilon ui} & \Omega_{\varepsilon i} \end{bmatrix}$$

is the long-run covariance matrix of $\left(u_{it}, \varepsilon'_{it}\right)'$, also $\Delta_i = \begin{bmatrix} \Delta_{ui} & \Delta_{u\varepsilon i} \\ \Delta_{\varepsilon ui} & \Delta_{\varepsilon i} \end{bmatrix}$ is the one-sided long-run covariance. For example, when $z_{it} = \{\mu_i\}$, we get

$$\sqrt{N}T\left(\widehat{\beta}_{OLS} - \beta\right) - \sqrt{N}\delta_{NT} \Rightarrow N\left(0, 6\Omega_{\varepsilon}^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{u.\varepsilon i}\Omega_{\varepsilon i}\right)\Omega_{\varepsilon}^{-1}\right), \tag{12.32}$$

where $\Omega_{\varepsilon} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Omega_{\varepsilon i}$ and

$$\delta_{NT} = \left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)'\right]^{-1}\frac{1}{N}\left[\sum_{i=1}^{N}\Omega_{\varepsilon i}^{1/2}\left(\int \widetilde{W}_i dW_i'\right)\Omega_{\varepsilon i}^{-1/2}\Omega_{\varepsilon ui} + \Delta_{\varepsilon ui}\right] \tag{12.33}$$

This shows that $\widehat{\beta}_{OLS}$ is inconsistent using panel data. This is in sharp contrast with the consistency of $\widehat{\beta}_{OLS}$ in time series under similar circumstances. In order to deal with this bias, Kao and Chiang (2000) suggest a fully modified (FM) and DOLS estimators in a cointegrated regression and show that their limiting distribution is normal. Phillips and Moon (1999) and Pedroni (2000) also obtained similar results for the FM estimator. The reader is referred to the cited papers for further details. Kao and Chiang also investigated the finite sample properties of the OLS, FM, and DOLS estimators. They found that (i) the OLS estimator has a non-negligible bias in finite samples, (ii) the FM estimator does not improve much over the OLS estimator, while (iii) the DOLS estimator may be more promising than OLS or FM estimators in estimating the cointegrated panel regressions.

Kao, Chiang and Chen (1999) apply the asymptotic theory of panel cointegration developed by Kao and Chiang (2000) to the Coe and Helpman (1995) international R&D spillover regression. Using a sample of 21 OECD countries and Israel, they re-examine the effects of domestic and foreign R&D capital stocks on total factor productivity of these countries. They find that OLS with bias correction, the fully modified (FM), and the dynamic OLS (DOLS) estimators produce different predictions about the impact of foreign R&D on total factor productivity (TFP) although all the estimators support the result that domestic R&D is related to TFP. Kao, Chiang and Chen (1999) empirical results indicate that the estimated coefficients in the Coe and Helpman's regressions are subject to estimation bias. Given the superiority of the DOLS over FM as suggested by Kao and Chiang (2000), Kao, Chiang and Chen (1999). leaned toward rejecting the Coe and Helpman hypothesis that international R&D spillovers are trade related. FM and DOLS can be performed with EViews; see Tables 12.6 and 12.7 and empirical example 2 in Sect. 12.7.

Kauppi (2000) developed a new joint limit theory where the panel data may be cross-sectionally heterogeneous in a general way. This limit theory builds upon the concepts of joint convergence in probability and in distribution for double-indexed processes by Phillips and Moon (1999) and develops new versions of the law of large numbers and the central limit theorem that apply in panels where the data may be cross-sectionally heterogeneous in a fairly general way. Kauppi demonstrates how this joint limit theory can be applied to derive asymptotics for a panel regression where the regressors are generated by a local to unit root with heterogeneous localizing coefficients across cross-sections. Kauppi discusses issues that arise in the estimation and inference of panel cointegrated regressions with near-integrated regressors. Kauppi shows that a bias corrected pooled OLS for a common cointegrating parameter has an asymptotic normal distribution centered on the true value irrespective of whether the regressor has near or exact unit root. However, if the regression model contains individual effects and/or deterministic trends, then Kauppi's bias corrected pooled OLS still produces asymptotic bias. Kauppi also shows that the panel FM estimator is subject to asymptotic bias regardless of how individual effects and/or deterministic trends are contained if the regressors are nearly rather than exactly integrated. This indicates that much care should be taken in interpreting empirical results achieved by the panel cointegration methods that assume exact unit roots when near unit roots are equally plausible.

Choi (2002) studied instrumental variable estimation for an error component model with stationary and nearly nonstationary regressors. In contrast to the time-series literature, Choi (2002) shows that IV estimation can be used for panel data with endogenous and nearly nonstationary regressors. To illustrate, consider the simple panel regression

$$y_{it} = \alpha + \beta x_{it} + u_{it},$$

where x_{it} is nearly nonstationary, u_{it} is $I(0)$, and z_t is an instrumental variable yielding the panel IV (Within) estimator

$$\widehat{\beta}_{IV} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x}_{i.})(z_{it}-\overline{z}_{i.})\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\overline{y}_{i.}\right)(z_{it}-\overline{z}_{i.})\right]$$

Choi (2002) shows $\sqrt{N}T(\widehat{\beta}_{IV}-\beta)$ has the weak limit as $T \to \infty$ of a standardized sum (over $i = 1, \ldots, N$) of zero mean random variables divided by a standardized sum of random variables. Thus when N is large, and proper conditions hold, the central limit theorem can be applied which leads to the asymptotic normality result for the panel estimator. In time series, standard hypothesis testing cannot be performed based on the corresponding IV estimator for β. The same intuition holds for Within-IV-OLS, IV-GLS, and Within-IV-GLS estimators discussed in Choi (2002). For panel regressions that allow for cross-section correlation, one can use the SUR approach in the panel unit root test for fixed N, see Mark, Ogaki and Sul (2005) who adopt this approach and show that the dynamic GLS estimator is most efficient. In fact, Mark, Ogaki and Sul (2005) show that dynamic SUR is feasible for balanced panels where N, the number of cointegrating relationships, is *much smaller* than T. They show that it is applicable for both heterogeneous as well as homogeneous cointegrating vectors. They apply this approach to the estimation of long-run correlations between national investment and saving for a small panel of 12 OECD countries observed over 100 quarters 1970.1–1995.4. They show that the long-run slope coefficients in the saving–investment regressions are very close to one for most countries and hence they do not reject the hypothesis that the solvency condition is not violated. They also apply this method to analyzing the cointegrating regressions of the future spot exchange rate on the current forward exchange rate. Their data are spot and 30-day forward exchange rates for the Pound, Deutschmark, and Yen relative to the U.S. dollar from January 1975 to December 1996. They find that the slope coefficient in this cointegrating regression to be insignificantly different from one. Hence they conclude that the evidence for non-stationarity of excess returns is less compelling.

Breitung (2005) proposed a two-step estimation procedure for the estimation of a common cointegrating vector across individuals. He considers a panel VAR setup where the long-run relationships are *identical* for all cross-section units. In the first step, the parameters are estimated individually as in LLC, and in the second step, the common long-run parameters are estimated from a pooled regression. A likelihood ratio test for the long-run parameters as well as a test for the number of cointegrating relationships is suggested. Monte Carlo results show that this parametric approach is more effective in reducing the small sample bias than the FMOLS of Pedroni (2000) and Phillips and Moon (1999) or the DOLS of Kao and Chiang (2000).

Westerlund (2005a) develops two variance ratio tests that extend Breitung's time-series tests to the panel case. These tests are nonparametric and allow for individual-specific short-run dynamics, individual-specific intercept and trend terms, as well as individual- specific slope parameters. Both tests are designed to test the null hypothesis of no cointegration. One test is designed to test the alternative hypothesis that the panel is cointegrated as a whole, while the other one is designed to test the alternative hypothesis that the fraction of cointegrated units is positive. These tests are compared in a Monte Carlo setup with tests proposed by Pedroni (2004). The results indicate that these tests have small size distortions and good power against highly autoregressive alternatives. The Westerlund (2005a) cointegration test is available in Stata using the command *xtcointest*. This also provides Kao's (1999) and Pedroni's (1999, 2004) cointegration tests. Westerlund (2005c) examines the small-sample performance of several information-based criteria that can be employed to facilitate data-dependent endogeneity correction in the estimation of cointegrated panel regressions. The Monte Carlo evidence suggests that these information-based criteria generally perform well when $T > 50$. The Schwarz Information Criterion and the Posterior Information Criterion perform the best in terms of the correct selection frequency. Also, the evidence suggests that the criterion with best small-sample performance also leads to the best performing estimator.

Wagner and Hlouskova (2010) also compare the performance of both single equation and system panel cointegration estimators. The estimators considered include those developed by Phillips and Moon (1999), Pedroni (2000), Kao and Chiang (2000), Mark and Sul (2003), Pedroni (2001), Breitung (2005). In the case of one-dimensional cointegrating spaces, the DOLS estimator outperforms all other estimators, both single equation and system estimators, even for large samples. The DOLS estimator is also the least sensitive estimator with respect to the stable root approaching the unit circle, $I(2)$ component, cross-sectional correlation, and cross-unit cointegration. For small values of $T \leq 25$ and small $N \leq 10$, the system estimators are in many cases outperformed by the single equation estimators. Hlouskova and Wagner present finite T mean and variance correction factors and corresponding response surface regressions for the panel cointegration tests presented in Pedroni (1999, 2004), Westerlund (2005b), Larsson, Lyhagen and Löthgren (2001) and Breitung (2005).

In many applications, one is not only interested in the hypothesis of cointegration, but also whether the individual cointegration parameters can be regarded as equal. Mark and Sul (2003) suggest testing the null hypothesis of equal parameters versus the general heterogeneous alternative by means of a simple Wald test. Westerlund and Hess (2011) argue that the Wald test has a tendency to be size distorted, rejecting the null hypothesis too frequently. They suggest a new poolability test based on a Hausman (1978) test comparing two estimators of the cointegration parameters—one individual and one pooled. They test the monetary exchange rate model studied by Rapach and Wohar (2004). Their results suggest that, although there is evidence of homogeneity across a majority of the countries, the subpanels considered by Rapach and Wohar (2004) are not suitable for pooling. They also reject the monetary model when fitted to those countries for which the null of poolability was not rejected.

12.7 Empirical Examples

12.7.1 Example 1: Purchasing Power Parity

Banerjee, Marcellino and Osbat (2005) survey the empirical literature on the validity of the purchasing power parity (PPP). The strong version of PPP tests whether the real exchange rate is stationary. A common finding is that PPP holds when tested in panel data, but not when tested on a country-by-country basis. The usual explanation is that panel tests for unit roots are more powerful than their univariate counterparts. Banerjee, Marcellino and Osbat (2005) offer an alternative explanation. Their results indicate that this mismatch may be due simply to the over-sizing that is present when cointegrating relationships link the countries of the panel together. Existing panel unit root tests assume that cross-unit cointegrating relationships among the countries are not present. Banerjee, Marcellino and Osbat (2005) show through simulations that when this assumption is violated, the empirical size of the tests is substantially higher than the nominal level and the null hypothesis of a unit root is rejected too often when it is true. They demonstrate this using quarterly data on real exchange rates for the period 1975:1–2002:4 for 18 OECD countries. Computing the ADF test on a country-by-country basis using both US and Germany in turn as numeraire, Banerjee, Marcellino and Osbat (2005) fail to reject the null hypothesis of a unit root for each country at any choice of lag length except for France and Korea, when Germany is the numeraire. The panel unit roots (assuming no cross-country cointegration) on the other hand reject the null of unit root in 13 out of 16 cases with the US as numeraire. These 16 cases correspond to four tests and four different lag-length choices. The four tests include two versions of the IPS test on ($\bar{t}$ and $\overline{LM}$), the LLC test, and the Maddala and Wu (1999) Fisher test. If Germany is the numeraire, the corresponding rejections are in 12 out of 16 cases. Using critical values adjusted for the presence of cross-country cointegration, these rejections decrease. For example, with 14 bivariate cointegrating relationships, the unit root hypothesis is rejected in only 2 out of 16 cases with the US as the numeraire and never with Germany as the numeraire. The authors conclude that this finding warns against the "automatic" use of panel methods for testing for unit roots in macroeconomic time series.

Table 12.1 performs panel unit root tests on the Banerjee, Marcellino and Osbat (2005) data on real exchange rates with Germany as the numeraire. This is done using EViews. This data was kindly provided by Chiara Osbat. The EViews options allow for the choice of exogenous variables, in this case, the inclusion of individual effects, also, the automatic selection of maximum lags, or the choice of a user-specified lag. In fact, Table 12.2 performs these panel unit root tests with a user-specified lag of 1. Note that EViews performs the LLC, Breitung, IPS, and Fisher-type tests of Maddala and Wu (1999) and Choi (2001) using ADF and Phillips–Perron- type individual unit root tests. Both Tables 12.1 and 12.2 confirm the results in Banerjee, Marcellino and Osbat (2005), i.e., that all panel unit root tests that assume cross-

Table 12.1 Panel Unit Root Test (Automatic lag) for Real Exchange Rates: Germany as a Numeraire

Pool unit root test: Summary
Sample: 1975Q1 2002Q4
Series: RER_AUSTRIA, RER_BELGIUM, RER_CANADA
RER_DENMARK, RER_FINLAND, RER_FRANCE
RER_GREECE, RER_ITALY, RER_JAPAN, RER_KOREA
RER_NETHERLANDS, RER_NORWAY, RER_PORTUGAL
RER_SPAIN, RER_SWEDEN, RER_SWITZ, RER_UK, RER_US
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic selection of lags based on SIC: 0 to 8
NeweyWest bandwidth selection using Bartlett kernel

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-1.83839	0.0330	18	1970
Breitung t-stat	-3.06048	0.0011	18	1952
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-3.42675	0.0003	18	1970
ADF - Fisher Chi-square	63.6336	0.0030	18	1970
PP - Fisher Chi-square	58.1178	0.0112	18	1998
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	9.43149	0.0000	18	2016

** Probabilities for Fisher tests are computed using an asympotic Chi -square distribution. All other tests assume asymptotic normality.

Table 12.2 Panel Unit Root Test (lag = 1) for Real Exchange Rates: Germany as a Numeraire
Pool unit root test: Summary
Sample: 1975Q1 2002Q4
Series: RER_AUSTRIA, RER_BELGIUM, RER_CANADA, RER_DENMARK, RER_FINLAND, RER_FRANCE, RER_GREECE, RER_ITALY, RER_JAPAN, RER_KOREA, RER_NETHERLANDS, RER_NORWAY, RER_PORTUGAL, RER_SPAIN, RER_SWEDEN, RER_SWITZ, RER_UK, RER_US
Exogenous variables: Individual effects
User specified lags at: 1
NeweyWest
bandwidth selection using Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-1.71432	0.0432	18	1980
Breitung t-stat	-2.86966	0.0021	18	1962
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-3.04702	0.0012	18	1980
ADF - Fisher Chi-square	59.3350	0.0085	18	1980
PP - Fisher Chi-square	58.1178	0.0112	18	1998
Null: No unit root (assumes common unit root process)				
Hadri Z-stat	9.43149	0.0000	18	2016

** Probabilities for Fisher tests are computed using an asympotic Chi -square distribution. All other tests assume asymptotic normality.

section independence and include individual effects reject the null hypothesis of a common unit root. EViews also computes the Hadri (2000) residual-based LM test which reverses the null hypothesis. In this case, it rejects the null hypothesis of no unit root in any of the series in the panel in favor of a common unit root in the panel. However, using EViews, Pesaran's (2007) CIPS test assuming cross-section dependence yields a t-stat of -2.08 which has a p-value larger than 0.10. This test does not reject the null hypothesis of common unit root. Also, the Bai and Ng (2004) unit root test allowing for common factors among the 18 countries does not reject he null hypothesis of no cointegration among the 18 countries. Problem 12.3 asks the reader to replicate these results and check their sensitivity to user-specified lags as well as the choice of Germany or the U.S. as the numeraire

Drine and Rault (2008) apply panel cointegration techniques to examine the robustness of PPP using a sample of 80 developed and developing countries. They classify these countries according to three criteria: The development level and the geographic zone, the nature of the exchange rate regime (fixed versus more flexible), and the level of inflation (low versus high). They implement several panel unit root tests including IPS, Hadri (2000), Choi (2015b) and Moon and Perron (2004), as well as Pedroni (2001) panel cointegration tests. They find that strong PPP is verified for OECD countries and weak PPP for the Middle East and North African countries. However, in African, Asian, Latin American, and Central and Eastern European countries, PPP does not seem relevant to characterize the long-run behavior of the real exchange rate.

12.7.2 Example 2: International R&D Spillover

Coe and Helpman (1995) studied the international R&D spillover phenomenon using a sample of 21 OECD countries and Israel, observed over the period 1971–1990. Kao, Chiang and Chen (1999) re-examine the effects of domestic and foreign R&D capital stocks, (denoted by RD and FRD) on total factor productivity (TFP) of these countries using panel cointegration estimation methods. The data can be downloaded from https://sites.google.com/site/chihwakao/programs. Using EViews, it is easy to verify that panel unit root tests with individual effects for log(TFP), log(RD), and log (FRD) do not reject the common panel unit roots hypothesis for all 3 series considered. However, one can also show that these results are sensitive to the inclusion of both individual effects and individual linear trends. This is explored in more detail in Exercise 12.3 in Baltagi (2009). Table 12.3 shows that the panel cointegration test of Kao (1999) rejects the null hypothesis of no cointegration. Table 12.4 shows that the panel cointegration tests of Pedroni (2000) reject the null hypothesis of no

Table 12.3 Kao Residual Cointegration Test: LTFP LRD LFRD

Kao Residual Cointegration Test
Series: LTFP LRD LFRD

Sample: 1971 1990
Included observations: 440
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
Automatic lag length selection based on SIC with a max lag of 4
NeweyWest automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-2.305756	0.0106
Residual variance	0.000543	
HAC variance	0.000662	

Table 12.4 Pedroni Residual Cointegration Test: LTFP LRD LFRD
Pedroni Residual Cointegration Test
Series: LTFP LRD LFRD
Sample: 1971 1990
Included observations: 440
Crosssections included: 22
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
Automatic lag length selection based on SIC with a max lag of 3
NeweyWest automatic bandwidth selection and Bartlett kernel

Alternative hypothesis: common AR coefs. (within-dimension)

	Statistic	Prob.	Weighted Statistic	Prob.
Panel v-Statistic	0.429844	0.3337	-0.588507	0.7219
Panel rho-Statistic	-0.587120	0.2786	-0.679037	0.2486
Panel PP-Statistic	-2.011444	0.0221	-2.208595	0.0136
Panel ADF-Statistic	-4.878134	0.0000	-4.717842	0.0000

Alternative hypothesis: individual AR coefs. (between-dimension)

	Statistic	Prob.
Group rho-Statistic	1.710431	0.9564
Group PP-Statistic	-0.709493	0.2390
Group ADF-Statistic	-5.068813	0.0000

cointegration in 5 out of 11 tests using the 5% significance level. Table 12.5 shows that the Fisher panel Johansen cointegration trace test (and maximum eigenvalue test) using individual effects reject the null of no cointegration as well as the nulls of at most 1 or 2 cointegrating relationships. Table 12.6 gives the FM OLS results estimating the cointegrating relationship reported in Table 4, column (i) of Kao, Chiang and Chen (1999, p. 701). All coefficients estimates are positive and statistically significant. Table 12.7 gives the DOLS results estimating the cointegrating relationship reported in Table 5, column (i) of Kao, Chiang and Chen (1999, p. 702). Only domestic R&D is statistically significant.

12.7.3 Example 3: OECD Health Care Expenditures

Hansen and King (1996) studied the stationarity properties of real per capita health care expenditures (HCE) and real per capita gross domestic product (GDP) for 20 OECD countries over the period 1960–1987. All variables were expressed in logarithms. This was done on a country-by-country basis. Their conclusion was that these variables are nonstationary, and inference based on regressions relating HCE

Table 12.5 Johansen Fisher Panel Cointegration Test: LTFP LRD LFRD
Series: LTFP LRD LFRD
Sample: 1971 1990
Included observations: 440
Trend assumption: Linear deterministic trend
Lags interval (in first differences): 1 1
Unrestricted Cointegration Rank Test (Trace and Maximum Eigenvalue)

Hypothesized No. of CE(s)	Fisher Stat.* (from trace test)	Prob.	Fisher Stat.* (from max-eigen test)	Prob.
None	220.0	0.0000	156.6	0.0000
At most 1	108.4	0.0000	83.83	0.0003
At most 2	96.41	0.0000	96.41	0.0000

Table 12.6 Fully Modified Least Squares (FMOLS): International R&D Spillover.
Dependent Variable: LTFP
Method: Panel Fully Modified Least Squares (FMOLS)
Sample (adjusted): 1972 1990
Periods included: 19)
Crosssections included: 22
Total panel (balanced) observations: 418
Panel method: Pooled estimation
Cointegrating equation deterministics: C
Coefficient covariance computed using default method
Longrun covariance estimates (Bartlett kernel, User bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LRD	0.082285	0.017282	4.761208	0.0000
LFRD	0.114299	0.029055	3.933886	0.0001

R-squared	0.608057	Mean dependent var	-0.016188
Adjusted R-squared	0.585177	S.D. dependent var	0.031833
S.E. of regression	0.020503	Sum squared resid	0.165622
Durbin-Watson stat	0.286816	Long-run variance	0.001347

to GDP is misleading and spurious. McCoskey and Selden (1998) challenged this finding by applying the IPS panel unit root test to this data finding the hypothesis of panel unit roots is rejected when individual effects are included, but not rejecting this hypothesis when individual effects and individual time trends are included. Problem 12.5 asks the reader to verify these results. Table 12.8 shows that the panel cointegration tests of Pedroni (2000) including both individual effects and individual linear

Table 12.7 Dynamic Least Squares (DOLS): International R&D Spillover
Dependent Variable: LTFP
Method: Panel Dynamic Least Squares (DOLS)
Sample (adjusted): 1974 1989
Periods included: 16)
Crosssections included: 22
Total panel (balanced) observations: 352
Panel method: Pooled estimation
Cointegrating equation deterministics: C
Fixed leads and lags specification (lead=1, lag=2)
Coefficient covariance computed using default method
Longrun variances (Bartlett kernel, Newey-West fixed bandwidth) used for coefficient covariances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LRD	0.109353	0.023067	4.740719	0.0000
LFRD	0.047674	0.037756	1.262690	0.2082

R-squared	0.933006	Mean dependent var	-0.018867
Adjusted R-squared	0.851464	S.D. dependent var	0.034316
S.E. of regression	0.013225	Sum squared resid	0.034632
Long-run variance	0.000156		

trends reject null hypothesis of no cointegration between HCE and GDP in 3 out of 11 tests at the 5% significance level. Table 12.9 shows that the Fisher panel Johansen cointegration trace test (and maximum eigenvalue test) including both individual effects and individual linear trends reject the null hypothesis of zero or at most 1 cointegrating relationships between HCE and GDP.

12.8 Further Reading

Pesaran, Shin and Smith (1999) derived the asymptotics of a pooled mean group (PMG) estimator. The PMG estimation constrains the long-run coefficients to be identical, but allows the short run and adjustment coefficients as the error variances to differ across the cross-sectional dimension. Binder, Hsiao and Pesaran (2005) considered estimation and inference in panel vector autoregressions (PVARs) with fixed effects when T is finite and N is large. A quasi-maximum likelihood estimator as well as unit root and cointegration tests are proposed based on a transformed likelihood function. This QMLE is shown to be consistent and asymptotically normally distributed irrespective of the unit root and cointegrating properties of the PVAR model. The tests proposed are based on standard chi-square and normal distributed

Table 12.8 Pedroni Residual Cointegration Test: GDP HCE
Pedroni Residual Cointegration Test
Series: HCE GDP
Sample: 1960 1987
Included observations: 560
Cross-sections included: 20
Null Hypothesis: No cointegration
Trend assumption: Deterministic intercept and trend
Automatic lag length selection based on SIC with a max lag of 5
Newey-West automatic bandwidth selection and Bartlett kernel

Alternative hypothesis: common AR coefs. (within-dimension)

	Statistic	Prob.	Weighted Statistic	Prob.
Panel v -Statistic	1.048933	0.1471	0.901342	0.1837
Panel rho -Statistic	-0.791082	0.2144	-0.198255	0.4214
Panel PP -Statistic	-2.017957	0.0218	-1.192514	0.1165
Panel ADF -Statistic	-3.627153	0.0001	-2.552095	0.0054

Alternative hypothesis: individual AR coefs. (between-dimension)

	Statistic	Prob.
Group rho -Statistic	1.457425	0.9275
Group PP -Statistic	0.153874	0.5611
Group ADF -Statistic	-1.284771	0.0994

Table 12.9 Johansen Fisher Panel Cointegration Test: GDP HCE
Series: GDP HCE
Sample: 1960 1987
Included observations: 560
Trend assumption: Linear deterministic trend
Lags interval (in first differences): 1 1
Unrestricted Cointegration Rank Test (Trace and Maximum Eigenvalue)

Hypothesized No. of CE(s)	Fisher Stat.* (from trace test)	Prob.	Fisher Stat.* (from max-eigen test)	Prob.
None	140.7	0.0000	103.3	0.0000
At most 1	117.2	0.0000	117.2	0.0000

statistics. Binder, Hsiao and Pesaran (2005) also show that the conventional GMM estimators based on standard orthogonality conditions break down if the underlying time series contain unit roots. Monte Carlo evidence is provided which favors MLE over GMM in small samples.

Granger and Hyung (1999) consider the problem of estimating a dynamic panel regression model when the variables in the model are strongly correlated with individual-specific size factors. For a large N cross-country panel with small T, the size variable could be country-specific like its area or time-varying like population or total income. They show that if the size is not explicitly taken into account, one gets a spurious regression. In particular, they show that implementing unit root tests is likely to lead to the wrong decision. Moreover, if the size variable is slightly varying over time or its distribution has thick tails (such as a panel of countries including Luxembourg and Cyprus as well as China and India), post-sample predictions will be biased. A pooling model appears to fit well in sample, but forecast poorly out-of-sample if the individual-specific size factor has a fat-tailed distribution. A panel model with individual-specific effects could be problematic if the panel series has a very short time-dimension. Since individual constant terms are estimated poorly, the forecasts based on them are poor. These problems may be more serious if the individual-specific factor is not constant but time-varying.

Hall, Lazarova and Urga (1999) proposed an approach based on principal components analysis to test for the number of common stochastic trends driving the nonstationary series in a panel data set. The test is consistent even if there is a mixture of $I(0)$ and $I(1)$ series in the sample. This makes it unnecessary to pretest the panel for unit root. It also has the advantage of solving the problem of dimensionality encountered in large panel data sets.

Lazarova, Trapani and Urga (2007) consider the case of nonstationary heterogeneous panels where N is finite and where each unit cointegrates. In this case, a large number of conditions have to be satisfied for cointegration to be preserved in the aggregate relationship. These conditions are not likely to hold in practice. If cointegration does not carry through the aggregation process, the macro-estimates are not consistent and the information provided by the macro-summary is meaningless. However, if these conditions are mildly violated, the aggregate relationship is said to be "approximately cointegrated", in the sense that the aggregate data may only have small nonstationary components, and the strictly speaking spurious macro-relationship is observationally equivalent to a cointegration equation. Lazarova, Trapani and Urga (2007) derive a measure of the degree of non-cointegration of the aggregate estimates, and explore its asymptotic properties for finite N and large T.

Hecq, Palm and Urbain (2000) extend the concept of serial correlation common features analysis to nonstationary panel data models. This analysis is motivated both by the need to study and test for common structures and co-movements in panel data with autocorrelation present and by an increase in efficiency due to pooling. The authors propose sequential testing procedures and test their performance using a small-scale Monte Carlo. Concentrating upon the fixed effects model, they define homogeneous panel common feature models and give a series of steps to implement these tests. These tests are used to investigate the liquidity constraints model for 22

OECD and G7 countries. The presence of a panel common feature vector is rejected at the 5% nominal level.

Murray and Papell (2000) propose a panel unit roots test in the presence of structural change. In particular, they propose a unit root test for non-trending data in the presence of a one-time change in the mean for a heterogeneous panel. The date of the break is endogenously determined. The resultant test allows for both serial and contemporaneous correlation, both of which are often found to be important in the panel unit roots context. Murray and Papell conduct two power experiments for panels of non-trending, stationary series with a one-time change in means and find that conventional panel unit root tests generally have very low power. Then they conduct the same experiment using methods that test for unit roots in the presence of structural change and find that the power of the test is much improved.

Bai (2010) studies the problem of structural change for panel data with an unknown common break point. Some examples of common breaks in panel data include the following: A credit crunch or debt crisis that may affect every company's stock returns or country's GDP growth, and an oil price shock may impact every country's output. A tax policy change which may alter each firm's investment incentive. A fad or fashion which can influence consumption habits. A health scare which can affect people's exercise, smoking, and drinking habits. Other examples include an emergence of new technology, a discovery of a new medicine, or the implementation of a new governmental program. While it may be difficult to identify a break point with a single series, Bai (2010) shows that it is much easier to locate the common break point using a number of series together. This panel data approach to the estimation of break point allows for heterogeneous means for each series. For example, the effect of an oil price shock on economic growth varies from country to country, depending on whether an economy is oil importing or exporting as well as on the extent of its oil consumption. The magnitude of change in the mean growth rate can be positive for some countries and negative for some others. In the univariate time-series case, the break point cannot be consistently estimated, no matter how large the sample. Bai (2010) shows that with panel data it is possible to obtain consistent estimates, as the number of series N goes to infinity. In a univariate setting, it is impossible to identify the break point when a regime has a single observation, because the change can be mistaken with an unusual realization of the disturbance term. With panel data, Bai (2010) shows that consistency is attainable even when a regime has a single observation. This property is especially useful when the objective is to locate as quickly as possible the onset of a new regime or the turning point, without the need of waiting for many observations from the new regime.

Im, Lee and Tieslau (2005) propose a panel unit-root LM test whose asymptotic distribution is not affected by the presence of structural shifts. This result holds under a mild condition that $N/T \to k$, where k is any finite constant. They apply their test to the purchasing power parity (PPP) hypothesis allowing for a maximum of two structural shifts for each time series. For a panel of 6, 12, 15, and 21 countries, with monthly and quarterly real exchange rates from April 1973 to December 1999, they find in all cases, strong evidence for PPP.

Westerlund (2005b) proposes a residual-based test of the null hypothesis of panel cointegration that allows for mixtures of cointegrated and spurious alternatives. The test is an extension of the CUSUM test proposed in the time-series context, which is based on measuring the fluctuation in the regression residuals. The intuition behind the test is that if the series are cointegrated, then the residuals should be stable and their fluctuations reflect only equilibrium errors. Thus, the null hypothesis of cointegration should be rejected whenever there is excessive fluctuation in the residual series. The proposed test is shown to be asymptotically normal under the null hypothesis that is free of nuisance parameters and it is robust to heteroskedasticity. He applies the CUSUM test to the international R&D spillover regressions of Coe and Helpman (1995). He finds no evidence of homogenous cointegration, but rather that total factor productivity is heterogeneously cointegrated with foreign and domestic R&D capital stocks. Westerlund (2006b) also proposes four simple tests for the null hypothesis of no panel cointegration in the presence of a single unknown level break for each individual regression. The tests are general enough to allow for endogenous regressors, serial correlation, and heterogeneous breaks of unknown timing. Using sequential limit arguments, the distributions of these tests are found to be normal and free of nuisance parameter dependencies. Critical values for up to five regressors are provided and a small Monte Carlo study is conducted to investigate the finite sample properties of these tests. The results show that these tests have small size distortions and good power. Westerlund (2007) builds on the result in time series that failure of the common factor restriction can cause significant loss of power for residual-based cointegration tests. In fact, he proposes four panel tests to test the null hypothesis of no cointegration by testing whether the error correction term in a conditional error correction model is equal to zero. If the null hypothesis of no error correction is rejected, then the null hypothesis of no cointegration is also rejected. These tests are able to accommodate individual-specific short-run dynamics, including serially correlated error terms and non-strictly exogenous regressors, individual-specific intercept and trend terms, as well as individual-specific slope parameters. A bootstrap procedure is also proposed to handle applications with cross-sectionally dependent data. Westerlund re-examines the evidence relating international health care expenditures and GDP using a panel consisting of 20 OECD countries over the period 1970–2001. He finds that the two series are cointegrated once the possibility of an invalid common factor restriction has been accounted for.

Choi and Chue (2007) study subsampling hypothesis tests for panel data that may be nonstationary, cross-sectionally correlated, and cross-sectionally cointegrated. The subsampling approach to hypothesis testing allows the regressors to be stationary or nonstationary with unit roots, or they may be a mixture of both types. It also allows for cross-sectional correlation that need not be estimated. This implies that there is less chance of size distortions due to misspecification, say from procedures assuming factor structures. Cross-sectional cointegration is also allowed without requiring knowledge of the cointegration coefficients and ranks. The subsampling approach provides approximations to the finite sample distributions of the tests without estimating nuisance parameters. The tests include panel unit root and cointegration tests as special cases. The number of cross-sectional units is assumed to be finite and that of time-series observations infinite. It is shown that subsampling provides asymptotic distributions that are equivalent to the asymptotic distributions of the panel tests. In addition, the tests using critical values from subsampling are shown to be consistent.

Kapetanios (2007) adopts the factor-based cross-sectional dependence paradigm of Bai and Ng (2004) to handle panel unit root tests, but suggests alternative factor extraction methods. These include the dynamic principal component method and the parametric state-space dynamic approach. A Monte Carlo study of these methods for multiple and persistent factors is undertaken. Previous simulation work in the literature has mainly focused on single serially uncorrelated factors to introduce cross-sectional dependence in panel data sets which may be restrictive. Simulation results suggest that the presence of multiple factors with persistent dynamics pose a significant problem for the available factor-based panel unit root tests. The actual rejection probabilities exceed the nominal significance level and in some cases it is preferable not to correct at all for cross-sectional dependence. The number of factors is of importance as well. Multiple factors are less easily extracted and have a further adverse impact on inference. Kapetanios suggests that a reasonable alternative to the factor-based tests seems to be the use of the Pesaran (2007) test in a number of cases.

Pesaran (2006) suggests that linear combinations of unobserved factors can be well approximated by cross-section averages of the dependent variable and the observed regressors. This leads to the Common Correlated Effects (CCE) estimator that can be computed by running standard panel regressions augmented with the cross-section averages of the dependent and independent variables. The CCE procedure does not require the number of unobserved factors to be smaller than the number of observed cross-section averages. This can be implemented with Stata using the *xtmg* command with option *cce*. Kapetanios, Pesaran and Yamagata (2011) extend this work to the case where the unobservable common factors follow unit root processes. Their analysis does not require a priori knowledge of the number of unobserved factors. It only requires that the number of unobserved factors remains fixed as the sample size increases. Their Monte Carlo experiments show that the CCE estimator is robust to a wide variety of data generation processes.

Serlenga and Shin (2007) develop a generalized Hausman–Taylor (HT) estimator for a heterogeneous panel with unobserved common time- specific factors. Specifically, this paper extends the correlated common effect pooled (CCEP) estimator of Pesaran (2006) in order to deal with time invariant as well as country invariant regressors. This is applied to the estimation of a gravity equation of bilateral trade among 15 EU member countries over the period 1960–2001. Empirical results show that this heterogeneous approach yields more sensible results than assuming homogeneous fixed time effects.

Moon and Perron (2007) study non-stationarities in a panel of 25 monthly Canadian and U.S. interest rates of different maturities and risk, spanning the period January 1985 to April 2004. They find significant cross-sectional correlation among the series in the panel, and model this cross-sectional dependence as a linear dynamic factor model. They then decompose the panel into common and idiosyncratic components, and analyze these in turn. Moon and Perron find that interest rates are characterized by a single nonstationary factor and some stationary idiosyncratic components, and conclude that the series are cointegrated. The results also suggest that the dominant factor in the interest rate panel is a level factor that is highly correlated with all rates and could be the result of inflationary expectations. The second factor has an interpretation as a slope factor, that is the differential between a long rate and a short rate since it affects short and long rates differently and might be a measure of the business cycle.

Sarafidis and Robertson (2009) consider the impact of error cross-sectional dependence (modeled as a factor structure) on FE and GMM estimators in the context of a dynamic panel data model. They show that the standard moment conditions used by these estimators are invalid, and as a result these estimators are inconsistent as $N \to \infty$ for a fixed T. Transforming the data in terms of deviations from time-specific averages helps to reduce the asymptotic bias of the estimators, unless the factor loadings have mean zero. Monte Carlo results suggest that the bias of these estimators can be severe to the extent that the standard FE estimator is not generally inferior anymore in terms of root median square error. Time-specific demeaning alleviates the problem, although the effectiveness of this transformation decreases when the variance of the factor loadings is large.

Westerlund (2008) develops two new panel cointegration tests of the null hypothesis of no cointegration that can be applied under very general conditions. These tests are based on a Hausman-type test comparing two estimators of a unit root in the residuals of an estimated regression. These tests are based on defactored residuals correcting for factors that are common across units. The asymptotic distributions of these tests are shown to be normal. Results from a small Monte Carlo study suggest that the tests have small size distortions and greater power than other popular panel cointegration tests. Using quarter panel data for 20 OECD countries observed over the period 1980–2004, Westerlund (2008) finds that the Fisher effect, stating that inflation and nominal interest rates should cointegrate with a unit slope on inflation, cannot be rejected once the panel evidence on cointegration has been taken into account.

Bai (2009) considers a panel data regression with large N and T that has unobservable multiple interactive effects, which are correlated with the regressors. The disturbance term is given by

$$u_{it} = \lambda_i' F_t + \varepsilon_{it} \qquad i = 1, \ldots N, t = 1, \ldots, T.$$

where λ_i is an $(r \times 1)$ vector of factor loadings and F_t is an $(r \times 1)$ vector of common factors. The emphasis here is on the *heterogeneous* impact of the common macro-shocks F_t on each country. When the common shocks have homogeneous effects, i.e., $\lambda_i = \lambda$ for all i, the model collapses to the usual fixed time effects. Similarly, in a Mincer wage equation, the interactive effects could be the result of changing prices for a vector of unmeasured skills. If the prices are constant over time, the standard individual fixed-effects model is obtained. Bai derives an interactive effects estimator that is $\sqrt{NT}$consistent. He also suggests a Hausman test for testing additive versus interactive effects.

Baltagi, Kao and Liu (2008) study the asymptotic properties of standard panel data estimators including OLS, FE and FD, and GLS estimators when both T and N are large. This is done in the context of a simple panel regression model with error component disturbances where both the regressor and the remainder disturbance term are assumed to be autoregressive and possibly nonstationary. They show that all the estimators have asymptotic normal distributions and have different convergence rates dependent on the non-stationarity of the regressors and the remainder disturbances. In fact, when the error term is $I(0)$ and the regressor is $I(1)$, the FE estimator is asymptotically equivalent to the GLS estimator and OLS is less efficient than GLS (due to a slower convergence speed). However, when the error term and the regressor are $I(1)$, GLS is more efficient than the FE estimator since GLS is $\sqrt{NT}$consistent, while FE is $\sqrt{N}$ consistent. This implies that GLS is the preferred estimator under both cases (i.e., regression error is either $I(0)$ or $I(1)$). Monte Carlo experiments show that the loss in efficiency of the OLS, FE, and FD estimators relative to true GLS can be substantial.

Demetrescu, Hassler and Tarcolea (2006) suggest a modification of Choi (2001) inverse normal combination test given as Z in Sect. 12.2.4 that relaxes the assumption of independent individual time series. They show that under certain conditions, this test performs better than the Maddala and Wu (1999) combination test given in (12.11) and Chang (2002) nonlinear IV unit root test. Also, Sheng and Yang (2013) employ a truncated product statistic for combination tests which truncate some large p_i-values (they used $p_i < 0.1$). This truncation reduces the impact of large p_i-values which may affect the power of the usual combination tests. They show that their truncated product test outperforms the Pesaran's (2007) test and the Demetrescu, Hassler and Tarcolea (2006) test.

For seasonal panel unit root tests, see Ottero, Smith and Giulietti (2007). This test allows for cross-sectional dependence using bootstrapping and Peasran's (2007) approach.

The reader is referred to Chaps. 12 and 15 of the Oxford Handbook of Panel Data. Chapter 12 is entitled panel cointegration by Choi (2015b) and Chap. 15 is entitled The analysis of macroeconomic panel data by Breitung (2015). For a thorough and up to date treatment of unit root testing, see Choi (2015a), especially Chap. 7 on panel unit roots. We close this chapter with a quote from Choi (2015a, p.222): "There seems to be no consensus yet on how best we can test for unit roots and stationarity for dependent panels. Each method developed thus far works well within its model specification, but it becomes less appealing under other specifications".

12.9 Notes

1. Early special issues on nonstationary panels include Banerjee (1999) in the *Oxford Bulletin of Economics and Statistics*, Baltagi, Fomby and Hill (2000) in the *Advances in Econometrics.* See also two special issues of the *Journal of Applied Econometrics,* discussing cross-section dependence in panel data models, edited by Baltagi and Pesaran (2007) and the other by Bai, Baltagi and Pesaran (2016)

and one special issue of *Econometric Reviews* edited by Baltagi and Maasoumi (2013).

2. The Levin, Lin and Chu (2002) paper has its origins in a Levin and Lin working paper in 1992, and most early applications in economics were based on the latter paper. In fact, this panel unit root test was commonly cited as the Levin–Lin test.

12.10 Problems

12.1 *A simple linear trend model with error components.* This is based on problem 97.2.1 in *Econometric Theory* by Baltagi and Krämer (1997). Consider the following simple linear trend: model

$$y_{it} = \alpha + \beta t + u_{it} \qquad i = 1, 2, \ldots, N, \text{ and } t = 1, 2, \ldots, T,$$

where y_{it} denotes the gross domestic product of country i at time t. The disturbances follow the one-way error component model given by

$$u_{it} = \mu_i + \nu_{it},$$

where $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ denote the random country (time-invariant) effects and $\nu_{it} \sim \text{IID}(0, \sigma_\nu^2)$ denote the remainder effects. These error components are assumed to be independent of each other and among themselves. Our interest focuses on the estimates of the trend coefficient β, and the estimators to be considered are ordinary least squares (OLS), first difference (FD), the fixed effects (FE) estimator, assuming the μ_i's are fixed effects, and the generalized least squares estimator (GLS), knowing the true variance components, which is the best linear unbiased estimator in this case.

(a) Show that the OLS, GLS, and FE estimators of β are identical and given by $\widehat{\beta}_{GLS} = \widehat{\beta}_{OLS} = \widetilde{\beta}_{FE} = \sum_{i=1}^N \sum_{t=1}^T y_{it}(t - \bar{t})/N \sum_{t=1}^T (t - \bar{t})^2$ where $\bar{t} = \sum_{t=1}^T t/T$.

(b) Show that the variance of the OLS, GLS, and FE estimators of β is given by $\text{var}(\widehat{\beta}_{GLS}) = \text{var}(\widehat{\beta}_{OLS}) = \text{var}(\widetilde{\beta}_{FE}) = 12\sigma_\nu^2/NT(T^2 - 1)$ and is therefore $O(N^{-1}T^{-3})$.

(c) Show that this simple linear trend model satisfies the necessary and sufficient condition for OLS to be equivalent to GLS.

(d) Show that the FD estimator of β is given by $\widehat{\beta}_{FD} = \sum_{i=1}^N (y_{iT} - y_{i1})/N(T - 1)$ with $\text{var}(\widehat{\beta}_{FD}) = 2\sigma_\nu^2/N(T - 1)^2$ of $O(N^{-1}T^{-2})$.

(e) What do you conclude about the asymptotic relative efficiency of FD with respect to the other estimators of β as $T \to \infty$? Hint: See solution 97.2.1 in *Econometric Theory* by Song and Jung (1998). Also, use the fact that $\sum_{t=1}^T t^2 = T(T + 1)(2T + 1)/6$ and $\sum_{t=1}^T t = T(T + 1)/2$.

12.2 *International R&D spillover*. Download the International R&D spillovers panel data set used by Kao, Chiang and Chen (1999) along with the GAUSS subroutines from https://sites.google.com/site/chihwakao/programs.

Using this data set, replicate the following results:
(a) Perform the panel unit root tests on total factor productivity, domestic R&D, and foreign R&D capital stocks. Show that the null hypothesis of panel unit roots is not rejected for all three variables.
(b) Perform the Kao (1999) and Pedroni (2000) panel cointegration tests on the regression relating total factor productivity to domestic and foreign R&D stocks. Show that the null hypothesis of no cointegration is rejected.
(c) Estimate the cointegrating relationship using FMOLS of Phillips and Moon (1999) and Pedroni (2000) and DOLS of Kao and Chiang (2000). This replicates Tables 4 and 5, columns (i) of Kao, Chiang and Chen (1999, pp. 701–702).

12.3 *Purchasing Power Parity*. Using the Banerjee, Marcellino and Osbat (2005) quarterly data set on real exchange rate for 18 OECD countries over the period 1975:1–2002:4.
(a) Replicate the panel unit root test in Table 12.1 with Germany as the numeraire. Check the sensitivity of these results to a user-specified lag of 1, 2, 3, and 4. Compare with Table 8 of Banerjee, Marcellino and Osbat (2005).
(b) Perform the panel unit root test as in Table 12.1 but now with the U.S. as the numeraire. Check the sensitivity of these results to a user-specified lag of 1, 2, 3, and 4. Compare with Table 8 of Banerjee, Marcellino and Osbat (2005).
(c) Perform the individual ADF unit root tests on a country-by- country basis for both parts (a) and (b). Compare with Table 7 of Banerjee, Marcellino and Osbat (2005). What do you conclude?
(d) Check the sensitivity of the results in parts (a) and (b) when both individual effects and individual linear trends are included.
(e) Perform the Pesaran (2007) CIPS test and the Bai and Ng (2004) unit root test. What do you conclude?

12.4 *Panel unit root tests: GDP of G7 countries*. Using the EViews G7 countries work file (Poolg7) containing the GDP of Canada, France, Germany, Italy, Japan, UK, and US.
(a) Perform the panel unit root tests using individual effects in the deterministic variables.
(b) Check the sensitivity of these results to a user-specified lag of 1, 2, 3, and 4. Show that all tests are in agreement about the possibility of a common unit root for all series.
(c) Check the sensitivity of the results in parts (a) and (b) when both individual effects and individual linear trends are included.

12.5 *Health care expenditures*. This problem is based on the Hansen and King (1996) data set and the replication by McCoskey and Selden (1998). Hansen and King (1996) studied the stationarity properties of real per capita health care expenditures (HCE) and real per capita gross domestic product (GDP) for 20 OECD countries over the period 1960–1987. All variables were expressed in logarithms. This was done on a country-by-country basis. Their conclusion was that these variables are nonstationary and inference based on regressions relating HCE to GDP are misleading and spurious. McCoskey and Selden (1998) challenged this finding by applying the IPS panel unit root test to this data finding the hypothesis of panel unit roots is rejected when individual effects are included, but not rejecting this hypothesis when individual effects *and* individual time trends are included. Using EViews, replicate the following results:

(a) Perform the panel unit root tests on HCE and GDP when individual effects are included.

(b) Check the sensitivity of these results when both individual effects and individual linear trends are included.

(c) Perform the Pedroni test including individual effects and individual linear trends in the deterministic variables.

(d) Perform the Fisher panel Johansen cointegration trace and maximum eigenvalue tests including individual effects and individual linear trends in the deterministic variables.

12.6 *Penn World Table*. Using the Penn World Table exchange rates in Stata (*webuse pennxrate*):

(a) Perform the LLC panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot llc lnrxrate if oecd, lags(aic 10) kernel(bartlett nwest) trend.*

(b) Perform the Harris and Tzavalis panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot ht lnrxrate if oecd, demean trend.*

(c) Perform the IPS panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot ips lnrxrate if oecd, lags(aic 3) trend.*

(d) Perform the Breitung panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot breitung lnrxrate if oecd, lags(3) robust trend.*

(e) Perform the combining p-values Fisher panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot fisher lnrxrate if oecd, dfuller lags(3) trend.*

(f) Perform the Hadri panel stationarity tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the command *xtunitroot hadri lnrxrate if oecd, kernel(bartlett) trend.*

(g) Compute the average correlation coefficients and Pesaran's CD test for ln(exchange rates) for OECD countries. This can be done using the command *xtcd lnxrate if oecd.*

(h) Perform the Maddala and Wu (1999) and Pesaran (2007) CIPS panel unit root tests for ln(exchange rates) for OECD countries with trend and without trend. This can be done using the Stata command *multipurt lnrxrate if oecd, lags(3).*

12.7 *Black markets for foreign exchange*. Luintel (2000) studies the behavior of real exchange rates (relative to the US dollar) using monthly data obtained from the black markets for foreign exchange of eight Asian developing countries. The sample period is 1958:1–1989:6. The data is available from the *Journal of Applied Econometrics* data archives.

(a) Plot this real exchange rate data for each country, i.e., replicate Fig. 1 of (Luintel, 2000, p. 166).
(b) Replicate Table III of (Luintel, 2000, p. 166). This performs ADF unit root tests on the real exchange rate for each country separately.
(c) Perform the panel unit root tests on the real exchange rate for these 8 countries. Compare with Table IV of (Luintel, 2000, p. 173). What do you conclude?

12.8 *Inflation rates*. Culver and Papell (1997) test for unit roots using the inflation rates of 13 OECD countries. With individual country tests, they find evidence of stationarity in only four of the thirteen countries. However, they reject the unit root hypothesis both for a panel of all thirteen countries and for a number of smaller panels consisting of as few as three countries. The inflation rate is calculated by differencing the logarithm of the consumer price index. This data was obtained from the *International Monetary Fund*'s *International Financial Statistics.* All series start in February 1957 and end in September 1994. The data is available from the *Journal of Applied Econometrics* data archives. (a) Replicate Table I on p. 437 of Culver and Papell (1997). Show that using the ADF test (with the lag levels provided in that table), the null hypothesis of unit root can be rejected at the 5% level for only three of the thirteen countries: France, Netherlands, and Japan, and can be rejected at the 10% level for Norway. (b) Perform the panel unit root tests (with individual country effects but no trend) provided by EViews. Show that the conclusion is basically the same as that reached by Culver and Papell (1997). Their results reported in Table III on p.442 indicate rejection of the null hypothesis of unit root for all countries.

References

Arellano, M., and O. Bover. 1995. Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics* 68: 29–51.

Bai, J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–1279.

Bai, J. 2010. Common breaks in means and variances for panel data. *Journal of Econometrics* 157: 78–92.

Bai, J., B.H. Baltagi, and M.H. Pesaran (eds.). 2016. Cross-sectional dependence in panel data models: A special issue. *Journal of Applied Econometrics* 31: 1–297.

Bai, J., and S. Ng. 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72: 1127–1177.

Baltagi, B.H. 2009. *A companion to econometric analysis of panel data*. Chichester: John Wiley.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2007. Panel unit root tests and spatial dependence. *Journal of Applied Econometrics* 22: 339–360.

Baltagi, B.H., T.B. Fomby, and R.C. Hill (eds.). 2000. *Nonstationary panels, panel cointegration, and dynamic panels, Advances in Econometrics*, vol. 15. Amsterdam: Elsevier Science.

Baltagi, B.H., and C. Kao. 2000. Nonstationary panels, cointegration in panels and dynamic panels: A survey. *Advances in Econometrics* 15: 7–51.

Baltagi, B.H., C. Kao, and L. Liu. 2008. Asymptotic properties of estimators for the linear panel regression model with individual effects and serially correlated errors: The case of stationary and non-stationary regressors and residuals. *Econometrics Journal* 11: 554–572.

Baltagi, B.H., and W. Krämer. 1997. A simple linear trend model with error components. Problem 97.2.1. *Econometric Theory* 13: 463.

Baltagi, B.H., and E. Maasoumi. 2013. An overview of dependence in cross-section, time-series and panel data. *Econometric Reviews* 32: 543–546.

Baltagi, B.H., and F. Moscone. 2010. Health care expenditure and income in the OECD reconsidered: Evidence from panel data. *Economic Modelling* 27: 804–811.

Baltagi, B.H., and M.H. Pesaran. 2007. Heterogeneity and cross section dependence in panel data models: Theory and applications introduction. *Journal of Applied Econometrics* 22: 229–232.

Banerjee, A. 1999. Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics* 61: 607–629.

Banerjee, A., M. Marcellino, and C. Osbat. 2004. Some cautions on the use of panel methods for integrated series of macro-economic data. *Econometrics Journal* 7: 322–340.

Banerjee, A., M. Marcellino, and C. Osbat. 2005. Testing for PPP: Should we use panel methods? *Empirical Economics* 30: 77–91.

Binder, M., C. Hsiao, and M.H. Pesaran. 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21: 795–837.

Breitung, J. 2000. The local power of some unit root tests for panel data. *Advances in Econometrics* 15: 161–177.

Breitung, J. 2005. A parametric approach to the estimation of cointegration vectors in panel data. *Econometric Reviews* 24: 151–173.

Breitung, J. 2015. The analysis of macroeconomic panel data, Chapter 15. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 453–492. Oxford: Oxford University Press.

Breitung, J., and M.H. Pesaran. 2008. Unit roots and cointegration in panels, Chapter 9. In *The econometrics of panel data: Fundamentals and recent developments in theory and practice*, (eds.) L. Matyas, and P. Sevestre, 279–322. Berlin: Springer.

Chang, Y. 2002. Nonlinear IV unit root tests in panels with cross sectional dependency. *Journal of Econometrics* 110: 261–292.

Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20: 249–272.

Choi, I. 2002. Instrumental variables estimation of a nearly nonstationary, heterogeneous error component model. *Journal of Econometrics* 109: 1–32.

Choi, I. 2015a. *Almost all about unit roots: Foundations, developments and applications*. New York: Camberidge University Press.

Choi, I. 2015b. Panel cointegration, Chapter 2. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, 46–75. Oxford: Oxford University Press.

Choi, I., and T.K. Chue. 2007. Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices. *Journal of Applied Econometrics* 22: 233–264.

Coe, D., and E. Helpman. 1995. International R&D spillovers. *European Economic Review* 39: 859–887.

Culver, S.E., and D. Papell. 1997. Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics* 12: 435–444.

De Blander, R., and G. Dhaene. 2012. Unit root tests for panel data with AR(1) errors and small T. *Econometrics Journal* 15: 101–124.

Demetrescu, M., U. Hassler, and A. Tarcolea. 2006. *Oxford Bulletin of Economics and Statistics* 68: 647–663.

Drine, I., and C. Rault. 2008. Purchasing power parity for developing and developed countries. What can we learn from non-stationary panel data models? *Journal of Economic Surveys* 22: 752–773.

Eberhardt, M., and F. Teal. 2011. Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys* 25: 109–155.

Elliott, G., T.J. Rothenberg, and J.H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836.

Gengenbach, C., F.C. Palm, and J.P. Urbain. 2006. Cointegration testing in panels with common factors. *Oxford Bulletin of Economics and Statistics* 68: 683–719.

Gengenbach, C., F.C. Palm, and J.P. Urbain. 2010. Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling. *Econometric Reviews* 29: 111–145.

Gerdtham, U.G., and M. Löthgren. 2000. On stationarity and cointegration of international health expenditure and GDP. *Journal of Health Economics* 19: 461–475.

Granger, C.W.J., and N. Hyung. 1999. Spurious stochastics in a short time-series panel data. *Annales D'Économie et de Statistique* 55–56: 299–315.

Groen, J.J.J., and F. Kleibergen. 2003. Likelihood-based cointegration analysis in panels of vector error correction models. *Journal of Business and Economic Statistics* 21: 295–318.

Gutierrez, L. 2003. On the power of panel cointegration tests: A Monte Carlo comparison. *Economics Letters* 80: 105–111.

Hadri, K. 2000. Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3: 148–161.

Hadri, K., and R. Larsson. 2005. Testing for stationarity in heterogeneous panel data where the time dimension is finite. *Econometrics Journal* 8: 55–69.

Hall, S., S. Lazarova, and G. Urga. 1999. A principal components analysis of common stochastic trends in heterogeneous panel data: Some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics* 61: 749–767.

Hansen, P., and A. King. 1996. The determinants of health care expenditure: A cointegration approach. *Journal of Health Economics* 15: 127–137.

Harris, R.D.F., and E. Tzavalis. 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–226.

Harris, R.D.F., and E. Tzavalis. 2004. Testing for unit roots in dynamic panels in the presence of a deterministic trend: Re-examining the unit root hypothesis for real stock prices and dividends. *Econometric Reviews* 23: 149–166.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hecq, A., F.C. Palm, and J.P. Urbain. 2000. Testing for common cyclical features in non-stationary panel data models. *Advances in Econometrics* 15: 131–160.

Hlouskova, J., and M. Wagner. 2006. The performance of panel unit root and stationarity tests: Results from a large scale simulation study. *Econometric Reviews* 25: 85–116.

Im, K.S., J. Lee, and M. Tieslau. 2005. Panel LM unit-root tests with level shifts. *Oxford Bulletin of Economics and Statistics* 67: 393–419.

Im, K.S., M.H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74.

Johansen, S. 1995. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Kao, C. 1999. Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90: 1–44.

Kao, C., and M.H. Chiang. 2000. On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics* 15: 179–222.

Kao, C., M.H. Chiang, and B. Chen. 1999. International R&D spillovers: An application of estimation and inference in panel cointegration. *Oxford Bulletin of Economics and Statistics* 61: 691–709.

Kapetanios, G. 2007. Dynamic factor extraction of cross-sectional dependence in panel unit root tests. *Journal of Applied Econometrics* 27: 313–338.

Kapetanios, G., M.H. Pesaran, and T. Yamagata. 2011. Panels with non-stationary multifactor error structures. *Journal of Econometrics* 160: 326–348.

Karlsson, S., and M. Löthgren. 2000. On the power and interpretation of panel unit root tests. *Economics Letters* 66: 249–255.

Kauppi, H. 2000. Panel data limit theory and asymptotic analysis of a panel regression with near integrated regressors. *Advances in Econometrics* 15: 239–274.

Larsson, R., and J. Lyhagen. 2007. Inference in panel cointegration models with long panels. *Journal of Business and Economic Statistics* 25: 473–483.

Larsson, R., J. Lyhagen, and M. Löthgren. 2001. Likelihood-based cointegration tests in heterogeneous panels. *Econometrics Journal* 4: 109–142.

Lazarova, S., L. Trapani, and G. Urga. 2007. Common stochastic trends and aggregation in heterogeneous panels. *Econometric Theory* 23: 89–105.

Levin, A., C.F. Lin, and C. Chu. 2002. Unit root test in panel data: Asymptotic and finite sample properties. *Journal of Econometrics* 108: 1–24.

Luintel, K.B. 2000. Real exchange rate behaviour: Evidence from black markets. *Journal of Applied Econometrics* 15: 161–185.

Maddala, G.S., and S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61: 631–652.

Maddala, G.S., S. Wu and P.C. Liu. 2000. Do panel data rescue purchasing power parity (PPP) theory?, Chapter 2 In *Panel data econometrics: Future directions*, (eds.) J. Krishnakumar and E. Ronchetti, 35–51. North-Holland, Amsterdam.

Mark, N.C., M. Ogaki, and D. Sul. 2005. Dynamic seemingly unrelated cointegrating regression. *Review of Economic Studies* 72: 797–820.

Mark, N.C., and D. Sul. 2003. Cointegration vector estimation by panel dynamic OLS and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65: 655–680.

McCoskey, S., and C. Kao. 1998. A residual-based test of the null of cointegration in panel data. *Econometric Reviews* 17: 57–84.

McCoskey, S., and T. Selden. 1998. Health care expenditures and GDP: Panel data unit root test results. *Journal of Health Economics* 17: 369–376.

Moon, H.R., and B. Perron. 2004. Testing for unit root in panels with dynamic factors. *Journal of Econometrics* 122: 81–126.

Moon, H.R., and B. Perron. 2007. An empirical analysis of nonstationarity in a panel of interest rates with factors. *Journal of Applied Econometrics* 22: 383–400.

Moon, H.R., and P.C.B. Phillips. 1999. Maximum likelihood estimation in panels with incidental trends. *Oxford Bulletin of Economics and Statistics* 61: 711–747.

Moon, H.R., and P.C.B. Phillips. 2000. Estimation of autoregressive roots near unity using panel data. *Econometric Theory* 16: 927–997.

Murray, C.J., and D.H. Papell. 2000. Testing for unit roots in panels in the presence of structural change with an application to OECD unemployment. *Advances in Econometrics* 15: 223–238.

O'Connell, P.G.J. 1998. The overvaluation of purchasing power parity. *Journal of International Economics* 44: 1–19.

Ottero, J., J. Smith, and M. Giulietti. 2007. Testing for seasonal unit roots in heterogeneous panels in the presence of cross-sectional dependence. *Economics Letters* 97: 179–184.

Pedroni, P. 1999. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* 61: 653–678.

Pedroni, P. 2000. Fully modified OLS for heterogeneous cointegrated panels. *Advances in Econometrics* 15: 93–130.

Pedroni, P. 2001. Purchasing power parity tests in cointegrated panels. *Review of Economics and Statistics* 83: 727–731.

Pedroni, P. 2004. Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory* 20: 597–625.

Pesaran, M.H. 2004. General Diagnostic Tests for Cross-section Dependence in Panels. Working paper, Trinity College, Cambridge.

Pesaran, M.H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012.

Pesaran, M.H. 2007. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* 27: 265–312.

Pesaran, M.H., and R. Smith. 1995. Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics* 68: 79–113.
Pesaran, M.H., Y. Shin, and R. Smith. 1999. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94: 621–634.
Phillips, P.C.B., and H. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67: 1057–1111.
Phillips, P.C.B., and H. Moon. 2000. Nonstationary panel data analysis: An overview of some recent developments. *Econometric Reviews* 19: 263–286.
Phillips, P.C.B., and S. Ouliaris. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58: 165–193.
Phillips, P.C.B., and D. Sul. 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6: 217–259.
Rapach, D.E., and M.E. Wohar. 2004. Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23: 867–895.
Sarafidis, V., and D. Robertson. 2009. On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal* 12: 62–81.
Serlenga, L., and Y. Shin. 2007. Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics* 22: 361–381.
Sheng, X., and J. Yang. 2013. Truncated product methods for panel unit root tests. *Oxford Bulletin of Economics and Statistics* 75: 624–636.
Shin, Y., and A. Snell. 2006. Mean group tests for stationarity in heterogeneous panels. *Econometrics Journal* 9: 123–158.
Smith, L.V., S. Leybourne, T.H. Kim, and P. Newbold. 2004. More powerful panel data unit root tests with an application to mean reversion in real exchange rates. *Journal of Applied Econometrics* 19: 147–170.
Song, S.H., and B.C. Jung. 1998. A simple linear trend model with error components, Solution 97.2.1. *Econometric Theory* 14: 286–289.
Sun, Y. 2004. Estimation of the long-run average relationship in nonstationary panel time series. *Econometric Theory* 20: 1227–1260.
Wagner, M., and J. Hlouskova. 2010. The performance of panel cointegration methods: Results from a large scale simulation study. *Econometric Reviews* 29: 182–223.
Wang, S., P. Wang, J. Yang, and Z. Li. 2010. A generalized nonlinear IV unit root test for panel data with cross-sectional dependence. *Journal of Econometrics* 157: 101–109.
Westerlund, J. 2005a. New simple tests for panel cointegration. *Econometric Reviews* 24: 297–316.
Westerlund, J. 2005b. A panel CUSUM test of the null of cointegration. *Oxford Bulletin of Economics and Statistics* 67: 231–262.
Westerlund, J. 2005c. Data dependent endogeneity correction in cointegrated panels. *Oxford Bulletin of Economics and Statistics* 67: 691–705.
Westerlund, J. 2006a. Reducing the size distortions of the panel LM test for cointegration. *Economics Letters* 90: 384–389.
Westerlund, J. 2006b. Testing for panel cointegration with a level break. *Economics Letters* 91: 27–33.
Westerlund, J. 2006c. Testing for panel cointegration with multiple structural breaks. *Oxford Bulletin of Economics and Statistics* 68: 101–132.
Westerlund, J. 2007. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69: 709–748.
Westerlund, J. 2008. Panel cointegration tests of the fisher effect. *Journal of Applied Econometrics* 23: 193–233.
Westerlund, J., and J. Breitung. 2013. Lessons from a decade of IPS and LLC. *Econometric Reviews* 33: 547–591.
Westerlund, J., and W. Hess. 2011. A new poolability test for cointegrated panels. *Journal of Applied Econometrics* 26: 56–88.

13 Spatial Panel Data Models

13.1 Introduction

In randomly drawn samples at the individual level, one does not usually worry about cross-section correlation. However, when one starts looking at a cross-section of countries, regions, states, counties, etc., these aggregate units are likely to exhibit cross-sectional correlation that have to be dealt with. There is an extensive literature using spatial statistics that deals with this type of correlation. These spatial dependence models are popular in regional science and urban economics. More specifically, these models deal with spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) primarily in cross-section data; see Anselin (1988) for a nice introduction to this literature. Spatial dependence models may use a metric of economic distance which provides cross-sectional data with a structure similar to that provided by the time index in time series. With the increasing availability of micro- as well as macro-level panel data, spatial panel data models are becoming increasingly attractive in empirical economic research. See Case (1991), Baltagi and Li (2004), Driscoll and Kraay (1998), Baltagi, Egger and Pfaffermayr (2007), for a few applications. Economists are interested in spillover effects and externalities. Spatial models allow simple econometric methods for modeling these spillover effects. For example, you spend more money on police in one neighborhood, you may increase the crime in an adjacent neighborhood. This externality is dependent on contiguity of the neighborhoods, their common borders, or the distance between these neighborhoods. The same idea can be applied for the analysis of welfare or trade. If California is generous in providing welfare to its residents, this may attract welfare recipients from adjacent states. Gravity models of trade use distance, common border, common language, culture and history, common colonizer, common currency, to see if these things enhance trade. These may be interpreted as distances that are economic, historic, or cultural in nature. In sum, these metrics can be used in a spatial economic model to explain crime or trade or dependency on welfare. See Elhorst (2014) and Lee and Yu (2015) for recent surveys of spatial panel models.

B. H. Baltagi, *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics, https://doi.org/10.1007/978-3-030-53953-5_13

13.2 Spatial Error Component Regression Model

One can model the *spatial correlation* as well as the *heterogeneity* across countries using a spatial error component regression model:

$$y_{ti} = X'_{ti}\beta + u_{ti}, \ i = 1, .., N; t = 1, \cdots, T, \tag{13.1}$$

where y_{ti} is the observation on the ith country for the tth time period, X_{ti} denotes the $k \times 1$ vector of observations on the non-stochastic regressors, and u_{ti} is the regression disturbance. In vector form, the disturbance vector of (13.1) is assumed to have random country effects as well as spatially autocorrelated remainder disturbances; see Anselin (1988):

$$u_t = \mu + \epsilon_t \tag{13.2}$$

with

$$\epsilon_t = \lambda W_N \epsilon_t + \nu_t \tag{13.3}$$

where $\mu' = (\mu_1, \cdots, \mu_N)$ denote the vector of random country effects which are assumed to be IIN$(0, \sigma^2_\mu)$. λ is the scalar spatial autoregressive coefficient with $|\lambda| < 1$. W_N is a known $N \times N$ spatial weight matrix whose diagonal elements are zero. W_N also satisfies the condition that $(I_N - \lambda W_N)$ is nonsingular. $\nu'_t = (\nu_{t1}, \cdots, \nu_{tN})$, where ν_{ti} is assumed to be IIN$(0, \sigma^2_\nu)$ and also independent of μ_i. One can rewrite (13.3) as

$$\epsilon_t = (I_N - \lambda W_N)^{-1}\nu_t = B^{-1}\nu_t \tag{13.4}$$

where $B = I_N - \lambda W_N$ and I_N is an identity matrix of dimension N. The model (13.1) can be rewritten in matrix notation as

$$y = X\beta + u \tag{13.5}$$

where y is now of dimension $NT \times 1$, X is $NT \times k$, β is $k \times 1$, and u is $NT \times 1$. X is assumed to be of full column rank and its elements are assumed to be bounded in absolute value. Equation (13.2) can be written in vector form as follows:

$$u = (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})\nu \tag{13.6}$$

where $\nu' = (\nu'_1, \cdots, \nu'_T)$. Under these assumptions, the variance–covariance matrix for u is given by

$$\Omega = \sigma^2_\mu(J_T \otimes I_N) + \sigma^2_\nu(I_T \otimes (B'B)^{-1}) \tag{13.7}$$

This matrix can be rewritten as

$$\Omega = \sigma^2_\nu\Big[\bar{J}_T \otimes (T\phi I_N + (B'B)^{-1}) + E_T \otimes (B'B)^{-1}\Big] = \sigma^2_\nu \Sigma \tag{13.8}$$

where $\phi = \sigma^2_\mu/\sigma^2_\nu$, $\bar{J}_T = J_T/T$ and $E_T = I_T - \bar{J}_T$. Using results in Wansbeek and Kapteyn (1982), Σ^{-1} is given by

$$\Sigma^{-1} = \bar{J}_T \otimes (T\phi I_N + (B'B)^{-1})^{-1} + E_T \otimes B'B. \tag{13.9}$$

Also, $|\Sigma| = |T\phi I_N + (B'B)^{-1}| \cdot |(B'B)^{-1}|^{T-1}$. Under the assumption of normality, the log-likelihood function for this model was derived by Anselin (1988, p. 154) as

$$\begin{aligned} L &= -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \\ &= -\frac{NT}{2}\ln 2\pi\sigma_\nu^2 - \frac{1}{2}\ln[|T\phi I_N + (B'B)^{-1}|] + \frac{(T-1)}{2}\ln|B'B| \\ &\quad -\frac{1}{2\sigma_\nu^2}u'\Sigma^{-1}u \end{aligned} \tag{13.10}$$

with $u = y - X\beta$. For a derivation of the first-order conditions of MLE as well as the LM test for $\lambda = 0$ for this model, see Anselin (1988). As an extension to this work, Baltagi, Song and Koh (2003) derived the joint LM test for spatial error correlation as well as random country effects. Additionally, they derived conditional LM tests, which test for random country effects given the presence of spatial error correlation. Also, spatial error correlation given the presence of random country effects. These conditional LM tests are an alternative to the one-directional LM tests that test for random country effects ignoring the presence of spatial error correlation or the one-directional LM tests for spatial error correlation ignoring the presence of random country effects. Extensive Monte Carlo experiments are conducted to study the performance of these LM tests as well as the corresponding Likelihood Ratio tests. Baltagi et al. (2007) generalize the Baltagi, Song and Koh (2003) paper by allowing for serial correlation over time for each spatial unit and spatial dependence across these units at a particular point in time. In addition, the model allows for heterogeneity across the spatial units through random effects. Testing for any one of these symptoms ignoring the other two is shown to lead to misleading results. For R programs implementing these LM spatial panel tests derived by Baltagi et al. (2007), see Millo and Piras (2012) and their spatial panel linear model (splm) package. These tests are applied to the Munnell (1990) data set used in example 3 of Chap. 2 but now with spatial correlation across states.

Baltagi, Song and Kwon (2009) extend these LM statistics to a panel data regression model with heteroskedastic as well as spatially correlated disturbances. A joint LM test for homoskedasticity and no spatial correlation is derived. In addition, a conditional LM test for no spatial correlation given heteroskedasticity, as well as a conditional LM test for homoskedasticity given spatial correlation, are also derived. These LM tests are compared with marginal LM tests that ignore heteroskedasticity in testing for spatial correlation, or spatial correlation in testing for homoskedasticity. Monte Carlo results show that these LM tests as well as their LR counterparts, perform well even for small N and T. However, misleading inference can occur when using marginal rather than joint or conditional LM tests when spatial correlation or heteroskedasticity is present. Using Monte Carlo experiments, Baltagi and Pirotte (2010) show that test of hypothesis based on the standard panel data estimators that ignore spatial dependence of the SAR(1) or SMA(1) type can lead to misleading inference, especially when the spatial coefficients are large.

As an alternative to the MLE, generalized method of moments have been proposed for spatial cross-section models by Conley (1999) and Kelejian and Prucha (1999). Frees (1995) derives a distribution-free test for spatial correlation in panels. This is based on Spearman-rank correlation across pairs of cross-section disturbances. Driscoll and Kraay (1998) show through Monte Carlo simulations that the presence of even modest spatial dependence can impart large bias to OLS standard errors when N is large. They present conditions under which a simple modification of the standard nonparametric time-series covariance matrix estimator yields estimates of the standard errors that are robust to general forms of spatial and temporal dependence as $T \rightarrow \infty$. However, if T is small, they conclude that the problem of consistent nonparametric covariance matrix estimation is much less tractable. Parametric corrections for spatial correlation are possible only if one places strong restrictions on their form, i.e., knowing W_N. For typical micro-panels with N much larger than T, estimating this correlation is impossible without imposing restrictions, since the number of spatial correlations increases at the rate N^2, while the number of observations grow at rate N. Even for macro-panels where $N = 100$ countries observed over $T = 20$–30 years, N is still larger than T and prior restrictions on the form of spatial correlation are still needed.

ML estimation, even in its simplest form entails substantial computational problems when the number of cross-sectional units N is large. Kelejian and Prucha (1999) suggested a generalized moments (GM) estimation method which is computationally feasible even when N is large. Kapoor, Kelejian and Prucha (2007) generalized this GM procedure from cross-section to panel data and derived its large sample properties when T is fixed and $N \rightarrow \infty$.

The basic regression model is the same as in (13.5), however, the disturbance term u follows the first-order spatial autoregressive process

$$u = \lambda(I_T \otimes W_N)u + \epsilon \tag{13.11}$$

with

$$\epsilon = (\iota_T \otimes I_N)\mu + \nu \tag{13.12}$$

where μ, ν, and W_N were defined earlier. This is different from the Anselin (1988) specification described in (13.2) and (13.3) since it also allows the individual country effects μ to be spatially correlated. While the two data generating processes look similar, they imply different spatial spillover mechanisms. For example, consider the question of cross-country dependence. Some countries were colonized by the British and therefore speak English and their financial institutions, laws, and governance may have been influenced by the British. This is a transmission of spatial effects through the individual country time-invariant effect. Countries are also affected by common global factors like macro-shocks and financial crisis and Tsunamis. Whereas the Anselin model assumes that spillovers are inherently time-varying, the KKP process assumes the spillovers to be time-invariant as well as time-variant. For example, firms located in the neighborhood of highly productive firms may get time-invariant permanent spillovers affecting their productivity in addition to the time-variant spillovers as in the Anselin model.

Defining $\bar{u} = (I_T \otimes W_N)u$, $\bar{\bar{u}} = (I_T \otimes W_N)\bar{u}$ and $\bar{\epsilon} = (I_T \otimes W_N)\epsilon$, Kapoor, Kelejian and Prucha (2007) suggest a GM estimator based on the following six moment conditions

$$\begin{aligned}
&E[\epsilon' Q\epsilon/N(T-1)] = \sigma_\nu^2 \\
&E[\bar{\epsilon}' Q\bar{\epsilon}/N(T-1)] = \sigma_\nu^2 \operatorname{tr}(W_N' W_N)/N \\
&E[\bar{\epsilon}' Q\epsilon/N(T-1)] = 0 \\
&E(\epsilon' P\epsilon/N) = T\sigma_\mu^2 + \sigma_\nu^2 = \sigma_1^2 \\
&E(\bar{\epsilon}' P\bar{\epsilon}/N) = \sigma_1^2 \operatorname{tr}(W_N' W_N)/N \\
&E(\bar{\epsilon}' P\epsilon/N) = 0
\end{aligned} \tag{13.13}$$

From (13.11), $\epsilon = u - \lambda\bar{u}$ and $\bar{\epsilon} = \bar{u} - \lambda\bar{\bar{u}}$, substituting these expressions in (13.13) we obtain a system of six equations involving the second moments of u, $\bar{u}$, and $\bar{\bar{u}}$. Under the random effects specification considered, the OLS estimator of β is consistent. Using $\widehat{\beta}_{OLS}$, one gets a consistent estimator of the disturbances $\widehat{u} = y - X\widehat{\beta}_{OLS}$. The GM estimator of σ_1^2, σ_ν^2, and λ is the solution of the sample counterpart of the six equations in (13.13).

Kapoor, Kelejian and Prucha (2007) suggest three GM estimators. The first involves only the first three moments in (13.13) which do not involve σ_1^2 and yield estimates of λ and σ_ν^2. The fourth moment condition is then used to solve for σ_1^2 given estimates of λ and σ_ν^2. Kapoor, Kelejian and Prucha (2007) give the conditions needed for the consistency of this estimator as $N \to \infty$. The second GM estimator is based upon weighing the moment equations by the inverse of a properly normalized variance–covariance matrix of the sample moments evaluated at the true parameter values. A simple version of this weighting matrix is derived under normality of the disturbances. The third GM estimator is motivated by computational considerations and replaces a component of the weighting matrix for the second GM estimator by an identity matrix. They perform Monte Carlo experiments comparing MLE and these three GM estimation methods. They find that on average, the RMSE of ML and their weighted GM estimators are quite similar. However, the first unweighted GM estimator has a RMSE that is 14–17% larger than that of the weighted GM estimators. For an application of this GM estimator to foreign direct investment (FDI), see Baltagi, Egger and Pfaffermayr (2007). Fingleton (2008) extends the GM estimator of Kapoor, Kelejian and Prucha (2007) to the Spatial Moving Average panel data model. The generalized moments estimator has the advantage that is computationally less demanding than MLE, especially as N gets large.

While the Anselin model seems restrictive in that it does not allow permanent spillovers through the individual firm effects, the KKP approach is restrictive in the sense that it does not allow for a differential intensity of spillovers of the permanent and transitory shocks. Baltagi, Egger and Pfaffermayr (2013) suggested a generalized spatial panel model which encompasses the Anselin (1988) and Kapoor, Kelejian and Prucha (2007) models and allows for spatial correlation in the individual and remainder error components that may have different spatial autoregressive parameters. They derive the maximum likelihood estimator (MLE) for this more general spatial panel model when the individual effects are assumed to be random.

This in turn allows the researcher to test whether this generalized model reduces to (i) the Anselin model, (ii) the Kapoor, Kelejian and Prucha model, or (iii) a simple random effects model that ignores the spatial correlation in the residuals. Baltagi, Egger and Pfaffermayr (2013) derive the corresponding LM and LR tests for these three hypotheses and compare their size and power performance using Monte Carlo experiments.

In fact, Baltagi, Egger and Pfaffermayr (2013) consider the following generalized spatial error components model:

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\
\mathbf{u} &= \mathbf{Z}_{\mu}\mathbf{u}_1 + \mathbf{u}_2 \\
\mathbf{u}_1 &= \rho_1 \mathbf{W}_N \mathbf{u}_1 + \boldsymbol{\mu} \\
\mathbf{u}_2 &= \rho_2 \mathbf{W}\mathbf{u}_2 + \boldsymbol{\nu}.
\end{aligned}$$

This is a balanced panel linear regression model given in (13.5). $\mathbf{Z}_{\mu} = \boldsymbol{\iota}_T \otimes \mathbf{I}_N$ denotes the design matrix for the $(N \times 1)$ vector of random individual effects $\mathbf{u}_1$. $\boldsymbol{\iota}_T$ is a $(T \times 1)$ vector of ones and $\mathbf{I}_N$ is an identity matrix of dimension N. The vector of individual effects $\boldsymbol{\mu}$ is assumed to be $i.i.d.N(0, \sigma_{\mu}^2 \mathbf{I}_N)$, while the $(n \times 1)$ vector of remainder disturbances $\boldsymbol{\nu}$ is assumed to be $i.i.d.N(0, \sigma_{\nu}^2 \mathbf{I}_n)$. If $\rho_1 = 0$ and $\rho_2 = \lambda$, this reverts to the Anselin (1988) random effects spatial panel model given in (13.2) and (13.3). If $\rho_1 = \rho_2 = 0$, this reverts to the random effects panel model with no spatial correlation. If $\rho_1 = \rho_2$, this reverts to the Kapoor, Kelejian and Prucha (2007) model given in (13.11) and (13.12). Of course, the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are assumed to be independent of each other. In this generalized spatial error components model, both $\mathbf{u}_1$ and $\mathbf{u}_2$ are spatially correlated involving the same spatial weight matrix $\mathbf{W}_N$ for each time period, but with different spatial autocorrelation parameters ρ_1 and ρ_2, respectively. $\mathbf{W}_N$ exhibits zero diagonal elements, the remaining entries are usually assumed to decline with distance. The eigenvalues of $\mathbf{W}_N$ are bounded and smaller than 1 in absolute value. The latter assumption holds for the row normalized $\mathbf{W}_N$. It also holds for the maximum-row normalized spatial weights matrices. This assumption also implies that all row and column sums of $\mathbf{W}_N$ are uniformly bounded in absolute value. In addition, we assume that $|\rho_r| < 1$ for $r = 1, 2$. The data are ordered such that $i = 1, ..., N$ is the fast index and $t = 1, ..., T$ is the slow one. The spatial weights matrix for the panel is then given by $\mathbf{W} = \mathbf{I}_T \otimes \mathbf{W}_N$, which is block diagonal and of dimension $(n \times n)$.

Baltagi, Egger and Pfaffermayr (2013) consider the following test of hypotheses:

$$\begin{aligned}
&H_0^A : \rho_1 = \rho_2 = 0 \text{ vs. } H_1^A : \text{at least one of the } \rho_1 \text{ or } \rho_2 \neq 0 \\
&H_0^B : \rho_1 = \rho_2 \text{ vs. } H_1^B : \rho_1 \neq \rho_2 \\
&H_0^C : \rho_1 = 0 \text{ vs. } H_1^C : \rho_1 \neq 0
\end{aligned} \tag{13.14}$$

First, they test H_0^A; $\rho_1 = \rho_2 = 0$, to see whether there is no spatial correlation in the error term. If H_0^A is not rejected, the pretest estimator reverts to the random effects MLE. In case H_0^A is rejected, they test H_0^B; $\rho_1 = \rho_2$. If H_0^B is not rejected, the pretest estimator reverts to the KKP MLE. Otherwise, $\rho_1 \neq \rho_2$. Next, they test H_0^C; $\rho_1 = 0$. In case H_0^C is not rejected, the pretest estimator reverts to the Anselin MLE. If H_0^C

is rejected, the pretest estimator reverts to the MLE of the general model considered by Baltagi, Egger and Pfaffermayr (2013). In other words,

$$\begin{aligned}
\widehat{\beta}_{pretest} &= \widehat{\beta}_{RE,MLE} \text{ if } H_0^A \text{ is not rejected} \\
&= \widehat{\beta}_{KKP,MLE} \text{ if } H_0^A \text{ is rejected, and } H_0^B \text{ is not rejected} \\
&= \widehat{\beta}_{Anselin,MLE} \text{ if } H_0^A \text{ and } H_0^B \text{ are rejected, and } H_0^C \text{ is not rejected} \\
&= \widehat{\beta}_{General,MLE} \text{ if } H_0^A \text{ and } H_0^B \text{ and } H_0^C \text{ are rejected.}
\end{aligned} \tag{13.15}$$

It has to be emphasized that the pretest estimator becomes the MLE of the general model when all three hypotheses are rejected. Also, it is the MLE of the RE model when H_0^A is not rejected. Hence changing the sequence of tests for H_0^B and H_0^C will not affect the number of times the pretest estimator reverts to the MLE of the RE or General model. This affects only the number of times the pretest estimator reverts to the Anselin or KKP ML estimators. In using the same data set to select the estimator to use based on a series of tests makes the statistical properties of the resulting pretest estimator difficult to derive. Given that the researcher does not know the true model, Baltagi, Egger and Pfaffermayr (2008) recommend the pretest estimator which performed well in Monte Carlo experiments no matter what the true underlying model. In fact this pretest estimator was a close second in MSE performance to the true MLE. Additionally, the Monte Carlo experiments shed some light on the performance of the Anselin MLE when the true model is KKP, and vice versa. Ignoring spatial correlation in panel data and performing RE MLE leads to considerable loss in MSE efficiency. When the true model is a general spatial panel model with $\rho_1 \neq \rho_2 \neq 0$, both KKP and Anselin MLE impose wrong restrictions on the ρ parameters, which in turn, introduce bias and lead to bad MSE performance of the resulting MLEs. Fortunately, this does not translate fully into bad MSE performance for the regression coefficients. The pretest estimator of the regression coefficients always performs better than the misspecified MLE and is recommended in practice.

For Stata programs implementing the Generalized Spatial Panel Random Effects model of Baltagi, Egger and Pfaffermayr (2013), see *xsmle* by Belotti, Hughes and Piano Mortari (2017) with an application to residential demand for electricity covering the 48 states in the continental United States plus the district of Columbia for the period 1990–2010. Note that Belotti, Hughes and Piano Mortari (2017) use this empirical example to illustrate the fixed effects spatial panel model specifications. We will use their data set to illustrate the Generalized Spatial Panel Random Effects model.

Empirical Example: *Residential Demand for Electricity*. The dependent variable is the log of residential electricity sales, and it is modeled as a function of log real per-capita income, log of real average residential price of electricity, log of housing units per capita, log of cooling degree and heating degree days. The Stata data set is available as *state_spatial_dbf.dta*. Using a rook W matrix for spatial contiguity of the 48 states in the continental United States plus the district of Columbia, *xsmle* allows us to estimate the Generalized Spatial Panel Random Effects model with the option *(model (gspre))* and *error(1)* giving $\rho_1 \neq \rho_2 \neq 0$. These are reported as

Table 13.1 Generalized spatial panel random effects model

```
       Residential Demand for Electricity
.  xsmle ln_sales_rpop  ln_rinc_cap ln_gprice_res ln_hunit_pop ln_degday_cool
 ln_degday_heat, wmatrix(WN_rook) ematrix(WN_rook) error(1) model(gspre)
SEM with spatial autoregressive random-effects       Number of obs =    1078
Group variable: state_id                          Number of groups = 49
Time variable: year                                   Panel length = 22

R-sq:    within = 0.8148
         between = 0.2553
         overall = 0.3000
Log-likelihood = 1879.1900
----------------------------------------------------------------------------
ln_sales_r~p |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
Main         |
 ln_rinc_cap |   .3720624   .0199573    18.64   0.000     .3329468     .411178
ln_gprice_~s |  -.2769048   .0144779   -19.13   0.000    -.3052811   -.2485286
ln_hunit_pop |   .8132593   .0675438    12.04   0.000     .6808758    .9456427
ln_degday_~l |   .0720415   .0072095     9.99   0.000     .0579112    .0861718
ln_degday_~t |   .1407485   .0189852     7.41   0.000     .1035382    .1779588
       _cons |  -5.318856   .4219833   -12.60   0.000    -6.145928   -4.491784
-------------+--------------------------------------------------------------
Spatial      |
         phi |   .4528639   .1623722     2.79   0.005     .1346202    .7711075
      lambda |   .4000918   .0425615     9.40   0.000     .3166728    .4835108
-------------+--------------------------------------------------------------
Variance     |
    sigma_mu |   .2241658    .023527     9.53   0.000     .1780538    .2702779
     sigma_e |   .0355355   .0008006    44.38   0.000     .0339663    .0371047
----------------------------------------------------------------------------
```

$\phi = \rho_2$ and $\lambda = \rho_1$. Table 13.1 gives the Generalized Spatial Panel Random Effects results where both ρ_1 and ρ_2 are significant with estimated values of 0.40 and 0.45. All regression coefficients are significant with the right sign. Table 13.2 gives the Anselin (1988) Spatial Autoregressive Random Effects results where ρ_2 is set to zero, so that there is no spatial correlation in the individual effects, and ρ_1 is significant with an estimated value of 0.40. Again, all regression coefficients are significant with the right sign. This is done with the option *(model (gspre))* and *error(3)*. Table 13.3 gives the KKP Spatial Autoregressive Random Effects results where $\rho_1 = \rho_2$ with an estimated value of 0.40. Again, all regression coefficients are significant with the right sign. This is done with the option *(model (gspre))* and *error(4)*. Note that this is maximum likelihood estimation and not GMM as KKP intended. Based on the log-likelihood values reported, one can compute the Likelihood Ratio tests. The results do not reject the KKP model vs. the Generalized model. However, it does reject the Anselin model vs. the Generalized model. So one can conclude that there is spatial correlation in the error across the states of magnitude 0.40, and one cannot reject that this spatial effect is of the same magnitude in the individual effects as well as in the remainder disturbances. The income elasticity of residential demand for electricity is 0.37 and the price elasticity is -0.28, both are significant.

Lee and Yu (2012) extend the Generalized Spatial Panel model by Baltagi, Egger and Pfaffermayr (2008, 2013) to include a spatial lag on the dependent variable as well as serial correlation in the remainder term disturbances. This encompasses a lot of panel and spatial models considered in the literature. We deal with the spatial lag on the dependent variable in the next Sect. 13.3. Baltagi and Liu (2016) consider the random effects case of this generalized model and derive the

Table 13.2 Anselin (1988) spatial autoregressive model with random effects

```
      Residential Demand for Electricity
. xsmle ln_sales_rpop  ln_rinc_cap ln_gprice_res ln_hunit_pop ln_degday_cool
ln_degday_heat, ematrix(WN_rook) error(3) model(gspre)
SEM with spatial autoregressive random-effects       Number of obs =      1078
Group variable: state_id                          Number of groups = 49
Time variable: year                                   Panel length = 22
R-sq:    within = 0.8143
         between = 0.2707
         overall = 0.3127
Log-likelihood = 1876.0367
----------------------------------------------------------------------------
ln_sales_r~p |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
Main         |
 ln_rinc_cap |   .3718932   .0200667    18.53   0.000     .3325631    .4112233
ln_gprice_~s |  -.2770843   .0144894   -19.12   0.000    -.3054829   -.2486857
ln_hunit_pop |   .8075186   .0678482    11.90   0.000     .6745385    .9404987
ln_degday_~l |   .0724319   .0072202    10.03   0.000     .0582806    .0865832
ln_degday_~t |   .1361191   .0190763     7.14   0.000     .0987302     .173508
       _cons |  -5.282289   .4219371   -12.52   0.000    -6.109271   -4.455308
-------------+--------------------------------------------------------------
Spatial      |
      lambda |   .4036953   .0426081     9.47   0.000     .3201849    .4872057
-------------+--------------------------------------------------------------
Variance     |
    sigma_mu |    .244208   .0254556     9.59   0.000     .1943159    .2941001
     sigma_e |   .0355311   .0008007    44.37   0.000     .0339618    .0371005
----------------------------------------------------------------------------
```

Table 13.3 KKP (2007) spatial autoregressive model with random effects

```
      Residential Demand for Electricity
. xsmle ln_sales_rpop  ln_rinc_cap ln_gprice_res ln_hunit_pop ln_degday_cool
 ln_degday_heat, wmatrix(WN_rook) ematrix(WN_rook) error(4) model(gspre)

SEM with spatial autoregressive random-effects       Number of obs =      1078
Group variable: state_id                          Number of groups = 49
Time variable: year                                   Panel length = 22
R-sq:    within  = 0.8146
         between = 0.2580
         overall = 0.3023
Log-likelihood =  1879.1425
----------------------------------------------------------------------------
ln_sales_r~p |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
Main         |
 ln_rinc_cap |   .3715805   .0199787    18.60   0.000     .3324229     .410738
ln_gprice_~s |  -.2772241   .0144426   -19.19   0.000     -.305531   -.2489171
ln_hunit_pop |   .8106253   .0670875    12.08   0.000     .6791361    .9421144
ln_degday_~l |   .0720378   .0072166     9.98   0.000     .0578935    .0861821
ln_degday_~t |   .1399189    .018829     7.43   0.000     .1030147     .176823
       _cons |  -5.304763    .420321   -12.62   0.000    -6.128577   -4.480949
-------------+--------------------------------------------------------------
Spatial      |
         phi |   .4036936   .0409352     9.86   0.000     .3234621    .4839252
-------------+--------------------------------------------------------------
Variance     |
    sigma_mu |    .225535   .0233563     9.66   0.000     .1797576    .2713124
     sigma_e |   .0355226   .0007991    44.45   0.000     .0339564    .0370889
----------------------------------------------------------------------------
```

Goldberger (1962) BLUP term for forecasting s periods ahead for the i-th cross-sectional unit. Forecasting with spatial panels will be discussed in Sect. 13.4.

Baltagi and Pirotte (2010) extend the Kapoor, Kelejian and Prucha (2007) methodology to the SUR error component panel data model with SAR disturbances. They compare the performance of the proposed fixed and random spatial SUR estimators to the maximum likelihood (ML) estimator for the SUR spatial panel model using Monte Carlo experiments. Baltagi and Bresson (2011) propose maximum likelihood estimators for the panel seemingly unrelated regressions with both spatial lag and spatial error components. They also derive joint and conditional Lagrange multiplier tests for spatial autocorrelation and random effects for this spatial SUR panel model. The small sample performance of the proposed estimators and tests are examined using Monte Carlo experiments. An empirical application to hedonic housing prices in Paris illustrates these methods. The proposed specification uses a system of three SUR equations corresponding to three types of flats within 80 districts of Paris over the period 1990–2003. They test for spatial effects and heterogeneity and find reasonable estimates of the shadow prices for housing characteristics.

Baltagi, Egger and Kesina (2016) extend the Hausman and Taylor (1981) approach studied in Chap. 7 to allow for a saptial SAR(1) error term. Unlike the spatial fixed effects (SFE) estimator, the spatial Hausman and Taylor (SHT) estimator does not wipe out the time-invariant variables. Instead, it uses the Between variation of the time-varying exogenous variables to instrument for endogenous time-invariant regressors. It also uses the GM estimation methodology of Kapoor, Kelejian and Prucha (2007). This estimation method is used to assess the role of intra-sectoral spillovers in total factor productivity across Chinese producers in the chemical industry. It uses a rich panel data set of 12,552 firms observed over the period 2004–2006. The paper finds evidence of positive spillovers across chemical manufacturers and a large and significant detrimental effect of public ownership on total factor productivity. The paper also presents variants of the SHT estimator in the spirit of Amemiya and MaCurdy (1986) as well as Breusch, Mizon and Schmidt (1989) studied in Chap. 7.

In a companion paper, Baltagi, Egger and Kesina (2012) study the small sample performance of various estimators applied to this *spatial* Hausman–Taylor model using Monte Carlo experiments. This paper generalizes the pretest estimator suggested by Baltagi, Bresson and Pirotte (2003), and studied in Chap. 4, to account for spatial correlation. This pretest estimator reverts to the spatial RE estimator if the standard Hausman test based on the SFE versus the spatial RE estimators is not rejected. It reverts to the SHT estimator if the choice of strictly exogenous regressors is not rejected by a second Hausman over-identification test based on the difference between the SFE and SHT estimators. If both tests are rejected, then the pretest estimator reverts to the SFE estimator. Monte Carlo experiments show that the spatial pretest estimator is a viable estimator and performs reasonably well in RMSE but should not be used for simple test of hypothesis. The SFE estimator is a consistent estimator, but its disadvantage is that it does not allow the estimation of the coefficients of the time-invariant regressors. When there is endogeneity among the regressors, there is substantial bias in spatial OLS and SRE estimators and both yield misleading inference.

13.3 Spatial Lag Panel Data Regression Model

Baltagi and Liu (2011) considered the following spatial lag panel data model:

$$\begin{aligned} y &= \lambda W y + X\beta + u, \\ u &= Z_\mu \mu + \nu, \end{aligned} \tag{13.16}$$

where y is of dimension $NT \times 1$, X is $NT \times k$, β is $k \times 1$, and u is $NT \times 1$. X is assumed to be exogenous of full column rank and its elements are assumed to be asymptotically bounded in absolute value. Ordering the data first by time (with index $t = 1, ..., T$) and then by individual units (with index $i = 1, ..., N$), we get $W = I_T \otimes W_N$, where the $(N \times N)$ spatial weight matrix W_N has zero diagonal elements and is row normalized with its entries usually declining with distance. This in turn results in row and column sums of W_N that are uniformly bounded in absolute value. In addition the column sums of W_N, as well as the row and column sums of $(I_N - \lambda W_N)^{-1}$ are bounded uniformly in absolute value by some finite constant. We also assume that λ is bounded in absolute value, i.e., $|\lambda| < 1$. The random error component structure is as described in Chap. 2, but now $Z_\mu = \iota_T \otimes I_N$ denoting the selector matrix for the $(N \times 1)$ random vector of individual effects μ which is assumed to be i.i.d. $(0, \sigma_\mu^2 I_N)$. Recall that ι_T is a vector of ones of dimension T and I_N is an identity matrix of dimension N. ν is a vector of $NT \times 1$ remainder disturbances which is assumed to be i.i.d. $(0, \sigma_\nu^2 I_{NT})$. Also, μ and ν are independent of each other and the regressor matrix X.

This is a panel data version of the cross-section spatial lag model considered by Kelejian and Prucha (1998) and Lee (2003). Let $A = I_T \otimes A_N$ where $A_N = I_N - \lambda W_N$, then one can write

$$y = A^{-1}(X\beta + u).$$

Note that

$$E\left[Wyu'\right] = E\left[WA^{-1}(X\beta + u)u'\right] = WA^{-1}\Omega \neq 0,$$

where $\Omega = E\left(uu'\right)$. The spatially lagged dependent variable Wy is correlated with the disturbance u. Therefore, the OLS estimator will be inconsistent. Let $Z = (X, Wy)$ and $\delta = (\beta, \lambda)'$, the model in Eq. (13.16) can be written as

$$y = Z\delta + u. \tag{13.17}$$

In the cross-section spatial autoregressive model, Kelejian and Prucha (1998) suggested a two-stage least square estimator (2SLS) based on feasible instruments like $H = \left(X, WX, W^2X\right)$ which yield

$$\hat{\delta}_{2SLS} = \left[Z'P_H Z\right]^{-1} Z'P_H y, \tag{13.18}$$

where $P_H = H\left(H'H\right)^{-1} H'$ denotes the projection matrix using H. They show that this 2SLS estimator is consistent under some general regularity conditions for this model. Let $\bar{J}_T = J_T/T$, where J_T is a matrix of ones of dimension T. Also, let $E_T = I_T - \bar{J}_T$, and define P to be the projection matrix on Z_μ, i.e., $P = \bar{J}_T \otimes I_N$, and $Q = I_{NT} - P = E_T \otimes I_N$. Premultiply Eq. (13.17) by Q to obtain

$$\widetilde{y} = \lambda W\widetilde{y} + \widetilde{X}\beta + \widetilde{u}. \tag{13.19}$$

This follows from the fact that $QW = (E_T \otimes I_N)(I_T \otimes W_N) = (E_T \otimes W_N) = (I_T \otimes W_N)(E_T \otimes I_N) = WQ$. Applying 2sls to this Q transformed panel spatial lag model, we get the fixed effects spatial 2SLS (FE-S2SLS) estimator of δ based upon $\widetilde{H} = \left(\widetilde{X}, W\widetilde{X}, W^2\widetilde{X}\right) = \left(QX, QWX, QW^2X\right) = QH$. We denote this by $\hat{\delta}_{FE-S2SLS}$. Note that the s^2 of this FE-S2SLS provides a consistent estimate $\hat{\sigma}_\nu^2$ of σ_ν^2. Similarly, one could premultiply Eq. (13.17) by P to obtain

$$\overline{y} = \lambda W\overline{y} + \overline{X}\beta + \overline{u}. \tag{13.20}$$

This follows from the fact that $PW = (\bar{J}_T \otimes I_N)(I_T \otimes W_N) = (\bar{J}_T \otimes W_N) = (I_T \otimes W_N)(\bar{J}_T \otimes I_N) = WP$. Applying 2sls to this P transformed panel spatial lag model, we get the Between effects spatial 2SLS (BE-S2SLS) estimator of δ based upon $\overline{H} = \left(\overline{X}, W\overline{X}, W^2\overline{X}\right) = \left(PX, PWX, PW^2X\right) = PH$. We denote this by $\hat{\delta}_{BE-S2SLS}$. Note that the s^2 of this BE-S2SLS provides a consistent estimate $\hat{\sigma}_1^2$ of $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$. As shown in Chap. 2, the variance–covariance matrix of u is $\Omega = E\left(uu'\right) = \sigma_1^2 P + \sigma_\nu^2 Q$, and $\Omega^{-1/2} = \left(\sigma_1^{-1}P + \sigma_\nu^{-1}Q\right)$. Left multiply Eq. (13.17) by $\Omega^{-1/2}$, we get

$$y^* = Z^*\delta + u^*, \tag{13.21}$$

where $y^* = \Omega^{-1/2}y$, $u^* = \Omega^{-1/2}u$, and $Z^* = \Omega^{-1/2}Z = \Omega^{-1/2}(X, Wy) = (X^*, \Omega^{-1/2}Wy) = \left(X^*, W\Omega^{-1/2}y\right) = (X^*, Wy^*)$. The last expression follows from the fact that $\Omega^{-1/2}W = \left(\sigma_1^{-1}PW + \sigma_\nu^{-1}QW\right) = \left(\sigma_1^{-1}WP + \sigma_\nu^{-1}WQ\right) = W\Omega^{-1/2}$. This means that Eq. (13.21) can also be written as

$$y^* = \lambda Wy^* + X^*\beta + u^*. \tag{13.22}$$

Applying 2sls to this $\Omega^{-1/2}$ transformed panel spatial lag model, we get the random effects spatial 2SLS estimator (RE-S2SLS) of δ given by

$$\hat{\delta}_{RE-S2SLS} = \left[Z^{*\prime}P_{H^*}Z^*\right]^{-1} Z^{*\prime}P_{H^*}y^* \tag{13.23}$$

where

$$\begin{aligned} H^* &= \left(X^*, WX^*, W^2X^*\right) = \left(\Omega^{-1/2}X, W\Omega^{-1/2}X, W^2\Omega^{-1/2}X\right) \\ &= \left(\Omega^{-1/2}X, \Omega^{-1/2}WX, \Omega^{-1/2}W^2X\right) = \Omega^{-1/2}H. \end{aligned}$$

Using the results in Chap. 7, one can similarly derive a spatial error component 2SLS estimator (SEC-2SLS) as follows: Left multiply Eq. (13.19) by $\tilde{H}'$, and (13.20) by $\bar{H}'$, and stack the system of two equations recognizing that they estimate the same δ, we get

$$\begin{pmatrix} \tilde{H}'\widetilde{y} \\ \bar{H}'\overline{y} \end{pmatrix} = \begin{pmatrix} \tilde{H}'\widetilde{Z} \\ \bar{H}'\overline{Z} \end{pmatrix}\delta + \begin{pmatrix} \tilde{H}'\widetilde{u} \\ \bar{H}'\overline{u} \end{pmatrix},$$

where $E\begin{pmatrix} \tilde{H}'\widetilde{u} \\ \bar{H}'\overline{u} \end{pmatrix} = 0$ and $Var\begin{pmatrix} \tilde{H}'\widetilde{u} \\ \bar{H}'\overline{u} \end{pmatrix} = \begin{bmatrix} \sigma_\nu^2\tilde{H}'\tilde{H} & 0 \\ 0 & \sigma_1^2\bar{H}'\bar{H} \end{bmatrix}$. This follows from the fact that $Q\Omega = \sigma_\nu^2 Q$, and $P\Omega = \sigma_1^2 P$, with $QP = 0$. Performing GLS on this two equation system, we get the spatial error component two-stage least squares (SEC-2SLS) estimator:

$$\hat{\delta}_{SEC-2SLS} = \left(\frac{\widetilde{Z}' P_{\tilde{H}} \widetilde{Z}}{\sigma_\nu^2} + \frac{\overline{Z}' P_{\bar{H}} \overline{Z}}{\sigma_1^2} \right)^{-1} \left(\frac{\widetilde{Z}' P_{\tilde{H}} \widetilde{y}}{\sigma_\nu^2} + \frac{\overline{Z}' P_{\bar{H}} \overline{y}}{\sigma_1^2} \right)$$
$$= \left(\sigma_\nu^{-2} Z' P_{\tilde{H}} Z + \sigma_1^{-2} Z' P_{\bar{H}} Z \right)^{-1} \left(\sigma_\nu^{-2} Z' P_{\tilde{H}} y + \sigma_1^{-2} Z' P_{\bar{H}} y \right), \tag{13.24}$$

using the fact that Q and P are idempotent. Using a similar argument as in Chap. 7, one can show that $\hat{\delta}_{SEC-2SLS}$ is a matrix weighted combination of $\hat{\delta}_{FE-S2SLS}$ and $\hat{\delta}_{BE-S2SLS}$ weighting each by the inverse of their variance–covariance matrix and such that the weights add up to the identity matrix. A feasible SEC-2SLS estimator can be obtained by substituting $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_1^2$ into Eq. (13.24). Following a similar argument as in Cornwell, Schmidt and Wyhowski (1992), one can show that $\hat{\delta}_{SEC-2SLS}$ can also be obtained as 2SLS from the $\Omega^{-1/2}$ transformed equation in (13.21) using $B = \left(\tilde{H}, \bar{H}\right)$ as instruments. To show this, note that $P_B = P_{\tilde{H}} + P_{\bar{H}}$ using the fact that $\tilde{H}$ and $\bar{H}$ are orthogonal to each other. This means that $Z^{*\prime} P_B Z^* = Z^{*\prime} P_{\tilde{H}} Z^* + Z^{*\prime} P_{\bar{H}} Z^* = \sigma_\nu^{-2} Z' P_{\tilde{H}} Z + \sigma_1^{-2} Z' P_{\bar{H}} Z$, since $Q\Omega^{-1/2} = \sigma_\nu^{-1} Q$, and $P\Omega^{-1/2} = \sigma_1^{-1} P$. Similarly, $Z^{*\prime} P_B y^* = Z^{*\prime} P_{\tilde{H}} y^* + Z^{*\prime} P_{\bar{H}} y^* = \sigma_\nu^{-2} Z' P_{\tilde{H}} y + \sigma_1^{-2} Z' P_{\bar{H}} y$. Therefore, $\hat{\delta}_{SEC-2SLS}$ given by Eq. (13.24) can be obtained as 2SLS on (13.21) using $B = \left(\tilde{H}, \bar{H}\right)$ as instruments.

Lee (2003) argued that in the cross-section spatial model, the optimal instruments for estimating δ in Eq. (13.17) is

$$E(Z) = E[X, Wy] = \left(X, WA^{-1}X\beta\right).$$

Therefore, a Lee (2003) type optimal instruments for estimating δ in the $\Omega^{-1/2}$ transformed panel autoregressive spatial model in Eq. (13.21) is

$$E\left(Z^*\right) = E\left(\Omega^{-1/2} Z\right) = E\left[\Omega^{-1/2}(X, Wy)\right] = \left(\Omega^{-1/2} X, \Omega^{-1/2} W A^{-1} X\beta\right),$$

and the resulting best spatial 2SLS estimator is given by Baltagi and Liu (2011)

$$\hat{\delta}_{BS-2SLS} = \left(H_b^{*\prime} Z^*\right)^{-1} H_b^{*\prime} y^*, \tag{13.25}$$

where $H_b^* = \left(\Omega^{-1/2} X, \Omega^{-1/2} W A^{-1} X\beta\right)$. A feasible version of this estimator is based on consistent estimators of σ_1^2, σ_ν^2, λ, and β, respectively.

Similarly, one can extend Baltagi (1981) error component 2SLS estimator to this panel spatial lag model using H_b^* as instruments. Left multiply Eq. (13.21) by $\left(\tilde{H}_b^*, \bar{H}_b^*\right)'$, where $\tilde{H}_b^* = QH_b^* = \sigma_\nu^{-1}\left(QX, QWA^{-1}X\beta\right)$, and $\bar{H}^* = PH_b^* = \sigma_1^{-1}\left(PX, PWA^{-1}X\beta\right)$, we get the two equation system recognizing that they estimate the same δ:

$$\begin{pmatrix} \tilde{H}_b^{*\prime} y^* \\ \bar{H}_b^{*\prime} y^* \end{pmatrix} = \begin{pmatrix} \tilde{H}_b^{*\prime} Z^* \\ \bar{H}_b^{*\prime} Z^* \end{pmatrix} \delta + \begin{pmatrix} \tilde{H}_b^{*\prime} u^* \\ \bar{H}_b^{*\prime} u^* \end{pmatrix},$$

where $E\begin{pmatrix}\tilde{H}_b^{*\prime}u^* \\ \bar{H}_b^{*\prime}u^*\end{pmatrix} = 0$ and $Var\begin{pmatrix}\tilde{H}_b^{*\prime}u^* \\ \bar{H}_b^{*\prime}u^*\end{pmatrix} = \begin{bmatrix}\tilde{H}_b^{*\prime}\tilde{H}_b^* & 0 \\ 0 & \bar{H}_b^{*\prime}\bar{H}_b^*\end{bmatrix}$. Performing GLS on this system, we obtain the spatial error component best 2SLS estimator (SEC-B2SLS):

$$\begin{aligned}\hat{\delta}_{SEC-B2SLS} &= \left(Z^{*\prime}P_{\tilde{H}_b^*}Z^* + Z^{*\prime}P_{\bar{H}_b^*}Z^*\right)^{-1}\left(Z^{*\prime}P_{\tilde{H}_b^*}y^* + Z^{*\prime}P_{\bar{H}_b^*}y^*\right) \\ &= \left(\sigma_\nu^{-2}Z'P_{\tilde{H}_b^*}Z + \sigma_1^{-2}Z'P_{\bar{H}_b^*}Z\right)^{-1}\left(\sigma_\nu^{-2}Z'P_{\tilde{H}_b^*}y + \sigma_1^{-2}Z'P_{\bar{H}_b^*}y\right).\end{aligned} \tag{13.26}$$

The second equality uses the fact that $Q\Omega^{-1/2} = \sigma_\nu^{-1}Q$ and $P\Omega^{-1/2} = \sigma_1^{-1}P$. This SEC-B2SLS estimator can also be obtained from the $\Omega^{-1/2}$ transformed equation in (13.21) using $B = \left(\tilde{H}_b^*, \bar{H}_b^*\right)$ as instruments. In fact, $\tilde{H}_b^*$ and $\bar{H}_b^*$ are orthogonal to each other, since $QP = 0$. Hence, $P_B = P_{\tilde{H}_b^*} + P_{\bar{H}_b^*}$. This also implies that $Z^{*\prime}P_BZ^* = Z^{*\prime}P_{\tilde{H}_b^*}Z^* + Z^{*\prime}P_{\bar{H}_b^*}Z^*$ and $Z^{*\prime}P_By^* = Z^{*\prime}P_{\tilde{H}_b^*}y^* + Z^{*\prime}P_{\bar{H}_b^*}y^*$. Therefore, $\hat{\delta}_{SEC-B2SLS}$ given by Equation (13.26) is the same as 2SLS on (13.21) using $B = \left(\tilde{H}_b^*, \bar{H}_b^*\right)$ as instruments.

Note that these 2SLS estimators are easy to apply using standard software. In fact, one can easily extend the *ec2sls* procedure in Stata for panels to the spatial panel case. Note also that SEC-2SLS and SEC-B2SLS use twice the instruments used by their counterparts RE-S2SLS and BS-2SLS. Although the set of instruments in the former is completely spanned by those for the latter estimators, these may yield smaller empirical standard errors in small samples.

Also, it is important to note that the above estimation procedures can be easily extended to handle right hand side endogenous regressors in the spatial lag model described in Eq. (13.16); see Baltagi and Deng (2015). In addition, Baltagi and Deng (2015) derive spatial EC3SLS estimators for a two equation simultaneous model with spatial lag and random effects in each equation. Spatial EC3SLS estimators can handle endoegeneity, spatial lag dependence, heterogeneity as well as cross equation correlation. This is done by utilizing the Kelejian and Prucha (1998) and Lee (2003) type instruments from the cross-section spatial autoregressive literature and marrying them to the EC3SLS estimator derived by Baltagi (1981) for a system of simultaneous panel data equations; see Chap. 7. Monte Carlo experiments are conducted to study the small sample properties of spatial EC2SLS as well as spatial EC3SLS estimators. The results indicate that, for the single equation spatial EC2SLS estimators, there is a slight gain in efficiency when Lee (2003) type rather than Kelejian and Prucha (1998) instruments are used. However, there is not much difference in efficiency between these instruments for spatial EC3SLS estimators.

Debarsy and Ertur (2010) derive LM and LR tests designed to discriminate between spatially autocorrelated disturbances versus a spatially lagged dependent variable in the context of fixed effects spatial panel data model. They combine a spatial lag model with spatially autocorrelated disturbances of order one, labeled SARAR, in a fixed effects spatial panel data setting. They derive joint, marginal as

Table 13.4 Fixed effects SARAR for residential electricity demand

```
. xsmle ln_sales_rpop  ln_rinc_cap ln_gprice_res ln_hunit_pop ln_degday_cool
ln_degday_heat,  wmatrix(WN_rook) ematrix(WN_rook) model(sac) fe effects
vceeffects(dm) robust
SAC with spatial fixed-effects                       Number of obs =      1078
Group variable: state_id                          Number of groups =        49
Time variable: year                                   Panel length =        22
R-sq:    within  = 0.8246
         between = 0.1440
         overall = 0.2030
Mean of fixed-effects = -3.0342
Log-pseudolikelihood =  2318.0700
                                (Std. Err. adjusted for 49 clusters in
state_id)
------------------------------------------------------------------------------
               |               Robust
 ln_sales_rpop |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+--------------------------------------------------------------
Main           |
   ln_rinc_cap |   .2068726   .0588692     3.51   0.000      .091491   .3222542
 ln_gprice_res |   -.240526   .0474242    -5.07   0.000    -.3334757  -.1475763
  ln_hunit_pop |   .7481685   .1103195     6.78   0.000     .5319462   .9643908
ln_degday_cool |   .0572585   .0124676     4.59   0.000     .0328225   .0816945
ln_degday_heat |   .1409366   .0277695     5.08   0.000     .0865094   .1953637
---------------+-------------------------------------------------------------
Sp-atial        |
           rho |   .3592947   .1074325     3.34   0.001     .1487308   .5698586
        lambda |   .0180619   .1722153     0.10   0.916    -.3194739   .3555976
---------------+--------------------------------------------------------------
Variance       |
      sigma2_e |   .0011869   .0001472     8.06   0.000     .0008984   .0014754
---------------+--------------------------------------------------------------
```

well as conditional LM and LR tests, under the assumption of normality of the disturbances. They show that these tests perform well using Monte Carlo experiments. For Stata programs that implement maximum likelihood methods for the fixed and random effects spatial panel model, see the *xsmle* command by Belotti, Hughes and Piano Mortari (2017). We used it in the Empirical Example to illustrate the Generalized Spatial Panel Random Effects model of Baltagi, Egger and Pfaffermayr (2013). Here, we continue with this example of residential electricity demand and apply it to a SARAR model with fixed effects and illustrate the Debarsy and Ertur (2010) LR tests for this empirical study.

Empirical Example (continued): *Residential Electricity Demand. xsmle* can estimate a spatially autocorrelated error model (SEM) described as the Anselin (1988) model in (13.1)–(13.3). Also, a spatial autoregressive model (SAR) which we presented in Sect. 13.3 as a spatial lag model described in (13.6). When we have both a spatial lag as well as a spatial error model, Debarsy and Ertur (2010) called it (SARAR). The model is given by (13.16) (with spatial lag coefficient ρ) but with error given by (13.2) and (13.3) (with spatial error coefficient λ). Belotti, Hughes and Piano Mortari (2017) call it the spatial autocorrelated model (SAC). This can be done for both random effects as well as fixed effects. Table 13.4 gives the SARAR or SAC model with fixed effects. The results are obtained by specifying the *(model (sac))* option and using the same rook W matrix for the error as well as the spatial lag. This replicates the SAC fixed effects results in Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). Note that ρ is the spatial lag coefficient and it is estimated at 0.36 and is significant, while λ is the spatial coefficient on the spatial error and it is estimated at 0.02 and is insignificant. One can also get estimates of the direct and indirect effects; see LeSage and Pace (2009). Problem 13.3 asks the reader to replicate Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174) which gives the fixed effects estimates without any spatial correlation, the SEM, SAR, and SAC, all with fixed effects. Given these results, one can perform the LR tests of Debarsy and Ertur (2010) with fixed effects. The log-likelihood values are given in the *xsmle* output. The marginal LR for ($H_0^a : \rho = 0$) is based on the log-likelihood of FE-SAR vs. the log-likelihood of FE: $-2(2029.91 - 2108.54) = 157.3$ which is χ_1^2. The Marginal LR for ($H_0^b : \lambda = 0$) is based on the log-likelihood of FE-SEM vs. the log-likelihood of FE: $-2(2029.91 - 2071.08) = 82.34$ which is χ_1^2. The Joint LR for ($H_0^c : \rho = \lambda = 0$) is based on the log-likelihood of FE-SAC vs. the log-likelihood of FE: $-2(2029.91 - 2108.56) = 157.3$ which is χ_2^2. The Conditional LR for ($H_0^d : \lambda = 0/\rho$) is based on the log-likelihood of FE-SAR vs. the log-likelihood of FE-SAC: $-2(2108.54 - 2108.56) = 0.04$ which is χ_1^2. The Conditional LR for ($H_0^e : \rho = 0/\lambda$) is based on the log-likelihood of FE-SEM vs. the log-likelihood of FE-SAC: $-2(2071.08 - 2108.56) = 74.96$ which is χ_1^2. The results for the fixed effects specification reject that $\rho = 0$, but do not reject that $\lambda = 0$ using the conditional LR test.

Mutl and Pfaffermayr (2010) propose a spatial Hausman test to compare the fixed effects and random effects SARAR panel model. A small Monte Carlo study shows that this test works well even in small panels. Debarsy (2012) extends the Mundlak (1978) approach considered in Chap. 7 to the spatial panel data model. This adds an auxiliary regression for the random individual effects that is a function of the explanatory variables averaged over time plus their spatial weighted average. A likelihood ratio (LR) test that assesses the significance of the correlation between regressors and individual effects is proposed, and its properties are investigated using Monte Carlo simulations.

Note that the SAR spatial models can allow for a spatial weighted term on some or all of the exogenous variables. This is dubbed the spatial Durbin model (SDM). This does not introduce any new complications for estimation and testing, but will affect the direct and indirect effects of a change in one unit of the exogenous variable on the dependent variable; see LeSage and Pace (2009). Problem 13.3 part (e) asks the reader to apply the SDM with fixed effects to the residential electricity demand example.

Spatial panel models can be generalized to include the lagged value of the dependent variable, therefore making it dynamic. In addition one can add a spatial weighted term for the lagged dependent variable. In Stata, *xsmle* allows the estimation of a dynamic specification by implementing the bias-corrected maximum likelihood approach described in Yu, de Jong and Lee (2008). Briefly, Yu, de Jong and Lee (2008) study the asymptotic properties of quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both the number of individuals N and the number of time periods T are large. They cover both the stationary and nonstationary cases. When the roots in the DGP are not all unitary, the estimators' rates of convergence will be the same as the stationary case, and the estimators can be asymptotically normal. In fact, for the distribution of the common parameters, when T is asymptotically large relative to N, the estimators are $\sqrt{NT}$ consistent and asymptotically normal, with the limiting distribution centered around 0. When N is asymptotically proportional to T, the estimators are $\sqrt{NT}$ consistent and asymptotically normal, but the limiting distribution is not centered around 0. When N is large relative to T, the estimators are consistent with rate T, and have a degenerate limiting distribution. Compared to the stationary case, the estimators' rate of convergence will be the same, but the asymptotic variance matrix will be driven by the nonstationary component and it is singular. Consequently, a linear combination of the spatial and dynamic effects can converge at a higher rate. They also propose a bias correction which performs well when T grows faster than $N^{1/3}$. Problem 13.3 part (f) asks the reader to estimate a dynamic SAR and dynamic SDM with fixed effects for the residential electricity demand example. This adds the lagged value of the dependent variable to the model with option *dlag(1)* in *xsmle*. Problem 13.3 part (g) checks the sensitivity of the results in part (d) by adding both the lagged value of the dependent variable and the spatial lag value of the dependent variable with option *dlag(3)* in *xsmle*. Table 13.5 gives the results of a dynamic SAR with only the lagged dependent variable. This replicates the third column labeled dynamic SAR in Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). This shows that lagged residential electricity sales is significant, while log real per-capita income is now insignificant.

Table 13.5 Dynamic SAR with fixed effects for residential electricity demand

```
. xsmle ln_sales_rpop  ln_rinc_cap ln_gprice_res ln_hunit_pop ln_degday_cool
ln_degday_heat, wmatrix(WN_rook) model(sar) dlag(1) fe effects vceeffects(dm)
robust
Dynamic SAR with spatial fixed-effects                     Number of obs =      1029
Group variable: state_id                                Number of groups =        49
Time variable: year                                         Panel length =        21
R-sq:     within  = 0.8927
          between = 0.9062
          overall = 0.8949
Mean of fixed-effects = -1.5651
Log-pseudolikelihood =  2313.0397
                                (Std. Err. adjusted for 49 clusters in
state_id)
-------------------------------------------------------------------------------
               |               Robust
 ln_sales_rpop |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+---------------------------------------------------------------
Main           |
 ln_sales_rpop |
           L1. |   .5537216   .0365993    15.13   0.000     .4819883   .6254548
               |
   ln_rinc_cap |   .0327113   .0279775     1.17   0.242    -.0221236   .0875462
 ln_gprice_res |  -.1436063   .0123956   -11.59   0.000    -.1679013   -1193114
  ln_hunit_pop |   .1496735   .0620334     2.41   0.016     .0280903    2712566
ln_degday_cool |   .0727512   .0122657     5.93   0.000     .0487109    0967916
ln_degday_heat |   .1460299   .0210817     6.93   0.000     .1047106    1873492
---------------+---------------------------------------------------------------
Spatial        |
           rho |   .2613619   .0423223     6.18   0.000     .1784118    .344312
---------------+---------------------------------------------------------------
Variance       |
      sigma2_e |   .0006695   .0000497    13.48   0.000     .0005722   .0007668
---------------+---------------------------------------------------------------
```

13.4 Forecasts Using Panel Data with Spatial Error Correlation

The literature on forecasting is rich with time-series applications, but this is not the case for spatial panel data applications. Exceptions are Baltagi and Li (2004, 2006) with applications to forecasting sales of cigarette and liquor per capita for U.S. states over time. In order to explain how spatial autocorrelation may arise in the demand for cigarettes, we note that cigarette prices vary among states primarily due to variation in state taxes on cigarettes. Border effect purchases not included in the cigarette demand equation can cause spatial autocorrelation among the disturbances. In forecasting sales of cigarettes, the spatial autocorrelation due to neighboring states and the individual heterogeneity across states is taken explicitly into account. Baltagi and Li (2004) derive the best linear unbiased predictor for the random error component model with spatial correlation using a simple demand equation for cigarettes based on a panel of 46 states over the period 1963–92. They compare the performance of several predictors of the states demand for cigarettes for one year and five years ahead. The estimators whose predictions are compared include OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random effects GLS estimator ignoring spatial correlation, and random effects estimator accounting for the spatial correlation. Based on the RMSE criteria, the fixed effects and the random effects spatial estimators gave the best out of sample forecast performance.

Using Goldberger's (1962) results on best linear unbiased predictor (BLUP) studied in Chap. 2, Baltagi and Li (2004, 2006) derived the BLUP correction term when both error components and spatial autocorrelation are present and ϵ_t follows a SAR process; see (13.1)–(13.3). The predictor for the SAR is given by

$$\begin{aligned}\widehat{y}_{i,T+\tau} &= X_{i,T+\tau}\widehat{\beta}_{MLE} + \phi\left(\iota_T' \otimes l_i' C_1^{-1}\right)\widehat{u}_{MLE} \\ &= X_{i,T+\tau}\widehat{\beta}_{MLE} + T\phi\sum_{j=1}^{N} c_{1,j}\overline{u}_{j.,MLE} \qquad (13.27)\end{aligned}$$

where $\phi = \sigma_\mu^2/\sigma_\nu^2$, c_{1j} is the jth element of the ith row of C_1^{-1} with $C_1 = \left[T\phi I_N + \left(B'B\right)^{-1}\right]$ and $\overline{u}_{j.,MLE} = \sum_{t=1}^{T}\widehat{u}_{tj,MLE}/T$. In other words, the BLUP of $y_{i,T+\tau}$ adds to $X_{i,T+\tau}\widehat{\beta}_{MLE}$ a weighted average of the MLE residuals for the N individuals averaged over time. The weights depend upon the spatial matrix W_N and the spatial autoregressive coefficient λ. To make these predictors operational, we replace ϕ and λ by their estimates from the RE-spatial MLE with SAR. When there are no random individual effects, so that $\sigma_\mu^2 = 0$, then $\phi = 0$ and the BLUP prediction terms drop out completely from Eq. (13.27). In these cases, Ω reduces to $\sigma_v^2\left[I_T \otimes \left(B'B\right)^{-1}\right]$ for SAR, and the corresponding MLE for these models yield the pooled spatial MLE with SAR remainder disturbances. This result can be extended to the spatial moving average model (SMA); see Baltagi, Bresson and Pirotte (2012).

For the Kapoor, Kelejian and Prucha (2007) model described in (13.11) and (13.12), the BLUP of $y_{i,T+\tau}$ for the SAR-RE also modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith individual. More specifically, the predictor is given by

$$\begin{aligned}\widehat{y}_{i,T+\tau} &= X_{i,T+\tau}\widehat{\beta}_{FGLS} + \left(\frac{\sigma_\mu^2}{\sigma_1^2}\right) b_i \left(\iota_T' \otimes B_N\right)\widehat{u}_{FGLS} \\ &= X_{i,T+\tau}\widehat{\beta}_{FGLS} + \left(\frac{\sigma_\mu^2}{\sigma_1^2}\right)\left(\iota_T' \otimes l_i'\right)\widehat{u}_{FGLS} \qquad (13.28)\end{aligned}$$

where b_i is the ith row of the matrix B_N^{-1}. This holds because $b_i\left(\iota_T' \otimes B_N\right) = (1 \otimes b_i)\left(\iota_T' \otimes B_N\right) = \left(\iota_T' \otimes l_i'\right)$ where l_i' is the ith row of I_N as defined above. $B_N^{-1}B_N = I_N$ and therefore $b_i B_N = l_i'$. This proof applies to both the Kapoor, Kelejian and Prucha (2007) SAR-RE specification and the Fingleton (2008) SMA-RE specification. Therefore, the BLUP of $y_{i,T+\tau}$ for the SAR-RE and the SMA-RE, like the usual RE model with no spatial effects, modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith individual. While the predictor formula is the same, the MLEs for these specifications yield different estimates which in turn yield different residuals and hence different forecasts.

The results of the Monte Carlo study by Baltagi, Bresson and Pirotte (2012) find that when the true DGP is RE with a SAR or SMA remainder disturbances, estimators that ignore heterogeneity/spatial correlation perform badly in RMSE forecasts. Accounting for heterogeneity improves the forecast performance by a big margin and accounting for spatial correlation improves the forecast but by a smaller margin. Ignoring both, leads to the worst forecasting performance. Heterogeneous estimators based on averaging perform worse than homogeneous estimators in forecasting performance. This performance improves with a larger sample size and seems robust to the type of spatial error structure imposed on the remainder disturbances. These Monte Carlo experiments confirm earlier empirical studies that report similar findings.

Baltagi, Fingleton and Pirotte (2014) focus on the estimation and predictive performance of several estimators for the dynamic and autoregressive spatial lag panel data model with spatially correlated disturbances. A dynamic spatial generalized method of moments estimator is proposed based on Kapoor, Kelejian and Prucha (2007) and Arellano and Bond (1991). The main idea is to mix non-spatial and spatial instruments to obtain consistent estimates of the parameters. Monte Carlo experiments find that when the true model is a dynamic first-order spatial autoregressive specification with SAR-RE disturbances, estimators that ignore the endogeneity of the spatial lag and the endogeneity of the lagged dependent variable perform badly in terms of bias and RMSE. Accounting for heterogeneity and endogeneity improves the forecast performance by a big margin. Accounting for spatial correlation in the disturbances also improves the forecast, but by a smaller margin.

13.5 Panel Unit Root Tests and Spatial Dependence

Baltagi, Bresson and Pirotte (2007) studied the performance of panel unit root tests when spatial effects are present that account for cross-section correlation. Monte Carlo simulations show that there can be considerable size distortions in panel unit root tests when the true specification exhibits spatial error correlation.

Panel data unit root tests have been proposed as alternative more powerful tests than those based on individual time-series unit roots tests; see Chap. 12. One of the advantages of panel unit root tests is that their asymptotic distribution is standard normal. This is in contrast to individual time-series unit roots which have non-standard asymptotic distributions. But these tests are not without their critics. The first generation panel unit root tests assumed cross-section independence. These tests include the one proposed by Levin, Lin and Chu (2002), hereafter denoted by LLC, where the null hypothesis is that each individual time series contains a unit root against the alternative that each time series is stationary. As Maddala pointed out, the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Im, Pesaran and Shin (2003), hereafter denoted by IPS, allow for heterogeneous panels and propose panel unit root tests which are based on the average of the individual ADF unit root tests computed from each time series. The null hypothesis is that each individual time series contains a unit root while the alternative allows for some but not all of the individual series to have unit roots. One major criticism of both the LLC and IPS tests is that they require cross-sectional independence. This is a restrictive assumption given the cross-section correlation and spillovers across countries, states and regions.

Maddala and Wu (1999) and Choi (2001) proposed combining the p-values from the individual unit root ADF tests applied to each time series. Once again, these tests follow a standard normal limiting distribution. They have the advantage that N, the number of cross-sections, can be finite or infinite; the time series can be of different length; and the alternative allows some groups to have unit roots while others may not.

Second generation panel unit root tests that try to account for cross-sectional dependence in panels include the following: Chang (2002) who explored the nonlinear IV methodology to solve the inferential difficulties in the panel unit root testing which arise from the intrinsic heterogeneities and dependencies of panel models. Chang (2002) suggests an average of individual nonlinear IV t-ratio statistics of the autoregressive coefficient obtained from using an integrable transformation of the lagged level as instrument. These methods assume cross-sectional correlation in the innovation terms driving the autoregressive processes. Choi (2002), on the other hand, generalizes the three unit root tests (inverse chi-square, inverse normal, and logit) to the case where the cross-sectional correlation is modeled by error component models. The tests are formulated by combining p-values from the ADF test applied to each individual time series whose stochastic trend components and cross-sectional correlations are eliminated using GLS-demeaning and GLS-detrending. Choi (2002) shows that the combination tests have a standard normal limiting distributions under the sequential asymptotics $T \to \infty$ and $N \to \infty$.

To avoid the restrictive nature of cross-section demeaning procedure, Bai and Ng (2004), and Phillips and Sul (2003), among others, propose dynamic factor models by allowing the common factors to have differential effects on cross-section units. Phillips and Sul's model is a one-factor model where the factor is independently distributed across time. They propose a moment-based method to eliminate the common factor which is different from principal components. More specifically, in the context of a residual one-factor model, Phillips and Sul (2003) provide an orthogonalization procedure which in effect asymptotically eliminates the common factors before preceding to the application of standard unit root tests. Pesaran (2007) suggests a simple way of getting rid of cross-sectional dependence that does not require the estimation of factor loading. His method is based on augmenting the usual ADF regression with the lagged cross-sectional mean and its first difference to capture the cross-sectional dependence that arises through a single-factor model.

Baltagi, Bresson and Pirotte (2007) run Monte Carlo simulations to compare the empirical size of panel unit root tests with and without spatial error dependence. The structure of the dependence is based on some commonly used spatial error processes: the spatial autoregressive (SAR) and the spatial moving average (SMA) error process and the spatial error components model (SEC). For each experiment, they perform nine panel unit root test statistics: the Levin, Lin and Chu (2002) test, the Breitung (2000) test, the Im, Pesaran and Shin (2003) test, the Maddala and Wu (1999) test, the Choi (2001, 2002) test with and without cross-sectional correlation, the Chang (2002) IV test, the Phillips and Sul (2003) test and the Pesaran (2007) test. The experiments include a case of no spatial correlation as well as four types of spatial correlation (SAR, SMA, SEC1, and SEC3), with two values of the parameters indicating weak versus strong spatial dependence. They also consider ten weight matrices, differing in their degree of sparseness, four pairs of (N, T) and two models including individual effects and individual deterministic trends. Even with this modest design, the total number of experiments considered is 1600. They find that ignoring spatial dependence when present can seriously bias the size of panel unit root tests.

13.6 Panel Data Tests for Cross-Sectional Dependence

Consider the heterogeneous panel data model:

$$y_{it} = x_{it}'\beta_i + u_{it}, \text{ for } i = 1, ..., N;\; t = 1, ..., T, \tag{13.29}$$

where i indexes the cross-sectional units and t the time-series observations. y_{it} is the dependent variable and x_{it} denotes the exogenous regressors of dimension $k \times 1$ with slope parameters β_i that are allowed to vary across i. u_{it} is allowed to be cross-sectionally dependent but is uncorrelated with x_{it}. Let $U_t = (u_{1t}, \cdots, u_{Nt})'$. The $N \times 1$ vectors $U_1, U_{2,} \cdots, U_T$ are assumed iid $N(0, \Sigma_u)$ over time. Let σ_{ij} be the (i, j)th element of the $N \times N$ matrix Σ_u. The errors u_{it} $(i = 1, ..., N; t = 1, ..., T)$

are cross-sectionally dependent if Σ_u is non-diagonal, i.e., $\sigma_{ij} \neq 0$ for $i \neq j$. The null hypothesis of cross-sectional independence can be written as

$$H_0 : \sigma_{ij} = 0 \text{ for } i \neq j,$$

or equivalently as

$$H_0 : \rho_{ij} = 0 \text{ for } i \neq j, \tag{13.30}$$

where ρ_{ij} is the correlation coefficient of the errors with $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$. Under the alternative hypothesis, there is at least one nonzero correlation coefficient ρ_{ij}, i.e., $H_a : \rho_{ij} \neq 0$ for some $i \neq j$.

The OLS estimator of y_{it} on x_{it} for each i, denoted by $\hat{\beta}_i$, is consistent. The corresponding OLS residuals $\hat{u}_{it}$ defined by $\hat{u}_{it} = y_{it} - x'_{it}\hat{\beta}_i$ are used to compute the sample correlation $\breve{\rho}_{ij}$ as follows:

$$\breve{\rho}_{ij} = \left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{-1/2} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{-1/2} \sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}. \tag{13.31}$$

In the fixed N case and as $T \to \infty$, the Breusch and Pagan's (1980) *LM* test can be applied to test for the cross-sectional dependence in heterogeneous panels. In this case it is given by

$$LM_{BP} = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \breve{\rho}_{ij}^2.$$

This is asymptotically distributed under the null as a χ^2 with $N(N-1)/2$ degrees of freedom.

Stata has programmed the Breusch and Pagan (1980) test for cross-section correlation after performing *xtreg, fe* or *xtgls*. The command is *xttest2*, and it provides estimates of the pairwise correlation between the residuals of any pair of cross-sections i and $j = 1, .., N$; as well as the LM statistic. It requires N to be smaller than T. Note that it is not based on OLS residuals of each cross-section as LM_{BP} but it is based on the fixed effects residuals of the pooled model. Table 13.6 gives the results for the Grunfled data with $N = 10$ and $T = 20$. The test statistic is 246.329 and is distributed as χ^2 with $N(N-1)/2$ degrees of freedom. In this case, it is χ^2_{45}, and the null is rejected.

However, this Breusch-Pagan LM test statistic is not applicable when $N \to \infty$. In this case, Pesaran (2004) proposes a scaled version of the LM_{BP} test given by

$$CD_{lm} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(T\breve{\rho}_{ij}^2 - 1\right). \tag{13.32}$$

Pesaran (2004) shows that CD_{lm} is asymptotically distributed as $N(0, 1)$, under the null, with $T \to \infty$ first, and then $N \to \infty$. However, as pointed out by Pesaran (2004), for finite T, $E[T\breve{\rho}_{ij}^2 - 1]$ is not correctly centered at zero, and with large N, the incorrect centering of the CD_{lm} statistic is likely to be accentuated. Thus, the standard normal distribution may be a bad approximation of the null distribution of

Table 13.6 The Breusch-Pagan (1980) test for zero correlation for the Grunfeld data

. xttest2

Correlation matrix of residuals:

	__e1	__e2	__e3	__e4	__e5	__e6	__e7	__e8	__e9	__e10
__e1	1.0000									
__e2	-0.0533	1.0000								
__e3	-0.2173	-0.0935	1.0000							
__e4	-0.1912	0.1540	0.2084	1.0000						
__e5	-0.4555	-0.3997	0.6075	0.0490	1.0000					
__e6	-0.2489	-0.1310	0.6277	0.1334	0.8320	1.0000				
__e7	-0.2234	-0.5074	0.5516	0.0291	0.9318	0.7896	1.0000			
__e8	-0.4390	-0.1120	0.8832	0.2432	0.8081	0.7542	0.7451	1.0000		
__e9	-0.3939	-0.3290	0.7952	0.0675	0.8772	0.7229	0.8399	0.9051	1.0000	
__e10	0.4196	0.3937	-0.1534	-0.0030	-0.6206	-0.3208	-0.5452	-0.4038	-0.5352	1.0000

Breusch-Pagan LM test of independence: chi2(45) = 246.329, Pr = 0.0000
Based on 20 complete observations

the CD_{lm} statistic in finite samples, and using the critical values of a standard normal may lead to big size distortion. Using finite sample approximations, Pesaran, Ullah and Yamagata (2008) rescale and recenter the CD_{lm} test. The new LM test, denoted as PUY's LM test, is given by

$$\text{PUY's } LM = \sqrt{\frac{2}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{(T-k)\check{\rho}_{ij}^2 - \mu_{Tij}}{\sigma_{Tij}}, \tag{13.33}$$

where

$$\mu_{Tij} = \frac{1}{T-k} tr[E(M_i M_j)]$$

is the exact mean of $(T-k)\check{\rho}_{ij}^2$ and

$$\sigma_{Tij}^2 = \{tr[E(M_i M_j)]\}^2 a_{1T} + 2tr\{E[(M_i M_j)^2]\} a_{2T}$$

is its exact variance. Here,

$$a_{1T} = a_{2T} - \frac{1}{(T-k)^2},$$

$$a_{2T} = 3\left[\frac{(T-k-8)(T-k+2)+24}{(T-k+2)(T-k-2)(T-k-4)}\right]^2,$$

$M_i = I - X_i(X_i'X_i)^{-1}X_i'$, where $X_i = (x_{i1}, \cdots, x_{iT})'$ contains T observations on the k regressors for the i-th individual regression. PUY's LM is asymptotically distributed as $N(0, 1)$, under the null, with $T \to \infty$ first, and then $N \to \infty$. Moscone and Tosetti (2009) survey cross-sectional dependence tests in panels. They examine tests based on the sample pairwise correlation coefficient, as well as tests based on the theory of spacing. Using Monte Carlo experiments, they show that tests based on the average of pairwise correlation coefficients work well when the alternative hypothesis is a factor model with nonzero mean loadings. Tests based on spacing are powerful in identifying various forms of strong cross-section dependence, but have low power when they are used to capture spatial correlation.

Baltagi, Feng and Kao (2012) consider the fixed effects homogeneous panel data model

$$y_{it} = \alpha + x_{it}'\beta + \mu_i + v_{it}, \text{ for } i = 1, ..., N; \; t = 1, ..., T \tag{13.34}$$

where μ_i denotes the time-invariant individual effect. The $k \times 1$ regressors x_{it} could be correlated with μ_i, but are uncorrelated with the idiosyncratic error v_{it}. This is a standard model in the applied panel data literature and differs from (13.29) in that the β_i's are the same, and heterogeneity is introduced through the μ_i's. Using the following asymptotics:

Assumption 1 $\frac{N}{T} \to c \in (0, \infty)$ as $(N, T) \to \infty$, where c is a nonzero bounded constant.

and the following standard assumptions:

Assumption 2 (i) The $N \times 1$ vectors of idiosyncratic disturbances $V_t = (v_{1t}, \cdots, v_{NT})'$, $t = 1, ..., T$, are iid $N(0, \Sigma_\nu)$ over time; (ii) $E[v_{it}|x_{i1}, ..., x_{iT}] = 0$ and $E[v_{it}|x_{j1}, ..., x_{jT}] = 0, i = 1, \cdots, N, t = 1, \cdots, T$; (iii) For the demeaned regressors $\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_{s=1}^{T} x_{is}$, $\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it}$, $\frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{jt}'$ are stochastic bounded for all $i = 1, \cdots, N$ and $j = 1, \cdots, N$, and $\lim_{(N,T)\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'$ exists and is nonsingular.

Under these assumptions, the Within estimator $\tilde{\beta}$ is $\sqrt{NT}$-consistent. This estimator is obtained by regressing $\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{s=1}^{T} y_{is}$ on $\tilde{x}_{it}$. The corresponding Within residuals given by $\widehat{v}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\tilde{\beta}$ are used to compute the sample correlation $\hat{\rho}_{ij}$ as follows:

$$\hat{\rho}_{ij} = \left(\sum_{t=1}^{T} \widehat{v}_{it}^2\right)^{-1/2} \left(\sum_{t=1}^{T} \hat{v}_{jt}^2\right)^{-1/2} \sum_{t=1}^{T} \hat{v}_{it}\hat{v}_{jt}. \tag{13.35}$$

The scaled version of the LM_{BP} test suggested by Pesaran (2004) but now applied to the fixed effects model is given by

$$LM_P = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(T\hat{\rho}_{ij}^2 - 1\right). \tag{13.36}$$

This replaces $\breve{\rho}_{ij}$ with $\hat{\rho}_{ij}$, and it now tests the null given in (13.30) only applied to the remainder disturbance v_{it}. This LM_P test, for the fixed effects model (13.36), suffers from the same problems discussed by Pesaran (2004) for the corresponding CD_{lm} statistic (13.32) for the heterogeneous panel model. Baltagi, Feng and Kao

(2012) show that it will exhibit substantial size distortions due to incorrect centering when N is large. Unlike the finite sample adjustment in Pesaran, Ullah and Yamagata (2008), this paper derives the asymptotic distribution of the LM_P statistic under the null as $(N, T) \to \infty$, and proposes a bias-corrected LM test. The asymptotics are done using high dimensional inference from the statistics literature. Baltagi, Feng and Kao (2012) find that in a fixed effects panel data model, after subtracting a constant that is a function of N and T, the LM_P test is asymptotically distributed, under the null, as a standard normal. Therefore, a bias-corrected LM test is proposed.

Theorem 1 *Under Assumptions 1, 2 and the null hypothesis of no cross-section dependence*

$$LM_P - \frac{N}{2(T-1)} \xrightarrow{d} N(0, 1).$$

The asymptotics are derived under the joint asymptotics of $(N, T) \to \infty$ *with* $N/T \to c \in (0, \infty)$.

Based on this result, Baltagi, Feng and Kao (2012) propose a bias-corrected LM test statistic given by:

$$LM_{BC} = LM_P - \frac{N}{2(T-1)} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(T\hat{\rho}_{ij}^2 - 1\right) - \frac{N}{2(T-1)}. \tag{13.37}$$

Under the null, the limiting distribution of the bias-corrected LM test is standard normal.

Monte Carlo simulations are used to examine the empirical size and power of our bias-corrected LM test defined in (13.37) for a static panel data model. They compare its performance with that of Pesaran (2004) CD test given by

$$\text{Pesaran's } CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \breve{\rho}_{ij},$$

and PUY's LM test given in (13.33). The sample correlations $\breve{\rho}_{ij}$ are computed using OLS residuals; see (13.31). They also include the John (1972) test for the null of sphericity discussed in Baltagi, Feng and Kao (2011).

Note that Hoyos and Sarafidis (2006) programmed Pesaran's CD test in Stata but applied to fixed effects or random effects residuals, i.e., after performing *xtreg*, fe, or re. Since this is based on the sum rather than the sum of squares of the correlations across any pairs of cross-section residuals, the command *xtcsd* reports the average of the absolute values of these estimated correlations. This is performed in Table 13.7 for the Grunfeld data. The average absolute value is 0.439 and indicates cross-section correlation for the fixed effects residuals across the firms. Pesaran's CD test on the fixed effects residuals rejects zero correlation across any pair of firms.

Table 13.7 The Pesaran CD test using fixed effects residuals for the Grunfeld data

```
. xtcsd, pesaran abs

Pesaran's test of cross sectional independence =     4.661, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.439
```

Sphericity means that Σ_ν is proportional to the identity matrix. The John (1972) test statistic is given by

$$J = \frac{T(\frac{1}{N}tr\hat{S})^{-2}\frac{1}{N}tr(\hat{S}^2) - T - N}{2} - \frac{1}{2} - \frac{N}{2(T-1)}$$

where $\hat{S} = \frac{1}{T}\sum_{t=1}^{T}\hat{V}_t\hat{V}_t'$ is the $N \times N$ sample covariance matrix computed using the Within residuals $\hat{V}_t = (\hat{v}_{1t}, ..., \hat{v}_{Nt})'$. $tr\hat{S}$ is the trace of the matrix $\hat{S}$. Under normality and homoskedasticity, the John test can be used to test for cross-sectional dependence. However, John's test is not robust to heteroskedasticity, and rejection of the null hypothesis using the John test could be due to heteroskedasticity or cross-sectional dependence. For this reason, the John test is included in the experiments only under the homoskedastic case. The John test is not recommended for testing cross-section dependence when heteroskedasticity is present. The simulation results show that the bias-corrected LM test successfully controls for size distortions as N gets large relative to T. It also maintains reasonable power under the alternative of a factor model or spatial model. However, Baltagi, Feng and Kao (2012) find that the proposed LM test is not robust to slope heterogeneity. More importantly, the simulation results show that the bias-corrected LM test can be applied in typical micro-panel data case with large N and small T.

Eviews programmed the Breusch-Pagan LM statistic for no cross-section dependence as well as Pesaran's scaled LM version, the Bias-corrected scaled LM of Baltagi, Feng and Kao (2012), and Pesaran's CD test. Applying it to the Grunfeld fixed effects residuals, we get Table 13.8. All tests statistics reject the null of no cross-section dependence. Note that Pesaran CD generated by EViews is the same as that in Table 13.7 by Stata. Also, the Breusch-Pagan test is the same as Table 13.6 by Stata.

Baltagi and Moscone (2010) reconsider the long-run economic relationship between health care expenditure and income using a panel of 20 OECD countries observed over the period 1971–2004. In particular, the paper studies the nonstationarity and cointegration properties between health care spending and income. This is done in a panel data context controlling for both cross-section dependence and unobserved heterogeneity. Cross-section dependence is modeled through a common

Table 13.8 Cross-Section dependence tests using EViews for the Grunfeld fixed effects residuals

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQFE
Periods included: 20
Cross-sections included: 10
Total panel observations: 200
Cross-section effects were removed during estimation

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	246.3288	45	0.0000
Pesaran scaled LM	21.22192		0.0000
Bias-corrected scaled LM	20.95876		0.0000
Pesaran CD	4.661192		0.0000

factor model and through spatial dependence. The average correlation coefficient for the first difference of the logarithm of health care expenditure and income are 0.48 and 0.55, respectively. The corresponding CD_{lm} statistics are 73.8 and 97.3, respectively. Both of which are statistically significant. The IPS, Breitung, and CIPS panel unit root tests are applied and the null hypothesis of panel unit roots is not rejected for the variables under study. Heterogeneity is handled through fixed effects in a panel homogeneous model and through a panel heterogeneous model. FE, Spatial MLE, and CCEP estimation are employed. The last method is the Common Correlated Effects (CCE) method suggested by Pesaran (2006). Here, the cross-section averages of the dependent variable and regressors are included in the heterogeneous panel regression. The CCE pooled (CCEP) estimator is the average of the slope coefficients. The FE estimate of the income elasticity is 0.899 compared to 0.896 for spatial MLE and 0.674 for CCEP. These results corroborate the hypothesis that health care is a necessity good.

Baltagi, Kao and Peng (2016) modify Pesaran's CD test to account for serial correlation of an unknown form in the error term. They derive the limiting distribution of this test as $(N, T) \to \infty$. The test statistic is distribution free and allows for unknown forms of serial correlation in the errors. Monte Carlo simulations show that existing tests for cross-sectional correlation encounter size distortions with serial correlation in the errors, but the proposed test has good size and power for large panels with serial correlation in the errors.

Sarafidis, Yamagata and Robertson (2009) propose a new testing procedure for detecting error cross-section dependence after estimating a linear dynamic panel data model with regressors using FD-GMM and System GMM. The test is valid when N is large relative to T. Finite sample simulation-based results suggest that their tests perform well, particularly the version based on system GMM. This approach allows one to examine whether any error cross-section dependence remains after including time dummies (or after transforming the data in terms of deviations from time-specific averages). This transformation will wipe out common effects, unless their impact differs across cross-sectional units (heterogeneous cross-section dependence).

Pesaran and Tosetti (2011) study large panel data sets where even after conditioning on common observed effects the cross-section units might remain dependently distributed. This could be due to unobserved common factors and/or spatial effects. They introduce the concepts of time-specific *weak and strong cross-section dependence* and show that the commonly used spatial models are examples of weak cross-section dependence. Pesaran (2006) common correlated effects (CCE) estimator for panel data models with a multifactor error structure continues to provide consistent estimates of the slope coefficient, even in the presence of spatial error processes. For an extensive survey of cross-section dependence in large panels, see Chap. 1 of the Oxford Handbook of Panel Data, by Chudik and Pesaran (2015).

13.7 Computational Note

For R programs implementing both maximum likelihood and generalized moments estimators in the context of fixed as well as random effects spatial panel data models, see the splm package by Millo and Piras (2012). These spatial panel estimation methods are applied to the Munnell (1990) data set used in example 3 of Chap. 2 but now with spatial correlation across states. For Stata programs (using the command *xsmle*), one gets maximum likelihood methods applied to various static and dynamic spatial panel models with fixed and random effects, see Belotti, Hughes and Piano Mortari (2017), with an application to residential demand for electricity for a panel of 48 states plus the district of Columbia over the period 1990–2010. These were used in the Empirical Example in this chapter. Stata also has an *spxtreg* command which is applied to the homicide data of Messner et al. (2000). This data covers 1412 counties for the census years 1960–1990. For matlab programs for spatial panel models, see Elhorst (2014) at https://spatial-panels.com/software/. These programs are demonstrated using the cigarette data example used in Chapter 8. Also, for Bayesian spatial methods not covered in this textbook, see LeSage and Pace (2009) for a textbook treatment of this subject. These are programmed with Matlab at www.spatial-econometrics.com.

13.8 Problems

13.1 *Prediction in the spatially autocorrelated error component model*. This is based on problem 99.2.4 in Econometric Theory by Baltagi and Li (1999). Consider the panel data regression model described in (13.1) with random country effects and spatially autocorrelated remainder disturbances described by (13.2) and (13.3). Using the Goldberger (1962) best linear unbiased prediction results discussed in Sect. 2.5, Eq. (2.37), derive the BLUP of $y_{i,T+S}$ for the ith country at period $T + S$ for this spatial panel model. Hint: see solution 99.2.4 in Econometric Theory by Song and Jung (2000).

13.2 *Random effects and spatial autocorrelation with equal weights*. This is based on Baltagi (2006). Consider the panel data regression model described in (13.1) with random individual effects and spatially autocorrelated remainder disturbances described by (13.2) and (13.3). In this special case, W is an $N \times N$ weighting matrix with zero elements across the diagonal, and equal elements $(1/(N-1))$ off the diagonal. In other words, the disturbance for each unit is related to an average of the $(N-1)$ disturbances of the remaining units. Such a weighting matrix would naturally arise if all units are neighbors to each other and there is no other reasonable or observable measure of distance between them. W can be written as $W = \frac{J_N}{(N-1)} - \frac{I_N}{(N-1)}$.

(a) Show that GLS on this model can be obtained using an OLS regression of $y_{ti}^* = (y_{ti} - \theta_1 \overline{y}_{t.} - \theta_2 \overline{y}_{.i} + \theta_3 \overline{y}_{..})$ on X^* similarly defined. Here, $\overline{y}_{t.}$ denotes the sample average over individuals; $\overline{y}_{.i}$ denotes the sample average over time; and $\overline{y}_{..}$ denotes the average over the entire sample. The $\theta's$ are scalars which depend on N, λ, and the variance components σ_μ^2 and σ_ν^2.

(b) Show that if there is no spatial autocorrelation, i.e., $\lambda = 0$, then y_{ti}^* reduces to $(y_{ti} - \theta_2 \overline{y}_{.i})$. This is the familiar Fuller and Battese (1973) random effects transformation that was obtained in Chap. 2; see problem 2.4. Show that if there are no random effects, i.e., $\sigma_\mu^2 = 0$, and $N \to \infty$, then $y_{ti}^* = (y_{ti} - \lambda \overline{y}_{t.})$

(c) Show that in the cross-section spatial regression model with $(T = 1)$ and equal weight matrix, OLS is equivalent to GLS as long as there is a constant in the regression.

(d) For the spatial panel regression with equal weights, show that two special cases where OLS is equivalent to GLS, are the following: (i) the trivial case where $\sigma_\mu^2 = 0$ and $\lambda = 0$; and (ii) when the matrix of regressors X is invariant across time. Baltagi (2006) showed that these results for the equal weight matrix hold whether we use the spatial autoregressive (SAR) specification for the disturbances, or the spatial moving average (SMA) specification described in Anselin (1988), or the spatial error components (SEC) specification described in Kelejian and Robinson (1995), or the Kapoor, Kelejian and Prucha (2007) panel data regression model with spatially correlated error components.

13.3 *Residential Demand for Electricity*. Belotti, Hughes and Piano Mortari (2017) estimated residential demand for electricity covering the 48 states in the continental United States plus the district of Columbia for the period 1990–2010.

(a) Replicate Tables 13.1, 13.2 and 13.3 using *xsmle* in Stata showing the maximum likelihood estimates of the Generalized Spatial Panel Random Effects model as well as its special cases of the Anselin and KKP random effects models.
(b) Using the log-likelihood values of these models, test the null that the KKP model restriction is satisfied. Test the null that the Anselin model restriction is satisfied. What do you conclude?
(c) Replicate some of the results in Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). In particular, the first column labeled FE which esimates the residential electricity demand model using fixed effects estimates without any spatial correlation. The second column labeled SAR which estimates a spatial lag model on the dependent variable only. The sixth column labeled SAC, which estimates a SARAR model. The seventh column labeled SEM which estimates a spatial error model as in Anselin (1988), all with fixed effects.
(d) Given the results in part (c), one can use the log-likelihood values to perform the LR tests of Debarsy and Ertur (2010). In particular, you are asked to compute the joint LR test for the spatial lag coefficient ρ as well as the spatial error coefficient λ are zero. This null hypothesis is represented by $(H_0^c : \rho = \lambda = 0)$ in the text. Also, the marginal LR tests for $(H_0^a : \rho = 0)$ as well as $(H_0^b : \lambda = 0)$. Finally, the conditional LR tests for $(H_0^d : \lambda = 0/\rho)$ as well as $(H_0^e : \rho = 0/\lambda)$. What do you conclude?
(e) Replicate the Spatial Durbin Model (SDM) results given in column 4 of Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). Test the extra spatial Durbin term using the LR test. This can be done by comparing the log-likelihood of SDM with that of SAR.
(f) Replicate the dynamic version of the Spatial Durbin Model (dynamic SDM) results given in column 5 of Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). Also, the dynamic version of the Spatial Autoregressive Model (dynamic SAR) results given in column 3 of Table 5 of Belotti, Hughes and Piano Mortari (2017, p. 174). This is Table 13.5 in this chapter. Test the dynamic terms using the LR test comparing the log-likelihood of SAR with that of dynamic SAR. Also, the LR test comparing SDM with dynamic SDM. Finally, test the spatial Durbin term, as in part (e) but now for the dynamic model using the LR test comparing the log-likelihood of dynamic SDM with that of dynamic SAR. These LR tests are reported in the first 3 rows of Table 6 of Belotti, Hughes and Piano Mortari (2017, p. 175). What do you conclude?
(g) Check the sensitivity of the results in part (f) by adding both the lagged value of the dependent variable and the spatial lag value of the dependent variable with option *dlag(3)* in *xsmle*. What do you conclude?

13.4 For the Grunfeld example, replicate Tables 13.6, 13.7 and 13.8, i.e., (i) obtain the Breusch and Pagan test based on the fixed effects residuals using Stata's command *xttest2*. (ii) obtain Pesaran's CD test based on fixed effects residuals using Stata's command *xtcsd*. (iii) Finally using EViews obtain the Breusch-

Pagan test, Pesaran's CD test, PUY's *LM* test which appears in EViews as Pesaran scaled LM test, also the Baltagi, Feng and Kao (2012) Bias-corrected scaled LM test, all based on fixed effects residuals. What do you conclude?

13.5 Test for cross-section dependence for the Gasoline example (as in problem 13.4). Do the same for the Public Capital example. What do you conclude?

References

Amemiya, T., and T.E. MaCurdy. 1986. Instrumental-variable estimation of an error components model. *Econometrica* 54: 869–881.

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Bai, J., and S. Ng. 2004. A panic attack on unit roots and cointegration. *Econometrica* 72: 1127–1177.

Baltagi, B.H. 1981. Simultaneous equations with error components. *Journal of Econometrics* 17: 189–200.

Baltagi, B.H. 2006. Random effects and spatial autocorrelation with equal weights. *Econometric Theory* 22: 973–984.

Baltagi, B.H., and G. Bresson. 2011. Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris. *Journal of Urban Economics* 69: 24–42.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2003. Fixed effects, random effects or Hausman-Taylor? A pretest estimator. *Economics Letters* 79: 361–369.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2007. Panel unit root tests and spatial dependence. *Journal of Applied Econometrics* 22: 339–360.

Baltagi, B.H., G. Bresson, and A. Pirotte. 2012. Forecasting with spatial panel data. *Computational Statistics and Data Analysis* 56: 3381–3397.

Baltagi, B.H., and Y. Deng. 2015. EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects. *Econometric Reviews* 34: 659–694.

Baltagi, B.H., P. Egger, and M. Kesina. 2012. Small sample properties and pretest estimation of a spatial Hausman-Taylor model. *Advances in Econometrics* 29: 219–241.

Baltagi, B.H., P. Egger, and M. Kesina. 2016. Firm-level productivity spillovers in China's chemical industry: A spatial Hausman-Taylor approach. *Journal of Applied Econometrics* 31: 214–248.

Baltagi, B.H., P. Egger, and M. Pfaffermayr. 2007. Estimating models of complex FDI: Are there third country effects? *Journal of Econometrics* 140: 260–281.

Baltagi, B.H., P. Egger, and M. Pfaffermayr. 2008. A Monte Carlo study for pure and pretest estimators of a panel data model with spatially autocorrelated disturbances. *Annales D'Économie et de Statistique* 87/88: 11–38.

Baltagi, B.H., P. Egger, and M. Pfaffermayr. 2013. A generalized spatial panel data model with random effects. *Econometric Reviews* 32: 650–685.

Baltagi, B.H., Q. Feng, and C. Kao. 2011. Testing for sphericity in fixed effects panel data models. *Econometrics Journal* 14: 25–47.

Baltagi, B.H., Q. Feng, and C. Kao. 2012. A Lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics* 170: 164–177.

Baltagi, B.H., B. Fingleton, and A. Pirotte. 2014. Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics* 76: 112–138.

Baltagi, B.H., C. Kao, and B. Peng. 2016. Testing cross-sectional dependence in large panel data models with serial correlation. *Econometrics* 4(4): 44.

Baltagi, B.H., and D. Li. 1999. Prediction in the spatially autocorrelated error component model, Problem 99.2.4. *Econometric Theory* 15: 259.
Baltagi, B.H. and D. Li. 2004. Prediction in the panel data model with spatial correlation. In *Advances in spatial econometrics: Methodology, tools and applications*, (eds.) L. Anselin, R.J.G.M. Florax, and S.J. Rey, Chapter 13, 283–295. Berlin: Springer.
Baltagi, B.H., and D. Li. 2006. Prediction in the panel data model with spatial correlation: The case of liquor. *Spatial Economic Analysis* 1: 175–185.
Baltagi, B.H., and L. Liu. 2011. Instrumental variable estimation of a spatial autoregressive panel model with random effects. *Economics Letters* 111: 135–137.
Baltagi, B.H., and L. Liu. 2016. Prediction in a generalized spatial panel data model with serial correlation. *Journal of Forecasting* 35: 573–591.
Baltagi, B.H., and F. Moscone. 2010. Health care expenditure and income in the OECD reconsidered: Evidence from panel data. *Economic Modelling* 27: 804–811.
Baltagi, B.H., and A. Pirotte. 2010. Panel data inference under spatial dependence. *Economic Modelling* 27: 1368–1381.
Baltagi, B.H., S.H. Song, B.C. Jung, and W. Koh. 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140: 5–51.
Baltagi, B.H., S.H. Song, and W. Koh. 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117: 123–150.
Baltagi, B.H., S.H. Song, and J.H. Kwon. 2009. Testing for heteroskedasticity and spatial correlation in a random effects panel data model. *Computational Statistics and Data Analysis* 53: 2897–2922.
Belotti, F., G. Hughes, and A. Piano Mortari. 2017. Spatial panel data models using Stata. *The Stata Journal* 17(1): 139–180.
Breitung, J. 2000. The local power of some unit root tests for panel data. *Advances in Econometrics* 15: 161–177.
Breusch, T.S., G.E. Mizon, and P. Schmidt. 1989. Efficient estimation using panel data. *Econometrica* 57: 695–700.
Breusch, T.S., and A.R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.
Case, A.C. 1991. Spatial patterns in household demand. *Econometrica* 59: 953–965.
Chang, Y. 2002. Nonlinear IV unit root tests in panels with cross sectional dependency. *Journal of Econometrics* 110: 261–292.
Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20: 249–272.
Choi, I. 2002. Instrumental variables estimation of a nearly nonstationary, heterogeneous error component model. *Journal of Econometrics* 109: 1–32.
Chudik, A. and M.H. Pesaran. 2015. Large panel data models with cross-sectional dependence: A survey. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, Chapter 1, 3–45. Oxford: Oxford University Press.
Conley, T.G. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* 92: 1–45.
Cornwell, C., P. Schmidt, and D. Wyhowski. 1992. Simultaneous equations and panel data. *Journal of Econometrics* 51: 151–181.
Debarsy, N. 2012. The Mundlak approach in the spatial Durbin panel data model. *Spatial Economic Analysis* 7: 109–131.
Debarsy, N., and C. Ertur. 2010. Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics* 40: 453–470.
Driscoll, J.C., and A.C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80: 549–560.
Elhorst, J.P. 2014. *Spatial econometrics: From cross-sectional data to spatial panels*. Heidelberg: Springer.
Fingleton, B. 2008. A generalized method of moments estimator for a spatial panel model with an endogeneous spatial lag and spatial moving average errors. *Spatial Economic Analysis* 3: 27–44.

Frees, E.W. 1995. Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69: 393–414.

Fuller, W.A., and G.E. Battese. 1973. Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68: 626–632.

Goldberger, A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57: 369–375.

Hausman, J.A., and W.E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49: 1377–1398.

Hoyos, R.E., and V. Sarafidis. 2006. Testing for cross-sectional dependence in panel-data models. *The Stata Journal* 6 (4): 482–496.

Im, K.S., M.H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74.

John, S. 1972. The distribution of a statistic used for testing sphericity of Normal distributions. *Biometrika* 59: 169–173.

Kapoor, M., H.H. Kelejian, and I.R. Prucha. 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140: 97–130.

Kelejian, H.H., and I.R. Prucha. 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17: 99–121.

Kelejian, H.H., and I. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509–533.

Kelejian, H.H., and D.P. Robinson. 1995. Spatial correlation: A suggested alternative to the autoregressive model. In *New directions in spatial econometrics*, (eds.) L. Anselin, and R.J. Florax, 75–95. Springer.

Lee, L.F. 2003. Best spatial two-stage least square estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22: 307–335.

Lee, L.F., and J. Yu. 2012. Spatial panels: Random components versus fixed effects. *International Economic Review* 53: 1369–1412.

Lee, L.F. and J. Yu. 2015. Spatial panel data models. In *Oxford handbook of panel data*, (ed.) B.H. Baltagi, Chapter 12, 363–401. Oxford: Oxford University Press.

LeSage, J.P., and R.K. Pace. 2009. *Introduction to spatial econometrics*. Boca Raton: Taylor and Francis.

Levin, A., C.F. Lin, and C. Chu. 2002. Unit root test in panel data: Asymptotic and finite sample properties. *Journal of Econometrics* 108: 1–24.

Maddala, G.S., and S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61: 631–652.

Messner, S.F., L. Anselin, D.F. Hawkins, G. Deane, S.E. Tolnay, and R.D. Baller. 2000. *An atlas of the spatial patterning of county-level homicide, 1960–1990*. Pittsburgh: National Consortium on Violence Research.

Millo, G., and G. Piras. 2012. splm: spatial panel data models in R. *Journal of Statistical Software* 47: 1–38.

Moscone, F., and E. Tosetti. 2009. A review and comparison of tests of cross-section independence in panels. *Journal of Economic Surveys* 23: 528–561.

Mundlak, Y. 1978. On the pooling of time series and cross-section data. *Econometrica* 46: 69–85.

Munnell, A. 1990. How does public infrastructure affect regional economic performance? *New England Economic Review* 3–22.

Mutl, J., and M. Pfaffermayr. 2010. The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* 10: 1–30.

Pesaran, M.H. 2004. General diagnostic tests for cross-section dependence in panels, working paper. Cambridge: Trinity College.

Pesaran, M.H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012.

Pesaran, M.H. 2007. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* 27: 265–312.

Pesaran, M.H., and E. Tosetti. 2011. Large panels with common factors and spatial correlations. *Journal of Econometrics* 161: 182–202.
Pesaran, M.H., A. Ullah, and T. Yamagata. 2008. A bias-adjusted LM test of error cross section independence. *Econometrics Journal* 11: 105–127.
Phillips, P.C.B., and D. Sul. 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6: 217–259.
Sarafidis, V., T. Yamagata, and D. Robertson. 2009. A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* 148: 149–161.
Song, S.H., and B.C. Jung. 2000. Prediction in the spatially autocorrelated error component model, Solution 99.2.4. *Econometric Theory* 16: 149–150.
Wansbeek, T.J., and A. Kapteyn. 1982. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11: 2105–2112.
Yu, J., R. de Jong, and L.F. Lee. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146: 118–134.