

密级：公开

中图分类号：F222



# 浙江工商大学

## 博士学位论文

论文题目： 半参数带固定效应的面板

数据模型的经验似然

作者姓名： 何帮强

学科专业： 统计学

研究方向： 经济统计

指导教师： 洪兴建 教授

提交日期： 2017年 05 月

## 半参数带固定效应的面板数据模型的经验似然

### 摘 要

面板数据是指一部分个体(个人、家庭、企业或国家等)在一段时期内某个变量的观测值所构成的多维数据集合。面板数据包含了两个现象的观测值,即对于每一个个体有多个时间观测值。从横截面看,面板数据是由若干个体在某一时刻构成的截面观测值。从个体看每个个体都是一个时间序列。时间序列数据和横截面数据是面板数据的特殊形式,它们仅仅是一维的。面板数据有时被认为是带有时间观测的横截面数据或者是混合的横截面时间序列数据。通过使用面板数据,经济学中的实证研究得到了极大的丰富。随着可用的面板数据的增加,面板数据的理论和应用在近年来变得更加流行。

面板数据集分析具有优于纯截面数据或时间序列数据集的各种优点,面板数据能够获得从不同环境采样的不同单元的个体特征,其中持久的个体特征是它的异质性,其中大部分是不可观测的。面板数据为研究者提供了灵活的方法来模拟横截面单位的异质性和可能随时间变化的结构。

我们研究的模型是半参数模型,半参数模型结合了非参数模型的灵活性和线性模型的简洁性。半参数模型在经济学、生物学和医

学领域的研究中有着广泛的应用。本学位论文研究带固定效应面板数据半参数模型的经验似然问题,主要运用Owen (1988,1990)提出的经验似然方法,经验似然方法在构造未知参数的置信区间(区域)有着很多突出的优点,如经验似然区域的形状完全有样本决定, Bartlett纠偏性,区域保持性和变换不变性等。因此,经验似然方法受到了许多统计学家和经济学家的广泛关注,许多研究者把经验似然方法应用到各种统计模型和领域。对于高维线性模型, Tang和Leng (2010), Leng和Tang (2012)建议惩罚经验似然进行变量选择和参数估计,并显示惩罚经验似然具有oracle特征。具体内容如下:

第一章是引言,首先阐述了本文的研究背景和研究意义,分别从模型,问题的角度论证选题的合理性和可行性。其次对有关半参数面板数据模型的研究现状进行了梳理和总结。最后介绍了本论文的研究内容,概括了本论文的研究方法,研究内容和主要创新点。

第二章,我们研究带固定效应部分线性面板数据模型的参数估计,为了解决组内相关性,我们应用块经验似然方法,并构造了未知参数的置信域。证明了调整的经验对数似然比函数在真参数点时是渐近卡方分布的。并且利用块经验似然比率函数,参数的最大经验似然估计被定义和正态逼近被显示。同时,模拟研究表明经验似然方法优于正态逼近方法。并且利用CD4数据也能实际验证我们建议

的方法更合适。

第三章, 我们考虑带固定效应面板数据部分线性误差变量模型的统计推断, 我们专注于一些协变量为附加误差情况下, 我们提出了一个修正的剖面最小二乘估计参数和非参数部分, 参数部分的渐近正态性和非参数部分的收敛速度被建立。一致估计的误差方差也被估计。此外, 我们介绍剖面似然比 (PLR) 检验并证明它在原假设条件下是一个渐进的卡方分布。我们进行模拟研究, 证明了我们所提出的方法更优于不考虑误差变量情况下的估计。我们也用英国联邦的气候数据验证我们建议的方法更合适。

第四章, 对于带固定效应面板数据的高维部分线性误差变量模型, 当误差方差为已知和未知情形下。在这一章, 我们建议一个修正的回归参数和最大经验似然比率回归参数。同时, 依据惩罚经验似然方法, 这个模型的参数估计和变量选择被调查, 这个建议的惩罚经验似然被证明具有神特征。同时, 我们建议的惩罚经验似然比率统计在原假设下具有渐近卡方分布, 其结果可以用来构造未知参数的置信域。模拟和实证结果用来评估经验似然方法的性能。

第五章, 考虑带固定效应面板数据的半变系数模型的经验似然, 提出了在协变量是 $\alpha$ -混合下情形下, 未知参数的经验对数似然比统计量, 证明了建议的对数似然比统计量是渐近卡方分布的, 我们也获得参数的最大似然估计量和证明在合适条件是渐近正态的。

由此,未知参数两个不同的渐近正确覆盖率的置信域被构造。模拟和实证结果用来评估经验似然方法的性能。

第六章,我们研究带固定效应面板数据部分线性时变系数模型的经验似然,提出了在协变量是 $\alpha$ -混合下情形下,未知参数的经验对数似然比统计量,证明了建议的对数似然比统计量是渐近标准卡方分布的,一个模拟表明对比正态逼近方法,建议的方法比正态逼近方法在概率覆盖率方面表现的更好。

第七章,简单总结全文的研究工作和主要的创新点,并指出需要进一步完善和深入研究的问题。

**关键词:** 面板数据; 固定效应; 部分线性回归模型; 半变系数模型; 误差变量; 经验似然; 惩罚经验似然; 剖面最小二乘

# Empirical likelihood for semiparametric panel data models with fixed effects

## **ABSTRACT**

Panel data refers to a portion of an individual (individual, family, business or Country, etc.) in a period of time a variable consisting of the observations dimensional data set. Panel data contains observations on two phenomena observed over multiple time periods for each individual. From the cross-section view, panel data is composed of a number of individuals in a certain point section observation. And from the individual to see each individual is a time series. Time series and cross-sectional data are special cases of panel data that are in one-dimension only. Panel data are sometimes treated as cross-sections over time or pooled cross-section time-series data. Empirical research in economics has been enriched by the availability of panel data. With the increased availability of panel da-

ta, both theoretical and applied work in panel data analysis have become more popular in the recent years.

Analysis of panel data sets has various advantages over that of pure cross sectional or time series data sets. Panel data usually cover individual units sampled from different backgrounds and with different individual characteristics so that an abiding feature of the data is its heterogeneity, much of which is simply unobserved. Panel data provide researchers a flexible way to model both heterogeneity among cross sectional units and possible structural changes over time.

The model we studied is a semiparametric model. Semiparametric model combines the flexibility of nonparametric regression and parsimony of linear regression. The models are important and applied widely in economic, biological and medical studies. In this thesis, we study the empirical likelihood for semiparametric panel data models with fixed effects. We mainly use the empirical likelihood method proposed by Owen (1988, 1990). There are several nice advantages for empirical likelihood. the empirical likelihood regions are shaped completely by the sample, Bartlett correctable, range preserving and transformation respecting. So, the empirical likelihood method has been paid great attention by statisticians and economists and has received widely researches and applica-

tions. Many researchers have applied the empirical likelihood method to various models and fields. For high-dimensional linear models, Tang and Leng (2010) and Leng and Tang (2012) propose a penalty empirical likelihood for parameter estimation and variable selection, and show that penalized empirical likelihood has the oracle property. The chapters are as follows:

The first chapter is introduction, we state research background and research significance. we demonstrate the rationality and feasibility of the subject selection from the perspective of models and problems. In addition, we briefly describing the research status of semi-parametric panel data models with fixed effects. At last, the main content of this dissertation are proposed. the research methods, research context and the main innovative points are summarized.

In chapter 2, we consider a partially linear panel data models with fixed effects. In order to accommodate the within-group correlation, we apply the block empirical likelihood procedure to partially linear panel data models with fixed effects, and prove a nonparametric version of Wilks' theorem which can be used to construct the confidence region for the parametric. By the block empirical likelihood ratio function, the maximum empirical likelihood estimator of the parameter is defined and the



asymptotic normality is shown. A simulation study and a real data application are undertaken to assess the finite sample performance of our proposed method. the confidence regions construction for the parameters of interest in the partially linear regression model with linear process errors under martingale difference is studied. It is shown that the adjusted empirical log-likelihood ratio at the true parameters is asymptotically chi-squared. A simulation study indicates that the adjusted empirical likelihood works better than a normal approximation-based approach.

In chapter 3, we consider the statistical inference for the partially linear panel data models with fixed effects. We focus on the case where some covariates are measured with additive errors. We propose a modified profile least squares estimator of the regression parameter and the nonparametric components. The asymptotically normality for the parametric component and the rate of convergence for the nonparametric component are established. Consistent estimations of the error variance are also developed. In addition, we introduce the profile likelihood ratio (PLR) test and then demonstrate that it follows an asymptotically  $\chi^2$  distribution under the null hypothesis. We conduct simulation studies to demonstrate the finite sample performance of our proposed method and we also present an illustrative empirical application.

In chapter 4, For the high-dimensional partially linear panel data models with fixed effects where covariates are measured with additive errors, we, in this chapter, propose a modified profile least squares estimator of the regression parameter and maximum empirical likelihood estimator of the regression parameter. At the same time, based on penalized empirical likelihood (PEL) approach, the parameter estimation and variable selection of the model are investigated, the proposed PEL estimators are shown to possess the oracle property. Also, we introduce the PEL ratio statistic to test a linear hypothesis of the parameter and prove it follows an asymptotically chi-square distribution under the null hypothesis. We conduct simulation studies to demonstrate the finite sample performance of our proposed method and we also present an illustrative empirical application.

In chapter 5, The empirical likelihood inference for semi-varying coefficient models for panel data with fixed effects is investigated in this paper. We propose an empirical log-likelihood ratio function for the regression parameters in the model under  $\alpha$ -mixing condition. The empirical log-likelihood ratio is proven to be asymptotically chi-squared. We also obtain the maximum empirical likelihood estimator of the parameters of interest, and prove that it is the asymptotically normal under some

suitable conditions. A simulation study and a real data application are undertaken to assess the finite sample performance of our proposed method.

In chapter 6, The empirical likelihood inference for time-varying coefficient models for panel data with fixed effects is investigated in this paper. We propose an empirical log-likelihood ratio function for the regression parameters in the model under  $\alpha$ -mixing condition. The empirical log-likelihood ratio is proven to be asymptotically standard chi-squared. A simulation study indicates that, compared with a normal approximation-based approach, the proposed method described herein works better in terms of coverage probabilities.

Finally, in chapter 7, the research and main innovative points are briefly summarized, and some problems which need to be further perfected and studied are pointed.

**Key Words:** Panel data; Fixed effect; Partially linear regression model; Semi-varying coefficient model; Errors-in-variables; Empirical likelihood; Penalized empirical likelihood; Profile least squares

## 目录

摘要	I
Abstract .....	V
目录 .....	XIV
第 1 章 引言 .....	1
1.1 研究背景和意义 .....	1
1.1.1 研究背景 .....	1
1.1.2 研究意义 .....	3
1.2 研究现状 .....	5
1.3 本文的工作 .....	8
1.3.1 研究方法 .....	8
1.3.2 研究内容和结构安排 .....	12
1.3.3 创新之处 .....	15
第 2 章 带固定效应部分线性面板数据模型的块经验似然 .....	17
2.1 引言 .....	17
2.2 模型和方法 .....	19

2.3 假定和主要结论.....	21
2.4 模拟研究与实证.....	23
2.4.1 Monte Carlo 模拟.....	23
2.4.2 基于艾滋病人的CD4数据的实证.....	24
2.5 主要结果的证明.....	26
<b>第3章 带固定效应的面板数据部分线性误差变量模型的统计推断</b> .....	<b>35</b>
3.1 引言.....	35
3.2 模型和方法.....	37
3.3 剖面似然比检验.....	40
3.4 假定和主要结论.....	41
3.5 模拟研究与实证.....	43
3.5.1 Monte Carlo 模拟.....	43
3.5.2 基于英国各地区气象数据的实证.....	44
3.6 主要结果的证明.....	46
<b>第4章 带固定效应的部分线性误差变量面板数据模型的惩罚经验似然</b> .....	<b>58</b>
4.1 引言.....	58

4.2 模型和方法 .....	60
4.3 假定和主要结论.....	63
4.4 计算方法 .....	65
4.5 模拟研究与实证.....	66
4.5.1 Monte Carlo 模拟.....	66
4.5.2 变量选择 .....	67
4.6 主要结果的证明.....	67
<b>第 5 章 带固定效应面板数据半变系数模型的经验似然 .....</b>	<b>78</b>
5.1 引言.....	78
5.2 模型和方法 .....	79
5.3 假定和主要结论.....	82
5.4 模拟研究与实证.....	83
5.4.1 Monte Carlo 模拟.....	83
5.4.2 基于艾滋病人的CD4数据的实证.....	86
5.5 主要结果的证明.....	87
5.6 本章补充的引理.....	95

第 6 章 带固定效应面板数据部分线性时变系数模型的经验似然 .....	98
6.1 引言 .....	98
6.2 模型和方法 .....	99
6.3 假定和主要结论 .....	101
6.4 Monte Carlo 模拟研究 .....	102
6.5 主要结果的证明 .....	103
第 7 章 结论与展望 .....	108
参考文献 .....	110
攻读博士学位期间发表和接受的学术论文与研究成果 .....	119
致谢 .....	120

## 第1章 引言

作为全文的引言部分,本章主要从以下三个方面展开论述:首先,对本文所研究问题的背景进行较为详细的说明,并指明本文研究问题的意义所在;其次,围绕着本文的研究问题,对相关面板数据和半参数模型进行文献综述,阐述研究现状,厘清研究脉络;最后,基于上述讨论,阐述我们的研究工作,介绍本文研究方法和研究内容,以及相应的结构安排和可能的创新之处。

### 1.1 研究背景和意义

面板数据是在不同时期跟踪由给定个体组成的样本而获取的数据,它包含样本中每个个体的多个观测值(Hsiao, 2003),随着经济现象的复杂化和经济学理论的发展需求,单纯应用截面数据或时间序列数据来说明经济问题已存在一定的不足。面板数据是在时间序列上取多个截面,在这些截面上同时选取样本观测值所构成的样本数据,可以克服时间序列分析受多重共线性的困扰,能够提供更多的样本信息、更大的自由度和更高的估计效率。面板数据是集合截面数据和时间序列数据优点的数据形式,能多方位的反映数据信息,提供更好的拟合模型。相对只利用截面数据或只利用时间序列数据进行经济分析而言,面板数据模型具有开拓性意义。

#### 1.1.1 研究背景

当今,无论是发达国家还是发展中国家,面板数据都很常见。譬如,美国有两个最著名的面板数据集:NLS(National Longitudinal Surveys of Labor Market Experience)数据集和密歇根大学PSID(Pane Study of Income Dynamics)数据集。NLS对数据的收集始于20世纪60年代中期,它有五个不同的纵列数据组成,涵括了劳动力的不同组成部分:1966年年龄在45~59岁的男子、1966年年龄在14~24岁的青年男子、1967年年龄在30~44岁的女子、1967年年龄在14~24岁的青年女子、1968年年龄在14~24岁的青年女子、1979年年龄在14~24岁的青年男女。1978年扩展后的NLS还包含对参加1979年调查的青年组女子所生的孩子的调查。调查的变量上千个,重点在于了解劳动力



市场的供给信息。PSID 从1968 年开始收集数据并持续至今,它在全国范围内收集具有代表性的6000多个家庭和15000多位个人的年度经济信息。该数据集有5000多个变量,包括就业、收入、人力资本、以及住房、是否乘车上班、流动性方面的信息。欧洲的许多国家有年度或更高频率的全国调查数据,如荷兰的SEP(Socio-Economic panel),德国的GSOEP (German Social Economics Panel),卢森堡的LSEP (Luxembourg Social Economic Panel),英国的BHPS (British Household Panel Survey) 等。面板数据越来越常见,获取原始调查数据对回答许多有意义的重大问题显得尤为重要,世界银行发出倡议并协助设计许多面板数据调查。

大多数面板数据来自非常复杂的日常经济生活,通常情况下,不同的个体可能受不同因素的影响。因此,在研究面板数据中,人们认为不同的横截面很可能具有异质性,这个异质性被认为是无法用已知的回归元观测的,同时异质性被假定依横截面不同而不同,但在不同时期却是稳定的,因此可以用横截面虚拟变量来控制横截面的异质性,此虚拟变量就是个体固定效应。如果异质性是发生在不同时期的,那么则用时期虚拟变量来控制,时期虚拟变量就是时间固定效应。固定效应的面板数据模型出现对于计量经济学的理论与应用研究产生了极其深远的影响,迅速成为了计量经济学前沿研究的重要方向,对计量经济学的发展具有里程碑的意义,其应用遍及微观经济学、宏观经济学等领域。

早在1986年,Hsiao 就写了关于面板数据模型方法研究的书籍,到2003年有了第二版,Wooldrige(2002)、Baltagi(2005)都对面板数据模型各方面给出了比较全面完整的说明。近年来对于面板数据模型的研究已在国内外掀起一股热潮,相关面板数据主题主要包括:静态面板数据模型、动态面板数据模型、非线性面板数据模型、非平稳面板数据模型等。Baltagi(2005)和Hsiao(2003)很好地概括了参数面板数据的和对经济统计推断学现象的解释,但在不少实际问题中,对截面和时间序列分析,当参数假定与实际背离时,其拟合情况就很差,对于解决这样问题,一些非参数面板数据模型已经被广泛的研究,因为非参数模型的函数形式可以任意,并有较大的适应性。Henderson 等(2008)研究固定效应非参数面板数据模型;Cai 和Li(2008)研究动态非参数面板数据模型。对于非参数处理,形式任意不能得到准确结论。还面对“维数灾祸”问题,为了克服“维数灾祸”问题,因而采用两者的混合,即采用半参数回归模型。在解决实

实际问题中, 实际工作者和学者们提出许多类型的半参数回归模型, 如部分线性回归模型、半参数变系数模型、单指标模型、部分线性单指标模型等。我们现在研究含有固定效应的面板数据下的半参数模型。国内外学者也已经做了这方面的工作, 如Su 和Ullah(2006)用剖面似然方法估计了带固定效应的部分线性面板数据模型, Li 等(2011)用非参数方法估计带固定效应的变系数面板数据模型, Chen 等(2013)估计了带固定效应的单指数线性面板数据模型。对含固定效应的半参数面板数据的研究不仅是半参数模型的深化, 而且更有现实的研究意义。

### 1.1.2 研究意义

在经济研究中, 与传统的横截面数据或时间序列数据相比, 面板数据具有多方面的优势。

首先, 面板数据通常能为研究人员分析大量用的数据点, 因此增加了数据的自由度并降低了解释变量间的共线性的程度, 故而提高了计量模型的有效性。面板数据能够分析大量用横截面数据与时间序列数据无法处理的重大经济问题。譬如, 我们考虑一个来自Ben-Porach (1973) 的案例: 假设从某已婚女士的横截面样本中发现样本中的女士平均每年有50% 的劳动参与率. 对该发现的一个极端解释是: 样本中的女士们来自同一个总体, 在任意给定的年份, 样本中的女士们来自不同的总体, 其中的50%一直在工作, 而另外50%从不工作。在第一种情形, 人们预期每位女士婚后一半的时间在劳动力市场之外度过, 而另一半时间花在劳动力市场内, 并以平均两年的工作期限频繁地变动工作。在第二种情形, 女士们没有工作的状态的变化, 她们的当前工作状态就是将来工作状态的完美预测。要对这两个模型进行区分, 我们需要利用个体的劳动力历史信息来估计在生命周期的不同时段参与劳动力市场的概率。这只有在我们对大量的个体又连续的观测时才可能实现。

第二, 面板数据模型可以从多层面分析经济问题。多层次样本形式既包括对样本单位在某一时间上多个特性的观察, 也包括对样本单位的这些特性在一段时间上的连续观察。面板数据集含有 $n \times T$ 个数据, 能提供更多的数据信息和更大的自由度。既可以分析个体之间的差异情况, 又可以描述个体的动态变化特征; 可以从不同角度反映已有数据的信息。在宏观经济领域, 它被广泛地应用于经济增长收敛理论的检验、产业结构的分析、技术创新的研

究、汇率决定理论研究和国际R&D溢出研究等领域;在微观经济领域,被大量应用于就业、家庭消费等领域。面板数据适合研究动态调整过程,例如,分析我国的结构性失业问题,它既受到各地区产业结构的影响,也受到国家在各个时期的宏观政策的影响。只利用截面数据,即选择同一时间上不同省市的数据作为样本观测值,可以分析各省市不同的产业结构对结构性失业的影响,但是不能分析国家的宏观政策对各省市结构性失业的影响;只利用时间序列数据,即选择同一省市或者全国整体在不同时间上的数据作为样本观测值,可以分析国家的宏观政策对某地区的结构性失业的影响,但是不能分析不同的产业结构对结构性失业的影响。如果采用面板数据模型,即在不同的时间上选择不同省市的数据作为样本观测值,既可以分析不同的产业结构对结构性失业的影响,也可以分析国家的宏观政策对结构失业的影响。还能清楚表明对经济政策变化的调整速度。

第三,面板数据的一个重要的优点是被认为能够控制个体的异质性。在面板数据中,人们认为不同的横截面很可能具有异质性,这个异质性被认为是无法用已知的回归元观测的,同时异质性被假定为依横截面不同而不同,但在不同时间点上却是稳定的,因此可以用横截面虚拟变量来控制横截面的异质性,此虚拟变量就是个体固定效应。如果异质性是发生在不同时期的,那么则用时期虚拟变量来控制,时期虚拟变量就是时间固定效应。而这些工作在只有横截面数据或时序数据时是无法完成的。对含固定效应的面板数据的研究不仅具有理论意义而且更有实际价值。

多年以来,统计学领域最具挑战性的议题,一直是如何创造一些方法以从非实验数据中进行有效的因果推断。而该议题最难的问题,是如何从统计上控制无法观测的变量。固定效应模型使得我们有可能对那些没有或无法被观测的变量进行控制。基本的思想其实非常的简单:用每个个体作为自身的控制(因素)。“固定效应模型”这一概念经常与“随机效应模型”形成对照。以传统的观点来看,固定效应模型将个体未被观察的差异作为一固定的参数,它们要么可以被直接估计出来,要么可以在估计方程中被抵消掉。而在随机效应模型中,未被观察的差异被处理成为具有特定概率分布的随机变量。在更新近的框架下(Wooldrige, 2002),未被观测的差异通常被当作随机变量。这时,将这两种模型区别开来的是已观测的变量与未被观测的变量之间的相关结构。在随机效应模型中,未被观测的变量被假定与所有观测变量之

间不相关(或者,更严格的说,在统计上相互独立)。而在固定效应模型中,允许未被观测的变量与任何已观测的变量之间存在相关(这实际上相当于将未被观测的变量当做固定的参数)。除非你允许这样的相关,否则就不能真正控制这些未被观测的变量的作用。正因为如此,固定效应模型才有吸引力。

半参数模型既含有参数部分又有非参数部分,其优点就在于它集中了主要部分(即参数分量)的信息,又不忽略干扰项(非参数分量)的作用,一方面克服了非参数方法所存在的“维数祸根”问题,另一方面解决单纯的非参数模型与参数模型难以解决的问题,增强了其适应性。目前,越来越多的学者开始利用该模型研究各类问题,如:环境污染、能源消耗以及经济发展等问题。在半参数回归模型中,现有大多数文献都是假设协变量能够直接观测的。而我们知道人们在收集数据的过程中经常会有所谓的测量误差(Errors-in-Variables, 简记为EV)产生。如当调查工资收入时,由于种种原因,人们往往不愿意把真实的工资告诉你,这时候所得到的数据就带有测量误差。测量误差可在包括经济学、流行病学、工程学等在内的几乎所有的应用领域出现,而这些误差在进行数据分析时常常被忽略。Fan和Truong(1993)研究了EV回归模型的非参数估计,Cui和Chen(2003)将经验似然方法应用到EV回归模型的统计推断。因此,研究半参数带固定效应面板模型在具有测量误差框架下的统计推断问题将更贴近于实际,更具有应用前景。

## 1.2 研究现状

在解决实际问题中,实际工作者和学者们提出许多类型的半参数回归模型,如部分线性回归模型、单指标模型、部分线性单指标模型、半参数变系数模型等。对于半参数回归模型研究不断向各方面发展,如对各种复杂数据,在纵向数据、面板数据、删失数据、测量误差数据和相依混合数据等情形下的研究都取得可喜的研究成果。

(1) 带固定效应面板数据的部分线性模型:Engel(1986)等提出部分线性模型,用于讨论关于气温与供电及电力价格关系提出来的,受到统计学家和经济学家广泛关注。其一般形式: $Y_i = Z_i^T \beta + g(U_i) + \epsilon$ ,  $i = 1, \dots, n$ . Heckman(1986), Speckman(1988), Chen(1988)分别基于不同的光滑思想研究线性模型的估计问题。Severiniti和Staniswalis(1994)研究了广义部分线性模

型；Zhang等(1998)利用该建模方法研究纵向数据；Wang与Jing(1999)利用经验似然方法考虑了具有固定设计的部分线性模型，构造了模型中未知参数的置信域；面板数据部分线性回归模型最初由Zeger和Diggle(1994)提出并被广大学者所研究，如Li和Ullah(1998)研究了部分线性面板数据模型。Xue和Zhu(2007a)将经验似然方法应用到纵向数据下的半参数回归模型；Qin, Zhu和Fung(2009)研究了纵向数据部分线性模型协方差参数的稳健估计；Bai, Fung和Zhu(2010)在纵向数据下给出了广义线性模型的加权经验似然推断，结果表明不论加权矩阵(Weight Matrix)如何选取，其提出的加权经验似然比总收敛于标准卡方分布；Qin, Bai和Zhu(2012)讨论了纵向数据下广义部分线性模型的稳健经验似然推断；Xue和Zhu(2006)构造纵向数据部分线性模型中未知参数的置信域和非参数的逐点置信带。You等(2006)考虑了纵向数据下半参数部分线性模型，提出了分组的经验对数似然比统计量。但是，对于部分带固定效应部分线性面板数据模型的经验似然很少有文献报道，这种问题的解决具有重要的理论意义和应用价值。

同时，对于协变量是高维时，Fan和Li(2001)提出惩罚似然的变量选择。对于高维线性模型，Tang和Leng(2010)提出惩罚经验似然，得到惩罚似然估计的oracle属性；然后应用惩罚似然比率测试假设和建立未知参数的有限维子集的置信区域。但是，对于部分带固定效应部分线性面板数据线性高维模型的统计推断还没有文献报道，同样考虑能否将惩罚经验似然方法应用带固定效应部分高维线性面板模型中去，这种问题的解决同样具有重要的理论和实践意义。

(2) 带固定效应面板数据半参数变系数模型：半参数变系数模型一般形式： $Y_i = Z_i^T \beta + X_i \alpha(U_i) + \varepsilon$ ,  $i = 1, \dots, n$ . 对于变系数建模思想早已存在，形式如下： $Y = \beta(u)X + \varepsilon$ , Hastie和Tibshirani(1993)提出利用局部加权以及光滑样条方法拟合变系数模型；对半参数回归模型的各种估计以及估计量的大样本性质系统的介绍可参见柴根象和洪圣岩的专著《半参数回归模型》(1995)，以及Härdle, Liang和Gao(2000)。如今，各种新的半参数变系数回归模型因具有比半参数回归模型更容易解释、应用性更广泛等特点而备受广大学者的青睐，见王启华，史宁中，耿直(2010)，Fan和Zhang(1999)基于局部多项式方法提出了两步估计；Cai, Fan和Li(2000)研究了广义变系数模型；Cai, Fan和Yao(2000)将该模型用到了非线性时间序列上；在纵向数据处理上，该模型

可以很好的揭示各协变量对因变量影响的变化趋势。如Xue和Zhu(2007a)将经验似然方法应用到纵向数据下的半参数回归模型；Zhang等(2002)，Fan和Huang利用剖面最小二乘估计参数并研究它的渐近正态性。Xue和Zhu(2007)将经验似然方法应用到纵向数据下的变系数回归模型；You和Zhou(2006)把经验似然方法应用于半参数变系数部分线性模型的统计推断。然而，对于带固定效应面板数据半参数变系数模型还没有文献报道。因此，研究带固定效应半参数变系数面板模型，以及协变量是误差变量时的估计问题更贴近实际生活。

(3) 带固定效应单指标部分线性面板数据模型：单指标部分线性模型一般形式： $Y_i = Z_i^T \beta + g(\theta_0^T X_i) + \varepsilon$ ,  $i = 1, \dots, n$ . 由于指标 $\theta_0^T X_i$ 合并了 $X$ 的维数，把多维协变量降维到一元指标，从而使单指标部分线性模型避免了多元非参数回归中出现的“维数灾祸”问题。该模型是在20世纪80年代中后期发展起来的，在金融经济和生物医学等领域具有广泛的应用背景。研究该模型一大任务就是对联系函数 $g(\cdot)$ 和指标系数 $\theta_0$ 进行估计，Ichimura(1993)和Härdle等(1993)提出了半参数最小二乘估计方法。该方法采用参数回归模型最小二乘估计思想，在指标系数给定的情况下，用去掉一点的N-W估计方法来估计联系函数 $g(\cdot)$ ，所得估计是指标系数 $\theta_0$ 的函数然后用非线性最小二乘思想极小化残差平方和，可得指标系数 $\theta_0$ 的估计。但计算较为复杂。Powell等(1989)使用了加权平均导数估计而且使用了局部常数核估计得到指标参数的条件均值的估计，进而得到指标参数 $\theta_0$ 的估计；Xue和Zhu(2006)利用局部线性方法和经验似然方法构造了单指标部分线性回归模型中指标参数的置信区域。Zhang, Huang和Lv(2010)利用局部线性方法研究了单指标模型中指标参数的广义似然比检验问题。Huang和Zhang(2011)给出了单指标模型参数中单个元素的有效经验似然推断。Huang(2012)在右删失数据下给出了单指标变系数模型指标参数的经验似然置信域。李高荣等(2010)分别研究纵向数据下单指标模型与纵向数据下线性单指标模型的经验似然置信域。然而，对于带固定效应单指标部分线性面板数据模型的经验似然还没有文献报道。因此，研究带固定效应单指标部分线性面板模型的经验似然，并研究协变量为误差变量时的估计问题具有重要的理论意义和实际应用价值。

(4) EV回归模型：过去20多年来，EV回归模型已被广大学者进行了大量的研究。比如Glessner(1990)，Fuller(1987)，Amemiya和Fuller(1984)等人利用

极大似然方法、广义最小二乘法给出EV回归中未知量的估计，并建立了估计量的大样本性质；Fan和Truong(1993)研究了EV回归模型的非参数估计；梁华等(1999)估计了半参数部分线性EV模型；Cui和Chen(2003)将经验似然方法应用到EV回归模型的统计推断；Cui和Kong(2006)将经验似然方法应用到部分线性EV回归模型的统计推断；Zhao和Xue(2009)在纵向数据下研究了半参数变系数部分线性EV回归模型的经验似然推断；Fan和Liang(2010)讨论了线性过程误差下半参数回归模型中未知参数的经验似然估计；Wang等(2011)研究了变系数部分线性EV回归模型的经验似然估计。但到目前为止，经验似然都很少运用在含误差变量带固定效应面板模型中。因此，在具有测量误差框架下研究半参数带固定效应面板模型的经验似然估计问题将更贴近于实际，更具有应用前景。

## 1.3 本文的工作

### 1.3.1 研究方法

#### 一、经验似然方法(Empirical Likelihood, 简记为EL)

经验似然是Owen(1988)在独立同分布样本下对总体均值提出的一种非参数统计推断方法，经验似然方法的本质是在约束条件下求非参数似然比的极大值。它有类似于Bootstrap的抽样特性。该方法的提出主要是基于经验分布函数是样本所在分布的非参数极大似然估计这一点。这实际上是Wilks(1938)的参数似然比理论的非参数扩展版本。这一方法与经典的或现代的统计方法相比较有很多突出的优点，如：用经验似然方法构造未知参数的置信区域时，除有域保持性、变换不变性和置信域的形状由数据自行决定等诸多优点外，还有Bartlett纠偏性及无需构造轴统计量等优点。此外，应用经验似然进行统计推断不需要估计方差(协方差)，而方差(协方差)的估计通常是一个不容易的问题。因此经验似然方法成为众多统计学家和经济学家研究的热点。目前已经有大量的文献使用这一方法研究实际问题，具体可参见Owen(2001)以及王启华(2004)的综述。下面我们介绍一下什么是经验似然。

设 $X_1, X_2, \dots, X_n$ 是 $R^d$ 中独立同分布(i.i.d.)的样本，具有共同的分布函

数 $F$ ,那么 $F$ 的非参数似然是

$$L(F) = \prod_{i=1}^n F(\{X_i\}) = \prod_{i=1}^n \{F(X_i) - F(X_{i-})\},$$

这里 $F(\{X_i\})$ 是分布 $F$ 在 $X_i$ 处的概率质量,  $i = 1, \dots, n$ . 经验分布函数 $F_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ 使上式达到最大, 其中 $\delta_x(A) = I[x \in A]$ . 也就是说 $F_n$ 是 $F$ 的非参数极大似然估计. 在参数推断中, 人们利用参数似然比进行假设检验与置信区间估计. 类似地, 在分布完全未知的情况下非参数似然比

$$R(F) = L(F)/L(F_n)$$

也可以用于统计推断. 不像参数似然比, 非参数似然比中不包含未知参数, 一个自然的问题是如何使用它对参数作统计推断. 注意到一些参数 $\theta$ 是总体分布的泛函, 即 $\theta = T(F) \in R^p$ , 其中 $T(\cdot)$ 是分布 $F$ 的某泛函,  $F$ 属于某分布类 $\mathcal{F}$ , 如总体均值及分位点等就是有上述形式泛函的例子. 为了对 $T(F) = \theta$ 作检验, Owen(1988)定义如下经验似然比检验统计量

$$\mathcal{R}(\theta) = \sup_F \{R(F) | T(F) = \theta, F \in (F)\}.$$

很显然, 经验似然比实际上是一种截面非参数似然比函数, 它要求 $F$ 在满足约束条件 $T(F) = \theta$ 下使非参数似然比达到极大(在无约束条件下, 极大非参数似然比是1, 只需取 $F = F_n$ 即可). 而参数 $\theta$ 由这一约束条件引入这一极大似然比中, 从而得到关于参数 $\theta$ 的极大截面非参数似然比函数, 用这一非参数似然比作假设检验、区间估计或进行其它统计推断, 这一方法就是所谓的经验似然方法. 如果 $\mathcal{R}_\theta < r_0$ , 经验似然假设检验拒绝 $H_0: T(F) = \theta_0$ , 而经验似然置信域为

$$\{\theta: \theta \geq r_0\},$$

其中 $r_0$ 是某临界值. 注意到 $R(\theta)$ 是关于非参数似然比在限制条件下关于 $F$ 求极大, 这种形式的定义显然不利于进行具体的计算, 考虑到这一点, 需要给出与上面定义等价但容易计算的形式. 由 $R(\theta)$ 的定义, 注意到只有那些在观察点有正概率质量的离散分布才有可能使似然比达到极大, 于是为求 $R(\theta)$ , 我们只需要限制在这样的离散分布类中求极大: 首先假设数据没有“结”(当有“结”时, 根据Owen(1988), 所求的经验似然是相同的), 在没有“结”时, 我们设分布 $F$ 在观察值 $X_l$ 有概率质量 $p_l > 0$ ,  $l = 1, 2, \dots, n$ , 并记这样的分布函数 $F$ 为 $F_p$ , 即 $F_p = \sum_{i=1}^n p_i \delta_{X_i}$ . 注意到 $L(F_n) = n^{-n}$ , 因而, 计算 $R(\theta)$ 就成为



关于 $p_i, i = \dots, n$ 求极大, 即

$$\theta = \sup \left\{ \prod_{i=1}^n (np_i) \mid \sum_{i=1}^n p_i = 1, T(F_p) = \theta \right\},$$

这是带有约束条件的极值问题, 可以应用Lagrange方法计算 $R(\theta)$ 。

经验似然应用于推断的另一个问题是如何确定临界值 $r_0$ , 这样才方便我们构造未知参数的置信区间和做出估计。这一问题实际上归结为求 $R(\theta)$ 的渐近分布。我们知道参数似然比的一个重要特性是似然比的对数是渐近卡方分布的, 即所谓的Wilks定理, 幸运的是经验似然比也有相同的特性, 这一特性形成了经验似然统计推断的基础。

经验似然方法已被众多学者广泛应用于线性回归模型, 广义线性模型, 部分线性回归模型, 非参数回归等等。最近, Cui 和Kong (2006) 及Li 和Xue (2008) 在误差为独立情况下将经验似然应用到部分线性EV回归模型中。Li 等(2008)及Xue 和Zhu(2008) 将经验似然方法应用到误差为独立假设时的纵向数据部分线性回归模型中。

## 二、惩罚方法

变量选择是现代统计分析的重要课题。传统的变量选择方法又称为子集选择方法(subset selection), 即对协变量集合所有子集进行比较分析, 通过假设检验或诸如AIC和BIC信息准则, 从中选出一个最优的子集拟合回归模型。这些方法的一个共性是根据已有样本用一个准则选择出变量子集, 然后再基于这一样本来估计回归参数。但是, 由于真实的相关变量往往是不知道的, 而选择变量过程也会产生一些偏差。很难评价最终模型中回归系数估计的精度。另外, 当协变量的维数 $p$ 较大时, 对所有的 $2^p - 1$ 个子集进行假设检验, 计算量相当的大, 而实际应用也难实现。因此, 利用惩罚方法进行变量选择受到统计学家的重视。惩罚方法基本思想就是在进行参数估计的同时, 把较小的估计系数压缩为0, 从而达到变量选择的目的。而该方法可以进行参数估计和变量选择, 从而大大减少计算量, 并且克服了传统变量选择的不稳定性。惩罚方法广泛地应用于各种统计模型的变量选择问题。尤其在生存分析等领域, 这个优点更加突出。目前系数压缩变量选择方法研究已经开始从最初变量个数固定的线性模型转为变量个数发散的高维复杂统计模型。

惩罚函数: 假设 $p$ 维协变量 $X_i$ 中有些分量不是统计显著的, 模型参数 $\beta$

是 $p$ 维的,  $l(\beta)$  为给定的二次损失函数. 惩罚最小二乘定义为

$$\mathcal{L}(\beta) = l(\beta) + n \sum_{i=1}^p \lambda_i p_i(|\beta_i|),$$

其中 $p_i(\cdot)$  为惩罚函数,  $\lambda_i$  为调整参数, 它是用来控制模型的复杂度, 并且可以用诸如交叉验证 (CV) 的数据方法来选择. 通过极小化 $\mathcal{L}(\beta)$ , 一些系数被估为0, 那么删除相应的变量, 否则保留该变量. 这个方法可以同时选择变量和估计系数. 所得到的估计量称为惩罚最小二乘估计. 惩罚函数 $p_i(\cdot)$  和调整参数 $\lambda_i$  不必要对所有的 $i$  相同. 这就允许我们通过使用不同的惩罚函数或去 $\lambda_i$  的不同值来合并未知系数的先验信息. 为表达方便, 假定惩罚函数和调整参数对所有系数都相同.

许多惩罚函数已经被用到各种参数模型的惩罚最小二乘和惩罚似然之中. 例如,  $L_q (q \geq 0)$  惩罚类:  $L_0$  惩罚 $p_\lambda(|\theta|) = 0.5\lambda^2 I(|\theta| \neq 0)$  相当于熵惩罚;  $L_1$  惩罚 $p_\lambda(|\theta|) = \lambda|\theta|$  导致Tibshirani (1996) 提出的LASSO惩罚;  $L_2$  惩罚 $p_\lambda(|\theta|) = \lambda|\theta|^2$  叫岭回归; 桥回归 (Frank and Friedman, 1993) 相当于 $p_\lambda(|\theta|) = \lambda|\theta|^q (0 < q < 1)$ . Antoniadis (1997) 考虑了硬门限惩罚函数 $p_\lambda(|\theta|) = \lambda|\theta|^2 - (|\theta| - \lambda)^2 I(|\theta| \leq \lambda)$ .

在惩罚最小二乘中, 一个关键的问题是如何选择惩罚函数. Antoniadis 和Fan (2001) 与Fan 和Li (2001) 对如何选择惩罚函数提出深刻的见解. 他们主张一个好的惩罚函数将导致具有三个性质的估计量.

(1) 无偏性: 当真参数很大时, 所的估计量几乎无偏的, 以避免不必要的建模误差.

(2) 稀疏性: 所得到的估计量是一个门限值, 自动将小的估计系数设置为0, 以减少模型复杂性.

(3) 连续性: 所得到的估计量是数据点处是连续的, 以避免模型预测中的不稳定性.

Fan 和Li (2001) 讨论了具有上述性质的惩罚函数所满足的条件, 得到结论: 大的真参数估计具有无偏性的充要条件是对大的 $|\theta|$  有 $p'_\lambda(|\theta|) = 0$ ; 具有稀疏性的充分条件是 $\min_{\theta} \{|\theta| + p'_\lambda(|\theta|)\} > 0$ ; 具有连续性的条件是 $\arg \min_{\theta} \{|\theta| + p'_\lambda(|\theta|)\} > 0$ .  $L_q (q \geq 0)$  惩罚都不满足上述三个性质. 为克服上述方法缺陷, Fan 和Li (2001) 提出SCAD惩罚方法, 该方法得到的估计量具有无偏性, 稀疏性和连

续性. SCAD惩罚函数定义为

$$p_{\lambda}(|\theta|) = \begin{cases} \lambda|\theta|, & |\theta| \leq \lambda \\ \frac{(a^2-1)\lambda^2 - (|\theta| - a\lambda)^2}{2(a-1)}, & \lambda < |\theta| \leq a\lambda \\ \frac{1}{2}(a+1)\lambda^2, & |\theta| > \lambda \end{cases}$$

由于SCAD变量选择方法具有很好的性质, 目前该方法已被广泛使用于半参数模型的变量选择之中。

### 三、剖面似然方法 (Profile Likelihood)

剖面似然方法在许多半参数模型中得到广泛使用。其基本思想如下: 固定模型中的有限维参数, 我们可以估计模型中的非参数部分, 把它表达成有限维参数的函数形式; 再把上述非参数部分估计量代入一个有待最小(或最大)化的目标函数, 并求出其最优解。这样一来, 我们便可估计出我们所感兴趣的所有参数或非参数部分。如果在第一步中使用局部线性估计, 我们还可得到对非参数部分的一阶导数的一致性估计。

### 1.3.2 研究内容和结构安排

本文研究了带固定效应的半参数面板数据模型的参数估计问题, 主要考虑下面四种模型的经验似然:

第一, 带固定效应面板数据部分线性模型, 模型为:

$$Y_{it} = X_{it}^T \beta + g(U_{it}) + v_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

其中 $Y_{it}$ 是响应变量,  $(X_{it}^T, U_{it}) \in R^p \times R$ 是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^T$ 是一个 $p$ -维未知参数, 上角 $\tau$ 表示一个向量的转置,  $g(U_{it}) = (g_1(U_{it}), \dots, g_q(U_{it}))^T$ 是一个 $q$ -维未知参数,  $v_i$ 是一个不可观察的个体效应。表示 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$ 是 $i$ 个体随机误差向量和 $\{\varepsilon_i, i = 1, \dots, n\}$ 是相互独立且 $E(\varepsilon_i | X_{it}, U_{it}) = 0$ 。我们利用局部多项式函数估计非参数部分, 建立未知参数的经验似然, 获得他们的渐近分布, 并证明参数估计的相合性和收敛性, 并进行蒙特卡洛模拟和探索在生活实际中的应用。

第二, 带固定效应面板数据部分线性误差变量模型, 模型为:

$$\begin{cases} Y_{it} = X_{it}^T \beta + m(Z_{it}) + \mu_i + \varepsilon_{it}, \\ W_{it} = X_{it} + \nu_{it}, \end{cases} \quad i = 1, \dots, n; \quad t = 1, \dots, T.$$

其中 $Y_{it}$ 是响应变量,  $(X_{it}^T, Z_{it}) \in R^p \times R$ 是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^T$ 是一个 $p$ -维未知参数, 上角 $\tau$ 表示一个向量的转置,  $m(Z_{it}) = (m_1(Z_{it}), \dots, m_q(Z_{it}))^T$ 是一个 $q$ -维未知参数,  $\mu_i$ 是一个不可观察的个体效应。表示 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$ 是 $i$ 个体随机误差向量和 $\{\varepsilon_i, i = 1, \dots, n\}$ 是相互独立且 $E(\varepsilon_i | X_{it}, Z_{it}) = 0$ . 协变量 $X_{it}$ 被测量伴随测量误差且不能直接被观察到. 代替 $X_{it}$ 我们观察到 $W_{it}$ . 我们建议一个修改的剖面最小二乘估计, 由此产生的估计是一致的和渐近正态。一致估计的误差的方差也能被考虑。

同时我们还考虑在协变量是高维的时候, 我们将利用惩罚似然方法进行变量选择。首先, 利用EL方法构建未知参数的置信区域并建立最大经验的渐近正态性参数的似然 (MEL) 估计; 其次, 采用PEL方法建立简约和鲁棒的模型获取PEL估计的oracle属性; 最后, 应用PEL比率测试假设和建立未知参数的有限维子集的置信区域。

第三, 带固定效应半参数变系数面板数据模型, 模型为

$$Y_{it} = Z_{it}^T \beta + X_{it}^T \alpha(U_{it}) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

其中 $Y_{it}$ 是响应变量,  $(Z_{it}, X_{it}) \in R^p \times R^q$ 和 $U_{it} \in R$ 是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^T$ 是一个 $p$ -维未知参数, 和上标 $\tau$ 表示一个向量的转置.  $\alpha(U_{it}) = (\alpha_1(U_{it}), \dots, \alpha_q(U_{it}))^T$ 是一个 $q$ -维向量的未知函数和 $\mu_i$ 是不能观察的个体效应,  $v_{it}$ 是随机误差。这儿, 我们假定 $\mu_i$ 是独立同分布的均值为0和有限的方差 $\sigma^2 > 0$ . 我们允许 $\mu_i$ 是相关的与 $Z_{it}, X_{it}$ , 和 $U_{it}$ 伴随一个的相关结构。我们将研究数据是 $\alpha$ -混合的情况, 利用Fan和Hang (2005) 提出剖面最小二乘估计和局部多项式估计, 建立未知参数的经验似然, 获得他们的渐近分布, 利用经验似然方法得到未知参数的置信域。建立参数的置信域, 并进行数据模拟和探索在实际中的应用。

第四, 带固定效应面板数据部分线性时变系数模型, 模型为

$$Y_{it} = Z_{it}^T \beta + X_{it}^T \alpha(t/T) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

其中 $Y_{it}$ 是响应变量,  $(Z_{it}, X_{it}) \in R^p \times R^q$ ,  $\beta = (\beta_1, \dots, \beta_p)^T$ 是一个 $p$ -维未知参数, 和上标 $\tau$ 表示一个向量的转置.  $\alpha(t/T) = (\alpha_1(t/T), \dots, \alpha_q(t/T))^T$ 是一个 $q$ -维向量的未知函数和 $\mu_i$ 是不能观察的个体效应,  $v_{it}$ 是随机误差。这儿, 我们假定 $\mu_i$ 是独立同分布的均值为0和有限的方差 $\sigma^2 > 0$ . 我们允许 $\mu_i$ 是相关的

与 $Z_{it}$ ,  $X_{it}$  伴随一个的相关结构。我们试利用剖面最小二乘估计和局部多项式估计, 建立未知参数的经验似然, 建立未知参数和非参数部分的估计, 获得他们的渐近分布, 利用经验似然方法得到未知参数的置信域。

全文由七章组成, 文章的内容安排如下:

第二章, 我们研究带固定效应部分线性面板数据模型的参数估计, 为了解决组内相关性, 我们应用块经验似然方法, 并构造了未知参数的经验对数似然比统计量。在一定条件下, 证明了调整的经验对数似然比函数在真参数点时是渐近卡方分布的。该结果可以用来构造模型中未知参数的置信域。模拟研究置信域构造方面我们提出的经验似然方法优于正态逼近方法。并且利用艾滋病感染者CD4数据实际验证我们建议的方法更合适。

第三章, 我们考虑带固定效应面板数据部分线性误差变量模型的统计推断, 我们专注于一些协变量为附加误差情况下, 我们提出了一个修正的剖面最小二乘估计参数和非参数部分, 参数部分的渐近正态性和非参数部分的收敛速度被建立。一致估计的误差方差也被估计。此外, 我们介绍的剖面似然比(PLR)检验并证明它在原假设条件下是一个渐进的卡方分布。我们进行模拟研究, 证明了我们所提出的方法更优于不考虑误差变量情况下的估计。我们也用英国联邦的气候数据验证我们建议的方法更合适。

第四章, 对于带固定效应面板数据的高维部分线性误差变量模型, 当误差方差为已知和未知情形下. 在这一章, 我们建议一个修正的回归参数和最大经验似然比率回归参数。同时, 依据惩罚经验似然方法, 这个模型的参数估计和变量选择被调查, 我们建议的惩罚经验似然被证明具有神(oracle)特征, 同时, 我们建议的惩罚经验似然比率统计在原假设下具有渐近卡方分布, 其结果可以用来构造未知参数的置信域。模拟结果用来评估惩罚经验似然方法的性能。

第五章, 考虑带固定效应面板数据的半变系数模型的经验似然, 提出了在协变量是 $\alpha$ -混合下情形下, 未知参数的经验对数似然比统计量, 证明了建议的对数似然比统计量是渐近卡方分布的, 我们也获得参数的最大似然估计量和证明在合适条件是渐近正态的。由此, 未知参数两个不同的渐近正确覆盖率的置信域被构造。模拟和实证结果用来评估经验似然方法优于正态逼近方法。

第六章, 我们研究带固定效应面板数据部分线性时变系数模型的经验似然, 提出了在协变量是 $\alpha$ -混合下情形下, 未知参数的经验对数似然比统计量,

证明了建议的对数似然比统计量是渐近卡方分布的,我们也获得参数的最大似然估计量和证明在合适条件是渐近正态的。由此,未知参数两个不同的渐近正确覆盖率的置信域被构造。模拟和实证结果用来评估经验似然方法的优越性。

第七章是本文的主要结论展望。简单总结全文的研究工作和主要的创新点,并指出需要进一步完善和深入研究的问题。

### 1.3.3 创新之处

一、将经验似然方法应用于带固定效应面板数据线性回归模型中,这是一个基础性强且比较前沿的课题。现有文献还很少有用经验似然方法考虑带固定效应面板数据线性回归模型,对这一问题的解决,将有助于丰富统计理论的内容,并且为这些模型在计量经济学等领域中的应用奠定基础。我们应用块经验似然方法,构造了未知参数的调整经验对数似然比统计量。在一定条件下,证明了调整的经验对数似然比函数是渐近卡方分布的,该结果可以用来构造模型中未知参数的置信域。

二、研究协变量有测量误差时带固定效应面板数据线性回归模型的统计推断问题,现有的文献还很少涉及到这方面的研究。由于带固定效应面板数据模型带测量误差,使得实际观察数据之间的关系不容易得到,由于模型的复杂性,建立它们的统计推断问题将变得非常困难,这是一个具有挑战性的研究课题。我们现用局部多项式方法估计非参数部分,然后提出了一个修正的剖面最小二乘估计参数,参数部分的渐近正态性和非参数部分的收敛速度被建立。一致估计的误差方差也被估计。此外,我们介绍的剖面似然比(PLR)检验在原假设条件下是一个渐进的卡方分布。

三、带固定效应面板数据线性回归模型的变量是高维时,怎样进行变量选择,探索利用惩罚经验似然解决高维带固定效应面板数据半参数回归模型,比较与一般的高维的半参数回归模型的区别,我们将惩罚经验似然运用带固定效应面板数据半参数回归模型的变量选择问题。证明了在一定条件下,惩罚似然估计具有参数高维时发散时的最优收敛速度,同时,该估计具有Oracle性质。最后,通过数据模拟和实例分析对估计的有限样本表现进行验证。

四、将经验似然方法应用于带固定效应面板数据半变系数数回归模型中,考虑在 $\alpha$ -混合条件下带固定效应的面板数据的半变系数模型,由于 $\alpha$ -混合的

广泛实用性及应用背景,且在诸多条件下是最弱的。我们先用局部多项式估计非参数部分,然后提出未知参数的修正的对数似然比统计量,证明了建议的对数似然比统计量是渐近卡方分布的。我们首次研究了在 $\alpha$ -混合条件下带固定效应的面板数据的半变系数模型的经验似然。同样该结果可以用来构造模型中未知参数的置信域。

五、将经验似然方法应用于带固定效应面板数据线性时变系数数回归模型中,时变系数模型已经获得了很多关注,因为它们在经济和金融方面有许多应用。我们首次研究了在 $\alpha$ -混合条件下带固定效应的面板数据线性时变系数数模型的经验似然。假设在 $\alpha$ -混合条件下,我们将对时变函数部分进行局部多项式估计,然后提出参数的修正对数似然比统计量,证明了建议的修正对数似然比统计量是渐近卡方分布的。

## 第2章 带固定效应部分线性面板数据模型的块经验似然

### 2.1 引言

面板数据分析得到极大关注在最近几十年, 由于面板数据可以应用于许多学科, 比如经济、金融、生物、工程和社会科学。因为面板数据存在两个维度: 时间和横截面, 可以让研究者估计复杂模型, 那是很难完成单纯依靠截面数据或者时间序列数据。已经有大量的关于参数与非参数面板数据模型的著作。如关于参数面板数据模型的统计推断和经济分析, 我们参考Baltagi (2005) 和Hsiao (2003)的书。为了避免对于参数面板数据的强约束性, 一些非参数方法已经用在面板数据的参数估计和假设检验(例如, Hjellvik 等, 2004; Henderson 和Ullah, 2008; Cai 和Li, 2008)。但非参数用于探索隐藏结构和减少模型误差, 它们是大灵活, 以至不能得到精确的结论。当有大量的协变量时, 非参数估计要面对“维数灾祸”, 为克服这些缺点, 我们利用半参数方法, 那融合了非参数方法与参数方法。

考虑下面的带固定效应的部分线性面板数据模型:

$$Y_{it} = X_{it}^T \beta + g(U_{it}) + v_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2-1)$$

其中 $Y_{it}$  是响应变量,  $(X_{it}^T, U_{it}) \in R^p \times R$  是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^T$  是一个 $p$ -维未知参数, 上角 $\tau$  表示一个向量的转置,  $g(U_{it}) = (g_1(U_{it}), \dots, g_q(U_{it}))^T$  是一个 $q$ -维未知参数,  $v_i$  是一个不可观察的个体效应。表示 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$  是 $i$ 个体随机误差向量和 $\{\varepsilon_i, i = 1, \dots, n\}$  是相互独立且 $E(\varepsilon_i | X_{it}, U_{it}) = 0$ 。模型(2-1)被叫固定效应模型, 如果 $v_i$  是与 $X_{it}$  和(或) $U_{it}$  相关的关于一个未知的相关结构。模型(2-1)被叫随机效应模型, 如果 $v_i$  与 $X_{it}$  和 $U_{it}$  是不相关的。固定效应的样本稳定性要优于随机效应样本(Baltagi, 2005; Horowitz 和Lee, 2004 等)。同时, 分析带固定效应的面板数据具有更多的挑战, 因为样本增长的参数, 容易产生著名的Neyman 和Scott (1948) 问题。在此, 我们只考虑固定效应情况。

显然, 模型(2-1)包括许多常用的参数, 非参数和半参数回归模型。例如, 当 $v_i = 0$ , 模型(2-1)缩减为部分线性面板数据模型。许多研究者已经研究了



部分线性面板数据模型(Roy, 1996; Li 和Ullah, 1998; You 和Zhou, 2006等)。Roy (1996) 已经应用部分线性面板数据模型去研究了卡路里与收入的关系, 他利用的是印度南部的农村两年的面板数据。当 $v_i = 0$  且 $g = 0$ , 模型(2-1) 转变为参数面板数据模型, 这种模型已经广泛应用于经济学中(Ahn 和Schmidt, 2000; Hsiao, 2003)。当 $v_i = 0$  且 $\beta = 0$ , 这模型转变为非参数面板数据模型, 那已经被Ruckstuhl 等(2000)等研究。

对于部分线性面板数据模型, Li 和Ullah (1998) 建立一可行的半参数广义最小二乘方法估计参数和推导了这个建议的估计参数的渐进正态性。对于模型(2-1), Su 和Ullah (2006) 采用一个局部线性哑变量方法去消除未知的固定效应。在这一章, 我们推导带固定效应的面板数据部分线性模型的参数 $\beta$  的统计推断。顺从Li 和Ullah (1998), Su 和Ullah (2006)建议估计方法, 这个 $\beta$  的最小二乘估计能被建立。依据这个, 我们能建立估计参数的一个正太逼近的置信区间。但是建立这样的置信区间不是很简单的, 因为包括估计这个参数复杂渐进协方差。为解决这个问题, 我们建议利用经验似然方法去建立未知参数 $\beta$  的置信区间, 根据经验似然方法, 我们能很快建立这个未知回归参数渐进置信区间。我们写这一章的一个动机是经验似然方法做推断时不需要得到这个参数的渐进协方差矩阵, 那是更加复杂为带固定效应的面板数据的部分线性模型。我们写这一章的另一个动机是依据经验似然方法建立的置信区间不需要预先强加区间形状的约束, 且我们估计的置信区间形状与方向完全由数据决定。因此, 经验似然方法已经被许多研究者运用, 比如Shi 和Lau (2000), Wang 和Jing (1999), You 和Zhou (2006), Fan 和Liang (2012) 等等。

然而, 我们通常用的经验似然方法不能被应用于带固定效应的面板数据的部分线性模型, 由于存在组内相关性。对于解决组内相关性的问题, 我们将利用You 等(2006)建议的块经验似然对于模型(2-1), 建立一个块经验对数似然比率为这个参数, 和推导一个非参数版本的Wilks 定理, 我们可以用来建立估计参数的块经验似然置信区间。

剩下的这一章组成是这样的: 2.2节介绍方法和建立参数 $\beta$  的经验对数似然函数的比率。假定和主要结论被给在2.3节, 一些模拟和实证安排在2.4节, 主要结果的证明安排在2.5节。

## 2.2 模型和方法

为介绍我们的估计, 我们假定模型(2-1) 在这个约束条件下  $\sum_{i=1}^n v_i = 0$ 。令  $v = (v_2, \dots, v_n)^\tau$  和  $v_0 = (-\sum_{i=2}^n v_i, v_2, \dots, v_n)^\tau$ , 我们写模型(2-1)的矩阵形式如下

$$Y = X\beta + g(U) + Mv + \varepsilon, \quad (2-2)$$

其中  $M = [-i_{n-1} \quad I_{n-1}]^\tau \otimes i_T$  是一个  $nT \times (n-1)$  矩阵,  $I_n$  表示  $n \times n$  单位矩阵,  $i_n$  表示  $n \times 1$  数字1的向量。有许多方法估计  $\{\beta_j, j = 1, \dots, p\}$  和这个函数  $\{g_i(\cdot), i = 1, \dots, q\}$ 。我们方法的主要思想是剖面最小二乘估计方法被建议Fan 和Huang (2005): 假定我们有一个随机样本  $\{(U_{it}, X_{it1}, \dots, X_{itp}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T\}$  来自模型(2-1)。令  $\theta = (v^\tau, \beta^\tau)^\tau$ 。指定  $\theta$ , 我们能够运用一个局部线性回归技术估计非参数部分  $\{g_j(\cdot), j = 1, \dots, q\}$  在模型(2-2), 对于  $U_{it}$  是在  $u$  的一个小领域内, 我们能估计  $g_j(U_{it})$  局部线性函数

$$g_j(U_{it}) \approx g_j(u) + g'_j(u)(U_{it} - u) \equiv a_j + b_j(U_{it} - u), \quad j = 1, \dots, q,$$

其中  $g'_j(u) = \partial g_j(u) / \partial u$ 。这导致下面的权局部最小二乘问题: 发现  $\{(a_j, b_j), j = 1, \dots, q\}$  去最小化

$$\sum_{i=1}^n \sum_{t=1}^T \left\{ \left( Y_{it} - X_{it}^\tau \beta - v_i \right) - \sum_{j=1}^q \left[ a_j + b_j(U_{it} - u) \right] \right\}^2 K_h(U_{it} - u), \quad (2-3)$$

其中  $K_h(\cdot) = K(\cdot/h)/h$ ,  $K(\cdot)$  是一个核函数和  $h$  是一系列接近于0的正数, 叫窗宽。令

$$X = \begin{pmatrix} X_{11}^\tau \\ \vdots \\ X_{1T}^\tau \\ \vdots \\ X_{nT}^\tau \end{pmatrix}, \quad D_u = \begin{pmatrix} 1 & \frac{U_{11}-u}{h} \\ \vdots & \vdots \\ 1 & \frac{U_{1T}-u}{h} \\ \vdots & \vdots \\ 1 & \frac{U_{nT}-u}{h} \end{pmatrix}, \quad Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ \vdots \\ Y_{nT} \end{pmatrix},$$

和  $W_u = \text{diag}(K_h(U_{11} - u), \dots, K_h(U_{1T} - u), \dots, K_h(U_{nT} - u))$ 。

简单计算得到

$$(\hat{g}_1(u), \dots, \hat{g}_q(u), h\hat{g}'_1(u), \dots, h\hat{g}'_q(u))^\tau = (D_u^\tau W_u D_u)^{-1} D_u^\tau W_u (Y - X\beta - \lambda v),$$

因此我们能估计 $g(u)$  通过

$$\hat{g}(u, \beta) = (I_q, 0_{q \times q})(D_u^r W_u D_u)^{-1} D_u^r W_u (Y - X\beta - Mv).$$

令 $N = nT$ ,  $\Omega = (I_q, 0_{q \times q})(D_u^r W_u D_u)^{-1} D_u^r W_u$ , 和 $\Omega$  的第 $[(i-1)T + t]$ 个元素是 $\omega_{it}(u)$ 。因此通过Chiou 和Müller (1999), 我们得到

$$\omega_{it}(u) = \frac{\frac{1}{N} K_h(U_{it} - u) [A_{N2}(u) - (U_{it} - u) A_{N1}(u)]}{A_{N0}(u) A_{N2}(u) - A_{N1}^2(u)},$$

其中 $A_{Ns}(u) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T K_h(U_{it} - u) (U_{it} - u)^s$ ,  $s = 0, 1, 2$ .

现在我们考虑一种方法去消除固定效应, 在参数面板数据分析哑变量模型中, 通过一个最小平方和, 即我们解决下面的最优问题:

$$\hat{\theta} = \arg \min_{\theta} [Y - X\beta - Mv - \Omega(Y - X\beta - Mv)]^r [Y - X\beta - Mv - \Omega(Y - X\beta - Mv)], \quad (2-4)$$

假定 $\tilde{X} = (I_{nT} - \Omega)X$ ,  $\tilde{Y} = (I_{nT} - \Omega)Y$ ,  $\tilde{M} = (I_{nT} - \Omega)M$ , 我们得到 $\tilde{v} = (\tilde{M}^r \tilde{M})^{-1} \tilde{M}^r (\tilde{Y} - \tilde{X}\beta)$ . 令 $\tilde{H} = I_{nT} - \tilde{M}(\tilde{M}^r \tilde{M})^{-1} \tilde{M}^r$ , 我们能得到 $\tilde{H}\tilde{M}v = 0$ . 因此这个固定效应项 $Mv$  被消除在模型(2-4)。令 $e_{it}$  是这 $nT \times 1$  向量, 其中第 $\{(i-1)T + t\}$ 个元素是1 和其它是0。我们表示渐近误差项作为下面的:

$$\hat{r}_{it}(\beta) = e_{it}^r \tilde{H}(\tilde{Y}_{it} - \tilde{X}_{it}^r \beta).$$

相似Owen (1990) 和Shi 和Lau (2000), 我们能处理 $\{\hat{r}_{it}(\beta), i = 1, \dots, n; t = 1, \dots, T\}$  作为一个随机筛选渐近随机误差序列 $\{\varepsilon_{it}, i = 1, \dots, n; t = 1, \dots, T\}$ 。为了处置组内相关性, 我们利用被You 等(2006)建议的块经验似然方法, 相似于You 等(2006), 这剖面似然估计 $\beta$  被定义如下:

$$\mathcal{L}_n(\beta) = \sup \left\{ \prod_{i=1}^n n p_i \left| \sum_{i=1}^n p_i \sum_{t=1}^T \tilde{X}_{it} e_{it}^r \tilde{H}(\tilde{Y}_{it} - \tilde{X}_{it}^r \beta) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right. \right\}. \quad (2-5)$$

对于给定的 $\beta$ ,  $\mathcal{L}_n$  存在最大值, 假如0 是在点 $\psi_i(\beta) = \sum_{t=1}^T \tilde{X}_{it} e_{it}^r \tilde{H}(\tilde{Y}_{it} - \tilde{X}_{it}^r \beta)$ ,  $i = 1, \dots, n$  的凸包里, 经过拉格朗日算法, 式子(2-5)最大值能被发现。最优的值 $p_i$  满足式(2-5) 能满足 $p_i = \frac{1}{n[1 + \lambda^r \psi_i(\beta)]}$ 。我们定义相应的剖面块经验对数似然比率为

$$\mathcal{LR}_n(\beta) = - \sum_{i=1}^n \log \{1 + \lambda^r \psi_i(\beta)\}, \quad (2-6)$$

其中 $\lambda(\beta)$ 被下式决定

$$\frac{1}{n} \sum_{i=1}^n \frac{\psi_i(\beta)}{1 + \lambda^\tau \psi_i(\beta)} = 0. \quad (2-7)$$

我们现在如果设 $\beta$ 是正的向量参数,  $\mathcal{LR}_n(\beta)$ 是渐近卡方分布。最大化 $\{\mathcal{LR}_n(\beta)\}$ 能获得最大块经验似然估计 $\hat{\beta}$ 的 $\beta$ 。通过Qin和Lawless (1994),  $\hat{\beta}$ 也等价于解决这个估计方程

$$\frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^\tau e_{it}^\tau \tilde{H} (\tilde{Y}_{it} - \tilde{X}_{it}^\tau \beta) = 0. \quad (2-8)$$

这个估计方程的解式(2-8)用通常的剖面最小二乘估计。因此, 这最大块经验似然估计 $\hat{\beta}$ 是等价于剖面最小二乘估计, 那被指定于

$$\hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^\tau e_{it}^\tau \tilde{H} \tilde{X}_{it} \right)^{-1} \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^\tau e_{it}^\tau \tilde{H} \tilde{Y}_{it}.$$

## 2.3 假定和主要结论

在这一章, 我们将显示如果 $\beta$ 是正参数向量,  $\mathcal{LR}_n(\beta)$ 是渐近卡方分布。在得到主要结论前, 我们给出下面的假定

(A1) 随机变量 $U_{it}$ 有一个连续密度函数 $f(\cdot)$ 伴随一个紧支撑 $\mathcal{U}$ 在 $\mathbb{R}$ 上,  $0 < \inf_{u \in \mathcal{U}} f(u) \leq \sup_{u \in \mathcal{U}} f(u) < \infty$ 。

(A2) 令 $m(u) = E(X_{it}|U_{it} = u)$ . 函数 $m(u)$ 和 $g(u)$ 是二级连续可微函数在 $\mathcal{U}$ 上。

(A3)  $(v_i, X_i, U_i, \varepsilon_i)$ ,  $i = 1, \dots, n$ , 是i.i.d. 存在一些数 $\delta > 2$ 以至 $E\|X_{it}\|^{2+\delta} < \infty$ 和 $E|\varepsilon_i|^{2+\delta} < \infty$ , 其中 $\|a\| = (\sum_{i=1}^n a_i^2)^{1/2}$ 是通常的Euclidean norm的一个向量 $a$ 。

(A4) 核函数 $K(v)$ 是一个在连续可微的对称概率密度函数在紧支撑 $\mathcal{U}$ 上。

(A5)  $E|\check{X}_{it}|^{2+\delta} < \infty$ ,  $\Psi = \sum_{t=1}^T [\check{X}_{it}(\check{X}_{it} - \sum_{s=1}^T \frac{\check{X}_{is}}{T})^\tau]$ 是正的有限的, 其中 $\check{X}_{it} = X_{it} - E(X_{it}|U_{it})$ 。

(A6)  $E(Y_{it}|X_i, U_i, v_i) = E(Y_{it}|X_{it}, U_{it}, v_{it}) = X_{it}\beta + g(U_{it}) + v_i$ 。

(A7) 窗宽 $h$ 满足 $h \rightarrow 0$ ,  $nh^8 \rightarrow 0$ 和 $nh^2/(\log n)^2 \rightarrow \infty$ 作为 $n \rightarrow \infty$ 。

(A8) 当 $k \rightarrow \infty$ ,  $P\left[0 \in \text{ch}\left\{\sum_{t=1}^T \tilde{X}_{1t} e_{it}^\tau \tilde{H}(\tilde{Y}_{1t} - \tilde{X}_{1t}^\tau \beta), \dots, \sum_{t=1}^T \tilde{X}_{kt} e_{it}^\tau \tilde{H}(\tilde{Y}_{kt} - \tilde{X}_{kt}^\tau \beta)\right\}\right] \rightarrow 1$ 其中“ch”表示一个集合的凸包在 $R^p$ 里。

**注 2.1.** 假定(A1)-(A8)看起来有点多,事实上条件是平凡的和容易被满足的。(A1)-(A6)能被发现在文章Su和Ullah(2006)。(A7)已经被假定Hu(2014)。(A8)也已经备用许多作者(比如Shi和Lau, 2000; You等, 2006)。□

**定理 2.1.** 假定(A1)-(A8)条件满足。对于模型(2-1),如果 $\beta_0$ 是正的实值参数,则有 $\mathcal{LR}_n(\beta_0) \xrightarrow{d} \chi_p^2$ ,作为 $n \rightarrow \infty$ ,其中 $\chi_p^2$ 是一个标准自由度为 $p$ 卡方随机变量和 $\xrightarrow{d}$ 代表依分布收敛。

作为定理的一个结果,参数的 $\beta_0$ 的置信区间能得到。更准确的说,对于任何 $0 \leq \alpha < 1$ ,令 $c_\alpha$ 是 $\chi^2$ 分布的 $1 - \alpha$ 分位数以至 $P(\chi_p^2 > c_\alpha) \leq 1 - \alpha$ 。因此 $\ell_{EL}(\alpha) = \{\beta \in R^p : \mathcal{LR}_n(\beta) \leq c_\alpha\}$ 建立 $\beta_0$ 一个置信区间是渐近 $1 - \alpha$ 收敛的,因为 $\beta_0$ 属于 $\ell_{EL}(\alpha)$ 等价于 $\mathcal{LR}_n(\beta_0) \leq c_\alpha$ 。

最大化 $\{-\mathcal{LR}_n(\beta_0)\}$ 能获得 $\beta_0$ 的最大经验似然估计 $\hat{\beta}$ 。记

$$\hat{\Psi} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} (\tilde{X}_{it} - \sum_{s=1}^T \frac{\tilde{X}_{is}}{T})^\tau. \quad (2-9)$$

如果矩阵 $\hat{\Psi}$ 是可转置的,相似于Qin和Lawless(1994)的定理1和Xue和Zhu(2008)定理2的证明,因此 $\beta_0$ 的最大经验似然估计 $\hat{\beta}$ 能被表示为

$$\hat{\beta} = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^\tau e_{it}^\tau H \tilde{Y}_{it} + o_p(n^{-\frac{1}{2}}).$$

同样 $\hat{\beta}$ 的渐近正态性能被表达如下定理。

**定理 2.2.** 假定(A1)-(A8)条件满足,如果 $\beta_0$ 是正的实值参数,因此我们得到 $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$ ,其中 $V = \Psi^{-1} \Theta \Psi^{-1}$ ,  $\Theta = \sum_{t=1}^T \sum_{s=1}^T E\left\{ \left[ \tilde{X}_{it} (\tilde{X}_{is} - \sum_{l=1}^T \frac{\tilde{X}_{il}}{T})^\tau \right] (\varepsilon_{it} \varepsilon_{is}) \right\}$ 。

为了建立置信区间,我们能得到 $\hat{\Psi}^{-1} \hat{\Theta} \hat{\Psi}^{-1}$ 作为 $\Psi^{-1} \Theta \Psi^{-1}$ 的估计,其中 $\hat{\Psi}$ 被定义在式(2-9),  $\hat{\Theta} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \tilde{X}_{it} (\tilde{X}_{it} - \sum_{s=1}^T \frac{\tilde{X}_{is}}{T})^\tau \hat{\varepsilon}_{it} \hat{\varepsilon}_{is}$ 和 $\hat{\varepsilon}_{it} = \tilde{Y}_{it} - \tilde{X}_{it}^\tau \hat{\beta} - \hat{v}_i$ 。

通过定理2.2, 我们得到

$$[\widehat{\Psi}^{-1}\widehat{\Theta}\widehat{\Psi}^{-1}]^{-1/2}\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p), \quad (2-10)$$

或者

$$(\widehat{\beta} - \beta_0)^\tau n[\widehat{\Psi}^{-1}\widehat{\Theta}\widehat{\Psi}^{-1}]^{-1}(\widehat{\beta} - \beta_0) \xrightarrow{d} \chi_p^2. \quad (2-11)$$

因此,  $\beta_0$  的置信区间能被得到从式(2-10) 和式(2-11)。

## 2.4 模拟研究与实证

### 2.4.1 Monte Carlo 模拟

在这一节, 我们将做一些模拟阐明对于有限的样本我们建议的块经验似然方法优于渐近正态方法。

首先, 我们考虑下面的带固定效应的部分线性模型:

$$Y_{it} = X_{it}^\tau \beta + g(U_{it}) + v_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2-12)$$

其中  $\beta = (\beta_1, \beta_2)^\tau = (1, \sqrt{2})^\tau / \sqrt{3}$ ,  $g(U_{it}) = \sin(2\pi U_{it})$ ,  $U_{it} \sim U(0, 1)$ ,  $v_i = \rho \bar{U}_i + w_i$ , 其中  $\rho = 0.5, 1$  和  $w_i \sim N(0, 1)$  为  $i = 2, 3, \dots, n$ , 且  $v_i = -\sum_{i=2}^n v_i$ ,  $\bar{U}_i = \frac{1}{T} \sum_{t=1}^T U_{it}$ . 我们用  $\rho$  去控制  $v_i$  和  $\bar{U}_i$  的相关性。

现在我们做模拟, 选取的样本大小分别为  $(n, T) = (50, 4), (50, 6), (80, 4), (80, 6)$ , 和我们选取的核为 Epanechnikov 核  $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \leq 1\}$ . 采取去掉第  $i$  个观测的交错核实验验证(cross validation, CV) 方法的窗宽  $CV(h)$  将被获得用最小化下式

$$CV(h) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - X_{it}^\tau \hat{\beta}_{[i]} - \hat{g}_{[i]}(U_{it}))^2,$$

其中  $\hat{\beta}_{[i]}$  和  $\hat{g}_{[i]}(U_{it})$  分别是  $\beta$  和  $g(U_{it})$  的估计, 那计算所有的测量但不计算第  $i$  观测。我们将做 500 次重复在置信水平为  $1 - \alpha = 0.95$ . 实验的结果在表 2.1-2.3, 其中表 2.1 报道一些具有代表性的  $(\beta_1, \beta_2)$  的置信区间的覆盖率, 表 2.2 报道  $\beta_1$  和  $\beta_2$  的置信区间的覆盖率与区间长度, 和表 2.3 报道这个建议的最大块经验似然估计这些参数  $\beta_1$  和  $\beta_2$  平均偏差和标准差。

从表 2.1 和表 2.2, 我们可以看出块经验似然方法优于渐近正态分布方法,

因为块经验似然方法给出更高的覆盖率和更小的置信区间比渐近正态方法。并且我们还能看出覆盖率增大和置信区间变小随着样本容量 $(n, T)$ 的变大。同样我们能发现置信区间的覆盖率变小和置信区间长度变大随着 $\rho$ 变大。我们能从表3得到我们建议的参数估计是渐近无偏的和较小的标准差, 进一步, 它是合理的, 那平均误差和方差变小随着样本容量的增大。

表 2.1 CP 表示 $(\beta_1, \beta_2)$  的置信区间依据于块经验似然和渐近正态方法在置信水平是0.95

$\rho$	$1 - \alpha$	$(n, T)$	$CP_{BEL}$	$CP_{AN}$	$\rho$	$1 - \alpha$	$(n, T)$	$CP_{BEL}$	$CP_{AN}$
0.5	0.95	(50,4)	0.940	0.890	1	0.95	(50,4)	0.923	0.810
		(50,6)	0.945	0.912			(50,6)	0.935	0.855
		(80,4)	0.951	0.917			(80,4)	0.948	0.890
		(80,6)	0.956	0.935			(80,6)	0.950	0.912

注: “ $CP_{BEL}$  和  $CP_{AN}$ ” 表示CP分别依据于块经验似然和渐近正态方法

#### 2.4.2 基于艾滋病人的CD4数据的实证

我们现在把本章建议的方法用于某艾滋病研究中心的一组CD4数据的研究, 该数据集包括从1984~1991年感染HIV病毒的283个患者的HIV状态数据。该数据详细的描述可见文献Kaslow 等(1987)。这个数据集也被许多作者用来做模型拟合, 比如Xue 和Zhu (2007), Zhou 和Lin (2014), Huang 等 (2001), Fan 和Zhang (2000)。本数据分析的最基本的目标是描述整个研究期间艾滋病患者的CD4 百分比率随时间变化的趋势, 并评估患者的吸烟情况、HIV感染前的CD4 百分率和HIV 感染时患者感染HIV 后平均CD4 百分率的影响。最初的实验设计对所有患者进行每半年观测一次。但是, 由于大量的患者错过了其诊断时间并且在整个研究过程中HIV感染是随机发生的, 所以, 对不同患者的观测次数及观测时间是不同的, 因此该数据是一个不平衡的纵向数据, 我们在283个患者中选取了158个患者, 那他们的前六次观测数据都在, 因此我们获得对于每个个体相同的观测数据, 即是平衡的面板数据。

令 $t_{ij}$  是第 $i$ 个体在感染了HIV后在在观察的几年时间内第 $t$ 次测量观察的时间,  $Y_{it}$  是第 $i$ 个体的CD4 细胞cell 的百分比在时间 $t_{ij}$  时. 我们考虑两个协变量:  $X_{1it}$ , 第 $i$ 个体的preCD4 细胞cell 的百分比在时间 $t_{ij}$  时;  $X_{2it}$ , 第 $i$ 个体的抽烟状

表 2.2 CP 和 AL 表示置信区间与区间长度依据于块经验似然和渐近正态方法在置信水平是 0.95

<i>Para</i>	$\rho$	$(n, T)$	$CP_{BEL}$	$CP_{AN}$	$AL_{BEL}$	$AL_{AN}$
$\beta_1$	0.5	(50,4)	0.935	0.895	0.1246	0.2309
		(50,6)	0.942	0.903	0.1146	0.1851
		(80,4)	0.950	0.921	0.1078	0.1806
		(80,6)	0.957	0.928	0.0899	0.1418
$\beta_2$		(50,4)	0.938	0.670	0.1376	0.2363
		(50,6)	0.941	0.780	0.1231	0.1853
		(80,4)	0.946	0.823	0.1152	0.1803
		(80,6)	0.954	0.895	0.0956	0.1415
$\beta_1$	1	(50,4)	0.931	0.770	0.1279	0.2541
		(50,6)	0.939	0.798	0.1049	0.2018
		(80,4)	0.941	0.807	0.0992	0.1931
		(80,6)	0.949	0.865	0.0816	0.1518
		(80,6)	0.949	0.865	0.0816	0.1518
$\beta_2$		(50,4)	0.924	0.654	0.1532	0.2541
		(50,6)	0.937	0.712	0.1324	0.2034
		(80,4)	0.941	0.810	0.1267	0.1900
		(80,6)	0.947	0.853	0.1087	0.1512

注: “ $CP_{BEL}$  和  $CP_{AN}$ ” 表示 CP 分别依据于块经验似然和渐近正态方法 “ $AL_{BEL}$  和  $AL_{AN}$ ” 表示 AL 分别依据于块经验似然和渐近正态方法

表 2.3 参数  $\beta_1$  和  $\beta_2$  依据建议的最大块经验似然估计的平均偏差和标准差

$(n, T)$	$\beta_1$				$\beta_2$			
	$\rho = 0.5$		$\rho = 1$		$\rho = 0.5$		$\rho = 1$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
(50,4)	0.0019	0.0536	0.0021	0.0837	0.0026	0.0609	-0.0021	0.0571
(50,6)	-0.0013	0.0520	-0.0016	0.0523	0.0014	0.0531	0.0023	0.0528
(80,4)	0.0007	0.0461	0.0009	0.0471	0.0010	0.0452	0.0014	0.0508
(80,6)	0.0001	0.0404	0.0006	0.0426	0.0006	0.0418	0.0008	0.0438

态, 它的取值是 1 或 0, 通过看个体是否是吸烟的还是不吸烟的。我们考虑下面



的模型:

$$Y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + g(U_{it}) + v_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (2-13)$$

其中  $U_{it} = t_{ij}/T$  和  $v_i$  是一个规定的状态效应, 可能包括时间, 区域, 种族和教育.  $v_i = \rho\bar{U}_i + w_i$  伴随着  $\bar{U}_i = \frac{1}{T} \sum_{t=1}^T U_{it}$  和  $w_i \sim N(0, 1)$ .  $\beta_1$  和  $\beta_2$  表示感染前一期pre-CD4 百分比和抽烟的影响, 用块经验似然方法和渐近正态分布方法, 我们得到  $(\beta_1, \beta_2)$  的置信水平  $1 - \alpha = 0.95$  置信区间被表示在图2.1. 这些图显示这建议的块经验似然有更小的置信区间比渐近正态分布方法, 且块经验似然方法的图更偏离中心作为  $\rho = 0$  变成  $\rho = 0.5$ .

## 2.5 主要结果的证明

为了表示的方面, 我们简化符号, 令  $\vartheta_k = \int u^k K(u) dt$  和  $P = (I_{nT} - \Omega)^r (I_{nT} - \Omega)$ .

**引理 2.1.** 假设条件(A1)-(A8) 都拥有, 因此我们能得到

$$E|g(U_{it}) - \sum_{k=1}^n \sum_{l=1}^T \omega_{kl}(U_{it})g(U_{kl})|^2 = O(h^4). \quad (2-14)$$

**证明.** 记

$$\omega_{it}(u) = \frac{\frac{1}{N} K_h(U_{it} - u)[A_{N2}(u) - (U_{it} - u)A_{N1}(u)]}{A_{N0}(u)A_{N2}(u) - A_{N1}^2(u)} = \frac{V_{Nit}(u)}{V_N(u)}, \quad (2-15)$$

其中  $V_{Nit}(u) = \frac{1}{N} K_h(U_{it} - u)[A_{N2}(u) - (U_{it} - u)A_{N1}(u)]$ ,  $V_N(u) = A_{N0}(u)A_{N2}(u) - A_{N1}^2(u)$ . 令  $u = U_{kl}$ ,  $k = 1, \dots, n$ ;  $l = 1, \dots, T$ . 注意  $\sum_{k=1}^n \sum_{l=1}^T V_{Nit}(u)(U_{it} - u) = 0$  和  $\sum_{k=1}^n \sum_{l=1}^T V_{Nit}(u) = V_N(u)$ . 记  $G(U_{it}, u) = g(U_{it}) - g(u) + g'(u)(U_{it} - u)$ . 我们通过一个简单的计算可得到

$$g(U_{it}) - \sum_{k=1}^n \sum_{l=1}^T \omega_{kl}(U_{it})g(U_{kl}) = \frac{\sum_{k=1}^n \sum_{l=1}^T V_{Nit}(u)G(U_{it}, u)}{V_N(u)}. \quad (2-16)$$

同样令  $E_{U_{it}}(\chi_n)$  是  $\chi_n$  条件期望, 在指定的  $U_{it}$ , 我们表示  $\chi_n = O_r(a_n)$ , 如果  $E(|\chi_n|^r) = O(a_n^r)$ . 我们用Cauchy-Schwarz 不等式能得到

$$O_r(a_n)O_r(b_n) = O_{r/2}(a_n b_n), \quad (2-17)$$

$$\chi_n = E_{U_{it}}[\chi_n] + O_4((E|\chi_n - E_{U_{it}}[\chi_n]|^4)^{1/4}). \quad (2-18)$$

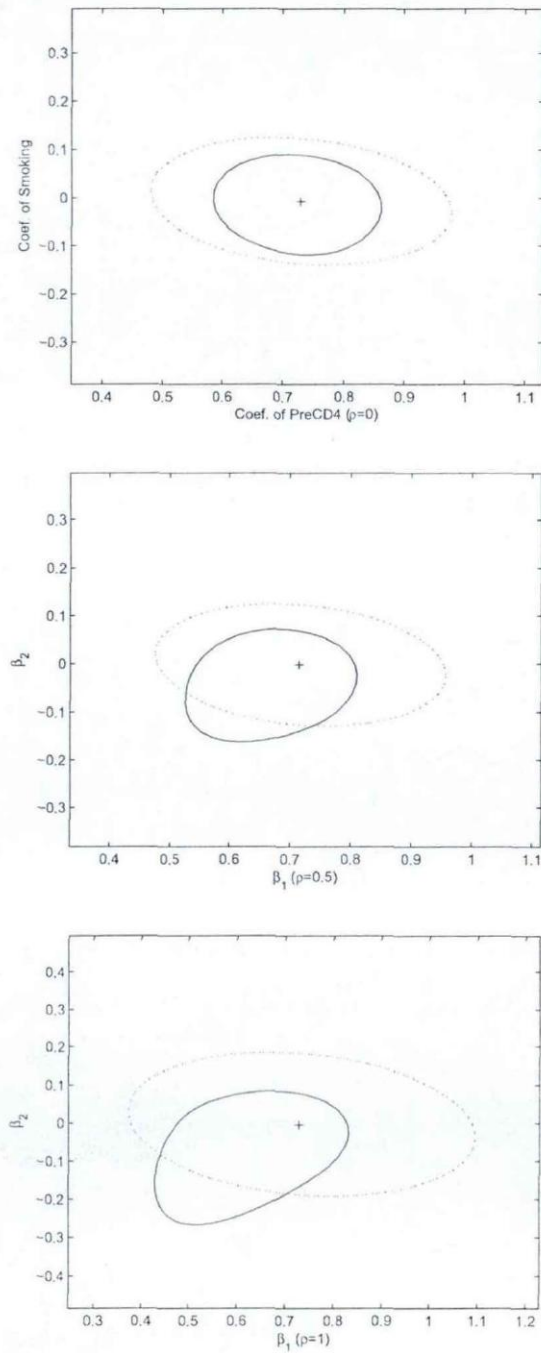


图 2.1 应用CD4 数据.  $\beta = (\beta_1, \beta_2)$ 的置信水平95% 的置信区间依据于渐近正态方法(dotted curve) 和块经验似然(solid curve)。

在条件(A1)下, 我们知道  $\sup_u f(u) < \infty$ . 因此  $\exists L > 0$ , 对于任意的  $y$  和  $u$ ,  $|f(y) - f(u)| \leq L|y - u|$ . 对于  $A_{Ns}(u)$ ,  $s = 0, 1, 2$ , 我们能得到

$$\begin{aligned} E_{U_{it}}[A_{Ns}(u)] &= E\left[\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^T (U_{it} - u)^s K_h(U_{it} - u)\right] \\ &= \int (U_{it} - u)^s K_h(U_{it} - u) f(U_{it}) dU_{it} \\ &= h^s \int u^s K(u) du [f(u) + O(h)] \\ &= h^s f(u) \vartheta_s [1 + O(h)], \end{aligned} \quad (2-19)$$

我们再有Yang (2001)的定理1, 我们容易得到

$$\begin{aligned} &E\{|A_{Ns}(u) - E_{U_{it}}[A_{Ns}(u)]|^4\} \\ &\leq \frac{c}{N^4} N h^{4s-3} \int |u^s K(u)|^4 du [E f(u) + Lh] \\ &\quad + \frac{c}{N^4} E\left\{N h^{2s-3} \int u^{2s} K^2(u) du [E f(u) + Lh]\right\} \\ &= O(N^{-3} h^{4s-3}) + O(N^{-2} h^{2s-2}). \end{aligned} \quad (2-20)$$

由式(2-18)和式(2-19), 我们得到

$$\begin{aligned} A_{Ns}(u) &= E_{U_{it}}[A_{Ns}(u)] + O_4(h^s (Nh)^{-1/2}) \\ &= h^s f(u) \vartheta_s [1 + O_4(h + (Nh)^{-1/2})], \quad s = 0, 1, 2. \end{aligned} \quad (2-21)$$

从式(2-17)和式(2-21), 我们能得到

$$V_N(u) = A_{N0}(u)A_{N2}(u) - A_{N1}^2(u) = h^2 f^2(u) \vartheta_s [1 + O_2(h + (Nh)^{-1/2})]. \quad (2-22)$$

令  $V_{Nh}(u) = \frac{1}{h^2} V_N(u)$ ,  $V(u) = f^2(u) \vartheta_s$ . 既然  $\mathcal{U}$  是一个紧集, 利用式(2-19)和Dvoretzky不等式, 和引用这个标准处置核密度估计方法, 我们能推出, 对于任意的  $\varepsilon > 0$ ,

$$P\{\sup_{u \in \mathcal{U}} |V_{Nh}(u) - V(u)| > \varepsilon\} \leq c \exp(-\gamma N h^2).$$

因为序列  $\sum_{N=1}^{\infty} \exp(-\gamma N h^2) < \infty$ , 对于  $\gamma > 0$ , 通过假设, 一个运用Borel-Cantelli引理可证明

$$\sup_{u \in \mathcal{U}} |V_{Nh}(u) - V(u)| \rightarrow 0, \quad a.s. \quad (2-23)$$

通过条件(A1), 我们知道  $\inf_{u \in \mathcal{U}} f(u) \geq c_0 > 0$ , 因此, 当  $n$  是足够大, 我们得到

$$\inf_{u \in \mathcal{U}} |V_{Nh}(u)| \geq \inf_{u \in \mathcal{U}} |V(u)| - \sup_{u \in \mathcal{U}} |V_{Nh}(u) - V(u)| > \vartheta_2 c_0^2 > 0, \quad a.s. \quad (2-24)$$

通过一个简单计算我们得到

$$\begin{aligned} \sum_{k=1}^n \sum_{l=1}^T V_{N_{it}}(u) G(U_{it}, u) &= \frac{1}{N} \sum_{k=1}^n \sum_{l=1}^T G(U_{it}, u) K_h(U_{it} - u) A_{N_2}(u) \\ &- \frac{1}{N} \sum_{k=1}^n \sum_{l=1}^T G(U_{it}, u) (U_{it} - u) K_h(U_{it} - u) A_{N_1}(u). \end{aligned} \quad (2-25)$$

相似式子(2-21)的证明, 对于  $s = 0, 1$ , 我们有

$$\begin{aligned} &\frac{1}{Nh^{2+s}} \sum_{k=1}^n \sum_{l=1}^T G(U_{it}, u) (U_{it} - u)^s K_h(U_{it} - u) \\ &= h^{-(2+s)} E_{U_{it}} [G(U_{it}, u) (U_{it} - u)^s K_h(U_{it} - u)] + O_4((Nh)^{-1/2}) \\ &= d_{Ns} + O_4((Nh)^{-1/2}), \end{aligned} \quad (2-26)$$

其中

$$|d_{Ns}| \leq c \int |u|^{s+2} K(u) du [f(u) + O(h)] = O(1).$$

合并式子(2-21), (2-25) 和(2-26), 我们得到

$$\sum_{k=1}^n \sum_{l=1}^T V_{N_{it}}(u) G(U_{it}, u) = h^4 f(u) \vartheta_2 d_{N0} + O_2(h + (Nh)^{-1/2}). \quad (2-27)$$

这合并方程(2-22), 我们推导如下结论

$$E \left| \frac{\sum_{k=1}^n \sum_{l=1}^T V_{N_{it}}(u) G(U_{it}, u)}{V_N(u)} \right|^2 \leq ch^{-4} E \left[ \sum_{k=1}^n \sum_{l=1}^T V_{N_{it}}(u) G(U_{it}, u) \right]^2 = O(h^4).$$

因此, 我们证明了引理2.1。

**引理 2.2.** 假设条件(A1)-(A8) 都拥有. 因此我们能得到

$$(a) \quad \omega_{it}(u) = o(1).$$

$$(b) \quad (M^T P M)^{-1} = (M^T M)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n)$$

$$\text{对于充分大的 } n, \text{ 其中 } \zeta_n = \frac{\sqrt{\ln n}}{nh} i_n i_n^T.$$

证明. 证(a). 类似于(2-27)的证明, 我们得到

$$E[V_{N_{it}}(U_{it})] = N^{-1} h^4 f(u) \vartheta_2 [1 + O_2(h + (Nh)^{-1/2})],$$

合并(2-22), 我们能证明(a)。

(b).  $M^T P M = M^T (I_{nT} - \Omega)^T (I_{nT} - \Omega) M = M^T M - M^T \Omega M - M^T \Omega^T M + M^T \Omega^T \Omega M = \Delta_{11} - \Delta_{12} - \Delta_{13} + \Delta_{14}$ . 利用(a), 我们能显示  $\Delta_{1l}$  的第  $(i, j)$  个元

素是 $[\Delta_{11}]_{ij} = o(1)$ ,  $l = 2, 3, 4$ . 类似于Su (2006)的引理A.2 的证明, 我们能证明 $(M^T P M)^{-1} = (M^T M)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n)$ .

**引理 2.3.** 假设条件(A1)-(A8) 都拥有. 因此我们能得到

$$\begin{aligned} (a) \quad & n^{-1} X^T P X \xrightarrow{p} \sum_{t=1}^T [(X_{it} - E(X_{it}|U_{it}))(X_{it} - E(X_{it}|U_{it}))^T]. \\ (b) \quad & n^{-1} X^T P M (M^T M)^{-1} M^T P X \xrightarrow{p} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T [(X_{it} - E(X_{it}|U_{it}))(X_{is} - E(X_{is}|U_{is}))^T]. \\ (c) \quad & \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T e_{it}^T \tilde{H} \tilde{X}_{it} \xrightarrow{p} \Psi. \end{aligned}$$

**证明.** (a),(b) 是Su 和Ullah (2006)的引理A.3 的直接结论。

(c) 利用(a), (b) 和引理2.2,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T e_{it}^T \tilde{H} \tilde{X}_{it} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T e_{it}^T (I_{nT} - \tilde{M}(\tilde{M}^T \tilde{M})^{-1} \tilde{M}^T) \tilde{X}_{it} \\ & = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T \tilde{X}_{it} - \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T \tilde{M}(\tilde{M}^T \tilde{M})^{-1} \tilde{M}^T \tilde{X}_{it} \\ & = n^{-1} X^T P X - n^{-1} X^T P M (M^T M)^{-1} M^T P X \xrightarrow{p} \Psi. \end{aligned}$$

**引理 2.4.** 假设条件(A1)-(A8) 都拥有. 因此我们能得到

$$\begin{aligned} (a) \quad & n^{-1/2} X^T P \varepsilon \xrightarrow{p} n^{-1/2} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - E(X_{it}|U_{it})) \varepsilon_{it} + o_p(1). \\ (b) \quad & n^{-1/2} X^T P M (M^T M)^{-1} M^T P \varepsilon \xrightarrow{p} n^{-1/2} \frac{1}{T} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T (X_{it} - E(X_{it}|U_{it})) \varepsilon_{is} + o_p(1). \end{aligned}$$

**证明.** (a),(b) 是Su 和Ullah (2006)的引理A.6 的直接结论。

**引理 2.5.** 假设条件(A1)-(A8) 都拥有. 因此我们能得到

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \psi_i(\beta) \xrightarrow{d} N(0, \Theta).$$

**证明.** 通过 $\psi_i(\beta)$  的定义, 我们得到

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T e_{it}^T \tilde{H} (I_{nT} - \Omega) \varepsilon_{it} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it}^T e_{it}^T \tilde{H} [g(U_{it}) \\ &\quad - \sum_{k=1}^n \sum_{l=1}^T \omega_{kl}(U_{it}) g(U_{kl})] = I_{1n} + I_{2n}. \end{aligned}$$

利用引理2.3, 它是容易得到  $E(I_{1n}) = 0$ ,  $Var(I_{1n}) = \Theta$ , 和  $I_{1n}$  满足 *Cramér-Wold* 设计的条件和 *Lindeberg* 条件。利用中心极限定理, 我们得到  $I_{1n} \xrightarrow{d} N(0, \Theta)$ 。相似与 *Su* 和 *Ullah* (2006) 的引理A.5的证明, 我们能得到  $\frac{1}{\sqrt{n}} \tilde{X}^\tau (I - \tilde{H}) \tilde{g}(U) = o_p(1)$ 。因此我们容易得到下面的等式

$$I_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} e_{it}^\tau \tilde{H} \tilde{g}(U_{it}) = o_p(1).$$

因此, 我们完成引理的证明。

**引理 2.6.** 在定理2.1的条件下, 令  $B_{it} = \tilde{X}_{it} e_{it}^\tau H$  我们得到

$$V = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau (\tilde{Y}_{it} - \tilde{X}_{it} \beta) (\tilde{Y}_{it} - \tilde{X}_{it} \beta)^\tau \xrightarrow{p} \Theta.$$

**证明.**

$$\begin{aligned} V &= \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau [\tilde{g}(U_{it}) + (I_{nT} - \Omega) \varepsilon_{it}] [\tilde{g}(U_{it}) + (I_{nT} - \Omega) \varepsilon_{it}]^\tau \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau [(I_{nT} - \Omega) \varepsilon_{it}] [(I_{nT} - \Omega) \varepsilon_{it}]^\tau \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau \tilde{g}(U_{it}) [(I_{nT} - \Omega) \varepsilon_{it}]^\tau \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau [(I_{nT} - \Omega) \varepsilon_{it}] \tilde{g}(U_{it})^\tau \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^\tau \tilde{g}(U_{it}) \tilde{g}(U_{it})^\tau \\ &= \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4. \end{aligned}$$

利用大数定律, 我们能推导出  $\Lambda_1 \xrightarrow{d} \Theta$ 。利用引理2.4 和 *Cauchy-Schwarz* 不等式, 我们能证明  $\Lambda_j \xrightarrow{d} \Theta$ ,  $j = 2, 3, 4$ , 引理2.6 证明完成。

**引理 2.7.** 在定理2.1的条件下, 我们得到

$$(i) \max_{1 \leq i \leq n} \|\psi_i(\beta)\| = o_p(n^{1/2}), \quad (2-28)$$

$$(ii) \lambda(\beta) = O_p(n^{-1/2}). \quad (2-29)$$

**证明.** 相似于 *Owen* (1998) 给的结论, 从式子(2-7)和引理2.5 和引理2.6, 我们能

够容易证明式子(2-29),我们现在证明式子(2-28)

$$\begin{aligned} \max_{1 \leq i \leq n} \|\psi_i(\beta)\| &\leq \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{X}_{it} e_{it}^{\tau} \tilde{H} (I_{nT} - \Omega) \varepsilon_{it} \right\| \\ &+ \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{X}_{it} e_{it}^{\tau} \tilde{H} [g(U_{it}) - \sum_{k=1}^n \sum_{l=1}^T \omega_{kl}(U_{it}) g(U_{kl})] \right\| \\ &= \Pi_1 + \Pi_2. \end{aligned}$$

对于 $\Pi_1$ , 注意

$$E \left[ \sum_{t=1}^T \sum_{s=1}^T \tilde{X}_{it} e_{it}^{\tau} \tilde{H} (I_{nT} - S) v_{it} \tilde{X}_{is} e_{is}^{\tau} \tilde{H} (I_{nT} - S) v_{is} \right] = \Theta_{s,s} + o(1) < \infty,$$

其中 $\Theta_{s,s}$ 表示矩阵 $\Theta$ 的第 $s$ 个对角元素。 $B_{it}$ 是相互独立同分布的对于每个截面 $i$ 指数和每个固定的 $t$ , 通过Owen (1990)的引理3, 我们得到 $\Pi_1 = o_p(n^{1/2})$ 。

现在看第二项 $\Pi_2$ , 通过Markov的不等式, 我们得到

$$\begin{aligned} P(\Pi_2 > n^{1/2}) &\leq \frac{1}{n} \sum_{i=1}^n E \left\| \sum_{t=1}^T \tilde{X}_{it} e_{it}^{\tau} \tilde{H} \tilde{g}(U_{it}) \right\|^2 \\ &= \frac{1}{n} \sum_{i=1}^n E \left\| \sum_{t=1}^T \sum_{s=1}^T \tilde{X}_{it} \tilde{X}_{is} e_{it}^{\tau} \tilde{H} \tilde{g}(U_{it}) e_{is}^{\tau} \tilde{H} \tilde{g}(U_{it}) \right\| \\ &\leq \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \left\{ E [\tilde{X}_{it} e_{it}^{\tau} \tilde{H} \tilde{g}(U_{it})]^2 \right\}^{1/2} \left\{ E [\tilde{X}_{is} e_{is}^{\tau} \tilde{H} \tilde{g}(U_{it})]^2 \right\}^{1/2} \\ &= o_p(1) \rightarrow 0. \end{aligned}$$

因此,  $\Pi_2 = o_p(n^{1/2})$  和  $\max_{1 \leq i \leq n} \|\psi_i(\beta)\| = o_p(n^{1/2})$ 。

证明定理2.1. 利用式子(2-7), 我们得到

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \frac{\psi_i(\beta)}{1 + \lambda^{\tau} \psi_i(\beta)} \\ &= \frac{1}{n} \sum_{i=1}^n \psi_i(\beta_0) - \frac{1}{n} \sum_{i=1}^n \psi_i(\beta) \psi_i^{\tau}(\beta) \lambda + \frac{1}{n} \sum_{i=1}^n \frac{\psi_i(\beta) [\lambda^{\tau} \psi_i(\beta)]^2}{1 + \lambda^{\tau} \psi_i(\beta)} = 0. \end{aligned} \quad (2-30)$$

应用泰勒展开式, 从(2-6), 我们能得到下式

$$\mathcal{LR}_n(\beta) = 2 \sum_{i=1}^n \left\{ \lambda^{\tau} \psi_i(\beta) - \frac{[\lambda^{\tau} \psi_i(\beta)]^2}{2} \right\} + o_p(1). \quad (2-31)$$

利用引理2.5-2.7, 我们得到

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^n \frac{\psi_i(\beta) [\lambda^{\tau} \psi_i(\beta)]^2}{1 + \lambda^{\tau} \psi_i(\beta)} \right\| \leq \frac{1}{n} \sum_{i=1}^n \frac{\|\psi_i(\beta)\|^3 \|\lambda\|^2}{|1 + \lambda^{\tau} \psi_i(\beta)|} \\ &\leq \|\lambda\|^2 \max_{1 \leq i \leq n} \|\psi_i(\beta)\| \frac{1}{n} \sum_{i=1}^n \|\psi_i(\beta)\|^2 \end{aligned}$$

$$= O_p(n^{-1})o_p(n^{1/2})O_p(1) = o_p(n^{-1/2}),$$

那合并式子(2-30), 得到 $\sum_{i=1}^n [\lambda^\tau \psi_i(\beta)]^2 = \sum_{i=1}^n \lambda^\tau \psi_i(\beta) + o_p(1)$  和

$$\lambda = \left[ \sum_{i=1}^n \psi_i(\beta) \psi_i^\tau(\beta) \right]^{-1} \sum_{i=1}^n \psi_i(\beta_0) + o_p(n^{-1/2}).$$

因此, 利用式子(2-31), 我们有

$$\mathcal{LR}_n(\beta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\beta) \right)^\tau \left( \frac{1}{n} \sum_{i=1}^n \psi_i(\beta) \psi_i^\tau(\beta) \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\beta) \right) + o_p(1),$$

那合并引理2.5 和引理2.6, 证明了定理2.1。

**证明定理2.2.** 顺延着Zhu and Xue (2008)的定理2的证明的相似结论, 我们有

$$\hat{\beta} - \beta_0 = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \psi_i(\beta) + o_p(n^{-1/2}).$$

利用引理2.3, 通过利用大数定律, 我们能证明 $\hat{\Psi} \xrightarrow{P} \Psi$ . 合并引理2.4 和Slutsky定理, 我们能够容易完成定理2.2的证明。





## 第3章 带固定效应的面板数据部分线性误差变量模型的统计推断

### 3.1 引言

面板数据记录每个个体单元在一段时间的的信息，面板数据包含丰富的信息能使研究人员估计复杂的模型并回答不可能单独使用时间序列或横截面数据问题。面板数据分析在最近二十年收到了很多的关注，已经应用于许多学科，如经济学，金融，生物学，工程和社会科学。Arellano (2003), Baltagi (2005) 和Hsiao (2003) 提供了极好的洞察关于参数面板数据模型的统计推断和经济分析，它是众所周知的半参数方法保持了非参数方法与完全参数方法的优点，半参数回归模型减少了完全参数模型的错误指定的高风险。也可以避免纯非参数的一些严重的缺陷，比如：维数灾祸，难以解释，缺乏外推能力。在最近二十年，各种半参数模型已经被发展。这些包括部分线性模型(Engle 等,1986)，变系数部分线性模型(Fan 和Huang, 2005)，单指数模型(Zhang 和Huang, 2009)和单指数面板数据模型(Zhu 等, 2014)。带固定效应的面板数据部分线性模型是一个用的工具对于经济分析。可以看Henderson 等(2008), Horowitz 和Lee (2004), Hu (2014), 和Li 等(2011)。

在这章，我们考虑下面的带固定效应的部分线性面板数据模型(Su 和Ullah, 2005):

$$Y_{it} = X_{it}^T \beta + m(U_{it}) + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (3-1)$$

其中 $Y_{it}$ 是响应变量,  $(X_{it}^T, U_{it}) \in R^p \times R$ 是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^T$ 是一个 $p$ -维未知参数, 上角 $\tau$ 表示一个向量的转置,  $m(U_{it}) = (m_1(U_{it}), \dots, m_q(U_{it}))^T$ 是一个 $q$ -维未知参数,  $\alpha_i$ 是一个不可观察的个体效应. 表示 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$ 是 $i$ 个体随机误差向量和 $\{\varepsilon_i, i = 1, \dots, n\}$ 是相互独立且 $E(\varepsilon_i | X_{it}, U_{it}) = 0$ 。我们允许 $\alpha_i$ 是与 $X_{it}$ , 和 $U_{it}$  相关的, 伴随一个未知的相关结构。因此模型(3-1)是一个固定效应模型。

然而，由于测量机制或环境的影响，协变量不总是能被直接地观察到，

而是含有误差的,简单地忽略测量误差(误差变量),将导致有偏见的估计。处理带有测量误差的协变量统计分析通常是一个挑战。在过去的二十年中,测量误差的回归分析有了很大的进展。一些细节的研究能被发现在Fuller (1987); Fan 和Truong (1993); Liang 等(1999); Fan 等(2013) 等专家的研究中。带有可加测量误差的模型(3-1)的统计分析,似乎仍然被忽略。本章的目的是填补这一空白。

具体而言,我们考虑以下带有固定效应的面板数据部分线性误差变量模型:

$$\begin{cases} Y_{it} = X_{it}^T \beta + m(U_{it}) + \alpha_i + \varepsilon_{it}, \\ Z_{it} = X_{it} + \eta_{it}, \end{cases} \quad i = 1, \dots, n; t = 1, \dots, T. \quad (3-2)$$

其中协变量 $X_{it}$ 被测量伴随测量误差且不能被观察到。代替 $X_{it}$ ,我们观察到 $Z_{it}$ ,含有的测量误差 $\eta_{it}$ 是独立同分布的,并且与 $(X_{it}, U_{it}, \varepsilon_{it})$ 独立,我们假定 $Cov(\eta_{it}) = \Sigma_\eta$ ,当 $\Sigma_\eta$ 是已知的,类似于Zhu 和Cui (2003), You 和Chen (2006), 和别的专家的文章。当 $\Sigma_\eta$ 是未知的,我们能估计通过重复测量 $Z_{it}$ ; 可以看Liang 等(1999)详细的描述。

在固定效应面板数据回归模型中,个体效应往往被视为干扰参数。由于发散的干扰参数的数目,估计参数和非参数的部分在带有固定效应的面板数据部分线性模型不再是简单的。在最近一些年,各种努力去发展合适的方法去估计。对于模型(3-1), Su 和Ullah (2006) 采用一个局部哑变量方法去消除未知的固定效应。在这章,我们对于带有固定效应的面板数据部分线性模型的参数 $\beta$ 的统计推断。顺从Fan 等(2005)的参数估计程序, $\beta$ 的剖面最小二乘估计能被获得。依据此方法,构造一个参数的正态区间。然而,由于先前提出的含有误差估计的衰减,这个剖面最小二乘估计(Fan 等2005)不再是一致估计,我们建议一个修正的剖面最小二乘估计,由此产生的估计是一致的和渐近正态。一致估计的误差的方差也能被考虑。

因为在半参数模型的参数部分常常是主要感兴趣的;因此,拟合后的模型(3-2),这是很自然的看是否某个参数部分统计中是显著的。如

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_l = 0, \quad l \leq p.$$

更一般的,我们考虑线性假设:

$$H_0 : A\beta = 0, \quad (3-3)$$

其中 $\Lambda$ 是一个指定的 $l \times p$ 满秩矩阵。这个检验是一个半参数假设相对于另一个半参数假设检验问题。我们证明在误差 $\varepsilon_{it}$ 分布的一般假设下，这个剖面似然比检验满足Wilks类型的结果。剩下的这一章布局是这样的：3.2节我们提出参数部分，非参数部分和误差的方差的估计，剖面似然检验问题被介绍在3.3节，假设条件和主要结论在3.4节，一些模拟和实证安排在3.5节，本章的结论和未来的研究在3.6节，主要结果的证明安排在3.7节。

### 3.2 模型和方法

为介绍我们的估计，我们假定模型(3-1)在这个约束条件下 $\sum_{i=1}^n \alpha_i = 0$ 。我们写模型(3-1)的矩阵形式如下

$$Y = X\beta + m(U) + B\alpha + \varepsilon, \quad (3-4)$$

其中 $B = I_n \otimes i_T$ 和 $\otimes$ 是Kronecker的积， $I_n$ 表示 $n \times n$ 单位矩阵，和 $i_n$ 表示 $n \times 1$ 数字1的向量。有许多方法估计 $\{\beta_j, j = 1, \dots, p\}$ 和这个函数 $m(\cdot)$ 。

我们方法的主要思想是剖面最小二乘估计方法(PLS)被建议Fan和Huang(2005): 假定我们有一个随机样本 $\{(U_{it}, X_{it1}, \dots, X_{itp}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T\}$ 来自模型(3-1)。令 $\theta = (v^\tau, \beta^\tau)^\tau$ 。指定 $\theta$ ，我们能够运用一个局部线性回归技术估计非参数部分 $\{m_j(\cdot), j = 1, \dots, q\}$ 在模型(3-1)。对于 $U_{it}$ 是在 $u$ 的一个小领域内，我们能估计 $m_j(U_{it})$ 局部线性函数

$$m(U_{it}) \approx m(u) + m'(u)(U_{it} - u) \equiv \mathbf{a} + \mathbf{b}(U_{it} - u).$$

其中 $m'(u) = \partial m(u) / \partial u$ 。这导致下面的权局部最小二乘问题：发现 $\mathbf{a}$ 和 $\mathbf{b}$ 去最小化

$$\sum_{i=1}^n \sum_{t=1}^T \left\{ (Y_{it} - X_{it}^\tau \beta - \alpha_i) - [\mathbf{a} + \mathbf{b}(U_{it} - u)] \right\}^2 K_h(U_{it} - u), \quad (3-5)$$

其中 $\mathbf{a}$ 和 $\mathbf{b}$ 是 $q$ -维列向量，和 $K_h(\cdot) = K(\cdot/h)/h$ ， $K(\cdot)$ 是一个核函数和 $h$ 是一系

列接近于0的正数, 叫窗宽。为方便起见, 令

$$X = \begin{pmatrix} X_{11}^T \\ \vdots \\ X_{1T}^T \\ \vdots \\ X_{nT}^T \end{pmatrix}, \quad D_u = \begin{pmatrix} 1 & \frac{U_{11}-u}{h} \\ \vdots & \vdots \\ 1 & \frac{U_{1T}-u}{h} \\ \vdots & \vdots \\ 1 & \frac{U_{nT}-u}{h} \end{pmatrix}, \quad Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ \vdots \\ Y_{nT} \end{pmatrix},$$

和  $W_u = \text{diag}(K_h(U_{11}-u), \dots, K_h(U_{1T}-u), \dots, K_h(U_{nT}-u))$ . 因此我们可以解方程(3-5)被下式

$$[\hat{a}, h\hat{b}] = (D_u^T W_u D_u)^{-1} D_u^T W_u (Y - X\beta - B\alpha).$$

在(3-5),  $m(U_{it})$  能被代替  $\hat{a}$ . 因此我们有

$$\hat{m}(U_{it}) = (1, 0)(D_{U_{it}}^T W_{U_{it}} D_{U_{it}})^{-1} D_{U_{it}}^T W_{U_{it}} (Y - X\beta - B\alpha). \quad (3-6)$$

令

$$\tilde{Y}_{it} = Y_{it} - (1, 0)(D_{U_{it}}^T W_{U_{it}} D_{U_{it}})^{-1} D_{U_{it}}^T W_{U_{it}} Y,$$

$$\tilde{X}_{it}^T = X_{it}^T - (1, 0)(D_{U_{it}}^T W_{U_{it}} D_{U_{it}})^{-1} D_{U_{it}}^T W_{U_{it}} X,$$

和

$$\tilde{B}_i^T = B_i^T - (1, 0)(D_{U_{it}}^T W_{U_{it}} D_{U_{it}})^{-1} D_{U_{it}}^T W_{U_{it}} B.$$

现在我们考虑一种方法去消除未知的固定效应, 动机是通过一个最小平方哑变量方法, 为此, 我们解决下面的最优问题

$$\hat{\theta} = \arg \min_{\theta} [Y - X\beta - B\alpha - S(Y - X\beta - B\alpha)]^T [Y - X\beta - B\alpha - S(Y - X\beta - B\alpha)],$$

假定  $\tilde{X} = (I_N - S)X$ ,  $\tilde{Y} = (I_N - S)Y$ ,  $\tilde{B} = (I_N - S)B$ , 其中  $N = nT$ ,

$$S = \begin{pmatrix} (1,0)(D_{U_{11}}^T W_{U_{11}} D_{U_{11}})^{-1} D_{U_{11}}^T W_{U_{11}} \\ \vdots \\ (1,0)(D_{U_{1T}}^T W_{U_{1T}} D_{U_{1T}})^{-1} D_{U_{1T}}^T W_{U_{1T}} \\ \vdots \\ (1,0)(D_{U_{nT}}^T W_{U_{nT}} D_{U_{nT}})^{-1} D_{U_{nT}}^T W_{U_{nT}} \end{pmatrix} = \begin{pmatrix} s_{11} \\ \vdots \\ s_{1T} \\ \vdots \\ s_{nT} \end{pmatrix}.$$

我们因此有

$$\tilde{Y} = \tilde{X}\beta + \tilde{B}\alpha + \varepsilon,$$

通过最小二乘方法和一个稍微复杂的计算我们有:

$$\begin{aligned} \hat{\beta} &= (\tilde{X}^T \tilde{X} - \tilde{X}^T \tilde{M}_{\tilde{B}} \tilde{X})^{-1} \tilde{X}^T (I - \tilde{M}) \tilde{Y}, \\ \hat{\alpha} &= (\tilde{B}^T \tilde{B})^{-1} \tilde{B}^T (\tilde{Y} - \tilde{X}\hat{\beta}), \end{aligned} \quad (3-7)$$

其中  $\tilde{M}_{\tilde{B}} = \tilde{B}(\tilde{B}^T \tilde{B})^{-1} \tilde{B}^T$  是  $\tilde{B}$  的投影矩阵。

然而,  $X_{it}$ 's 在我们的例中不能被观察和我们仅能获得  $Z_{it}$ 。如果我们忽略测量误差和代替  $X_{it}$  用  $Z_{it}$  在式(3-7), 式(3-7)可能被显示被推导的估计是不一致的。它是众所周知的线性回归模型或部分线性回归模型, 因为测量误差的不一致发生能被克服通过应用一种叫“误差修正”方法, 可以看 Fuller (1987) 和 Liang 等(1999)为更多的细节。在这部分线性回归, 我们介绍修正的剖面最小二乘方法去估计  $\beta$  作为下式:

$$\hat{\beta} = \{ \tilde{Z}^T (I - \tilde{M}_{\tilde{B}}) \tilde{Z} - n(T-1)\Sigma_{\eta} \}^{-1} \tilde{Z}^T (I - \tilde{M}_{\tilde{B}}) \tilde{Y}. \quad (3-8)$$

进一步, 注意  $E(m_j(U_{it})) = E[Y_{it} - X_{it}^T \beta - \alpha_i | U_{it}] = E[Y_{it} - Z_{it}^T \beta - \alpha_i | U_{it}]$  和 Fan 等(2005) 的结果表明剖面最小二乘估计非参数部分  $m = [m(U_{11}), \dots, m(U_{1T}), \dots, m(U_{nT})]^T$  能被得到通过下面的方程:

$$\begin{bmatrix} I_N & \tilde{M}_{\tilde{B}} \\ S & I_N \end{bmatrix} \begin{bmatrix} B\alpha \\ m \end{bmatrix} = \begin{bmatrix} \tilde{M}_{\tilde{B}} \\ S \end{bmatrix} (Y - Z\hat{\beta}),$$

其中  $\tilde{M}_{\tilde{B}} = B(I_N - i_N i_N^T / N)(B^T B)^{-1} B^T$ 。通过 Opsomer 和 Ruppert (1997) 的结

果,  $m$  的估计可以由下式得到

$$\hat{m} = \{I - (I - S\bar{M}_B)^{-1}(I - S)\}(Y - Z\hat{\beta}). \quad (3-9)$$

有时, 它是必要的去估计误差的方差  $\sigma^2 = E(\varepsilon_{it}^2)$ , 基于模型的检验, 模型的选择程序, 单一噪声比率决定等等。因此我们想建立  $\sigma^2 = E(\varepsilon_{it}^2)$  的置信区间, 从  $E[Y_{it} - X_{it}^T\beta - B_i^T\alpha_i - m(U_{it})]^2 = \sigma^2$  和  $E[Y_{it} - Z_{it}^T\beta - B_i^T\alpha_i - m(U_{it})]^2 = \sigma^2 + \beta^T\Sigma_\eta\beta$ , 我们定义  $\sigma^2$  的估计如下

$$\hat{\sigma} = \frac{T}{N(T-1)}(Y - Z\hat{\beta} - \hat{m})^T(I - M_B)(Y - Z\hat{\beta} - \hat{m}) - \hat{\beta}^T\Sigma_\eta\hat{\beta} \quad (3-10)$$

其中  $M_B = B(B^TB)^{-1}B$ 。

### 3.3 剖面似然比检验

为获得洞察剖面似然比检验的统计结构, 为简化, 假定  $\varepsilon_{it} \in N(0, \sigma^2)$ 。事实上, 正态假定不一定是必要的, 那已经被证明Fan等(2001)和我们证明在附录里。在模型(3-1), 对数似然函数为:

$$l = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - X_{it}^T\beta - m(U_{it}) - \alpha_i)^2 = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{RSS_{11}}{2\sigma^2}.$$

不幸的是,  $X_{it}$  不能被观察。这动机我们给出

$$\begin{aligned} l &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=1}^T [(Y_{it} - Z_{it}^T\beta - m(U_{it}) - \alpha_i)^2 - \beta^T\Sigma_\eta\beta] \\ &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{RSS_{12}}{2\sigma^2}. \end{aligned} \quad (3-11)$$

在原假设(3-3)下, 它的剖面似然估计是最大化(3-11)在约束条件(3-3)下。这个解可以有下式得到

$$\hat{\beta}_0 = \hat{\beta} - \Lambda^{-1}A^T(A\Lambda^{-1}A^T)^{-1}A\hat{\beta},$$

进一步,  $\hat{m}_0 = S(Y - Z\hat{\beta}_0 - B\alpha)$ 。其中  $\Lambda = \tilde{Z}^T(I - \tilde{M}_{\tilde{B}})\tilde{Z} - n(T-1)\Sigma_\eta$ 。在原假设下, 最大化(3-11)关于  $\beta$  和  $\sigma^2$  得到剖面似然估计

$$l(H_0) = -\frac{N}{2}\log\left(\frac{2\pi}{N}\right) - \frac{N}{2}\log(RSS_0) - \frac{N}{2}.$$

其中  $RSS_0 = \sum_{i=1}^n \sum_{t=1}^T [(Y_{it} - Z_{it}^T\hat{\beta}_0 - \hat{m}_0(U_{it}))^2 - \beta^T\Sigma_\eta\beta]$ 。另一方面, 表示  $RSS_1 = \sum_{i=1}^n \sum_{t=1}^T [(Y_{it} - Z_{it}^T\hat{\beta} - \hat{m}(U_{it}))^2 - \beta^T\Sigma_\eta\beta]$  和  $l(H_1) = -\frac{N}{2}\log\left(\frac{2\pi}{N}\right) - \frac{N}{2}\log(RSS_1) - \frac{N}{2}$ 。现在, 利用上面的剖面似然, 这剖面似然比统计推断被建立:

$$\lambda_n(\beta) = l(H_1) - l(H_0) = \frac{N}{2}\log\left(\frac{RSS_0}{RSS_1}\right) \approx \frac{N}{2} \frac{RSS_0 - RSS_1}{RSS_1}. \quad (3-12)$$

### 3.4 假定和主要结论

在这一章,我们将建立前面一章建议的估计的渐近特征,在得到主要结论前,我们给出下面的假定

- (A1) 随机变量 $U_{it}$ 有一个连续密度函数 $f(\cdot)$ 伴随一个紧支撑 $\mathcal{U}$ 在 $\mathbb{R}$ 上.  $0 < \inf_{u \in \mathcal{U}} f(u) \leq \sup_{u \in \mathcal{U}} f(u) < \infty$ .
- (A2) 令 $\Phi(u) = E(X_{it}|U_{it} = u)$ . 函数 $\Phi(\cdot)$ 和 $m(\cdot)$ 在 $\mathcal{U}$ 上是二阶连续可微有界的.
- (A3)  $(\alpha_i, X_{it}, U_{it}, \varepsilon_{it})$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ 是独立同分布的. 存在 $\delta > 2$ 以至 $E\|X_{it}\|^{2+\delta} < \infty$ 和 $E|\varepsilon_{it}|^{2+\delta} < \infty$ , 其中 $\|a\| = (\sum_{i=1}^n a_i^2)^{1/2}$ 是通常的向量 $a$ 的欧氏范数.
- (A4) 核函数 $K(v)$ 是一个在连续可微的对称概率密度函数在紧支撑 $\mathcal{U}$ 上.
- (A5)  $E(Y_{it}|X_i, U_i, \alpha_i) = E(Y_{it}|X_{it}, U_{it}, \alpha_{it}) = X_{it}^T \beta + m(U_{it}) + \alpha_i$ .
- (A6)  $E|\check{X}_{it}|^{2+\delta} < \infty$ ,  $\Sigma = E[\check{X}_{it} \check{X}_{it}^T]$ 是非奇异的, 其中 $\check{X}_{it} = X_{it} - E(X_{it}|U_{it})$ .
- (A7) 窗宽 $h$ 满足 $h \rightarrow 0$ ,  $Nh^8 \rightarrow 0$ 和 $Nh^2/(\log N)^2 \rightarrow \infty$ 作为 $n \rightarrow \infty$ .

**注3.1.** 假定(A1)-(A7)看起来有点多,事实上条件是平凡的和容易被满足的.(A1)-(A2)是标准局部线性多项式估计文献常用的条件, A5暗示 $E(\varepsilon_{it}|X_i, U_i, \mu_i) = E(\varepsilon_{it}|X_{it}, U_{it}, \mu_{it}) = 0$ . (A1)-(A5)能被发现在文章Su和Ullah (2006). (A6)和(A7)已经被用Zhou等(2010).  $\square$

**定理3.1.** 假定(A1)-(A7)条件满足,对于模型(3-4),如果 $\beta_0$ 是正实值参数,因此 $\beta$ 建议的估计 $\hat{\beta}$ 是渐近正态分布的. 令

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \frac{T}{N(T-1)} \Sigma^{-1} \Sigma_1 \Sigma^{-1}) \quad (3-13)$$

其中 $\Sigma_1 = [E(\varepsilon - \eta^T \beta)^2 \Sigma + \sigma^2 \Sigma_\eta + E[(\eta \eta^T - \Sigma_\eta) \beta]^{\otimes 2}]$ 和 $A^{\otimes 2}$ 表示 $AA^T$ .

进一步,  $\Sigma^{-1} \Sigma_1 \Sigma^{-1}$ 的一致估计是 $\hat{\Sigma}^{-1} \hat{\Sigma}_1 \hat{\Sigma}^{-1}$ , 其中 $\hat{\Sigma} = \frac{T}{N(T-1)} \tilde{Z}^T (I - \tilde{M}_{\tilde{B}}) \tilde{Z} - \Sigma_\eta$ 和 $\hat{\Sigma}_1 = \frac{1}{N} [\tilde{Z} \text{diag}(\tilde{Y}_{11} - \tilde{Z}_{11}^T \hat{\beta}, \dots, \tilde{Y}_{1T} - \tilde{Z}_{1T}^T \hat{\beta}, \dots, \tilde{Y}_{nT} - \tilde{Z}_{nT}^T \hat{\beta}) + l_N(\Sigma_\eta \hat{\beta})^T]^{\otimes 2}$ .



**定理 3.2.** 假定(A1)-(A7)条件满足, 因此 $\hat{m}$ 的最小二乘估计的损失风险如下:

$$MSE\{\hat{m}|U_{11}, \dots, U_{1T}, \dots, U_{nT}\} \leq \frac{T}{(\sqrt{T}-1)^2} \left\{ \frac{\mu_2^2 h^4}{4} E[m''(u)] + \frac{(\sigma^2 + \beta^T \Sigma_\eta \beta) \nu_0 |U|}{Nh} \right\} + o_p\left(h^4 + \frac{1}{Nh}\right),$$

其中

$$\mu_i = \int u^i K(u) du, \quad \nu_i = \int u^i K^2(u) du,$$

$$\hat{m} = [\hat{m}(U_{11}), \dots, \hat{m}(U_{1T}), \dots, \hat{m}(U_{nT})]^T,$$

$$MSE\{\hat{m}|U_{11}, \dots, U_{1T}, \dots, U_{nT}\} = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T E[\hat{m}(U_{it}) - m(U_{it}) | U_{11}, \dots, U_{1T}, \dots, U_{nT}]^2.$$

**定理 3.3.** 假定(A1)-(A7)条件满足, 因此得到

$$\sqrt{N}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, \Theta), \quad (3-14)$$

其中 $\Theta = E(\varepsilon - \eta^T \beta)^4 + (3 - T)/(T - 1)(\sigma^2 + \beta^T \Sigma_\eta \beta)^2$ . 定义 $\hat{\Psi} = (\hat{\psi}_{11}, \dots, \hat{\psi}_{1T}, \dots, \hat{\psi}_{nT})^T = (I - M_B)(Y - Z\hat{\beta} - M)$ 和

$$\hat{\Theta} = \left(1 + \frac{6}{T^2} - \frac{3}{T^3} - \frac{4}{T}\right)^{-1} \left[ \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \hat{\psi}_{it}^4 - \frac{T-1}{T^2} \left(6 - \frac{9}{T}\right) (\hat{\sigma}^2 + \hat{\beta}^T \Sigma_\eta \hat{\beta}) \right] + \frac{3-T}{T-1} (\hat{\sigma}^2 + \hat{\beta}^T \Sigma_\eta \hat{\beta}),$$

因此 $\Theta$ 的一致估计为 $\hat{\Theta}$ .

**定理 3.4.** 假定(A1)-(A7)条件满足, 因此在原假设(3-3)条件下, 我们得到

$$2\lambda_n(\beta) \xrightarrow{d} \sum_{i=1}^l \omega_i \chi_{i1}^2, \quad (3-15)$$

其中 $[\sigma^2 A(\Sigma + \Sigma_\eta)^{-1} A^T]^{-1} (A \frac{T}{N(T-1)} \Sigma^{-1} \Sigma_1 \Sigma^{-1} A^T)$ 的特征值是 $\omega_i$ , 并且 $\chi_{i1}^2$ 是自由度为1的中心卡方分布; 和在备则假设条件下,

$$2\lambda_n(\beta) - N\sigma^{-2} \beta^T A^T [A(\Sigma + \Sigma_\eta)^{-1} A^T]^{-1} A\beta \xrightarrow{d} \sum_{i=1}^l \omega_i \chi_{i1}^2. \quad (3-16)$$

### 3.5 模拟研究与实证

#### 3.5.1 Monte Carlo 模拟

在这一节，我们将做一些模拟实验去研究对于有限的样本我们在3.2节建议的估计 $\hat{\beta}$ ,  $\hat{m}$  和  $\hat{\sigma}^2$ 。首先，我们考虑下面的带固定效应的面板数据含误差变量部分线性模型：

$$\begin{cases} Y_{it} = X_{it}^T \beta + m(U_{it}) + \alpha_i + \varepsilon_{it}, \\ Z_{it} = X_{it} + \eta_{it} \end{cases} \quad i = 1, \dots, n; t = 1, \dots, T. \quad (3-17)$$

其中  $\beta = (\beta_1, \beta_2)^T = (1, \sqrt{2})^T / \sqrt{3}$ ,  $m(U_{it}) = \cos(2\pi U_{it})$ ,  $U_{it} \sim U(0, 1)$ ,  $X_{it} = (X_{it1}, X_{it2})^T \stackrel{i.i.d.}{\sim} N((1, 1)^T, \text{diag}(4, 4))$ ,  $\alpha_i = \rho \bar{X}_i + w_i$  with  $\rho = 0.5, 1$  和  $w_i \sim N(0, 1)$  for  $i = 1, 2, \dots, n$ , 和  $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it1}$ . 我们用  $\rho$  来控制  $\alpha_i$  和  $\bar{X}_i$  的相关系数。测量误差  $\eta_{it} \sim N(0, \Sigma_\eta)$ , 其中我们取  $\Sigma_\eta = 0.2^2 I_2$  和  $0.4^2 I_2$  去代表不同水平的测量误差。

现在我们做模拟，选取的样本大小分别为  $(n, T) = (50, 4), (50, 6), ((100, 4), (100, 6))$ , 和我们选取的核为 Epanechnikov 核  $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \leq 1\}$ 。采取去掉第  $i$  个观测的交错核实验验证 (cross validation, CV) 方法的窗宽  $CV(h)$  将被获得用最小化下式

$$CV(h) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - X_{it}^T \hat{\beta}_{[i]} - \hat{m}_{[i]}(U_{it}))^2,$$

其中  $\hat{\beta}_{[i]}$  和  $\hat{m}_{[i]}(U_{it})$  分别是  $\beta$  和  $m(U_{it})$  的估计，那计算所有的测量但不计算  $i$ th 观测。我们将做 500 次重复实验在置信水平为  $1 - \alpha = 0.95$ 。我们计算我们在 3.2 节建议的参数  $\beta_1$  和  $\beta_2$  和误差的方差  $\sigma^2$  的样本均值和样本标准差 (SD)，为了对比建议的修正的剖面最小二乘估计方法和已经存在的方法估计参数部分和误差的方差的估计。我们也计算修正的剖面最小二乘估计的样本均值和标准差和单纯的估计 (忽略误差变量)。 $(\beta_1, \beta_2)^T$  和  $\sigma^2$  的单纯的估计定义如下：

$$\tilde{\beta} = \{\tilde{Z}^T (I - \tilde{M}_{\tilde{\beta}}) \tilde{Z}\}^{-1} \tilde{Z}^T (I - \tilde{M}_{\tilde{\beta}}) \tilde{Y},$$

和

$$\hat{\sigma} = \frac{T}{N(T-1)}(Y - Z\hat{\beta} - \hat{m})'(I - M_B)(Y - Z\hat{\beta} - \hat{m}),$$

其中

$$\hat{m} = \{I - (I - SM_B)^{-1}(I - S)\}(Y - Z\hat{\beta}).$$

模拟的结果总结在下表3.1和表3.2, 从表3.1和表3.2, 我们做以下的观察:

1. 有一个明显的区别在单纯估计的这个标准差 (SD)与那我们建议的估计。我们想这个区别因为测量误差。
2. 修正的最小二乘估计估计参数和误差方差比单纯的估计有更小的偏差和更小的标准差。
3. 当 $\rho$ 增大时, 建议的估计比单纯的估计表现更好。标准差SD变的稍微有点大。

从图3.1, 我们能看出修正的剖面最小二乘估计估计非参数部分优于单纯的剖面最小二乘估计。后者是有偏的。

### 3.5.2 基于英国各地区气象数据的实证

为检查所提出的方法的性能, 我们分析了气候数据集来自英国气象局网站: <http://www.metoffice.gov.uk/climate/uk/>。这个数据集包括平均最高气温(摄氏度), 平均最低气温、降雨量(毫米), 总下雨量 $\geq 1.0mm$ 的天数, 总日照时数(小时)和空气霜冻天数(天)从覆盖全英国的37个气象站。这项研究的目的是描述的共同趋势, 平均最高气温系列的平均最高气温与总雨量和总日照的关系, 在一月/2006日至十二月/2015。从16个站的数据被选择根据数据的可用性(记录从不同的站在不同的时间开始)。这个数据集已经被研究李等(2011)。令 $Y_{it}$  是第 $i$ 个站在第 $t$ 月测量观察的对数转换的平均最高气温,  $X_{1it}$  是第 $i$ 个站在第 $t$ 月测量观察对数转换的总日照时数。  $X_{2it}$  是第 $i$ 个站在第 $t$ 月测量观察对数转换的总降水量,  $X_{3it}$  是第 $i$ 个站在第 $t$ 月测量观察对数转换的下雨 $\geq 1.0mm$ 的天数。我们考虑下面的模型:

表 3.1 修正的剖面最小二乘估计和单纯的估计估计的样本均值和标准差当  $\rho = 0.5$ .

$\Sigma_\eta$	$(n, T)$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\check{\beta}_1$	$\check{\beta}_2$	$\check{\sigma}^2$
$0.2^2 I_2$	(50, 4)	Mean	0.5797	0.8201	0.9852	0.5810	0.8221	0.9452
		SD	0.0199	0.0210	0.1156	0.0201	0.0217	0.1270
	(50, 6)	Mean	0.5791	0.8189	0.9912	0.5795	0.8206	0.9510
		SD	0.0173	0.0165	0.0991	0.0197	0.0186	0.1011
	(100, 4)	Mean	0.5782	0.8173	0.9927	0.5785	0.8193	0.9525
		SD	0.0154	0.0159	0.0814	0.0156	0.0161	0.0946
(100, 6)	Mean	0.5772	0.8165	0.9941	0.5716	0.8164	0.9539	
	SD	0.0109	0.0126	0.0636	0.0112	0.0128	0.0790	
$0.4^2 I_2$	(50, 4)	Mean	0.5840	0.8250	0.9848	0.5868	0.8234	0.8320
		SD	0.0247	0.0218	0.1210	0.0251	0.0247	0.2099
	(50, 6)	Mean	0.5834	0.8193	0.9875	0.5833	0.8220	0.8311
		SD	0.0196	0.0183	0.0997	0.0198	0.0192	0.1920
	(100, 4)	Mean	0.5816	0.81817	0.9913	0.5828	0.8254	0.8345
		SD	0.0174	0.0170	0.0868	0.0179	0.0177	0.1869
	(100, 6)	Mean	0.5794	0.8174	0.9928	0.5812	0.8256	0.8258
		SD	0.0125	0.0135	0.0650	0.0132	0.0135	0.1841

$$\begin{cases} Y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + X_{3it}\beta_3 + m(U_{it}) + \alpha_i + \varepsilon_{it}, \\ Z_{1it} = X_{1it} + \eta_{it}, \end{cases} \quad i = 1, \dots, n; t = 1, \dots, T. \quad (3-18)$$

其中  $U_{it} = t_{ij}/T$  和  $\alpha_i = \rho \bar{X}_i$ , 对于  $i = 1, 2, \dots, n$ , 和  $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it1}$ . 测量误差  $\eta_{it} \sim N(0, \Sigma_\eta)$ , 其中我们取  $\Sigma_\eta = 0.2^2 I_2$ . 一个自然的问题是, 是否总日照时数、总降水量和下雨  $\geq 1.0mm$  的天数是否是统计显著的. 对于回答这个问题, 我们建议的PLR检验方法被应用. 作为一个结果, 这个PLR检验统计量分别是  $\lambda_n(\beta_1) = 80.12$ ,  $\lambda_n(\beta_2) = 0.8490$  和  $\lambda_n(\beta_3) = 5.1266$ , 那提供明显的证明那变量  $X_{1it}$  和  $X_{3it}$  对于平均最高气温在置信水平0.9是显著的, 那也表明变量  $X_{2it}$  的系数是零. 我们利用在3.2节给的修正的最小二乘方法对于模型(3-17, 和获得参数  $\beta = (\beta_1, \beta_3)$  的估计  $\hat{\beta} = (0.6861, 0.2365)$  伴随  $SD = 0.040$ . 这表明下雨  $\geq 1.0mm$  的天数没有明显的影响对于平均气温, 但总日照时数有非常正

表 3.2 修正的剖面最小二乘估计和单纯的估计估计的样本均值和标准差当  $\rho = 1$ .

$\Sigma_\eta$	$(n, T)$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\check{\beta}_1$	$\check{\beta}_2$	$\check{\sigma}^2$
$0.2^2 I_2$	(50, 4)	Mean	0.5794	0.8173	0.9854	0.5797	0.8264	0.9552
		SD	0.0234	0.0268	0.1228	0.0255	0.0270	0.1356
	(50, 6)	Mean	0.5792	0.8184	0.9841	0.5806	0.8204	0.9670
		SD	0.0203	0.0217	0.1027	0.0204	0.0218	0.1125
	(100, 4)	Mean	0.5785	0.8176	0.9879	0.5770	0.8198	0.9479
		SD	0.0156	0.0162	0.0997	0.0157	0.0163	0.1059
(100, 6)	Mean	0.5764	0.8174	0.9933	0.5768	0.8194	0.9542	
	SD	0.0134	0.0143	0.0663	0.0146	0.0149	0.0805	
$0.4^2 I_2$	(50, 4)	Mean	0.5844	0.8250	0.9805	0.5825	0.8256	0.8302
		SD	0.0260	0.0243	0.1147	0.0268	0.0258	0.2049
	(50, 6)	Mean	0.5832	0.8247	0.9832	0.5837	0.8255	0.8340
		SD	0.0204	0.0216	0.0845	0.0217	0.0206	0.1908
	(100, 4)	Mean	0.5799	0.8160	0.9891	0.5794	0.8214	0.8390
		SD	0.0182	0.0168	0.0839	0.0172	0.0171	0.1860
(100, 6)	Mean	0.5792	0.8174	0.9910	0.5802	0.8236	0.8644	
	SD	0.0139	0.0144	0.0739	0.0155	0.0151	0.1778	

面的影响对于平均最高气温，换句话说，更长的日照时间得到更高的平均气温。并且， $m(\cdot)$  的估计显示在图3.2。从图3.2，我们能看见从2006年开始到2008年末，有一个上升趋势对于每月的平均最高气温，并且，从2009年开始到2011年年末有一个轻微的下降，并且从2013年开始到2015年年末有一个上升最高温度。同样从图3.2，我们也能看出，修正的剖面最小二乘估计非参数部分由于单纯的剖面最小二乘估计，后者有一个更大的波动。

### 3.6 主要结果的证明

为了表示的方便，我们简化符号，令  $\vartheta_k = \int u^k K(u) dt$ ,  $c_N = \{\log(1/h)/(Nh)\}^{1/2} + h^2$  和集合  $\Phi(U) = E(\mathbf{X}|U)$ .

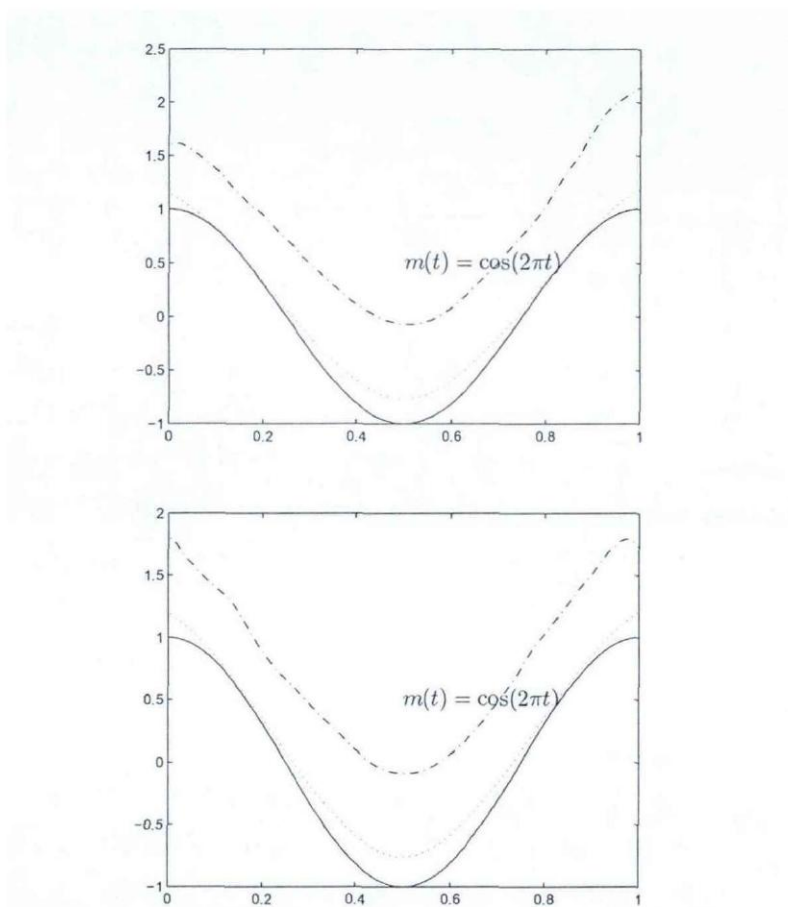


图 3.1 非参数  $m(\cdot)$  的估计当  $(n, T) = (50, 4)$  和  $\Sigma_\eta = 0.2$  (左图) 和  $\Sigma_\eta = 0.4$  (右图).  $\cos(2\pi t)$  (实线), 建议的估计(点线) 和单纯估计(点划线).

引理 3.1. 假设条件(A1)-(A5) 都拥有, 因此我们能得到

$$\sup_{U \in \mathcal{P}} \frac{1}{N} D_u^T W_u D_u = f(u) \begin{pmatrix} 1 & 0 \\ 0 & \vartheta_2 \end{pmatrix} + O_p(c_N),$$

$$\sup_{U \in \mathcal{P}} \frac{1}{N} D_u^T W_u X = f(u) E(X|U) (1 \ 0)^T + O_p(c_N).$$

证明. 注意

$$D_u^T W_u D_u = \begin{pmatrix} \sum_{i=1}^n \sum_{t=1}^T K_h(U_{it} - u) & \sum_{i=1}^n \sum_{t=1}^T \left(\frac{U_{it}-u}{h}\right) K_h(U_{it} - u) \\ \sum_{i=1}^n \sum_{t=1}^T \left(\frac{U_{it}-u}{h}\right) K_h(U_{it} - u) & \sum_{i=1}^n \sum_{t=1}^T \left(\frac{U_{it}-u}{h}\right)^2 K_h(U_{it} - u) \end{pmatrix}.$$

上面矩证的每一个元素是核回归的形式。相似于Fan 和Huang (2005)的引理2 证明, 我们能证明结论。

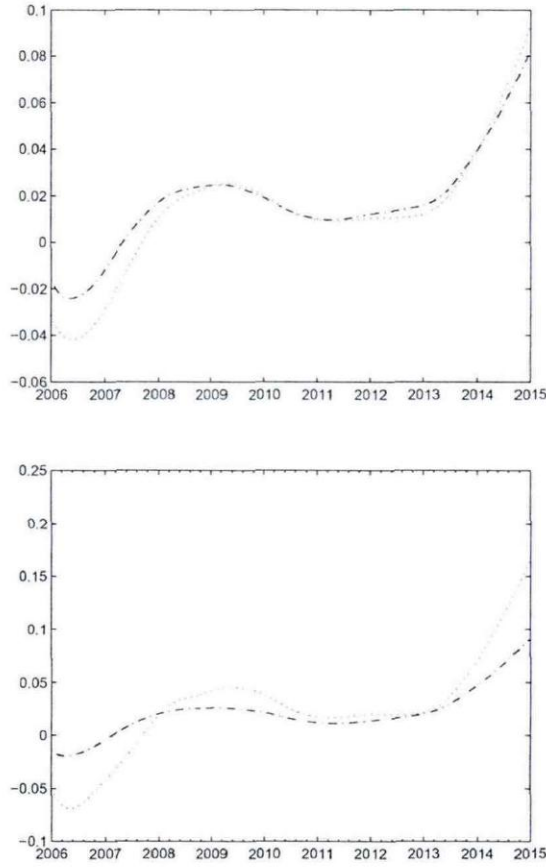


图 3.2 非参数部分  $m(\cdot)$  的估计伴随  $\Sigma_\eta = 0.2$  (左图) 和  $\Sigma_\eta = 0.4$  (右图). 我们建议的估计(dot-dashed curve) 和单纯的估计(dotted curve).

引理 3.2. 假设条件(A1)-(A7) 都拥有. 因此我们能得到

$$\frac{1}{N} \tilde{Z}^\tau (I - \tilde{M}_{\tilde{B}}) \tilde{Z} \xrightarrow{p} \frac{T-1}{T} (\Sigma + \Sigma_\eta),$$

其中  $\Sigma = E\{[X - E(X|U)]^\tau [X - E(X|U)]\}$ .

证明. 由引理3.1, 我们能得到

$$[I_q \ 0_q^\tau]^{-1} (D_U^\tau W_U D_U)^{-1} D_U^\tau W_U X = E(X|U) + O_p(c_N).$$

因此我们有

$$SX = [\Phi(U_{11}), \dots, \Phi(U_{1T}), \dots, \Phi(U_{nT})]^\tau + O_p(c_N),$$

$$\tilde{X} = [X_{11} - \Phi(U_{11}), \dots, X_{1T} - \Phi(U_{1T}), \dots, X_{nT} - \Phi(U_{nT})]^\tau + O_p(c_N),$$

和

$$\tilde{Z} = \tilde{X} + \eta + O_p(c_N) \triangleq A + O_p(c_N).$$

通过利用大数定律, 我们得到

$$\frac{1}{N} \tilde{Z}^{\tau} \tilde{Z} = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{ [X_{it} - \Phi(U_{it})]^{\tau} [X_{it} - \Phi(U_{it})] + \eta_{it}^{\tau} \eta_{it} \} + O_p(c_N) \xrightarrow{p} \Sigma + \Sigma_{\eta}. \quad (3-19)$$

因此, 为证明此引理, 我们考虑  $N^{-1} \tilde{Z}^{\tau} \tilde{M}_{\tilde{B}} \tilde{Z}$  极限. 它是容易去显示  $N^{-1} \tilde{Z}^{\tau} \tilde{M}_{\tilde{B}} \tilde{Z} = N^{-1} A^{\tau} \tilde{M}_{\tilde{B}} A + O_p(c_N)$ . 令  $(\tilde{M}_{\tilde{B}})_{e_{kl}e_{it}} \triangleq m_{e_{kl}e_{it}}$  和  $(A)_{it} \triangleq a_{it} = \tilde{Z}_{it}$ , 其中  $e_{kl} = (k-1)T + l$ . 因此

$$\begin{aligned} \frac{1}{N} A^{\tau} \tilde{M}_{\tilde{B}} A &= \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^n \sum_{l=1}^T a_{kl} m_{e_{kl}e_{it}} a_{it} = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T a_{it} m_{e_{it}e_{it}} a_{it} \\ &\quad + \frac{1}{N} \sum_{e_{kl} \neq e_{it}} \sum_{k=1}^n \sum_{l=1}^T a_{kl} m_{e_{kl}e_{it}} a_{it} \triangleq I_1 + I_2. \end{aligned}$$

对于项  $I_2$ , 我们有

$$EI_2^2 = \frac{1}{N^2} E \left[ \sum_{e_{kl} \neq e_{it}} \sum_{k=1}^n \sum_{l=1}^T \sum_{e_{rs} \neq e_{uv}} \sum_{r=1}^n \sum_{s=1}^T a_{kl} m_{e_{kl}e_{it}} a_{it} a_{rs} m_{e_{rs}e_{uv}} a_{uv} \right].$$

注意  $(X_{11}, U_{11}), \dots, (X_{nT}, U_{nT})$  是独立同分布的. 和  $E(a_{it}|U_{it}) = 0$ , when  $e_{kl} \neq e_{rs}$  和  $e_{it} \neq e_{uv}$ , 我们得到

$$\begin{aligned} E(a_{kl} m_{e_{kl}e_{it}} a_{it} a_{rs} m_{e_{rs}e_{uv}} a_{uv}) &= E[m_{e_{kl}e_{it}} m_{e_{rs}e_{uv}} E(a_{kl} a_{it} a_{rs} a_{uv} | U_{kl}, U_{it}, U_{rs}, U_{uv})] \\ &= E[m_{e_{kl}e_{it}} m_{e_{rs}e_{uv}} E(a_{it} a_{rs} a_{uv} | U_{it}, U_{rs}, U_{uv}) E(a_{kl} | U_{kl})] = 0. \end{aligned}$$

利用同样结论和  $m_{e_{kl}e_{it}} = m_{e_{it}e_{kl}}$ , 我们有

$$EI_2^2 = \frac{1}{N^2} \sum_{e_{kl} \neq e_{it}} \sum_{k=1}^n \sum_{l=1}^T E(a_{kl} m_{e_{kl}e_{it}} a_{it})^2 + \frac{1}{N^2} \sum_{e_{kl} \neq e_{it}} \sum_{k=1}^n \sum_{l=1}^T E(m_{e_{kl}e_{it}}^2 a_{kl} a_{it} a_{it} a_{kl}).$$

通过假定(A3), 我们得到

$$EI_2^2 \leq \frac{2c}{N^2} \sum_{e_{kl} \neq e_{it}} E(m_{e_{kl}e_{it}})^2 \leq \frac{1}{N^2} \text{tr}(\tilde{M}^2) \leq \frac{2c}{N}, \quad (3-20)$$

其中  $c$  是一个实数. 因此

$$EI_2 = o_p(1). \quad (3-21)$$

注意  $I_1$  能被分解为



$$I_1 = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T m_{e_{it}e_{it}} [a_{it}a_{it} - E(a_{it}a_{it})] + \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T m_{e_{it}e_{it}} E(a_{it}a_{it}) \triangleq \Pi_1 + \Pi_2.$$

通过 $S$ 的定义,它是容易得到

$$S = (S_{11}, \dots, S_{1T}, \dots, S_{nT})^T [I + \text{diag}[O_p(c_n)]],$$

其中

$$S_{it} = \left( \frac{K_h(U_{11} - U_{it})}{Nf(U_{it})}, \dots, \frac{K_h(U_{1T} - U_{it})}{Nf(U_{it})}, \dots, \frac{K_h(U_{nT} - U_{it})}{Nf(U_{it})} \right)^T.$$

令 $B_1$ 是 $B$ 第一列向量,因此我们有

$$\begin{aligned} B_1^T (I - S^T)(I - S) B_1 &= \left\{ T - 2 \sum_{t=1}^T \left[ \sum_{l=1}^T \frac{K_h(U_{1l} - U_{1t})}{Nf(U_{1t})} \right] \right. \\ &\quad \left. + \sum_{i=1}^n \sum_{t=1}^T \left[ \sum_{l=1}^T \frac{K_h(U_{1l} - U_{it})}{Nf(U_{it})} \right]^2 \right\} \{1 + O_p(c_N)\}. \end{aligned}$$

因为

$$\sum_{t=1}^T \left[ \sum_{l=1}^T \frac{K_h(U_{1l} - U_{1t})}{Nf(U_{1t})} \right] \{1 + O_p(c_N)\} = O_p\left(\frac{1}{Nh}\right),$$

和

$$\sum_{i=1}^n \sum_{t=1}^T \left[ \sum_{l=1}^T \frac{K_h(U_{1l} - U_{it})}{Nf(U_{it})} \right]^2 \{1 + O_p(c_N)\} = O_p\left(\frac{1}{Nh}\right),$$

我们得到

$$B_1^T (I - S^T)(I - S) B_1 = T \left[ 1 + O_p\left(\frac{1}{Nh}\right) \right].$$

考虑投影矩阵,对于 $i = 1, \dots, T$ ,我们得到

$$\begin{aligned} (\widetilde{M}_{\widetilde{B}_1})_{ii} &= (I - S) B_1 [B_1^T (I - S^T)(I - S) B_1]^{-1} B_1^T (I - S^T) \\ &= \frac{1}{T} \left[ 1 + O_p\left(\frac{1}{Nh}\right) \right] \left[ 1 - \sum_{l=1}^T \frac{K_h(U_{1l} - U_{1t})}{Nf(U_{1t})} \{1 + O_p(c_N)\} \right]^2 \\ &= \frac{1}{T} + O_p\left(\frac{1}{Nh}\right). \end{aligned}$$

因为 $\widetilde{B}_1$ 是 $\widetilde{B}$ 第一列向量,这很容易证明 $\widetilde{M}_{\widetilde{B}} \widetilde{M}_{\widetilde{B}_1} = \widetilde{M}_{\widetilde{B}_1} \widetilde{M}_{\widetilde{B}} = \widetilde{M}_{\widetilde{B}_1}$ .因此, $\widetilde{M}_{\widetilde{B}} - \widetilde{M}_{\widetilde{B}_1}$ 也是投影矩阵.因此 $\widetilde{M}_{\widetilde{B}} - \widetilde{M}_{\widetilde{B}_1} = (\widetilde{M}_{\widetilde{B}} - \widetilde{M}_{\widetilde{B}_1})^2 \geq 0$ .我们获得 $(\widetilde{M}_{\widetilde{B}})_{ii} \geq (\widetilde{M}_{\widetilde{B}_1})_{ii} = \frac{1}{T} + O_p\left(\frac{1}{Nh}\right)$ ,  $i = 1, \dots, T$ .通过一个类似的结论,我们可以证明 $(\widetilde{M}_{\widetilde{B}})_{ii} \geq \frac{1}{T} + O_p\left(\frac{1}{Nh}\right)$ ,  $i = T + 1, \dots, N$ .于是,我们有

$$1 \geq m_{e_{it}e_{it}} \geq \frac{1}{T} + O_p\left(\frac{1}{Nh}\right),$$

因此, 这很容易证明

$$\text{tr}(\widetilde{M}_{\widetilde{B}}) = \sum_{i=1}^n \sum_{t=1}^T m_{e_{it}e_{it}} \geq \frac{N}{T} + O_p\left(\frac{1}{Nh}\right).$$

于是, 我们有

$$\Pi_2 = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T m_{e_{it}e_{it}} E(a_{it}a_{it}) = \frac{1}{T}(\Sigma + \Sigma_\eta) + O_p\left(\frac{1}{Nh}\right). \quad (3-22)$$

利用(3-20), 这很容易证明

$$\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (m_{e_{it}e_{it}} - T^{-1}) = O_p\left(\frac{1}{Nh^2}\right),$$

利用大数定律,  $\Pi_1$  是有界的

$$\begin{aligned} \Pi_1 &= \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (m_{e_{it}e_{it}} - T^{-1})(a_{it}a_{it} - E(a_{it}a_{it})) + \frac{1}{NT} \sum_{i=1}^n \sum_{t=1}^T (a_{it}a_{it} - E(a_{it}a_{it})) \\ &\leq \frac{1}{N} \left[ \sum_{i=1}^n \sum_{t=1}^T (m_{e_{it}e_{it}} - T^{-1})^2 \right]^{1/2} + o_p(1) = o_p(1). \end{aligned} \quad (3-23)$$

由式子(3-22) 和式子(3-23), 我们有

$$I_1 = \frac{1}{T}(\Sigma + \Sigma_\eta) + o_p(1). \quad (3-24)$$

由式子(3-19), (3-21) 和(3-24), 这个引理的结论被证明。

**引理 3.3.** 假定(A1)-(A7) 条件满足, 我们得到

$$\frac{1}{N} \widetilde{Z}^\tau (I - \widetilde{M}_{\widetilde{B}})(I - S)m = O_p(c_N).$$

证明. 利用引理3.1 和引理3.2 的一个相似的证明, 我们能得到

$$(I - S)m = O_p(c_N).$$

定义  $\Omega = (\Omega_{11}, \dots, \Omega_{1T}, \dots, \Omega_{nT}) = (I - \widetilde{M}_{\widetilde{B}})(I - S)m$ , 我们得到

$$\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \Omega_{it}^2 \leq \frac{1}{N} \|(I - S)m\|^2 = O_p(c_N^2),$$

和

$$\begin{aligned} \frac{1}{N} \widetilde{Z}^\tau (I - \widetilde{M}_{\widetilde{B}})(I - S)m &= \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \Omega_{it} [X_{it} - \Phi(U_{it}) + \eta_{it} + O_p(c_N)] \\ &= \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \Omega_{it} [X_{it} - \Phi(U_{it}) + \eta_{it}] + O_p(c_N^2). \end{aligned}$$

我们现在处置第一项, 通过假定(A6), 它是容易的看出

$$E\left\{\frac{1}{N}\sum_{i=1}^n\sum_{t=1}^T\Omega_{it}[X_{it}-\Phi(U_{it})+\eta_{it}]\right\}^2\leq cN^{-2}\sum_{i=1}^n\sum_{t=1}^T\Omega_{it}=O(N^{-1}c_N^2)=o_p(1).$$

因此, 引理3.3被证明。

**引理 3.4.** 假定(A1)-(A7) 条件满足, 我们得到

$$\frac{1}{\sqrt{N}}\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(I-S)\varepsilon\stackrel{p}{\rightarrow}N(0,\frac{T-1}{T}\sigma^2(\Sigma+\Sigma_\eta)).$$

**证明.** 利用引理3.1, 我们得到 $S\varepsilon=O_p(c_N)$ . 相似于引理3.3 的证明和在假定(A7)的条件下, 我们得到 $\frac{1}{\sqrt{N}}\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})S\varepsilon=O(\sqrt{N}c_N^2)=o_p(1)$ . 因此, 由引理3.2 和中心极限定理, 我们得到:

$$\frac{1}{\sqrt{N}}\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(I-S)\varepsilon=\frac{1}{\sqrt{N}}\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\varepsilon+o_p(1)\stackrel{p}{\rightarrow}N(0,\frac{T-1}{T}\sigma^2(\Sigma+\Sigma_\eta)).$$

**证明定理3.1.** 既然 $\tilde{Y}=\tilde{B}\alpha+\tilde{X}\beta+(I-S)m+\tilde{\varepsilon}$  和 $(I-\tilde{M}_{\tilde{B}})\tilde{B}=0$ , 由 $\hat{\beta}$  的定义, 我们得到

$$\begin{aligned}\hat{\beta}-\beta&=[\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\tilde{Z}-n(T-1)\Sigma_\eta]^{-1}\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})[(\tilde{X}-\tilde{Z})\beta+(I-S)m+\tilde{\varepsilon}] \\ &+[\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\tilde{Z}-n(T-1)\Sigma_\eta]^{-1}n(T-1)\Sigma_\eta\beta \\ &=[\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\tilde{Z}-n(T-1)\Sigma_\eta]^{-1}\cdot[\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(\tilde{X}-\tilde{Z})\beta \\ &+\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(I-S)m+\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\tilde{\varepsilon}+n(T-1)\Sigma_\eta\beta].\end{aligned}$$

由引理3.2-3.4, 我们得到

$$\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(\tilde{X}-\tilde{Z})\beta=-\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})\eta\beta+O_p(c_N).$$

因此, 利用引理3.1和这个 $(A+aB)^{-1}\rightarrow A^{-1}$  作为 $a\rightarrow 0$  结论, 我们得到

$$\begin{aligned}\sqrt{N}(\hat{\beta}-\beta)&=[\frac{T-1}{T}\Sigma]^{-1}\cdot[\tilde{Z}^\tau(I-\tilde{M}_{\tilde{B}})(\varepsilon-\eta\beta)+n(T-1)\Sigma_\eta]+o_p(1) \\ &=[\frac{T-1}{T}\Sigma]^{-1}\cdot[(X-E(X|U))(I-\tilde{M}_{\tilde{B}})(\varepsilon-\eta\beta) \\ &+\eta^\tau(I-\tilde{M}_{\tilde{B}})\varepsilon+\{n(T-1)\Sigma_\eta-\eta^\tau(I-\tilde{M}_{\tilde{B}})\}\beta]+o_p(1) \\ &=[\frac{T-1}{T}\Sigma]^{-1}\cdot J+o_p(1).\end{aligned}$$

注意

$$\begin{aligned}Cov(J)&=E[(X-E(X|U))(I-\tilde{M}_{\tilde{B}})(\varepsilon-\eta\beta)]^{\oplus 2} \\ &+E[\eta^\tau(I-\tilde{M}_{\tilde{B}})\varepsilon]^{\oplus 2}+E[n(T-1)\Sigma_\eta-\eta^\tau(I-\tilde{M}_{\tilde{B}})\eta\beta]^{\oplus 2}.\end{aligned}$$

因此,

$$\lim_{n \rightarrow \infty} \text{Cov}\left(\frac{1}{\sqrt{N}}J\right) = \frac{T-1}{T} \{E(\varepsilon - \eta\beta)^2 \Sigma + \sigma^2 \Sigma_\eta + E[(\eta\eta^\tau - \Sigma_\eta)\beta]^{\otimes 2}\} = \frac{T-1}{T} \Sigma_1.$$

引用Slutsky定理和中心极限定理, 我们得到想要的结论。同样可以证明  $\Sigma^{-1}\Sigma_1\Sigma^{-1}$  的一致估计是  $\widehat{\Sigma}^{-1}\widehat{\Sigma}_1\widehat{\Sigma}^{-1}$ , 我们在这里省略证明的细节。

**引理 3.5.** 假定(A1)-(A7)条件满足, 我们得到

$$\lambda[(I - S\bar{M})(I - S\bar{M})^\tau] = \frac{\sqrt{T} - 1}{T} + O_p(c_N),$$

其中  $\lambda(A)$  表示向量  $A$  的特征值。

**证明.** 因为  $\bar{M} = B(I_N - i_N i_N^\tau / N)(B^\tau B)^{-1} B^\tau = \frac{1}{T} I_n \oplus (i_T i_T^\tau) - \frac{1}{N} i_N i_N^\tau$  和  $S i_N = i_N$ , 很容易证明

$$\begin{aligned} S\bar{M}S^\tau &= \frac{1}{T} S[I_n \oplus (i_T i_T^\tau)]S^\tau - \frac{1}{N} i_N i_N^\tau = \frac{1}{T} S(I - \frac{1}{N} i_N i_N^\tau)S^\tau + \frac{1}{T} S[I_n \oplus (i_T i_T^\tau) - I]S^\tau \\ &\quad - (1 - \frac{1}{T}) \frac{1}{N} i_N i_N^\tau \triangleq K_1 + K_2 + K_3. \end{aligned}$$

对于项  $K_2$ , 我们有

$$(K_2)_{ij} = \sum_{l=1}^N \sum_{k=1}^N \frac{1}{T} S_{il}[I_n \oplus (i_T i_T^\tau) - I]_{lk} S_{jk} = \frac{1}{T} \sum_{l=1}^T \sum_{k=1, k \neq l}^T \sum_{s=0}^{n-1} S_{i(sn+i)} S_{j(sn+k)}.$$

利用引理3.1, 我们有

$$\begin{aligned} \sum_{s=0}^{n-1} S_{i(sn+i)} S_{j(sn+k)} &= \left[ \sum_{s=0}^{n-1} \frac{k_h(U_{i(sn+i)} - u)}{Nf(u)} \frac{k_h(U_{j(sn+k)} - u)}{Nf(u)} \right] [1 + O_p(c_N)] \\ &= \frac{T}{N^2} [1 + O_p(c_N)] = \frac{1}{NT} [1 + O_p(c_N)], \end{aligned}$$

那对于  $i, j = 1, \dots, n$  是一致的。因此

$$(K_2)_{ij} = \frac{1}{T} \sum_{l=1}^T \sum_{k=1, k \neq l}^T \frac{1}{NT} [1 + O_p(c_N)] = \frac{T-1}{NT} [1 + O_p(c_N)],$$

或

$$(K_2) = \frac{T-1}{N} \frac{1}{N} i_N i_N^\tau [1 + O_p(c_N)] = -K_3 [1 + O_p(c_N)].$$

因此,

$$S\bar{M}S^\tau = K_1 + K_3 O_p(c_N).$$

它是明显的  $K_1$  的特征值满足  $0 \leq \lambda(K_1) \leq \frac{1}{T}$ 。相似于Fan等(2005)的引理A.5的证明, 我们有

$$\lambda[(I - S\bar{M})(I - S\bar{M})^\tau] \geq \frac{(\sqrt{T} - 1)^2}{T^2} + O_p(c_N).$$

证明定理3.2. 根据 $\hat{m}$ 的定义, 我们有

$$\begin{aligned}\hat{m} &= \{I - (I - S\bar{M}_B)^{-1}(I - S)\}(B\alpha + m + \varepsilon) \\ &= (I - S\bar{M}_B)^{-1}S(I - \bar{M}_B)B\alpha_i + \{I - (I - S\bar{M}_B)^{-1}(I - S)\}m \\ &\quad + \{I - (I - S\bar{M}_B)^{-1}(I - S)\}\varepsilon.\end{aligned}$$

因为 $\sum_{i=1}^n \alpha_i = 0$ , 容易证明 $(I - S\bar{M}_B)^{-1}S(I - \bar{M}_B)B\alpha_i = 0$ . 因此我们有

$$\hat{m} - m = -(I - S\bar{M}_B)^{-1}(I - S)m + \{I - (I - S\bar{M}_B)^{-1}(I - S)\}\varepsilon$$

和

$$MSE(\hat{m}) = \frac{1}{N} \|(I - S\bar{M}_B)^{-1}(I - S)m\|^2 + \frac{1}{N} \|E\{I - (I - S\bar{M}_B)^{-1}(I - S)\}\varepsilon\|^2 = L_1 + L_2.$$

相似于Fan等(2005)的定理3的证明, 我们得到

$$L_1 \leq \frac{T}{(\sqrt{T} - 1)^2} \frac{\mu_2^2 h^4}{4} E[m''(u)] + o_p(h^4) \quad (3-25)$$

和

$$\begin{aligned}L_2 &\leq \frac{T}{(\sqrt{T} - 1)^2} \frac{1}{N} E\|S(I - \bar{M}_B)\varepsilon\|^2 = \frac{T}{(\sqrt{T} - 1)^2} \frac{\sigma^2 + \beta^\tau \Sigma_\eta \beta}{N} \text{tr}[S(I - \bar{M}_B)S] \\ &\leq \frac{T}{(\sqrt{T} - 1)^2} \frac{\sigma^2 + \beta^\tau \Sigma_\eta \beta}{N} \text{tr}(SS^\tau).\end{aligned} \quad (3-26)$$

通过大数定律, 我们有

$$\text{tr}(SS^\tau) = \sum_{i=1}^n \sum_{i=1}^T \frac{\nu}{Nh f(U_{it})} + o_p\left(\frac{1}{Nh}\right) = \frac{\nu|\mathcal{U}|}{h} + o_p\left(\frac{1}{Nh}\right),$$

那合并式子(3-25)和(3-26), 定理3.2被证明。

证明定理3.3. 利用式子(3-10)和 $(I - M_B)B = 0$ ,  $\hat{\sigma}^2$ 能被分解为

$$\begin{aligned}\hat{\sigma}^2 &= \left[ \frac{T}{N(T-1)} (\varepsilon - \eta\hat{\beta})^\tau (I - M_B) (\varepsilon - \eta\hat{\beta}) - \hat{\beta}^\tau \Sigma_\eta \hat{\beta} \right] \\ &\quad + \frac{T}{N(T-1)} [X(\beta - \hat{\beta}) + m - \hat{m}]^\tau (I - M_B) [X(\beta - \hat{\beta}) + m - \hat{m}] \\ &\quad + \frac{2T}{N(T-1)} [X(\beta - \hat{\beta}) + m - \hat{m}]^\tau (I - M_B) (\varepsilon - \eta\hat{\beta}) = \Pi_1 + \Pi_2 + \Pi_3.\end{aligned}$$

因此,

$$\Pi_1 - \sigma^2 = \frac{T}{N(T-1)} (\varepsilon - \eta\hat{\beta})^\tau (I - M_B) (\varepsilon - \eta\hat{\beta}) - (\sigma^2 + \hat{\beta}^\tau \Sigma_\eta \hat{\beta})$$

和

$$\frac{T}{N(T-1)} (\varepsilon - \eta\hat{\beta})^\tau (I - M_B) (\varepsilon - \eta\hat{\beta})$$

$$\begin{aligned}
 &= \frac{T}{N(T-1)} (\varepsilon - \eta\hat{\beta})^\tau (I_N - \frac{1}{T} I_T \oplus l_n l_n^\tau) (\varepsilon - \eta\hat{\beta}) \\
 &= \frac{T}{N(T-1)} \sum_{i=1}^n \left\{ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta})^2 - \frac{1}{T} \left[ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^2 \right\} \\
 &= \frac{T}{N(T-1)} \sum_{i=1}^n \Delta.
 \end{aligned}$$

因为 $\varepsilon_{it}$ 和 $\eta_{it}$ 是独立同分布的,我们有

$$E\Delta = E\left\{ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta})^2 - \frac{1}{T} \left[ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^2 \right\} = (T-1)(\sigma^2 + \beta^\tau \Sigma_\eta \beta).$$

因此,通过简单的计算,我们得到

$$\begin{aligned}
 \text{Var}(\Delta) &= E(\Delta)^2 - (E(\Delta))^2 = E\left\{ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta})^2 - \frac{1}{T} \left[ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^2 \right\}^2 \\
 &\quad - (T-1)(\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2 = TE(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + T(T-1)(\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2 \\
 &\quad + \frac{1}{T} E(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + \frac{3T(T-1)}{T^2} (\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2 \\
 &\quad - \frac{2}{T} [T(\varepsilon_{it} - \eta_{it}\hat{\beta})^2 + T(T-1)(\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2] - (T-1)^2 (\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2 \\
 &= \frac{T^2 - 2T + 1}{T} E(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + (T-1) \frac{3-T}{T} (\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2.
 \end{aligned}$$

因此,

$$\text{Var}(\Pi_1 - \sigma^2) = \frac{1}{N} E(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + \frac{3-T}{N(T-1)} (\sigma^2 + \beta^\tau \Sigma_\eta \beta)^2.$$

通过Zhou等(2010)定理3证明的同样的结论,我们可以证明 $\Pi_i = o_p(N^{-\frac{1}{2}})$ ,  $i = 2, 3$ . 因此,式子(3-14)被获得.我们现在显示 $\hat{\Theta}$ 的一致估计.

简化表示,我们记

$$\Xi_{it} = X_{it}^\tau (\beta - \hat{\beta}) - \frac{1}{T} \sum_{t=1}^T X_{it}^\tau (\beta - \hat{\beta}) + m(U_{it}) - \hat{m}(U_{it}) - \frac{1}{T} \sum_{t=1}^T m(U_{it}) + \frac{1}{T} \sum_{t=1}^T \hat{m}(U_{it})$$

和 $\zeta_{it} = \varepsilon_{it} - \eta_{it}\hat{\beta} - \frac{1}{T} \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta})$ . 因此

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \hat{\psi}_{it}^4 &= \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \zeta_{it}^4 + \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \Xi_{it}^4 + 4 \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \zeta_{it}^3 \Xi_{it} + 6 \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \zeta_{it}^2 \Xi_{it}^2 \\
 &\quad + 4 \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \zeta_{it} \Xi_{it}^3 = J_1 + J_2 + J_3 + J_4 + J_5.
 \end{aligned}$$

对于 $J_1$ ,我们有

$$J_1 = \frac{1}{T} \sum_{t=1}^T E[\varepsilon_{it} - \eta_{it}\hat{\beta} - \frac{1}{T} \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta})]^4 + o_p(1)$$

$$\begin{aligned}
 &= E(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + \frac{1}{T^2} \left[ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^4 + \frac{6}{T^3} \sum_{t=1}^T E(\varepsilon_{it} - \eta_{it}\hat{\beta})^2 \left[ \frac{1}{T} \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^2 \\
 &\quad - \frac{4}{T^2} \sum_{t=1}^T E(\varepsilon_{it} - \eta_{it}\hat{\beta})^3 \left[ \frac{1}{T} \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right] - \frac{4}{T^2} \left[ \sum_{t=1}^T (\varepsilon_{it} - \eta_{it}\hat{\beta}) \right]^4 + o_p(1) \\
 &= \left(1 + \frac{6}{T^2} - \frac{3}{T^3} - \frac{4}{T}\right) E(\varepsilon_{it} - \eta_{it}\hat{\beta})^4 + \frac{3-T}{T-1} \left(6 - \frac{9}{T}\right) (\sigma^2 + \beta^T \Sigma_{\eta} \beta).
 \end{aligned}$$

此外, 通过Hölder不等式和这个事实  $\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \Xi_{it}^4 = o_p(1)$ , 我们能证明  $J_s = o_p(1)$  为  $s = 2, 3, 4, 5$ . 因此,  $\Theta$  的一致估计结果被得到.

**证明定理3.3.** 式子(3-15) 是一个特例为式子(3-16), 因此我们仅证明式子(3-16). 观察

$$\begin{aligned}
 RSS_1 &= \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - X_{it}^T \hat{\beta}_i - \hat{m}_i(U_{it}) - \alpha_i)^2 \\
 &= [(Y - Z\hat{\beta} - \hat{m} - B\alpha)^T (Y - Z\hat{\beta} - \hat{m} - B\alpha) - N\beta^T \Sigma_{\eta} \beta] \\
 &= [\tilde{Z}(\beta_0 - \hat{\beta}) - \tilde{m} + \varepsilon]^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) [\tilde{Z}(\beta_0 - \hat{\beta}) - \tilde{m} + \varepsilon] - N\beta^T \Sigma_{\eta} \beta \\
 &= \varepsilon^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \varepsilon + \Upsilon_1 - N\beta^T \Sigma_{\eta} \beta,
 \end{aligned}$$

其中

$$\begin{aligned}
 \Upsilon_1 &= (\beta_0 - \hat{\beta})^T \tilde{Z}^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \tilde{Z} (\beta_0 - \hat{\beta}) + \tilde{m}^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \tilde{m} \\
 &\quad + 2(\beta_0 - \hat{\beta})^T \tilde{Z}^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \tilde{m} + 2(\beta_0 - \hat{\beta})^T \tilde{Z}^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \varepsilon \\
 &\quad + 2\tilde{m}^T (I - \tilde{M}_{\tilde{B}})^T (I - \tilde{M}_{\tilde{B}}) \varepsilon.
 \end{aligned}$$

通过运用引理3.2, 引理3.3和引理3.4同样的结论, 我们有

$$RSS_1 = N\sigma^2(1 + o_p(1)) - N\beta^T \Sigma_{\eta} \beta.$$

另一方面,

$$\begin{aligned}
 RSS_0 &= [(Y - Z\hat{\beta}_0 - \hat{m}_0 - B\alpha)^T (Y - Z\hat{\beta}_0 - \hat{m}_0 - B\alpha) - N\hat{\beta}_0^T \Sigma_{\eta} \hat{\beta}_0] \\
 &= [(Y - Z\hat{\beta} - \hat{m} - B\alpha + \tilde{Z}(\hat{\beta} - \hat{\beta}_0))^T [(Y - Z\hat{\beta} - \hat{m} - B\alpha + \tilde{Z}(\hat{\beta} - \hat{\beta}_0))] - N\hat{\beta}_0^T \Sigma_{\eta} \hat{\beta}_0] \\
 &= RSS_1 + [\tilde{Z}(\hat{\beta} - \hat{\beta}_0)]^T [\tilde{Z}(\hat{\beta} - \hat{\beta}_0)] + 2(Y - Z\hat{\beta} - \hat{m} - B\alpha)^T [\tilde{Z}(\hat{\beta} - \hat{\beta}_0)] - N\hat{\beta}_0^T \Sigma_{\eta} \hat{\beta}_0 \\
 &= RSS_1 + \Gamma_1 + \Gamma_2 + \Gamma_3.
 \end{aligned}$$

作为  $\beta$  的估计在原假设和备择假设条件下有这个关系

$$\hat{\beta}_0 = \hat{\beta} - \Lambda^{-1} A^T (A \Lambda^{-1} A^T)^{-1} A \hat{\beta}.$$

$\Gamma_1$  能被记作

$$\Gamma_1 = [\tilde{Z}(\hat{\beta} - \hat{\beta}_0)]^T [\tilde{Z}(\hat{\beta} - \hat{\beta}_0)] = \hat{\beta}^T A^T [A(\tilde{Z}^T \tilde{Z})^{-1} A^T]^{-1} A \hat{\beta}.$$

通过条件(A.1), 我们有  $\frac{1}{N} \tilde{Z}^T \tilde{Z} \xrightarrow{p} \Sigma + \Sigma_\eta$ . 这合并  $\hat{\beta}$  的渐近正态性, 能得到

$$\hat{\beta}^T A^T [A(\tilde{Z}^T \tilde{Z})^{-1} A^T]^{-1} A \hat{\beta} - \beta^T A^T [A(\Sigma + \Sigma_\eta)^{-1} A^T]^{-1} A \beta \xrightarrow{d} \sum_{i=1}^l \omega_i \chi_{i1}^2.$$

容易证明  $\Gamma_2 = \Gamma_3 = 0$ . 因此,

$$RSS_0 - RSS_1 \xrightarrow{d} \sigma^2 \sum_{i=1}^l \omega_i \chi_{i1}^2.$$

因此通过Slutsky 定理, 由此得出结论

$$2\lambda_n(\beta) - N\sigma^{-2} \beta^T A^T [A(\Sigma + \Sigma_\eta)^{-1} A^T]^{-1} A \beta \xrightarrow{d} \sum_{i=1}^l \omega_i \chi_{i1}^2.$$



## 第4章 带固定效应的部分线性误差变量面板数据模型的 惩罚经验似然

### 4.1 引言

面板数据的分析是计量经济学作品中最活跃和最有创新的的主题。与纯粹的时间序列或横截面数据相比，面板数据可以为研究人员提供更大可能建模的扩展。它们控制更多的信息，如不可观察和时间不变的个体异质性，并且它们在变量之间包含更多的变化和更小的共线性。Arellano (2003), Baltagi (2005) and Hsiao (2003)极好的概述了参数面板数据模型的统计推断和计量分析。然而，错误指定的参数面板数据模型可能导致错误的推断。因此，因此经济学家和统计学家已经发展了一些灵活的非参数和半参数面板数据模型。例如，Su 和Ullah (2007)提出了一种对于具有随机效应的非参数面板数据模型的两步估计。Cai 和Li (2008)研究了未知变系数的动态非参数面板数据模型；Rodriguez-poo 和Soberon (2014)和Sun 等(2009)考虑带固定效应变系数面板数据模型，建立半参数估计，局部线性平滑估计和用于检验随机效应还是固定效应的统计推断。Chen 等(2013)分析了带固定效应的部分线性单指数面板数据模型，建议一个哑变量方法消除固定效应和建立半参数最小平均方差估计过程。Baltagi 和Li (2002)讨论带固定效应的部分线性面板数据模型，发展了序列估计方法和剖面似然估计技术。Hu (2014)带固定效应的半参变系数面板数据模型。非参数和半参数面板数据模型的优点之一是对模型的结构施加了很少的先前限制。

考虑下面的带固定效应的部分线性面板数据模型

$$Y_{it} = X_{it}^{\tau} \beta + m(Z_{it}) + \mu_i + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (4-1)$$

其中 $Y_{it}$ 是响应变量,  $(X_{it}^{\tau}, Z_{it}) \in R^p \times R$ 是严格外生变量,  $\beta = (\beta_1, \dots, \beta_p)^{\tau}$ 是一个 $p$ -维未知参数, 上角 $\tau$ 表示一个向量的转置,  $m(Z_{it}) = (m_1(Z_{it}), \dots, m_q(Z_{it}))^{\tau}$ 是一个 $q$ -维未知参数,  $\mu_i$ 是一个不可观察的个体效应. 表示 $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})$ 是 $i$ 个体随机误差向量和 $\{\varepsilon_i, i = 1, \dots, n\}$ 是相互独立且 $E(\varepsilon_i | X_{it}, Z_{it}) = 0$ 。模型(4-1)被叫固定效应模型, 如果 $\mu_i$ 是与 $X_{it}$ 和(或) $Z_{it}$ 相关的关于一个未知

的相关结构。模型(4-1)被叫随机效应模型, 如果 $\mu_i$ 与 $X_{it}$ 和 $Z_{it}$ 是不相关的。固定效应的样本稳定性要优于随机效应样本(Baltagi, 2005; Horowitz 和Lee, 2004 等)。然而, 分析带固定效应的面板数据具有更多的挑战, 因为样本增长的参数, 容易产生著名的Neyman 和Scott (1948) 问题。在此, 我们只考虑固定效应情况。

众所周知, 在许多领域, 如工程, 经济学, 生物学, 生物医学科学和流行病学, 观测都是有误差的。例如, 尿氯化钠水平 (Wang et al.1996), 暴露于污染物 (Tosteson et al.1989) 和血清胆固醇水平 (Carroll et al. 1995) 经常受到测量误差的影响。简单地忽略测量误差 (变量中的误差) 将导致得到有偏差估计量。在协变量中处理测量误差通常是统计分析的挑战。在过去二十年中, 伴随测量误差的回归分析有很大的进步。一个细节的研究能被发现在Fuller (1987); Fan 和Truong (1993); You 和Chen (2006); Liang 等(1999); Fan 等(2013) 等专家的研究中。因此, 我们考虑以下带固定效应的面板数据含误差变量部分线性模型:

$$\begin{cases} Y_{it} = X_{it}^T \beta + m(Z_{it}) + \mu_i + \varepsilon_{it}, \\ W_{it} = X_{it} + \nu_{it}, \end{cases} \quad i = 1, \dots, n; t = 1, \dots, T. \quad (4-2)$$

其中协变量 $X_{it}$  被测量伴随测量误差且不能被观察到。代替 $X_{it}$  我们观察到 $W_{it}$  拥有的测量误差 $\nu_{it}$  是独立同分布的, 并且独立于 $(X_{it}, Z_{it}, \varepsilon_{it})$ , 我们假定 $Cov(\nu_{it}) = \Sigma_\nu$  伴随着 $\Sigma_\nu$  假定是已知的, 类似于Zhu 和Cui (2003), You 和Chen (2006), 和别的专家的文章. 当 $\Sigma_\nu$  是未知的, 我们能估计通过重复测量 $Z_{it}$ ; 可以看Liang 等(1999)详细的描述。

然而, 上述所有文章都是具有固定数量的预测变量的情况。众所周知, 高维数据在许多当代统计研究中经常出现。高维数据的出现, 如表达基因值的微阵列和单核苷酸多态性数据, 给传统统计方法和理论带来挑战。高维数据的一个重要背景在回归下协变量的数量是发散的。当参数的数量随样本大小而增长时, 更多的作品一直致力于各种参数和半参数模型下的高维模型。例如, Fan 和Pen (2004)利用非凹惩罚似然方法研究了参数模型的发散数量的参数。Zhu 和Zhu (2010) 考虑稳定的方向恢复在单一指数模型的发散参数的预测; Wu 和Li (2015)研究发散数量参数充分降维估计时的渐近性质。Li 等(2012)利用经验似然方法对于发散参数的变系数部分线性模型的参数建立置信域。当维度发生偏差时, 通过正则化的变量选择被证明是有效的。已经开

发了用于变量选择的各种强大的惩罚方法。Fan和Li (2001) 提出了一种通过非凹惩罚最小二乘法同时自动选择变量和估计变量的系数的统一方法。该方法不仅保留了子集选择和岭回归的良好特征, 而且产生稀疏解, 确保所选模型的连续性并且对大系数具有无偏估计。Li和Liang (2008) 开发了非凹惩罚惩罚半参数回归模型中的变量选择方法。最近, Tang和Leng (2010) 首次提出的惩罚性经验似然 (PEL) 被用于分析多变量的均值向量和线性模型的发散数量回归系数。Tang和Leng (2010) 证实的PEL在来自非参数似然法的效率和适应性方面具有优点。另外, PEL 方法具有使用数据来确定置信区域的形状和取向的EL的相同优点并且不估计共协方差。据我们所知, 有一些与PEL方法相关的论文, 比如Tang 和Leng (2012) 将PEL方法应用于一般估计方程的参数估计和变量选择; Wang和Xiang (2016) 研究了带有发散协变量的稀疏可加危险回归模型的PEL推理。Fan 等 (2016) 利用惩罚经验似然方法对高维部分线性变系数含误差变量模型的参数进行估计。

值得指出的是, 当具有固定效应的部分线性误差变量面板数据模型中的协变量数量(1.2)是发散的。还没有相关文献研究结果, 在本章中, 我们关注这个模型。我们的贡献包括以下三个方面: (1) 利用EL方法构建未知参数的置信区域并建立最大经验的渐近正态性参数的似然 (MEL) 估计; (2) 采用PEL方法建立简约和鲁棒的模型获取PEL估计的oracle属性; (3) 应用PEL比率测试假设和建立未知参数的有限维子集的置信区域。

剩下的这一章组成是这样的: 4.2节介绍模型和方法, 假定和主要结论安排在4.3节, 模拟的计算方法在4.4节, 一些模拟研究和实证放在4.5节, 主要结果的证明安排在4.6节。

## 4.2 模型和方法

下面我们将用向量和矩的形式表示, 令  $Y = (Y_1^T, \dots, Y_n^T)^T$ ,  $X = (X_1^T, \dots, X_n^T)^T$ ,  $Z = (Z_1^T, \dots, Z_n^T)^T$ ,  $\mu_0 = (\mu_1^T, \dots, \mu_n^T)^T$  和  $\varepsilon = (\varepsilon_1^T, \dots, \varepsilon_n^T)^T$  是  $(nT) \times 1$  向量, 其中  $Y_i = (Y_{i1}, \dots, Y_{iT})^T$ ,  $X_i = (X_{i1}, \dots, X_{iT})^T$ ,  $Z_i = (Z_{i1}, \dots, Z_{iT})^T$ ,  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})^T$  和  $D_0 = I_n \otimes e_T$  用  $\otimes$  表示克罗内克积,  $I_n$  表示  $n \times n$  单位向量, 和  $e_n$  表示  $n \times 1$  个1的向量。我们写模型 (4-1) 的矩阵形式:

$$Y = X\beta + m(Z) + D_0\mu + \varepsilon, \quad (4-3)$$

为了识别的目的, 我们增加约束条件 $\sum_{i=1}^n \mu_i = 0$ . 令 $\mu = (\mu_2, \dots, \mu_n)^\tau$  和 $\mu_0 = (-\sum_{i=2}^n \mu_i, \mu_2, \dots, \mu_n)^\tau$ . 我们重新写模型(4-2) 为下面的矩阵形式:

$$Y = X\beta + m(Z) + D\mu_0 + \varepsilon, \quad (4-4)$$

其中 $D = [-i_{n-1} \quad I_{n-1}] \otimes i_T$  是一个 $nT \times (n-1)$  矩阵. 令 $K_h(\cdot) = K(\cdot/h)/h$ ,  $K(\cdot)$  是一个核函数和 $h$  是一系列接近于0的正数, 叫窗宽. 令对角矩阵

$$K_h(Z_i, z) = \begin{bmatrix} K_h(Z_{i1}, z) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(Z_{iT}, z) \end{bmatrix}, \quad W_h(u) = \begin{bmatrix} K_h(Z_1, z) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(Z_n, z) \end{bmatrix},$$

并且 $\theta = (\mu^\tau, \beta^\tau)^\tau$ . 当指定 $\theta$ , 我们能估计函数 $M(z) = (m(z), \{hm'(z)\}^\tau)^\tau$  通过
$$M_\theta(z) = \underset{\theta}{\operatorname{argmin}} [Y - X\beta - D\mu - G(z, h)M(z)]^\tau W_h(u) [Y - X\beta - D\mu - G(z, h)M(z)]. \quad (4-5)$$

其中

$G(z, h) = (G_1(z, h), \dots, G_n(z, h))^\tau$ ,  $G_i(z, h) = (G_{i1}(z, h), \dots, G_{iT}(z, h))^\tau$ ,  
 $G_{it}(z, h) = (1, \{h^{-1}(U_{it} - u)\}^\tau)^\tau$ ,  $m'(u) = \partial m(u)/\partial z$  和 $z = (z_1, \dots, z_q)^\tau$  是在 $Z_{it}$  的一个邻域. 因此模型(4-4) 的解的问题可以通过下式
$$M(z) = (G(z, h)^\tau W_H(u) G(z, h))^{-1} G(z, h)^\tau W_h(u) (Y - X\beta - D\mu) = S(z, h) (Y - X\beta - D\mu).$$
特别,  $m(z)$  的估计可以有下式得到

$$m(z) = s(z, h) (Y - X\beta - D\mu). \quad (4-6)$$

其中 $s(z, h) = e^\tau S(z, h)$ , 和 $e = (1, \dots, 1, 0, \dots, 0)^\tau$  是一个 $2q \times 1$  向量.

现在我们考虑一种方法消除固定效应动机是最小二乘哑变量模型在面板数据的参数分析中, 我们解决下式的最优问题:

$$\hat{\theta} = \arg \min_{\theta} [Y - X\beta - D\mu - S(Y - X\beta - D\mu)]^\tau [Y - X\beta - D\mu - S(Y - X\beta - D\mu)], \quad (4-7)$$

假定 $\tilde{X} = (I_{nT} - S)X$ ,  $\tilde{Y} = (I_{nT} - S)Y$ ,  $\tilde{D} = (I_{nT} - S)D$ , 我们得到 $\tilde{\mu} = (\tilde{D}^\tau \tilde{D})^{-1} \tilde{D}^\tau (\tilde{Y} - \tilde{X}\beta)$ . 令 $\Omega = I_{nT} - \tilde{D}(\tilde{D}^\tau \tilde{D})^{-1} \tilde{D}^\tau$ , 我们能得到 $\Omega \tilde{D}\mu = 0$ . 因此, 在(2.3) 这固定效应项 $D\mu$  被消除, 我们表达渐近误差作为下面:

$$\Lambda_i(\beta) = \sum_{t=1}^T \tilde{X}_{it} \Omega (\tilde{Y}_{it} - \tilde{X}_{it} \beta), \quad i = 1, \dots, n. \quad (4-8)$$

虽然,  $X_{it}$ 's 在我们的案例中不能被观察和我们仅能获得  $Z_{it}$ 。如果我们忽略测量误差和代替  $X_{it}$  用  $Z_{it}$  在式(4-7), 式(4-7)可能被显示被推导的估计是不一致的。它是众所周知的线性回归模型或部分线性回归模型, 因为测量误差的不一致发生能被克服通过应用一种叫“误差修正”方法, 可以看 Fuller (1987) 和 Liang 等(1999)为更多的细节。

$$\Gamma_i(\beta) = \sum_{t=1}^T \widetilde{W}_{it} \Omega(\widetilde{Y}_{it} - \widetilde{W}_{it} \beta) + \Sigma_v \beta, \quad i = 1, \dots, n. \quad (4-9)$$

注意  $E(\Lambda_i(\beta)) = 0$  如果  $\beta$  是实值参数。因此, 相似于 Li 和 Xue(2008), 我们定义  $\beta$  一个误差修正的经验似然比率如下

$$R_n(\beta) = \max \left\{ \prod_{i=1}^n n p_i | p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \Gamma_i(\beta) = 0 \right\}.$$

经过拉格朗日算法。我们能得到

$$\log(R_n(\beta)) = - \sum_{i=1}^n \log\{1 + \lambda^T \Gamma_i(\beta)\}, \quad (4-10)$$

其中  $\gamma$  被决定下式

$$\frac{1}{n} \sum_{i=1}^n \frac{\Gamma_i(\beta)}{1 + \lambda^T \Gamma_i(\beta)} = 0. \quad (4-11)$$

我们用惩罚经验似然方法 (PEL) 那是合并了经验似然方法和平滑修剪绝对偏差 (SCAD) 惩罚方法。我们定义惩罚经验似然(PEL)如下

$$\mathcal{L}_n(\beta) = \log(R_n(\beta)) - n \sum_{j=1}^p p_\lambda(|\beta_j|),$$

其中  $p_\lambda(\cdot)$  是惩罚函数和调节参数  $\lambda$ 。例子可以看 Fan 和 Li (2001) 对于此函数的描述。在这一章, 我们用平滑修剪绝对偏差 (SCAD) 惩罚方法, 它的定义通过它的一阶微分, 定义如下:

$$p'_\lambda(\theta) = \theta \left\{ I(\theta \leq \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\}, \quad \theta > 0, \lambda > 0,$$

对于  $a > 2$  的数和  $p'_\lambda(0) = 0$ 。顺从 Fan 和 Li (2001), 我们下面规定  $a = 3.7$ 。

最大化惩罚函数(4-12)是等价于最小化下式

$$\mathcal{L}_n(\beta, \lambda) = \sum_{i=1}^n \log\{1 + \lambda^T \Gamma_i(\beta)\} + n \sum_{j=1}^p p_\lambda(|\beta_j|), \quad (4-12)$$

令  $\mathcal{B} = \{j : \beta_{0j}\}$  是实值参数向量  $\beta_0$  的非零部分的集合和它的基数  $|\mathcal{B}| =$

$d$ , 其中  $d$  是允许增长伴随着  $n \rightarrow \infty$ . 不失一般性, 我们能分割参数向量为  $\beta = (\beta_1^T, \beta_2^T)^T$ , 其中  $\beta_1 \in \mathbb{R}^d$  和  $\beta_2 \in \mathbb{R}^{p-d}$ . 因此, 这实值参数  $\beta_0 = (\beta_{10}^T, 0^T)^T$  和我们记  $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$  叫 PEL 估计是最小化(4-12)。

### 4.3 假定和主要结论

在这一章, 我们首先扩展结果 Qin 和 Lawless (1994) 的事例到发散维数, 在高维经验似然, 这最大  $\|\lambda\|$  不再是  $O_p(n^{-1/2})$ , 令  $a_n = (p/n)^{1/2}$ ,  $\mathcal{D}_n = \{\beta : \|\beta - \beta_0\| \leq ca_n\}$ . 在得到主要结论前, 我们给出下面的假定

(A1) 随机变量  $U_{it}$  有一个连续密度函数  $f(\cdot)$  伴随一个紧支撑  $\mathcal{U}$  在  $\mathbb{R}$  上.  $0 < \inf_{u \in \mathcal{U}} f(u) \leq \sup_{u \in \mathcal{U}} f(u) < \infty$ .

(A2) 令  $m(u) = E(W_{it}|U_{it} = u)$ . 函数  $m(u)$  和  $g(u)$  是二级连续可微函数在  $\mathcal{U}$  上。

(A3)  $(v_i, X_i, U_i, \varepsilon_i)$ ,  $i = 1, \dots, n$ , 是 i.i.d. 存在一些数  $\delta > 2$  以至  $E\|W_{it}\|^{2+\delta} < \infty$  和  $E|\varepsilon_i|^{2+\delta} < \infty$ , 其中  $\|a\| = (\sum_{i=1}^n a_i^2)^{1/2}$  是通常的 Euclidean norm 的一个向量  $a$ 。

(A4) 核函数  $k(u)$  是一个在连续可微的对称概率密度函数在紧支撑  $\mathcal{U}$  上。

(A5)  $E|\check{W}_{it}|^{2+\delta} < \infty$ ,  $\Psi = \sum_{t=1}^T [\check{W}_{it}(\check{W}_{it} - \sum_{s=1}^T \frac{\check{W}_{is}}{T})^T]$  是正的有限的, 其中  $\check{W}_{it} = W_{it} - E(W_{it}|U_{it})$ 。

(A6)  $E(Y_{it}|W_i, U_i, v_i) = E(Y_{it}|W_{it}, U_{it}, v_{it}) = W_{it}\beta + m(U_{it}) + v_i$ 。

(A7) 窗宽  $h$  满足  $nh^6 \rightarrow 0$  和  $nh^3/(\ln n)^3 \rightarrow \infty$ 。

(A8) 存在正实数  $c$  和  $p^{2+4/(k-2)}/n \rightarrow 0$  作为  $n \rightarrow \infty$ , 都有  $E(\|\varpi_s/\sqrt{p}\|^r) < c$ ,  $s = 1, 2, 3$ . 其中  $\varpi_1 = \sum_{t=1}^T \frac{T-1}{T} (X_{it} - E(X_{it}|U_{it}))(\varepsilon_{it} - \nu_{it}\beta_0)$ ,  $\varpi_2 = \sum_{t=1}^T \frac{T-1}{T} \nu_{it}\varepsilon_{it}$ ,  $\varpi_3 = \sum_{t=1}^T \frac{T-1}{T} (\nu_{it}\nu_{it}^T - \Sigma_\nu)\beta_0$ 。

(A9) 函数  $p'_\lambda(\cdot)$  满足  $\max_{j \in \mathcal{B}} p'_\lambda(|\beta_{0j}|) = o\{(np)^{-1/2}\}$ ,  $\max_{j \in \mathcal{B}} p''_\lambda(|\beta_{0j}|) = o\{p^{-1/2+\delta}n^{-\delta}\}$ 。

注 4.1. 假定(A1)-(A8) 看起来有点多, 事实上条件是平凡的和容易被满足的, (A1)-(A6) 能被发现在文章 Su 和 Ullah (2006). (A7) 已经被假定 Hu (2014).

(A8) 也已经被许多作者作为条件(比如Shi 和Lau, 2000; You 等, 2006)。 □

最大化 $\{-\mathcal{L}_n(\beta_0)\}$ 能获得 $\beta_0$ 的最大经验似然估计 $\hat{\beta}$ 。记

$$\hat{\Psi} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} (\tilde{X}_{it} - \sum_{s=1}^T \frac{\tilde{X}_{is}}{T})^\tau. \quad (4-13)$$

如果矩阵 $\hat{\Psi}$ 是可转置的, 相似于Qin 和Lawless (1994)的定理1 和Xue 和Zhu (2008)定理2 的证明, 因此 $\beta_0$ 的最大经验似然估计 $\hat{\beta}$ 能被表示为

$$\hat{\beta}_{EL} = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{W}_{it}^\tau e_{it}^\tau \Omega \tilde{Y}_{it} + o_p(n^{-\frac{1}{2}}).$$

同样 $\hat{\beta}$ 的渐近正态性能被表达如下定理。

**定理 4.1.** 假定(A1)-(A8) 条件满足, 如果 $\beta_0$  是正的实值参数, 因此我们得到 $\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$ , 其中 $V = \Psi^{-1} \Theta \Psi^{-1}$ ,

$$\begin{aligned} \Theta = & \lim_{N \rightarrow \infty} \text{Var} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} [X_{it} - E(X_{it}|U_{it})] \varepsilon_{it} \right\} \\ & - \frac{T-1}{T} [X_{11} - E(X_{11}|U_{11})] \beta_0^\tau \Sigma_\nu \beta_0 \\ & + \frac{T-1}{T} \sigma^2 \Sigma_\nu + \frac{T-1}{T} E[(\nu \nu^\tau - \Sigma_\nu) \beta_0]^{\otimes 2} \end{aligned}$$

和 $\Psi = \frac{T-1}{T} [X_{11} - E(X_{11}|U_{11})][X_{11} - E(X_{11}|U_{11})]^\tau$ 。

为了建立置信区间, 我们能得到 $\hat{\Psi}^{-1} \hat{\Theta} \hat{\Psi}^{-1}$  作为 $\Psi^{-1} \Theta \Psi^{-1}$  的估计, 其中 $\hat{\Psi}$  被定义在式(4-13),  $\hat{\Theta} = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \{[\tilde{X}_{it} \Omega (\varepsilon_{it} - \nu_{it} \beta_0)] + \nu_{it} \Omega \varepsilon_{it} + (\Sigma_\nu - \nu_{it} \Omega \nu_{it}^\tau) \beta_0\}$  通过定理2.2, 我们得到

$$[\hat{\Psi}^{-1} \hat{\Theta} \hat{\Psi}^{-1}]^{-1/2} \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p), \quad (4-14)$$

或者

$$(\hat{\beta} - \beta_0)^\tau n [\hat{\Psi}^{-1} \hat{\Theta} \hat{\Psi}^{-1}]^{-1} (\hat{\beta} - \beta_0) \xrightarrow{d} \chi_p^2. \quad (4-15)$$

因此,  $\beta_0$  的置信区间能被得到从式(4-14) 和式(4-15)。

**定理 4.2.** 假定(A1)-(A8) 条件满足. 如果条件 $p^5/n \rightarrow 0$ , 因此 $\hat{\beta}$ 满足

(a).(稀疏性) :  $\hat{\beta}_2 = 0$ ;

(b).(渐近正态性):  $\sqrt{n} B_n \Sigma_p^{-1/2} (\hat{\beta}_1 - \beta_{10}) \xrightarrow{d} N(0, V)$ , 其中 $B_n B_n^\tau \rightarrow V$  和 $\Sigma_p = J(\beta_0)^{-1} \Psi_2^\tau (\Psi_2 J(\beta_0)^{-1} \Psi_2^\tau)^{-1} \Psi_2 J(\beta_0)$ 。

下一步我们考虑统计假设检验和建立置信区间为 $\beta$ 。考虑零假设在固定维数在下面的形式:

$$H_0 : L_n \beta_0 = 0, \quad H_1 : L_n \beta_0 \neq 0,$$

其中 $L_n \in R^{q \times d}$  以至 $L_n L_n^T = I_q$  对于固定的 $q$ , 一个非参数剖面似然比率统计被建造为:

$$\mathcal{LR}_n(\beta_0) = -2 \left\{ \mathcal{L}_n(\hat{\beta}, \lambda) - \min_{\beta_0, L_n \beta_0 = 0} \mathcal{L}_n(\beta, \lambda) \right\}$$

我们显示渐近特征比率下列定理。

**定理 4.3.** 假定(A1)-(A8)条件满足。对于模型(4-2), 如果 $\beta_0$  是正的实值参数, 则有 $\mathcal{LR}_n(\beta_0) \xrightarrow{d} \chi_q^2$ , 作为 $n \rightarrow \infty$ , 其中 $\chi_q^2$  是一个标准自由度为 $q$  卡方随机变量和 $\xrightarrow{d}$  代表依分布收敛。

作为定理的一个结果, 参数的 $\beta_0$  的置信区间能得到。更准确的说, 对于任何 $0 \leq \alpha < 1$ , 令 $c_\alpha$  是 $\chi^2$  分布的 $1 - \alpha$  分位数以至 $P(\chi_q^2 > c_\alpha) \leq 1 - \alpha$ 。因此 $\ell_{EL}(\alpha) = \{\beta \in R^p : \mathcal{LR}_n(\beta_0) \leq c_\alpha\}$  建立 $\beta_0$  一个置信区间是渐近 $1 - \alpha$  收敛的, 因为 $\beta_0$ 属于 $\ell_{EL}(\alpha)$  等价于 $\mathcal{LR}_n(\beta_0) \leq c_\alpha$ 。

## 4.4 计算方法

注意SCAD惩罚函数在原点是奇异的, 所以普通的梯度计算方法将不能直接应用。类似Fan 和Li(2001), 利用惩罚函数 $p_\lambda(\beta)$  进行二次逼近的方法来给出一个迭代计算程序。具体地, 假定我们有 $\beta$  当前第 $k$ 个迭代估计 $\hat{\beta}$ 。如果 $\hat{\beta}^{(k)}$  接近于零, 规定 $|\hat{\beta}_0^{(k)}| \leq \epsilon$ , 其中 $\epsilon$  是预先允许的公差, 因此 $\hat{\beta} = 0$ 。否则, 在任给一个非零 $\beta_j^{(k)}$  的某个小领域内, 惩罚函数 $p_\lambda(\beta)$  在 $\beta_0$ 点可以近似表示为

$$p_\lambda(|\beta|) \approx p_\lambda(|\beta_0|) + \frac{p'_\lambda(|\beta_0|)}{2|\beta_0|} (\beta^2 - \beta_0^2),$$

因此我们最小化惩罚经验似然式子(4-12), 消除不相关的实数, 我们推的

$$\hat{\beta}^{(k+1)} = \operatorname{argmin}_\beta \left\{ \sum_{i=1}^n \log \{1 + \lambda^{(k)\tau} \Gamma_i(\beta)\} + n \sum_{j=1}^p \frac{p'_\lambda(|\hat{\beta}_j^{(k)}|)}{2|\hat{\beta}_j^{(k)}|} (\hat{\beta}_j^2) \right\}, \quad (4-16)$$

我们插入 $\beta_j^{(k+1)}$ 到公式(4-11)和获得 $\lambda^{(k+1)}$  通过解下式

$$\frac{1}{n} \sum_{i=1}^n \frac{\Gamma_i(\beta_j^{(k+1)})}{1 + \lambda^\tau \Gamma_i(\beta_j^{(k+1)})} = 0. \quad (4-17)$$



对于一个指定的 $\lambda$ , 我们用最小化 $\mathcal{L}_n(\beta, \lambda)$  获得 $\hat{\beta}$  的初始值 $\beta_0$ , 和拉格朗日系数 $\lambda = 0$ , 可以得到如下迭代计算步骤。

步骤1 令 $k = 0$ , 给出初始值 $\hat{\lambda}^{(0)} = 0, \hat{\beta}^{(0)} = \beta_0$ 。

步骤2 通过式子(4-16)计算 $\hat{\beta}^{(k+1)}$ 。

步骤3 通过式子(4-17)计算 $\hat{\lambda}^{(k+1)}$ 。

步骤4 令 $k = k + 1$ , 迭代步骤2—步骤3直到收敛, 且记最终 $\beta_n$  和 $\lambda$ 的估计为 $\hat{\beta}_n$  和 $\hat{\lambda}_n$ 。因此我们获得

$$\mathcal{L}_n(\beta, \lambda) = \sum_{i=1}^n \log\{1 + \hat{\lambda}_n^\tau \Gamma_i(\hat{\beta}_n^{(k+1)})\} + n \sum_{j=1}^p p_\lambda(|\hat{\beta}_n^{(k+1)}|),$$

王等(2007)讨论当参数维数发散时, 如何选取调整参数 $\lambda$ , 提出BIC方法选择调整参数 $\lambda$ 。

$$BIC(\lambda) = 2\mathcal{L}_n(\beta, \lambda) + d_\lambda \log(n),$$

其中 $d_\lambda$ 表示估计的参数非零的个数, 因此最优的BIC方法选择调整参数是最小化BIC准则。

## 4.5 模拟研究与实证

### 4.5.1 Monte Carlo 模拟

在这一节, 我们将做一些模拟研究有限的样本我们建议的惩罚经验似然方法估计 $\hat{\beta}$  和最大似然估计 $\hat{\beta}_{EL}$ 。特别的, 在第一部分显示估计的一致性和渐近正态性。第二部分考察惩罚经验似然估计的稀疏性。

首先, 我们考虑下面的带固定效应的面板数据部分线性误差变量模型:

$$\begin{cases} Y_{it} = X_{it}^\tau \beta + m(Z_{it}) + \mu_i + \varepsilon_{it}, \\ W_{it} = X_{it} + \nu_{it}, \end{cases} \quad i = 1, \dots, n; t = 1, \dots, T. \quad (4-18)$$

其中 $\beta = (1, \sqrt{2}, 0, 0, 2, \dots, 0)^\tau$ ,  $m(U_{it}) = \sin(2\pi U_{it})$ ,  $U_{it} \sim U(0, 1)$ ,  $\nu_i = \rho \bar{U}_i + w_i$  其中 $\rho = 0.5, 1$  和 $w_i \sim N(0, 1)$  为 $i = 2, 3, \dots, n$ , 且 $\mu_i = -\sum_{i=2}^n \mu_i$ ,  $\bar{U}_i = \frac{1}{T} \sum_{t=1}^T U_{it}$ . 我们用 $\rho$  去控制 $\mu_i$  和 $\bar{U}_i$  的相关性。

现在我们做模拟, 我们选取 $p$ 作为 $10(3n)^{1/5} - 20$ 整数部分, 选取的样本大

小分别为 $(n, T) = (50, 4), (50, 6), (80, 4), (80, 6)$ , 和我们选取的核为Epanechnikov核 $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \leq 1\}$ . 采取去掉第 $i$ 个观测的交错核实验证(cross validation, CV)方法的窗宽 $CV(h)$ 将被获得用最小化下式

$$CV(h) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - W_{it}^T \hat{\beta}_{[i]} - \hat{m}_{[i]}(U_{it}))^2,$$

其中 $\hat{\beta}_{[i]}$ 和 $\hat{m}_{[i]}(U_{it})$ 分别是 $\beta$ 和 $m(U_{it})$ 的估计, 那计算所有的测量但不计算 $i$ th观测。为了显示建议的方法, 我们对比惩罚经验似然(PEL)和最大似然方法(MEL)估计, 和忽略测量误差的纯惩罚经验似然估计(NPEL)和纯最大似然估计(NMEL), 非0系数 $\beta_1$ 依据PEL, MEL, NPEL和NMEL方法计算平均误差(MEs)和 $\beta_1$ 的标准差(SDs), 我们将做500次重复实验, 模拟的结果在表4.1。

从表4.1, 我们可以看出PEL和MEL估计量的MEs和SDs随着样本大小增加而减小, 在所有相同的背景下PEL估计比MEL, NPEL和NMEL方法给出估计有更小的MEs和SDs。进一步看出, 具有较小测量误差情况下的估计有更好的性能。另一方面, 表4.1显示估计量的抽样分布合理地适合正态, 这种拟合在增加样本大小时更好除此以外。此外, 具有较小测量误差的估计量的分布更近似于正常值。

#### 4.5.2 变量选择

在这节, 我们评价变量选择为PEL和NPEL估计, 数据来源于(4-16), 每个背景下, 我们重复100次, 平均零参数成分的个数(表示为 $N_{zero}$ )和平均实际参数个数(表示为 $N_{true}$ )被概括在表4.2。

从表4.2, 我们能看出变量选择方法依据PEL选择都是3个实数预测和平均估计零的个数在所有背景下接近 $p - 3$ 个。在同样样本大小下越小的测量误差导致更好的估计, 同时能看出在同样大小的样本背景下, PEL估计要优于NPEL估计。这些暗示依据于PEL估计方法变量选择满足选取的模型更接近与实际模型。

### 4.6 主要结果的证明

为了表示的方面, 我们简化符号。令 $\vartheta_k = \int u^k K(u) dl$ ,  $a_n = \sqrt{p_n/n}$ 和 $c_N = \{\log n / (nh)\}^{1/2} + h^2$ 。  $c$ 表示随情况变化的有界正实数。

表 4.1 MEs 和  $\beta_1$  的 SDs 通过 PEL, MEL, NPEL 和 NMEL 估计

$\Sigma_v$	$\rho$	$(n, T)$	PEL		MEL		NPEL		NMEL	
			ME	SD	ME	SD	ME	SD	ME	SD
$0.2I_2$	0	(50,4)	0.288	0.177	0.780	0.284	0.344	0.285	0.980	0.241
		(100,2)	0.287	0.173	0.778	0.389	0.232	0.270	0.847	0.258
		(100,4)	0.185	0.275	0.790	0.506	0.246	0.296	0.819	0.328
	0.5	(50,4)	0.314	0.162	0.771	0.236	0.232	0.205	0.794	0.211
		(100,2)	0.282	0.161	0.758	0.279	0.225	0.197	0.771	0.259
		(100,4)	0.207	0.250	0.799	0.225	0.218	0.145	0.810	0.237
$0.5I_2$	0	(50,4)	0.295	0.207	0.986	0.328	0.334	0.268	0.847	0.376
		(100,2)	0.236	0.266	0.942	0.336	0.326	0.206	0.829	0.477
		(100,4)	0.272	0.281	0.926	0.457	0.245	0.190	0.973	0.338
	0.5	(50,4)	0.164	0.173	0.832	0.271	0.293	0.268	0.875	0.296
		(100,2)	0.155	0.142	0.820	0.260	0.192	0.260	0.815	0.239
		(100,4)	0.102	0.164	0.883	0.287	0.196	0.170	0.816	0.395

表 4.2 变量选择依据 PEL 和 NPEL 方法

$p$	$\rho$	$(n, T)$	PEL ( $\Sigma_v = 0.2$ )		MEL ( $\Sigma_v = 0.2$ )		NPEL ( $\Sigma_v = 0.4$ )		NMEL ( $\Sigma_v = 0.4$ )	
			$N_{zero}$	$N_{true}$	$N_{zero}$	$N_{true}$	$N_{zero}$	$N_{true}$	$N_{zero}$	$N_{true}$
10	0	(50,4)	5.98	3	5.790	3	5.44	3	5.24	3
		(100,2)	6.20	3	6.17	3	5.93	3	5.54	3
		(100,4)	6.28	3	6.21	3	5.96	3	5.39	3
	0.5	(50,4)	5.88	3	5.71	3	5.84	3	5.70	3
		(100,2)	6.12	3	6.07	3	5.92	3	6.09	3
		(100,4)	6.17	3	6.12	3	5.98	3	6.10	3
15	0	(50,4)	12.85	3	12.77	3	12.34	3	12.28	3
		(100,2)	13.13	3	12.82	3	12.66	3	12.32	3
		(100,4)	13.53	3	13.22	3	13.45	3	13.33	3
	0.5	(50,4)	12.64	3	12.21	3	12.30	3	12.21	3
		(100,2)	12.85	3	12.49	3	12.23	3	12.15	3
		(100,4)	13.12	3	12.93	3	12.36	3	12.20	3

引理 4.1. 假定条件 (A1)-(A5) 满足, 因此有

$$G(z, h)^{\tau} W_h(u) G(z, h) = N f(u) \times \begin{pmatrix} 1 & 0 \\ 0 & v_2 \end{pmatrix} \{1 + O_p(c_N)\},$$

$$G(z, h)^{\tau} W_h(u) X = N f(u) E(X|U) \times (1, 0)^{\tau} \{1 + O_p(c_N)\},$$

证明. 类似于 Fan 和 Huang (2005) 的引理 2 证明, 我们容易得到结论。

引理 4.2. 假设条件 (A1)-(A8) 都拥有. 因此我们能得到

$$E|m(U_{it}) - \sum_{k=1}^n \sum_{l=1}^T S_{kl} m(U_{kl})|^2 = O(h^4). \quad (4-19)$$

证明. 参见何 (2017) 引理 5.1 的证明。

引理 4.3. 假设条件 (A1)-(A8) 都拥有, 因此我们能得到

$$(\tilde{D}^{\tau} \tilde{D})^{-1} = (D^{\tau} D)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n)$$

$$\text{对于充分大的 } n, \text{ 其中 } \zeta_n = \frac{\sqrt{\ln n}}{n|H|} i_n i_n^{\tau}.$$

证明. 证(a).  $\tilde{D}^{\tau} \tilde{D} = D^{\tau} (I_{nT} - S)^{\tau} (I_{nT} - S) D = D^{\tau} D - D^{\tau} S D - D^{\tau} S^{\tau} D + D^{\tau} S^{\tau} S D = \Delta_{11} - \Delta_{12} - \Delta_{13} + \Delta_{14}$ .  $\Delta_{1l}$  的第  $(i, j)$  个元素是  $[\Delta_{1l}]_{ij} = o(1)$ ,  $l = 2, 3, 4$ . 类似于 Su (2006) 的引理 A.2 的证明, 我们能证明  $\tilde{D}^{\tau} \tilde{D} = (D^{\tau} D)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n)$ .

引理 4.4. 假设条件 (A1)-(A7) 都拥有. 因此我们能得到

$$\frac{1}{N} \tilde{X}^{\tau} \Omega \tilde{X} \xrightarrow{p} \frac{T-1}{T} \Sigma,$$

其中  $\Sigma = E\{[X_{it} - E(X_{it}|U_{it})]^{\tau} [X_{it} - E(X_{it}|U_{it})]\}$ .

证明.

$$[I_q \ 0_q^{\tau}]^{-1} (D_U^{\tau} X_U D_U)^{-1} D_U^{\tau} X_U X = E(X|U) + O_p(c_N).$$

因此我们有

$$\tilde{X} = [X_{11} - E(X_{11}|U), \dots, X_{1T} - E(X_{1T}|U), \dots, X_{nT} - E(X_{nT}|U)]^{\tau} + O_p(c_N),$$

和

$$\tilde{X} = X - E(X|U) + O_p(c_N) \triangleq A + O_p(c_N).$$

通过利用大数定律, 我们得到

$$\frac{1}{N} \tilde{X}^\tau \tilde{X} = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{ [X_{it} - E(X_{it}|U)]^\tau [X_{it} - E(X_{it}|U)] \} + O_p(c_N) \xrightarrow{P} \Sigma. \quad (4-20)$$

利用引理4.3和Su (2006) 的引理A.3, 我们得到:

$$\frac{1}{N} \tilde{W}^\tau \tilde{D} (\tilde{D}^\tau \tilde{D})^{-1} \tilde{D}^\tau \tilde{W} \xrightarrow{P} \frac{1}{T} \Sigma. \quad (4-21)$$

从式子(4-20)和(4-21), 我们容易证明定理结论。

**引理 4.5.** 假设条件(A1)-(A8) 都拥有. 因此我们能得到

$$(a) \quad N^{-1/2} X^\tau (I_{nT} - S)^\tau (I_{nT} - S) \varepsilon \xrightarrow{P} N^{-1/2} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - E(X_{it}|U_{it})) \varepsilon_{it} + o_p(1).$$

$$(b) \quad N^{-1/2} X^\tau (I_{nT} - S)^\tau (I_{nT} - S) D (D^\tau D)^{-1} D^\tau (I_{nT} - S)^\tau (I_{nT} - S) \varepsilon$$

$$\xrightarrow{P} N^{-1/2} \frac{1}{T} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T (X_{it} - E(X_{it}|U_{it})) \varepsilon_{is} + o_p(1).$$

证明. (a),(b) 是Su 和Ullah (2006)的引理A.6 的直接结论.

**引理 4.6.** 假设条件(A1)-(A8) 都拥有, 如果 $\beta_0$  是实数, 因此我们能得到

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \Gamma_i(\beta_0) \xrightarrow{d} N(0, \Theta).$$

证明. 由式子(4-9)的 $\Gamma_i(\beta)$  定义, 通过简单的计算我们得到

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \Gamma_i(\beta_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \{ \tilde{X}_{it} \Omega \tilde{m}_{it} - \tilde{X}_{it} \Omega \tilde{v}_{it} \beta_0 + \tilde{X}_{it} \Omega \tilde{\varepsilon}_{it}$$

$$+ \tilde{v}_{it} \Omega \tilde{m}_{it} - \tilde{v}_{it} \Omega \tilde{v}_{it}^\tau \beta_0 + \tilde{v}_{it} \Omega \tilde{\varepsilon}_{it} + \Sigma_v \beta_0 \}.$$

相似于Li等 (2011) 定理3.3的证明, 我们有

$$\sup [G(z, h)^\tau W_h(u) G(z, h)] G(z, h)^\tau W_h(u) \varepsilon = O_p\left(\left(\frac{\log n}{nh}\right)^{1/2}\right),$$

合并4.4, 我们容易建立

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{\varepsilon}_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} (X_{it} - E(X_{it}|U_{it})) \varepsilon_{it} \quad (4-22)$$

相似于式子(4-22), 我们能得到

$$\begin{aligned}\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{\nu}_{it} &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} (X_{it} - E(X_{it}|U_{it})) \nu_{it}, \\ \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{\nu}_{it} \Omega \tilde{\nu}_{it}^T &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} \nu_{it} \nu_{it}^T, \\ \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{\nu}_{it} \Omega \tilde{\varepsilon}_{it} &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} \nu_{it} \varepsilon_{it},\end{aligned}$$

利用引理4.2, 它是容易得到  $\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{m}_{it} = o_p(1)$  和  $\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \nu_{it} \Omega \tilde{m}_{it} = o_p(1)$ ,

因此我们容易得到下面的等式

$$\begin{aligned}& \frac{1}{\sqrt{N}} \sum_{i=1}^n \Gamma_i(\beta_0) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \{[\tilde{X}_{it} \Omega (\varepsilon_{it} - \nu_{it} \beta_0)] + \nu_{it} \Omega \varepsilon_{it} + (\Sigma_\nu - \nu_{it} \Omega \nu_{it}^T) \beta_0\} + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T H_{it}(\beta_0) + o_p(1).\end{aligned}$$

注意  $\nu_{it}$  是独立同分布的, 均值是零, 并且独立于  $(X_{it}, Z_{it}, \varepsilon_{it})$ , 因此  $E(H_{it}(\beta_0)) = 0$  和  $\lim_{N \rightarrow \infty} \text{Var}(\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T H_{it}(\beta_0)) = \Theta$ . 满足Cramér-Wold设计的条件和Lindeberg条件. 利用中心极限定理, 我们得到  $\frac{1}{\sqrt{N}} \sum_{i=1}^n \Gamma_i(\beta_0) \xrightarrow{d} N(0, \Theta)$ . 因此, 我们完成引理4.6的证明.

**引理 4.7.** 在定理4.1的条件下, 如果  $\beta_0$  是实数, 我们得到

$$\|\lambda(\beta_0)\| = O_p(a_n),$$

**证明:** 由引理4.6证明知道

$$\begin{aligned}\Gamma_i(\beta_0) &= \sum_{t=1}^T \{[\tilde{X}_{it} \Omega (\varepsilon_{it} - \nu_{it} \beta_0)] + \nu_{it} \Omega \varepsilon_{it} + (\Sigma_\nu - \nu_{it} \Omega \nu_{it}^T) \beta_0\} + o_p(1) \\ &= \sum_{t=1}^T \left\{ \frac{T-1}{T} (X_{it} - E(X_{it}|U_{it})) (\varepsilon_{it} - \nu_{it} \beta_0) + \frac{T-1}{T} \nu_{it} \varepsilon_{it} \right. \\ &\quad \left. + \left( \frac{T-1}{T} \nu_{it} \nu_{it}^T - \Sigma_\nu \right) \beta_0 \right\} + o_p(1) = \varpi_1 + \varpi_2 + \varpi_3 + o_p(1).\end{aligned}$$

我们首先来看  $\varpi_1$ , 定义  $\varpi_s^* = \max_{1 \leq i \leq n} \|\varpi_s\|$ ,  $s = 1, 2, 3$ , 相似Li等(2012)的 (A6) 证

明, 对于任意的  $\epsilon > 0$ ,

$$\begin{aligned} P\{\varpi_1^* \geq (p)^{1/2} n^{1/k} \epsilon\} &\leq \sum_{i=1}^n P\{\|\varpi_1\| \geq (p)^{1/2} n^{1/k} \epsilon\} \\ &\leq \frac{1}{np^{k/2} \epsilon^k} \sum_{i=1}^n E\|\varpi_1\|^k \\ &\leq \frac{1}{\epsilon^k} E\|\varpi_1/p^{1/2}\|^k. \end{aligned}$$

通过条件 (A8), 容易得到

$$\varpi_1^* = o_p(\sqrt{pn}^{1/k}) = o_p(\sqrt{n/pn^{2-k/(2k)}p}) = o_p(\sqrt{n/p}).$$

同样的方法可得到  $\varpi_2^* = o_p(\sqrt{n/p})$  和  $\varpi_3^* = o_p(\sqrt{n/p})$ . 因此,

$$\max_{1 \leq i \leq n} \|\Gamma_i(\beta_0)\| = o_p(\sqrt{n/p}), \quad (4-23)$$

用Li等 (2012)的定理1, 我们有  $\|\lambda(\beta_0)\| = O_p(a_n)$ .

**引理 4.8.** 假定在定理条件下, 对于  $\beta \in \mathcal{D}_n$  我们得到  $\|\lambda(\beta)\| = O_p(a_n)$ .

证明: 令  $\lambda(\beta) = \rho\theta$ , 其中  $\rho \geq 0, \theta \in R^p$ , 且  $\|\theta\| = 1$ . 集合

$$J(\beta) = \frac{1}{n} \sum_{i=1}^n \Gamma_i(\beta) \Gamma_i^T(\beta), \quad \bar{\Gamma}(\beta) = \frac{1}{n} \sum_{i=1}^n \Gamma_i(\beta), \quad \Gamma^*(\beta) = \max_{1 \leq i \leq n} \|\Gamma_i(\beta)\|.$$

从式子(4-11), 我们得到

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{\Gamma_i(\beta)}{1 + \lambda^T \Gamma_i(\beta)} = \frac{1}{n} \sum_{i=1}^n \theta^T \Gamma_i(\beta) - \rho \frac{1}{n} \sum_{i=1}^n \frac{(\theta^T \Gamma_i(\beta))^2}{1 + \rho \theta^T \Gamma_i(\beta)} \\ &\leq \theta^T \bar{\Gamma}(\beta) - \frac{\rho}{1 + \rho \Gamma^*} \theta^T J(\beta) \theta. \end{aligned}$$

因此

$$\rho[\theta^T J(\beta) \theta - \max_{1 \leq i \leq n} \|\Gamma_i(\beta)\| n^{-1} \sum_{i=1}^n \theta^T \Gamma_i(\beta)] \leq |\theta^T \bar{\Gamma}(\beta)|.$$

观察

$$\Gamma^*(\beta) \leq \Gamma^*(\beta_0) + \max_{1 \leq i \leq n} \left\| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{X}_{it} (\beta - \beta_0) \right\|$$

表示  $\mathcal{X}_{it} = \tilde{X}_{it} \Omega \tilde{X}_{it}$ . 通过引理4.4 和Minkowski 不等式, 我们得到

$$\text{Var}(\|\mathcal{X}_{it}\|^{r/2}) \leq E(\|\mathcal{X}_{it}\|^r) = E\left[\left(\sum_{j=1}^p \mathcal{X}_{itj}^2\right)^{r/2}\right] \leq \left\{\sum_{j=1}^p E[\mathcal{X}_{itj}^r]\right\}^{r/2} = O(p^{r/2}).$$

因此我们得到:

$$\begin{aligned} \max_{1 \leq i \leq n} \|\mathcal{X}_{it}\| &\leq \left[ \{\text{Var}(\|\mathcal{X}_{it}\|^{r/2})\}^{1/2} \max_{1 \leq i \leq n} \left\{ \frac{\|\mathcal{X}_{it}\|^{r/2} - E\|\mathcal{X}_{it}\|^{r/2}}{\{\text{Var}(\|\mathcal{X}_{it}\|^{r/2})\}^{1/2}} + E\|\mathcal{X}_{it}\|^{r/2} \right\}^{r/2} \right] \\ &\leq o_p(\sqrt{pn}^{1/r}) + O_p(\sqrt{p}) = o_p(\sqrt{pn}^{1/r}). \end{aligned}$$

因此有

$$\max_{1 \leq i \leq n} \|\mathcal{X}_{it}\| = o_p(\sqrt{pn}^{1/r}).$$

合并式子(4-23)

$$\Gamma^*(\beta) = o_p(n/p) \quad (4-24)$$

$$\bar{\Gamma}(\beta) = \bar{\Gamma}(\beta_0) + \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{X}_{it} (\beta - \beta_0)$$

$\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{X}_{it} = O_p(1)$ 。因此有  $(\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \tilde{X}_{it} \Omega \tilde{X}_{it})(\beta - \beta_0) = O_p(a_n)$ 。我们能容易得到

$$\theta^r \bar{\Gamma}(\beta) = O_p(a_n) \quad (4-25)$$

合并式子(4-24)和(4-25), 得到  $\max_{1 \leq i \leq n} \|\Gamma_i(\beta)\| n^{-1} |\sum_{i=1}^n \theta^r \Gamma_i(\beta)| = o_p(1)$ 。

类似于Li等(2012)和Fan等(2016), 我们能得到  $\theta^r J(\beta) \theta \xrightarrow{P} c$ 。因此, 我们得到  $\rho \leq c |\theta^r \bar{\Gamma}(\beta)| = O_p(a_n)$ , 因此  $\|\lambda(\beta)\| = O_p(a_n)$ 。

**证明定理4.1.** 顺从Zhu和Xue(2008)的定理2的证明的相似结论, 我们有

$$\hat{\beta} - \beta_0 = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \Gamma_i(\beta) + o_p(n^{-1/2}).$$

利用引理4.4, 通过利用大数定律, 我们能证明  $\hat{\Psi} \xrightarrow{P} \Psi$ 。合并引理4.6和Slutsky定理, 我们能够容易完成定理4.1的证明。

**证明定理4.2.** 通过引理, 我们注意在  $\mathcal{D}_n$ , 最小化  $\mathcal{L}_n(\beta, \lambda)$ , 考虑到  $\beta \in \mathcal{D}_n$ , 我们得到成分的每个元素

$$\frac{1}{n} \frac{\partial \mathcal{L}_n(\beta, \lambda)}{\partial \beta_i} = \frac{1}{n} \sum_{i=1}^n \frac{-\lambda \frac{1}{T} \sum_{t=1}^T \tilde{W}_{it} \tilde{H} \tilde{W}_{it} + \lambda_{\nu\nu}}{1 + \lambda^r \Gamma_i(\beta)} + p'_\lambda(|\beta_j|) \text{sign}(\beta_j) = I_j + II_j, \quad (4-26)$$

通过引理4.8,  $\lambda^r \Gamma_i(\beta) = o_p(1)$ , 对于  $\max_{j \in B} |I_j| \leq \lambda \Sigma_j (1 + o_p(1)) = O_p(a_n)$ , 其中  $\Sigma_j$  为  $\Sigma$  的第  $j$  个元素, 在条件  $\tau(n/p)^{1/2} \rightarrow \infty$ 。暗示  $Pr(\max_{j \in B} |I_j| > \tau/2) \rightarrow 0$ 。因此  $p'_\lambda(|\beta_j|) \text{sign}(\beta_j)$  是  $\frac{\partial \mathcal{L}_n(\beta, \lambda)}{\partial \beta_i}$  的主要部分当  $j \notin B$ , 作为  $n \rightarrow \infty$ , 对于  $j \notin B$ , 伴随以概率1趋向,

$$\frac{\partial \mathcal{L}_n(\beta, \lambda)}{\partial \beta_i} > 0, \beta_j \in (0, ca_n); \frac{\partial \mathcal{L}_n(\beta, \lambda)}{\partial \beta_i} < 0, \beta_j \in (0, -ca_n)$$

那暗示对于所有的  $j \notin B^c$ , 以概率1有  $\hat{\beta}_j = 0$ , 因此定理a部分成立。下一步, 我们来证明b部分成立, 令  $\Psi_1$  和  $\Psi_2$  是这样的矩, 满足  $\Psi_1 \beta = \beta_1$  和  $\Psi_2 \beta = \beta_2$ , 作



为  $n \rightarrow \infty$ ,  $Pr(\hat{\beta}_2 = 0) \rightarrow 1$ , 因此通过拉格朗日乘数法, 发现最小化  $\mathcal{LR}_n(\beta, \lambda)$  问题等价于下面最小化下面目的函数:

$$\frac{1}{n} \sum_{i=1}^n \log\{1 + \lambda^\tau \Gamma_i(\beta)\} + n \sum_{j=1}^p p_\lambda(|\beta_j|) + v^\tau \Psi_2 \beta, \quad (4-27)$$

其中  $v$  是  $p-s$  维列向量是拉格朗日算子, 定义

$$\begin{aligned} \tilde{Q}_{1n}(\beta, \lambda, v) &= \frac{1}{n} \sum_{i=1}^n \frac{\Gamma_i(\beta)}{1 + \lambda^\tau \Gamma_i(\beta)}, \\ \tilde{Q}_{2n}(\beta, \lambda, v) &= \frac{1}{n} \sum_{i=1}^n \frac{\lambda}{1 + \lambda^\tau \Gamma_i(\beta)} + b(\beta) + \Psi_2^\tau v, \end{aligned}$$

和  $\tilde{Q}_{3n}(\beta, \lambda, v) = \Psi_2 \beta$ . 其中

$$b(\beta) = \{p'_\lambda(|\beta_1|) \text{sign}(\beta_1), p'_\lambda(|\beta_2|) \text{sign}(\beta_2), \dots, p'_\lambda(|\beta_p|) \text{sign}(\beta_p), 0^\tau\}^\tau.$$

式子(4-27)中最小的  $(\beta, \lambda, v)$  满足  $\tilde{Q}_{in}(\beta, \lambda, v) (i = 1, 2, 3)$ . 既然在  $\beta \in B$  时,  $\|\lambda\| = O_p(a_n)$ . 我们由  $\tilde{Q}_{2n}(\beta, \lambda, v) = 0$ , 得到  $\|v\| = O_p(a_n)$ . 为了扩展  $\tilde{Q}_{in}(\beta, \lambda, v) (i = 1, 2, 3)$  在值周围  $(\beta_0, 0, 0)$ , 我们首先给出下面的微分:

$$\begin{aligned} \frac{\partial \tilde{Q}_{1n}(\beta, 0, 0)}{\partial \lambda} &= -J(\beta), \quad \frac{\partial \tilde{Q}_{1n}(\beta, 0, 0)}{\partial \beta} = \Omega(\beta), \quad \frac{\partial \tilde{Q}_{1n}(\beta, 0, 0)}{\partial v} = 0, \\ \frac{\partial \tilde{Q}_{2n}(\beta, 0, 0)}{\partial \lambda} &= \Omega(\beta), \quad \frac{\partial \tilde{Q}_{2n}(\beta, 0, 0)}{\partial \beta} = b'(\beta), \quad \frac{\partial \tilde{Q}_{2n}(\beta, 0, 0)}{\partial v} = \Psi_2^\tau, \\ \frac{\partial \tilde{Q}_{3n}(\beta, 0, 0)}{\partial \lambda} &= 0, \quad \frac{\partial \tilde{Q}_{3n}(\beta, 0, 0)}{\partial \beta} = \Psi_2, \quad \frac{\partial \tilde{Q}_{3n}(\beta, 0, 0)}{\partial v} = 0. \end{aligned}$$

因此通过泰勒展开式, 我们立即得到:

$$\begin{pmatrix} \tilde{Q}_{1n}(\beta, 0, 0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -J(\beta_0) & \Omega(\beta_0) & 0 \\ \Omega(\beta_0)^\tau & 0 & \Psi_2^\tau \\ 0 & \Psi_2 & 0 \end{pmatrix} \begin{pmatrix} \tilde{\lambda} \\ \tilde{\beta} \\ \tilde{v} \end{pmatrix} + R_n, \quad (4-28)$$

其中  $\Omega(\beta) = n^{-1} \sum_{i=1}^n \partial \Gamma_i(\beta) / \partial \beta$  和我们定义  $\Omega(\beta_0) = n^{-1} \sum_{i=1}^n \partial \Gamma_i(\beta_0) / \partial \beta$ .  $R_n = \sum_{l=1}^5 R_n^{(l)}$ ,  $R_n^{(1)} = (R_{1n}^{\tau 1}, R_{2n}^{\tau 1}, 0)^\tau$ .  $R_{jn}^{\tau 1} \in R^p$  和第  $k$  个成分  $R_{jn}^{\tau 1}$  对于  $j = 1, 2$  被给

$$R_{jn}^{(1)} = \frac{1}{2} (\hat{v} - v_0)^\tau \frac{\partial^2 \tilde{Q}_{jn,k}(\tilde{v})}{\partial v \partial v^\tau} (\hat{v} - v_0),$$

$\vartheta = (\beta, \lambda)^\tau$ ,  $\tilde{\vartheta} = (\tilde{\beta}, \tilde{\lambda})^\tau$  满足  $\|\tilde{\vartheta} - \vartheta_0\| \leq \|\hat{\vartheta} - \vartheta_0\|$ .  $R_n^{(2)} = (0, b^\tau(\beta_0), 0)^\tau$ ,  $R_n^{(3)} = [0, \{b'(\tilde{\vartheta}(\hat{\vartheta} - \vartheta_0))\}, 0]^\tau$  和  $R_n^{(4)} = [\{(J(\beta) - J(\beta_0))\hat{\lambda}\}^\tau + (\Omega(\beta) - \Omega(\beta_0))(\hat{\beta} -$

$\beta\}^r, 0, 0]^r$  和  $R_n^{(5)} = [0, \{(\Omega(\beta) - \Omega(\beta_0))\hat{\lambda}\}^r, 0]^r$ . 对于  $\|\Omega(\beta) - \Omega(\beta_0)\| = O_p(n^{-1/2})$ , 和条件A8, 因此我们有  $k = 1, \dots, 5, R_n^{(k)} = o_p(n^{-1/2})$ . 定义  $K_{11} = -J(\beta_0)$ ,  $K_{12} = [\Omega(\beta_0), 0]$  和  $K_{21} = K_{12}^T$ ,

$$K_{22} = \begin{pmatrix} 0 & \Psi_2^T \\ \Psi_2 & 0 \end{pmatrix}, K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

和令  $\kappa = (\beta^T, \lambda^T)^T$ , 因此对于式子的转置, 我们发现

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\kappa} - \kappa_0 \end{pmatrix} = K^{-1} \left\{ \begin{pmatrix} -\bar{Q}_{1n}(\beta, 0, 0) \\ 0 \end{pmatrix} + R_n \right\}, \quad (4-29)$$

运用块矩阵的逆的性质, 我们得到:

$$K^{-1} = \begin{bmatrix} K_{11}^{-1} + K_{11}^{-1}K_{12}A^{-1}K_{12}K_{11}^{-1} - K_{11}^{-1}K_{12}A^{-1} & \\ & A^{-1}K_{12}K_{11}^{-1} \\ & & A^{-1} \end{bmatrix},$$

其中  $A = K_{22} - K_{21}K_{11}^{-1}K_{12} = \begin{bmatrix} J(\beta_0)^{-1} & \Psi_2^T \\ \Psi_2 & 0 \end{bmatrix}$ , 因此

$$\hat{\kappa} - \kappa_0 = A^{-1}K_{21}K_{11}^{-1}\bar{Q}_{1n}(\beta, 0, 0) + o_p(n^{-1/2}).$$

$$A^{-1} = \begin{bmatrix} J(\beta_0)^{-1} + J(\beta_0)^{-1}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0) & J(\beta_0)^{-1}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1} \\ (\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)^{-1}\Psi_2^{-1} & -(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1} \end{bmatrix},$$

进一步, 我们有

$$\hat{\beta} - \beta_0 = [J(\beta_0)^{-1} - J(\beta_0)^{-1}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)](\Omega(\beta_0)J(\beta_0)^{-1} + o_p(n^{-1/2})). \quad (4-30)$$

因此扩展此式得  $\hat{\beta}_1$  的估计

$$\hat{\beta}_1 - \beta_0 = [\Psi_1J(\beta_{10})^{-1} + \Psi_1J(\beta_0)^{-1}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)](\Omega(\beta_0)J(\beta_0)^{-1} + o_p(n^{-1/2})). \quad (4-31)$$

运用林德伯格-福德中心极限定理, (b)部分证明完成。

证明定理4.3. 令  $\hat{\beta}$  是最小化式子(4-12), 令  $z_i = \hat{\lambda}^T \Gamma_i(\beta)$  我们利用泰勒展开式

$$\mathcal{LR}_n(\beta, \lambda) = \sum_{i=1}^n z_i - \sum_{i=1}^n z_i^2/2 + \sum_{i=1}^n z_i^3/3(1 + \xi_i)^4 + o_p(1),$$

其中 $|\xi_i| \rightarrow |z_i|$ 和 $o_p(1)$ 项是因为惩罚函数项。从式子(4-25), 我们得到 $\hat{\lambda} = [J(\beta_0)^{-1} + J(\beta_0)^{-1}\Omega(\beta_0)\{J(\beta_0)^{-1} - J(\beta_0)^{-1}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)\}\Omega(\beta_0)^TJ(\beta_0)^{-1}][\bar{\Gamma}(\beta) + o_p(n^{-1/2})]$ . 类似于Tang和Leng(2010), 代替展开式中的 $\hat{\lambda}$ 和 $\hat{\beta}$ 在定理4.2中式子(4-26)用 $z_i$ , 我们能显示

$$2\mathcal{LR}_n(\hat{\beta}, \hat{\lambda}) = n\bar{\Gamma}(\beta)^T\Psi_2^TJ(\beta_0)^{-1}(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)\bar{\Gamma}(\beta) + o_p(1) \quad (4-32)$$

在原假设条件下, 既然 $L_nL_n^T = I_q$ , 因此存在矩阵 $\tilde{\Psi}_2$ 以至 $\tilde{\Psi}_2\beta = 0$ 和 $\tilde{\Psi}_2\tilde{\Psi}_2^T = I_{p-d+q}$ . 用一个相似定理4.2的推导,  $\beta$ 的估计能够通过最小化下式:

$$\mathcal{LR}_n(\tilde{\beta}, \lambda) = \frac{1}{n} \sum_{i=1}^n \log\{1 + \lambda^T \Gamma_i(\beta)\} + n \sum_{j=1}^p p_\lambda(|\beta_j|) + v^T \tilde{\Psi}_2 \beta, \quad (4-33)$$

注意最小化式子(4-31), 通过利用 $(\tilde{\beta}, \tilde{\lambda}, \tilde{v})$ . 通过定理4.2的证明,  $\tilde{\beta} = 0$ 的概率倾向于1. 通过代替 $\Psi_2$ 利用 $\tilde{\Psi}_2$ , 我们得到

$$2\mathcal{LR}_n(\tilde{\beta}, \tilde{\lambda}) = n\bar{\Gamma}(\beta)^T\tilde{\Psi}_2^TJ(\beta_0)^{-1}(\tilde{\Psi}_2J(\beta_0)^{-1}\tilde{\Psi}_2^T)^{-1}\tilde{\Psi}_2J(\beta_0)\bar{\Gamma}(\beta) + o_p(1). \quad (4-34)$$

合并式子(4-32)和(4-34)

$$\mathcal{LR}_n = n\bar{\Gamma}(\beta)^TJ(\beta_0)^{-1/2}(P_1 - P_2)J(\beta_0)^{-1/2}\bar{\Gamma}(\beta) + o_p(1).$$

其中

$$P_1 = J(\beta_0)^{-1/2}\Psi_2^T(\Psi_2J(\beta_0)^{-1}\Psi_2^T)^{-1}\Psi_2J(\beta_0)^{-1/2},$$

和

$$P_2 = J(\beta_0)^{-1/2}\tilde{\Psi}_2^T(\tilde{\Psi}_2J(\beta_0)^{-1}\tilde{\Psi}_2^T)^{-1}\tilde{\Psi}_2J(\beta_0)^{-1/2},$$

作为 $P_1 - P_2$ 是一个级数为 $q$ 的幂等阵, 因此 $P_1 - P_2$ 能写为 $\Upsilon^T\Upsilon$ , 其中 $\Upsilon$ 是一个 $q \times p$ 矩阵以至 $\Upsilon^T\Upsilon = I_q$ , 进一步, 我们得到:

$$\sqrt{n}\Upsilon J(\beta_0)^{-1/2}\bar{\Gamma}(\beta) \xrightarrow{d} N(0, I_q).$$

因此

$$n\bar{\Gamma}(\beta)^TJ(\beta_0)^{-1/2}(P_1 - P_2)J(\beta_0)^{-1/2}\bar{\Gamma}(\beta) \xrightarrow{d} \chi_q^2.$$

因此定理证明被完成。



## 第5章 带固定效应面板数据半变系数模型的经验似然

### 5.1 引言

面板数据是在不同时期跟踪由给定个体组成的样本而获取的数据集。因此，它包含样本中每个个体的多个观测值。在面板数据中存在两种信息：反映在对象之间的差异中的横截面信息，以及反映在对象内随时间的变化的时间序列或对象内信息。伴随着可用的面板数据增长的，关于面板数据分析的作品近年越来越流行。Baltagi (2005) 和Hsiao (2003)提供了一个极好的对广泛使用的参数面板数据模型的统计推断与计量经济学解释综述。面板数据分析包括了截面数据和时间序列数据的分析，然而，参数面板数据可能被错误的指定，并且来自错误指定的估计通常是不一致的。为处理这个问题，一些非参数面板数据模型已经被介绍和研究，例如带固定效应非参数面板数据模型(Henderson 等, 2008), 随机效应非参数面板数据模型(Henderson 和Ullah, 2005)和动态非参数面板数据模型(Cai 和Li, 2008)。但非参数用于探索隐藏结构和减少模型误差，它们是太灵活，以至不能得到精确的结论。同时，当有大量的协变量时，非参数估计面对“维数灾祸”。怎样避免“维数灾祸”是一个重要的课题对于非线性时间序列和面板数据分析。许多有用的方法被发展去避免这样问题(可看书Fan 和Yao, 2003 为例子)和各种各样半参数面板数据模型已经被建议。例如，Hu 和Li (2011)研究纵向数据半参数变系数部分线性模型。Hu (2014)考虑半变系数面板数据模型。变系数模型是众所周知在统计作品和同样在经济和金融有许多应用(Cai 和Li, 2008; Hu 和Li, 2011; Fan 和Zhang, 1999 等)。一个主要优点是变系数模型是它允许系数依赖于一些信息变量，然后能平衡维数缩减和模型灵活性。半变系数模型包括两个重要的统计模型：变系数模型和半参数部分线性模型。相关的研究能被发现Fan 和Huang (2005), You 和Zhou (2006), Hu (2014)等等。Hu (2014)建议剖面似然方法估计带固定效应的面板数据的半变系数模型。

我们目标在这章应用经验似然 (EL)方法去建立 $\beta$  的置信区间对于带固定效应面板数据半变系数模型。首先，我们将建立一个经验对数似然比率函数为模型在 $\alpha$ -混合条件下的回归参数。根据这种方法，一个人能立即

为这个回归参数建立一个渐进置信区间。一个动机是经验似然方法推断不涉及估计量的渐近协方差，这对于具有在 $\alpha$ -混合条件下带固定效应的面板数据的半参数变系数模型的协方差将是更加复杂的结构。另一个动机是基于经验似然方法的置信区域不对区域形状强加先前的约束，并且置信区域的形状和方向完全由数据确定。经验似然方法已经被许多作者使用到现在，例如Shi和Lau (2000), Wang和Jing (1999), You和Zhou (2006), Hu和Li (2011), Fan和Liang (2010)等等。Fan和Liang (2010)等等。大多数与经验似然方法相关的上述论文总是假定数据是独立同分布的。然而，数据的独立性假设在应用中并不总是适合的，特别是对于有顺序收集的经济数据，其经常表现出明显的相依性，并且许多作者已经假设在混合条件下。例如，Fan, Liang和Huang (2012)采用经验似然方法来构建在相依条件下部分时变系数模型的置信区间。

对于本章，我们假定 $\{(X_{it}, Z_{it}, Y_{it}, U_{it}, v_{it}), t \geq 1\}$ 是平稳 $\alpha$ -混合的对于每个 $i$ 。回想一个序列 $\{\xi_k, k \geq 1\}$ 据说是 $\alpha$ -混合的， $\alpha$ -混合系数

$$\alpha(n) \stackrel{\text{def}}{=} \sup_{k \geq 1} \sup \{|P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A})P(\mathcal{B})| : \mathcal{A} \in \mathcal{F}_{n+k}^\infty, \mathcal{B} \in \mathcal{F}_1^k\}$$

收敛到零作为 $n \rightarrow \infty$ ，其中 $\mathcal{F}_a^b = \sigma\{\xi_i, a \leq i \leq b\}$ 表示 $\sigma$ -代数生成的 $\xi_a, \xi_{a+1}, \dots, \xi_b$ ，在各种混合条件下， $\alpha$ -混合条件是弱的，并且许多时间序列模型是 $\alpha$ -混合的。例如，Auestad和Tjøstheim (1990)提供了关于在非线性时间序列分析中模型识别的 $\alpha$ -混合的作用的有趣的讨论。进一步，Masry和Tjøstheim (1995)显示在一些温和的条件下，包括自回归条件异方差(ARCH)过程和伴随着外生变量可加自回归过程，都是稳定和 $\alpha$ -混合的。这些过程在金融方面都特别流行。

剩下的这一章组成是这样的：5.2节介绍方法和建立 $\beta$ 经验对数似然比率函数。假定和主要结论被给在5.3节，一些模拟和实证安排在5.4节，主要结果的证明安排在5.5节。

## 5.2 模型和方法

这一章考虑下面的带固定效应面板数据半变系数模型

$$Y_{it} = Z_{it}^T \beta + X_{it}^T \alpha(U_{it}) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (5-1)$$

其中 $Y_{it}$ 是响应变量， $(Z_{it}, X_{it}) \in R^p \times R^q$ 和 $U_{it} \in R$ 是严格外生变量， $\beta =$

$(\beta_1, \dots, \beta_p)^\tau$  是一个  $p$ -维未知参数, 和上标  $\tau$  表示一个向量的转置。  $\alpha(U_{it}) = (\alpha_1(U_{it}), \dots, \alpha_q(U_{it}))^\tau$  是一个  $q$ -维向量的未知函数和  $\mu_i$  是不能观察的个体效应,  $v_{it}$  是随机误差。 这儿, 我们假定  $\mu_i$  是独立同分布的均值为0 和有限的方差  $\sigma^2 > 0$ 。 我们允许  $\mu_i$  是相关的与  $Z_{it}, X_{it}$ , 和  $U_{it}$  伴随一个的相关结构。 为介绍我们估计的方法, 我们假定模型拥有下面的约束条件  $\sum_{i=1}^n \mu_i = 0$ 。 令  $\mu = (\mu_2, \dots, \mu_n)^\tau$  和  $\mu_0 = (-\sum_{i=2}^n \mu_i, \mu_2, \dots, \mu_n)^\tau$ 。 我们重新写模型(5-1) 成矩阵的形式

$$Y = Z\beta + X\alpha(U) + H\mu + V, \quad (5-2)$$

其中  $H = [-i_{n-1} \ I_{n-1}] \otimes i_T$  是一个  $nT \times (n-1)$  矩阵,  $I_n$  表示  $n \times n$  单位矩阵,  $i_n$  表示  $n \times 1$  数字1的向量。 有许多方法可以估计参数  $\{\beta_j, j = 1, \dots, p\}$  和变系数函数  $\{\alpha_i(\cdot), i = 1, \dots, q\}$ 。 我们方法的主要思想是剖面最小二乘估计方法被建议Fan 和Huang (2005): 假定我们有一个随机样本  $\{(U_{it}, Z_{it1}, \dots, Z_{itp}, X_{it1}, \dots, X_{itq}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T\}$  来自模型(5-1)。 令  $\theta = (\mu^\tau, \beta^\tau)^\tau$ 。 指定  $\theta$ , 我们能够运用一个局部线性回归技术估计变系数函数  $\{\alpha_j(\cdot), j = 1, \dots, q\}$  来自模型(5-1)。 对于  $U_{it}$  是在  $u$  的一个小领域内, 我们能估计  $\alpha_j(U_{it})$  局部线性函数

$$\alpha_j(U_{it}) \approx \alpha_j(u_0) + \alpha'_j(u_0)(U_{it} - u_0) \equiv a_j + b_j(U_{it} - u_0), \quad j = 1, \dots, q,$$

其中  $\alpha'_j(u) = \partial \alpha_j(u) / \partial u$ 。 这导致下面的权局部最小二乘问题: 发现  $\{(a_j, b_j), j = 1, \dots, q\}$  最小化

$$\sum_{i=1}^n \sum_{t=1}^T \left\{ \left( Y_{it} - Z_{it}^\tau \beta - \mu_i \right) - \sum_{j=1}^q \left[ a_j + b_j(U_{it} - u_0) \right] X_{it} \right\}^2 K_{h_1}(U_{it} - u_0), \quad (5-3)$$

其中  $K_h(\cdot) = K(\cdot/h)/h$ ,  $K(\cdot)$  是一个核函数和  $h$  是一系列接近于0的正数, 叫窗宽。 简单计算得到

$$(\hat{\alpha}_1(u), \dots, \hat{\alpha}_q(u), h\hat{a}'_1(u), \dots, h\hat{a}'_q(u))^\tau = (D_u^\tau W_u D_u)^{-1} D_u^\tau W_u (Y - Z\beta - H\mu),$$

其中

$$Z = \begin{pmatrix} Z_{11}^r \\ \vdots \\ Z_{1T}^r \\ \vdots \\ Z_{nT}^r \end{pmatrix}, \quad X = \begin{pmatrix} X_{11}^r \\ \vdots \\ X_{1T}^r \\ \vdots \\ X_{nT}^r \end{pmatrix}, \quad D_u = \begin{pmatrix} X_{11}^r & \frac{U_{11}-u}{h} X_{11}^r \\ \vdots & \vdots \\ X_{1T}^r & \frac{U_{1T}-u}{h} X_{1T}^r \\ \vdots & \vdots \\ X_{nT}^r & \frac{U_{nT}-u}{h} X_{nT}^r \end{pmatrix},$$

$Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{nT})^r$  和  $W_u = \text{diag}(K_h(U_{11} - u), \dots, K_h(U_{1T} - u), \dots, K_h(U_{nT} - u))$ . 参数  $\theta$  的剖面似然的估计被指定通过下式:

$$\hat{\theta} = \arg \min_{\theta} [Y - Z\beta - H\mu - S(Y - Z\beta - H\mu)]^r [Y - Z\beta - H\mu - S(Y - Z\beta - H\mu)], \quad (5-4)$$

其中光滑矩阵  $S$  是

$$S = \begin{pmatrix} (X_{11}^r \ 0_q^r)(D_{U_{11}}^r W_{U_{11}} D_{U_{11}})^{-1} D_{U_{11}}^r W_{U_{11}} \\ \vdots \\ (X_{1T}^r \ 0_q^r)(D_{U_{1T}}^r W_{U_{1T}} D_{U_{1T}})^{-1} D_{U_{1T}}^r W_{U_{1T}} \\ \vdots \\ (X_{nT}^r \ 0_q^r)(D_{U_{nT}}^r W_{U_{nT}} D_{U_{nT}})^{-1} D_{U_{nT}}^r W_{U_{nT}} \end{pmatrix} = \begin{pmatrix} s_{11} \\ \vdots \\ s_{1T} \\ \vdots \\ s_{nT} \end{pmatrix}. \quad (5-5)$$

令  $M = (m_1^r, \dots, m_n^r)^r$ ,  $m_i = (x_{i1}\alpha(U_{i1}), \dots, x_{iT}\alpha(U_{iT}))$ ,  $Z_i = (Z_{i1}^r, \dots, Z_{iT}^r)^r$ ,  $Y_i = (Y_{i1}, \dots, Y_{iT})^r$ ,  $\tilde{Z} = (I_{nT} - S)Z$ ,  $\tilde{H} = (I_{nT} - S)H$ ,  $\tilde{Y} = (I_{nT} - S)Y$ ,  $\tilde{\mu} = (\tilde{H}^r \tilde{H})^{-1} \tilde{H}^r (\tilde{Y} - \tilde{Z}\beta)$  和  $\Lambda = I_{nT} - \tilde{H}(\tilde{H}^r \tilde{H})^{-1} \tilde{H}^r$ . 我们引入下面的辅助随机向量:

$$\eta_i(\beta) = \tilde{Z}_i \Lambda (\tilde{Y}_i - \tilde{Z}_i \beta), \quad i = 1, \dots, n. \quad (5-6)$$

注意  $E(\eta_i(\beta)) = 0$  如果  $\beta$  是实值参数。因此, 相似于Owen (1990), 我们定义经验对数似然比率如下

$$\log \mathcal{L}_n(\beta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i \eta_i(\beta) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \quad (5-7)$$

利用拉格朗日乘数方法, 我们能得到  $p_i = \frac{1}{n[1 + \lambda^r \eta_i(\beta)]}$ , 和  $\log \mathcal{L}_n(\beta)$  能被表示如



下

$$\log \mathcal{L}_n(\beta) = 2 \sum_{i=1}^n \log \{1 + \lambda^\tau \eta_i(\beta)\}, \quad (5-8)$$

其中  $\lambda(\beta)$  被决定下式

$$\frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^\tau \eta_i(\beta)} = 0. \quad (5-9)$$

### 5.3 假定和主要结论

在这一章, 我们将显示如果  $\beta$  是实参数向量,  $\log \mathcal{L}_n(\beta)$  是渐近卡方分布。在得到主要结论前, 我们给出下面的一些假定

- (A1) 核函数  $K(v)$  是一个在连续可微的对称概率密度函数在紧支撑  $[-1, 1]$  上。
- (A2) 随机变量  $U$  在支撑  $\Omega$  上是有界的。矩阵  $\Gamma(U) \triangleq E(ZZ^\tau | U)$  是非奇异的对于每个  $U \in \Omega$ 。  $G(U) \triangleq E(XX^\tau | U)$ ,  $\Gamma^{-1}(U)$  和  $\Phi(U) \triangleq E(ZX^\tau | U)$  都是利普希茨连续的。
- (A3)  $\{\alpha_j(\cdot), j = 1, \dots, q\}$  是二级连续可微函数在  $U \in \Omega$ 。
- (A4) (i)  $\{X_{it}, Y_{it}, Z_{it}, U_{it}, v_{it}\}$  是独立同分布的对于截面第  $i$  个体为每一个固定时间  $t$ , 并且是严平稳的为  $t$  序列对于每个固定的  $i$ 。  
 (ii) 对于固定的  $i$ , 过程  $\{X_{it}, Y_{it}, Z_{it}, U_{it}, v_{it}\}$  是一个严平稳的  $\alpha$ -混合序列伴随这混合系数满足条件  $\alpha(k) = O(k^{-\kappa})$ ,  $\kappa = \frac{(2+\delta)(1+\delta)}{\delta}$  和  $\delta > 2$ 。
- (A5) 存在一些  $\delta > 2$  以致  $E\|X_i\|^{2+\delta} < \infty$ ,  $E\|Z_i\|^{2+\delta} < \infty$ ,  $E\|v_i\|^{2+\delta} < \infty$ , 其中  $\|\cdot\|$  是这个  $L_2$ -距离。
- (A6) 令  $N = nT$ , 窗框  $h$  满足  $Nh \rightarrow \infty$ ,  $Nh^6 \rightarrow 0$ ,  $(\log N)^{\kappa+1/2} N^{-((\kappa-1/2)-1/\delta)} T^{\kappa+1} h^{\kappa-3} \rightarrow 0$ 。其中  $\kappa$  和  $\delta$  被定义在上面的 A4(ii)。
- (A7)  $E|\check{Z}_{it}|^{2+\delta} < \infty$ ,  $\Psi = \sum_{t=1}^T [\check{Z}_{it}(\check{Z}_{it} - \sum_{s=1}^T \frac{\check{Z}_{is}}{T})^\tau]$  是正的有限的, 其中  $\check{Z}_{it} = Z_{it} - G^{-1}(u)\Phi(u)X_{it}$ 。

**注 5.1.** 假定(A1)-(A7) 看起来有点多, 事实上条件是平凡的和容易被满足的。(A1)-(A3) 能被发现在文章 Fan 和 Huang (2005)。(A4)-(A5) 已经被用被许多

作者(看Cai 和Li 2008; Chen 等2013 为例子)。(A7) 已经被假定Hu (2014)。技术条件(A6) 容易被满足, 例如, 当 $T \sim N^{1/5}$  和 $h \sim N^{-\theta}$ , 它能被证明 $Nh \rightarrow \infty, Nh^6 \rightarrow 0$ , 和 $(\log N)^{\kappa+1/2} N^{-((\kappa-\frac{1}{2})-\frac{1}{2\delta})} T^{\frac{\kappa+1}{\delta}} h^{\kappa-3} \rightarrow 0$  都能被满足当 $\frac{1}{6} < \theta < \frac{1}{5}$ 。 □

**定理 5.1.** 假定(A1)-(A7) 条件满足。对于模型(5-1), 如果 $\beta_0$  是正的实值参数, 则有 $\log \mathcal{L}_n(\beta) \xrightarrow{d} \chi_p^2$ , 作为 $n \rightarrow \infty$ , 其中 $\chi_p^2$  是一个标准自由度为 $p$  卡方随机变量和 $\xrightarrow{d}$  代表依分布收敛。

作为定理的一个结果, 参数的 $\beta_0$  的置信区间能被建立。更准确的说, 对于任何 $0 \leq \alpha < 1$ , 令 $c_\alpha$  是 $\chi^2$  分布的 $1 - \alpha$  分位数以至 $P(\chi_p^2 > c_\alpha) \leq 1 - \alpha$ 。因此 $\ell_{EL}(\alpha) = \{\beta \in R^p : \log \mathcal{L}_n(\beta) \leq c_\alpha\}$  建立 $\beta$  一个置信区间是渐近 $1 - \alpha$  收敛的, 因为 $\beta$  属于 $\ell_{EL}(\alpha)$  等价于 $\log \mathcal{L}_n(\beta) \leq c_\alpha$ 。

最大化 $\{-\log \mathcal{L}_n(\beta)\}$  能获得 $\beta$  的最大经验似然估计 $\hat{\beta}$ 。记 $\hat{\Psi} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it}^T (\tilde{Z}_{it} - \sum_{s=1}^T \frac{\tilde{Z}_{is}}{T})$ 。如果矩阵 $\hat{\Psi}$  是可转置的, 相似于Qin 和Lawless (1994)的定理1, 因此 $\beta$  的最大经验似然估计 $\hat{\beta}$ 能被表示为

$$\hat{\beta} = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \tilde{Z}_i^T \Lambda \tilde{Y}_i + o_p(n^{-\frac{1}{2}}).$$

同样 $\hat{\beta}$  的渐近正态性能被表达如下定理。

**定理 5.2.** 假定(A1)-(A7) 条件满足, 如果 $\beta$  是正的实值参数, 因此我们得到 $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$ , 其中 $V = \Psi^{-1} \Theta \Psi^{-1}$ ,  $\Theta = \sum_{t=1}^T \sum_{s=1}^T E \left\{ \left[ \tilde{Z}_{it} (\tilde{Z}_{is} - \sum_{t=1}^T \frac{\tilde{Z}_{it}}{T})^T \right] (v_{it} v_{is}) \right\}$ 。

## 5.4 模拟研究与实证

### 5.4.1 Monte Carlo 模拟

在这一节, 我们将做一些模拟阐明对于有限的样本我们建议的经验似然方法优于渐近正态方法。

首先, 我们考虑下面的带固定效应面板数据半变系数模型:

$$Y_{it} = Z_{it}^T \beta + X_{it,1} \alpha_1(U_{it}) + X_{it,2} \alpha_2(U_{it}) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (5-10)$$

其中  $\beta = (\beta_1, \beta_2)^\tau = (1, \sqrt{2})^\tau / \sqrt{3}$ ,  $\alpha_1(U_{it}) = \sin(2\pi U_{it})$ ,  $\alpha_2(U_{it}) = \cos(\pi U_{it})$ ,  $U_{it} \sim U(0, 1)$ ,  $\mu_i = \rho \bar{U}_i + w_i$  伴随着  $\rho = 0.5, 1$  和  $w_i \sim N(0, 1)$  对于  $i = 2, 3, \dots, n$ , 和  $\mu_i = -\sum_{i=2}^n \mu_i$ ,  $\bar{U}_i = \frac{1}{T} \sum_{t=1}^T U_{it}$ . 我们用  $\rho$  去控制  $\mu_i$  和  $\bar{U}_i$  的相关关系.  $X_{it}$  和  $Z_{it}$  是来自下面的AR(1)模型,

$$\begin{aligned} v_{it} &= 0.5v_{it-1} + \varepsilon_i, & \varepsilon_i &\stackrel{i.i.d}{\sim} N(0, 0.25^2), \\ X_{it,1} &= 0.6X_{it-1,1} + u_{i,1}, & X_{it,2} &= 0.4X_{it-1,2} + u_{i,2}, \\ Z_{it,1} &= 0.8Z_{it-1,1} + e_{i,1}, & Z_{it,2} &= 0.2Z_{it-1,2} + e_{i,2}, \end{aligned}$$

其中  $u_i = (u_{i,1}, u_{i,2})^\tau \stackrel{i.i.d}{\sim} N((0, 0)^\tau, \text{diag}(1, 1))$  和  $e_i = (e_{i,1}, e_{i,2})^\tau \stackrel{i.i.d}{\sim} N((0, 0)^\tau, \text{diag}(1, 1))$ . 容易证明  $\{X_{it}, Z_{it}, v_{it}\}$  是平稳和  $\alpha$ -混合的.

现在我们做模拟, 我们选取的核为  $K(u) = \frac{15}{16}(1 - u^2)^2 I\{|u| \leq 1\}$ . 采取去掉第  $i$  个观测主题的交错核实验验证(cross validation, CV)方法来获得窗宽  $CV(h)$  被用最小化下式

$$CV(h) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - Z_{it}^\tau \hat{\beta}_i - X_{it}^\tau \hat{\alpha}_i(U_{it}))^2,$$

其中  $(\hat{\beta}_1, \hat{\beta}_2)^\tau$  和  $\hat{\alpha}_j(U_{it})$  在  $j = 1, 2$  分别是  $(\beta_1, \beta_2)^\tau$  和  $\alpha_j(U_{it})$  在  $j = 1, 2$  的估计, 计算所有的测量但不计算第  $i$  个观测主题.

我们提出在定理5.2中的  $\hat{\beta}$  的协方差的一致估计. 一个一致估计的  $\Theta$  能被指定通过代替  $\beta$  用  $\hat{\beta}$ , 即,

$$\hat{\Theta} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \left[ \tilde{Z}_{it} (\tilde{Z}_{it} - \sum_{s=1}^T \frac{\tilde{Z}_{is}}{T})^\tau (\tilde{Y}_{it} - \tilde{Z}_{it} \hat{\beta}) (\tilde{Y}_{it} - \tilde{Z}_{it} \hat{\beta})^\tau \right]$$

和一个一致估计  $\hat{\Psi} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} (\tilde{Z}_{it} - \sum_{s=1}^T \frac{\tilde{Z}_{is}}{T})^\tau$ . 因此, 一个  $1 - \alpha$  置信水平的置信区间依据于正态逼近分布的是  $\ell_{NA}(\alpha) = (\hat{\beta} - z_{\frac{\alpha}{2}} \hat{\Psi} \sqrt{n^{-1} \hat{\Theta}}, \hat{\beta} + z_{\frac{\alpha}{2}} \hat{\Psi} \sqrt{n^{-1} \hat{\Theta}})$ .

我们考虑两种方法计算个体  $\beta_i (i = 1, 2)$  的置信区间的覆盖率(CPs)和平均长度(APs): 经验似然(EL)和正态逼近(NA), 样本大小  $(n, T)$  我们分别选取  $(50, 4)$ ,  $(100, 2)$  和  $(100, 4)$ , 我们将做500次重复在置信水平分别为  $1 - \alpha = 0.90$  和  $0.95$ , 一些有代表性的  $(\beta_1, \beta_2)$  的置信区间的覆盖率被报道在表5.1, 同时, 我们给出一些有代表性的  $\beta_1$  和  $\beta_2$  的置信水平分别为  $1 - \alpha = 0.90$  和  $0.95$  的置信区间的覆盖率(CPs)和平均长度(APs)在表5.2. 从表5.1和表5.2, 我们可以看出经验似然方法优于渐近正态分布方法因为经验似然方法给出更高的覆盖

率和更小的置信区间比渐近正态方法。并且我们还能看出覆盖率增大和置信区间变小随着样本容量 $(n, T)$ 的变大。同样我们能发现置信区间的覆盖率变小和置信区间长度变大随着 $\rho$ 变大。

表 5.1 CP 表示 $(\beta_1, \beta_2)$ 的置信区间依据于经验似然(EL)和正态逼近方法(NA)在置信水平是 $1 - \alpha = 0.90$ 和 $0.95$ 的覆盖率

$\rho$	$1 - \alpha$	$(n, T)$	$CP_{EL}$	$CP_{NA}$	$\rho$	$1 - \alpha$	$(n, T)$	$CP_{EL}$	$CP_{NA}$
0.5	0.90	(50,4)	0.881	0.683	1	0.90	(50,4)	0.874	0.693
		(100,2)	0.876	0.670			(100,2)	0.866	0.706
		(100,4)	0.892	0.716			(100,4)	0.882	0.713
	0.95	(50,4)	0.929	0.848		0.95	(50,4)	0.928	0.835
		(100,2)	0.926	0.832			(100,2)	0.920	0.810
		(100,4)	0.936	0.856			(100,4)	0.934	0.854

Note: “ $CP_{EL}$ 和 $CP_{NA}$ ”表示CP分别依据于经验似然和正态逼近方法.

表 5.2 CP 和AL 表示置信区间与区间长度依据于经验似然和正态逼近方法在置信水平是 $1 - \alpha = 0.90$ 和 $0.95$ 的覆盖率和区间长度

$\rho$	Para	$(n, T)$	$1 - \alpha = 0.90$				$1 - \alpha = 0.95$			
			CE	CN	AE	AN	CE	CN	AE	AN
0.5	$\beta_1$	(50,4)	0.885	0.770	0.2890	0.5844	0.944	0.895	0.3804	0.7471
		(100,2)	0.870	0.738	0.3786	0.5891	0.932	0.870	0.5427	0.7588
		(100,4)	0.895	0.795	0.2309	0.5006	0.946	0.916	0.3199	0.7282
	$\beta_2$	(50,4)	0.884	0.623	0.3713	0.8436	0.932	0.705	0.4943	1.0681
		(100,2)	0.882	0.610	0.4586	0.7579	0.925	0.697	0.5971	0.9659
		(100,4)	0.887	0.650	0.2993	0.7525	0.938	0.745	0.4102	0.9137
1	$\beta_1$	(50,4)	0.850	0.778	0.3286	0.9286	0.934	0.868	0.4847	1.2076
		(100,2)	0.836	0.766	0.4220	0.9361	0.926	0.906	0.6329	1.3677
		(100,4)	0.872	0.815	0.2672	0.9057	0.945	0.910	0.3373	1.1380
	$\beta_2$	(50,4)	0.864	0.732	0.4321	1.3571	0.930	0.868	0.6275	1.7960
		(100,2)	0.855	0.742	0.4982	1.3260	0.923	0.860	0.7315	1.6739
		(100,4)	0.882	0.764	0.3243	0.8871	0.936	0.870	0.4816	1.6095

Note: “CE和CN”表示CP分别依据于经验似然和正态逼近方法“AE和AN”表示AL分别依据于经验似然和正态逼近方法.

### 5.4.2 基于艾滋病人的CD4数据的实证

为了考察本章建议的方法的表现, 我们现在把本章建议的方法用于某艾滋病研究中心的一组CD4数据的研究, 该数据集包括吸烟状况, HIV感染者的岁数, HIV感染前的CD4百分率和重复测量的感染HIV后平均CD4百分率从1984~1991年感染HIV病毒的283个同性恋男子。该数据详细的描述艾滋病研究中心相关的设计, 方法和医药治疗可见文献Kaslow等(1987)。这个数据集也被许多作者用来做模型拟合, 比如Huang等(2001), Xue和Zhu(2007), Zhou和Lin(2014)。最初的实验设计对所有患者进行每半年观测一次。但是, 由于大量的患者错过了其诊断时间并且在整个研究过程中HIV感染是随机发生的, 所以, 对不同患者的观测次数及观测时间是不同的, 因此该数据是一个不平衡的纵向数据, 我们在283个患者中选取了158个患者, 那他们的前六次观测数据都在, 因此我们获得对于每个个体相同的观测数据, 即是平衡的面板数据。这个研究目的是描述整个研究期间患者的CD4百分比率随时间变化的趋势, 并评估患者的吸烟情况、HIV感染前的CD4百分率和HIV感染时患者感染HIV后平均CD4百分率的影响。 $Y_{it}$  是第*i*个体的CD4细胞cell的百分比在时间*t*时。 $Z_{it1}$  第*i*个体的抽烟状态, 它的取值是1或0, 通过看个体是否是吸烟的还是不吸烟的在HIV感染后。 $Z_{it2}$  是第*i*个体的preCD4细胞cell的百分比在时间*t*时。我们考虑下面的模型:

$$Y_{it} = Z_{it1}\beta_1 + Z_{it2}\beta_2 + X_{it,1}\alpha_1(U_{it}) + X_{it,2}\alpha_2(U_{it}) + \mu_i + v_{it},$$

$$i = 1, \dots, n, \quad t = 1, \dots, T.$$

其中 $X_{it,1}$ , 基准CD4百分比, 代表平均CD4细胞cell的百分比在感染后时间*t*时,  $X_{it,2}$  表示感染后的感染者的平均年龄,  $\beta_1$  和 $\beta_2$  分别表示抽烟和pre-infection CD4百分比的影响, 一个剖面最小二乘估计 $\beta$  是 $\hat{\beta}_T = (-0.0372, 0.7795)$ , 那表示HIV感染者吸烟没有什么影响对于感染者的CD4百分率, 但先前CD4百分率有非常高的正相关感染者的CD4百分率, 那是一致的与Huang等(2002), Xue和Zhu(2007)和Zhou和Lin(2014)等的发现。用经验似然方法和渐近正态分布方法, 我们得到 $(\beta_1, \beta_2)$  的置信水平 $1 - \alpha = 0.95$  置信区间被表示在图5.1。这些图显示这建议的块经验似然有更小的置信区间比渐近正态分布方法, 且置信区间变得更大作为 $\rho$  变得更大。

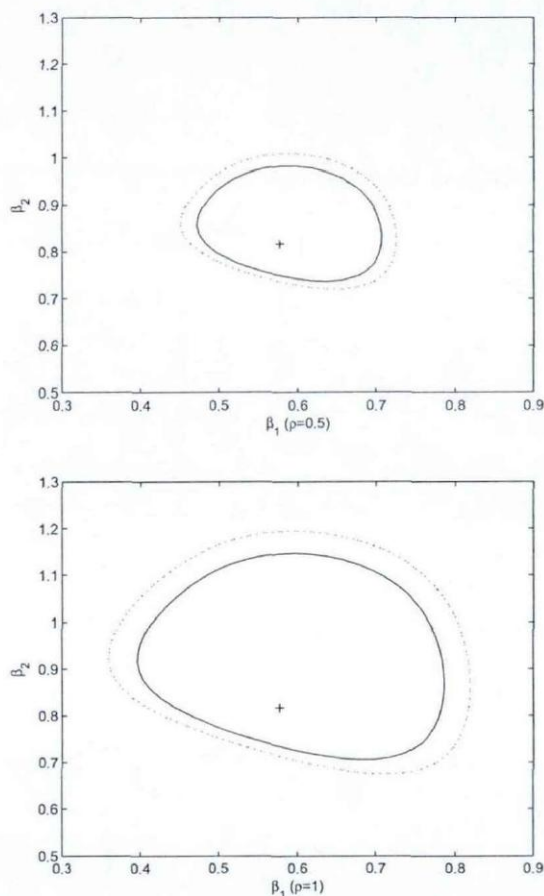


图 5.1 应用CD4 数据.  $\beta = (\beta_1, \beta_2)$ 的置信水平95% 的置信区间依据于正态逼近方法(dotted curve) 和经验似然(solid curve).

### 5.5 主要结果的证明

为了表示的方面，我们简化符号。令  $\varepsilon_N = \{(Nh)^{-1} \log N\}^{1/2}$ ,  $c_N = \varepsilon_N + h^2$ ,  $\vartheta_k = \int t^k K(t) dt$ ,  $v_k = \int t^k K^2(t) dt$  和  $W = (I_{nT} - S)^\tau (I_{nT} - S)$ . 注意  $S = (s_{11}, \dots, s_{1T}, \dots, s_{nT})^\tau$ . 表示一个典型的元素  $s(u)$  通过  $s_{it}(u) = (X_{it} \ 0)(D_u^\tau W_u D_u)^{-1} D_{it} K_h(U_{it} - u)$ . 定义  $N_{it}(u, j) = X_{it} X_{it}^\tau (\frac{U_{it}-u}{h})^j K_h(U_{it} - u)$ ,  $R_{it}(u, j) = X_{it} (\frac{U_{it}-u}{h})^j K_h(U_{it} - u) Y_{it}$ .

引理 5.1. 假设条件(A1)-(A7) 都拥有. 因此我们能得到

$$\sup_{u \in [-1, 1]} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T N_{it}(u, j) - G(u) f(u) \vartheta_j \right| = O_P(c_N), \quad (5-11)$$

$$\sup_{u \in [-1,1]} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T R_{it}(u, j) - \Phi(u) f(u) \vartheta_j \right| = O_P(c_N). \quad (5-12)$$

证明. 为证明(5-11), 它是充分条件是

$$\sup_{u \in [-1,1]} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(u, j) - E[N_{it}(u, j)]\} \right| = O_P[(Nh)^{-1} \log N]^{\frac{1}{2}} \quad (5-13)$$

和

$$\sup_{u \in [-1,1]} |E[N_{it}(u, j)] - G(u) f(u) \vartheta_j| = O_P(h^2). \quad (5-14)$$

利用条件(A2), 我们得到

$$\begin{aligned} & E\left[\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^{\tau} \left(\frac{U_{it} - u}{h}\right)^j K_h(U_{it} - u)\right] = E\left[X_{it} X_{it}^{\tau} \left(\frac{U_{it} - u}{h}\right)^j K_h(U_{it} - u)\right] \\ & = \int G(U_{it}) \left(\frac{U_{it} - u}{h}\right)^j K_h(U_{it} - u) f(U_{it}) dU_{it} = [G(u) + O(h)] \int u^j K(u) du [f(u) + O(h)] \\ & = G(u) f(u) \vartheta_j + O(h^2), \end{aligned}$$

那表明式子(5-14)成立. 让我们现在去证明式子(5-13). 我们仅需要证明

$$\sup_{u \in [-1,1]} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(u, j) - E[N_{it}(u, j)]\} \right| = o(l(N) \varepsilon_N), \quad (5-15)$$

对于任意的  $l(N) \rightarrow \infty$ , as  $N \rightarrow \infty$ . 主要思想去证明(5-15)将被考虑密集  $D = [-1, 1]$  的覆盖集, 通过一个有限的子覆盖集  $W_k$ , 其中它们的中心  $w_k$  的长度  $\xi_n = o(h^2)$ . 用  $\Gamma_n$  表示这样的集合总的数目. 因此  $\Gamma_n = O(\xi_n^{-1})$ . 那是容易证明

$$\begin{aligned} & \sup_{u \in [-1,1]} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(u, j) - E[N_{it}(u, j)]\} \right| \\ & = \max_{1 \leq k \leq \Gamma_n} \sup_{u \in W_k} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(u, j) - E[N_{it}(u, j)]\} \right| \\ & \leq \max_{1 \leq k \leq \Gamma_n} \sup_{u \in W_k} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(u, j) - N_{it}(w_k, j)\} \right| \\ & + \max_{1 \leq k \leq \Gamma_n} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{N_{it}(w_k, j) - E(N_{it}(w_k, j))\} \right| \\ & + \max_{1 \leq k \leq \Gamma_n} \sup_{u \in W_k} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \{E[N_{it}(w_k, j)] - E(N_{it}(u, j))\} \right| \\ & = \Pi_1 + \Pi_2 + \Pi_3. \end{aligned} \quad (5-16)$$

注意 $k(\cdot)$  是lipschitz 连续的在假定条件(A1) 和取 $\xi_n = O(\varepsilon_N h^2)$ , 我们得到

$$E(\Pi_1) \leq C \frac{\xi_n}{h^2} E[G(u)] = O(\varepsilon_N) = o(l(N)\varepsilon_N). \quad (5-17)$$

从式子(5-15), 我们立即能发现

$$E(\Pi_3) = o(l(N)\varepsilon_N). \quad (5-18)$$

对于 $\Pi_2$ , 我们运用截断方法, 定义

$$\bar{G}_{it} = X_{it} X_{it}^T \{ \|X_{it} X_{it}^T\| \leq CN^{\frac{1}{\delta}} \} \text{ and } \bar{G}_{it}^c = G_{it} - \bar{G}_{it}.$$

容易去查看

$$\begin{aligned} \Pi_2 &\leq \max_{1 \leq k \leq \Gamma_n} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (\bar{G}_{it} - E[\bar{G}_{it}]) K_h(U_{it} - w_k) \right| \\ &+ \max_{1 \leq k \leq \Gamma_n} \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (\bar{G}_{it}^c - E[\bar{G}_{it}^c]) K_h(U_{it} - w_k) \right| = \Pi_{21} + \Pi_{22}. \end{aligned} \quad (5-19)$$

通过利用

$$\begin{aligned} P(\Pi_{22} > l(N)\varepsilon_N) &\leq \sum_{i=1}^n \sum_{t=1}^T P(\|X_{it} X_{it}^T\| > l(N)N^{\frac{1}{\delta}}) \\ &= \sum_{i=1}^n \sum_{t=1}^T \frac{E\|X_{it} X_{it}^T\|^\delta}{[l(N)N^{\frac{1}{\delta}}]^\delta} = O\left(\frac{1}{l(N)^\delta}\right) = o(1). \end{aligned}$$

因此, 我们有

$$\Pi_{22} = o_p(l(N)\varepsilon_N). \quad (5-20)$$

观察

$$|(\bar{G}_{it} - E[\bar{G}_{it}]) K_h(U_{it} - w_k)| \leq \frac{2N^{\frac{1}{\delta}} l(N)}{h}.$$

通过取 $q = [T^{1+1/\delta} \varepsilon_N]$  和Bernstein 不等式对于 $\alpha$ -混合过程(看本章补充引理A.4), 我们得到, 对于任何

$$\begin{aligned} &P\left(\left|\frac{1}{T} \sum_{t=1}^T (\bar{G}_{it} - E[\bar{G}_{it}]) K_h(U_{it} - s_t)\right| > l(N)\varepsilon_N\right) \\ &\leq 4 \exp(-l(N)\varepsilon_N^2 Th) + C \left\{ \frac{2l(N)N^{\frac{1}{\delta}}}{[l(N)\varepsilon_N h]} \right\}^{\frac{1}{2}} [T^{1+\frac{1}{\delta}} \varepsilon_N] [T^{-\frac{1}{\delta}} \varepsilon_N^{-1}]^{-\kappa} \\ &= 4 \exp\left[-\frac{l(N)}{n} \log N\right] + C (\log N)^{\kappa+\frac{1}{2}} N^{-((\kappa+\frac{1}{2})-\frac{1}{2\delta})} T^{1+\frac{\kappa+1}{\delta}} h^{\kappa-1}. \end{aligned}$$

利用条件(A6), 那包涵

$$\begin{aligned} P(|\Pi_{21}| > l(N)\varepsilon_N) &\leq P\left(\max_{1 \leq k \leq \Gamma_n} \max_{1 \leq i \leq n} \left| \frac{1}{T} \sum_{t=1}^T (\bar{G}_{it} - E[\bar{G}_{it}]) K_h(U_{it} - w_k) \right| > l(N)\varepsilon_N\right) \\ &\leq C \Gamma_n N^{-l(N)} + C n \Gamma_n (\log N)^{\kappa+1/2} N^{-((\kappa+\frac{1}{2})-\frac{1}{2\delta})} T^{1+\frac{\kappa+1}{\delta}} h^{\kappa-1} \end{aligned}$$



$$= o(1) + O[(\log N)^{\kappa+1/2} N^{-((\kappa-\frac{1}{2})-\frac{1}{2\delta})} T^{\frac{\kappa+1}{\delta}} h^{\kappa-3}] = o(1).$$

因此, 我们得到

$$\Pi_{21} = o_p(l(N)\varepsilon_N). \quad (5-21)$$

运用式子(5-15)-(5-21), 我们得到式子(5-13)成立. 合并(5-13)和(5-14), 我们得到(5-11). 同样的方法能够证明式子(5-12). 因此, 这个引理5.1 证明被完成.

**引理 5.2.** 假设条件(A1)-(A6) 拥有. 因此我们能得到

$$(a) \quad s_{it} = O\left(\frac{1}{Nh}\right).$$

$$(b) \quad (H^T W H)^{-1} = (H^T H)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n).$$

$$\text{对于充分大的 } n, \text{ 其中 } \zeta_n = \frac{\sqrt{\ln n}}{nh} i_n i_n^T.$$

证明. (a). 通过大数定律和引理5.1,  $N^{-1} D_u^T W_u D_u = f(u)G(u) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \{1 + O_P(c_n)\}$ , 相应的结论:  $s_{it}(u) = N^{-1}(X_{it} \ 0)[N^{-1} D_u^T W_u D_u]^{-1} D_{it} K_h(U_{it} - u) = N^{-1} K_h(U_{it} - u) f^{-1}(u) \{1 + O_P(c_n)\} = O\left(\frac{1}{Nh}\right)$ .

(b).  $H^T W H = H^T (I_{nT} - S)^T (I_{nT} - S) H = H^T H - H^T S H - H^T S^T H + H^T S^T S H = \Omega_{11} - \Omega_{12} - \Omega_{13} + \Omega_{14}$ . 利用(a), 我们能显示  $\Omega_{1l}$  的第  $(i, j)$  个元素是  $[\Omega_{1l}]_{ij} = O\left(\frac{1}{Nh}\right)$ ,  $l = 2, 3, 4$ . 相似于 Su 和 Ullah (2006) 引理A.2 的证明. 我们容易证明  $(H^T W H)^{-1} = (H^T H)^{-1} + O(\zeta_n) = T^{-1} I_n + O(\zeta_n)$ .

**引理 5.3.** 假设条件(A1)-(A8) 都拥有. 令  $e_{it}$  是向量  $nT \times 1$  第  $it$  元素是1, 其余元素都是0. 因此

$$(a) \quad Z_{it} e_{it}^T W v_{it} = (Z_{it} - G^{-1}(u)\Phi(u)X_{it})v_{it} + o_P(1).$$

$$(b) \quad Z_{it} e_{it}^T W H (H^T H)^{-1} H^T W v_{it} = \frac{1}{T} \sum_{s=1}^T (Z_{it} - G^{-1}(u)\Phi(u)X_{it})v_{is} + o_P(1).$$

$$(c) \quad \text{令 } B_{it} = \tilde{Z}_{it} e_{it}^T \Lambda (I_{nT} - S) v_{it}, \text{ 因此 } \text{Var}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{s=1}^T B_{it}\right) \xrightarrow{d} \Delta.$$

证明. (a),(b) 是 Hu (2014) 的引理4.5 的直接结论. (c). 利用(a), (b), 我们得到  $Z_{it} e_{it}^T \Lambda (I_{nT} - S) v_{it} = \tilde{Z}_{it} \tilde{v}_{it} + o_p(1)$ . 明显的,  $E(B_{it}) = 0$  和

$$\text{Var}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T B_{it}\right) = \frac{1}{T} \sum_{t=1}^T \text{Var}(B_{it}) + \frac{2}{T} \sum_{t=1}^{T-1} (T-t) \text{cov}(B_{i1}, B_{i(t+1)}) = J_1 + J_2.$$

利用Hu (2014)的引理4.6,它是容易得到 $J_1 = \text{Var}(\tilde{Z}_{it}e_{it}^T\Lambda(I_{nT} - S)v_{it}) \xrightarrow{d} \Delta$ . 因此,为了证明(c),它的充分条件是 $J_2 = o(1)$ ,令 $d_n$ 是一系列正整数以至 $d_n h \rightarrow 0$ . 定义

$$J_{21} = \frac{1}{T} \sum_{l=1}^{d_n-1} |\text{Cov}(B_{i1}, B_{il})|, \quad J_{22} = \frac{1}{T} \sum_{l=d_n}^T |\text{Cov}(B_{i1}, B_{il})|.$$

从(A5)和(A7),通过选择 $d_n$ ,我们得到

$$J_{21} = \frac{2}{T} \sum_{l=1}^{d_n-1} |\text{Cov}(B_{i1}, B_{il})| \leq \frac{2d_n}{T} E|\tilde{Z}_{i1}\tilde{Z}_{il}\tilde{v}_{i1}\tilde{v}_{il}| \leq C\frac{2d_n}{T} = o(1).$$

下一步我们考虑 $J_{22}$ 的上限。为这个目的,利用补充引理A.2,我们得到

$$|\text{Cov}(B_{i1}, B_{il})| \leq C[E|B_{i1}|^{2+\delta}]^{\frac{1}{2+\delta}}[E|B_{il}|^{2+\delta}]^{\frac{1}{2+\delta}}[\alpha(l)]^{\frac{\delta}{2+\delta}}.$$

通过条件(A4)和(A5),通过选择 $d_n$ ,我们得到

$$J_{22} = \frac{2}{T} \sum_{l=d_n}^T |\text{Cov}(B_{i1}, B_{il})| \leq \frac{C}{T} \sum_{l=d_n}^{\infty} [\alpha(l)]^{\frac{\delta}{2+\delta}} \leq \frac{C}{T} \sum_{l=d_n}^{\infty} d_n^{-(1+\delta)} = o(1).$$

因此, $J_2 = o(1)$ 。(c)的证明被完成。

**引理 5.4.** 在定理5.1条件下,我们得到

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \eta_i(\beta) \xrightarrow{d} N(0, \Delta).$$

**证明.** 通过 $\eta_i(\beta)$ 的定义,我们得到

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^n \eta_i(\beta) &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it}e_{it}^T\Lambda(I_{nT} - S)v_{it} + \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it}e_{it}^T\Lambda[X_{it}\alpha(u_{it}) - s_{it}^T M] \\ &= I_{1n} + I_{2n}. \end{aligned}$$

首先,我们证明 $I_{1n} \xrightarrow{d} N(0, \Delta)$ . 鉴于Cramér-Wold设计的条件,这足以表明对于任何 $d \times 1$ 向量 $a = (a_1, a_2, \dots, a_d)^T \neq 0$ ,

$$a^T I_{1n} = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T a^T \tilde{Z}_{it}e_{it}^T\Lambda(I_{nT} - S)v_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{B}_{it} \rightarrow N(0, a^T \Delta a). \quad (5-22)$$

为证式子(5-22),我们运用Doob's小-块和大-块技术. 分割集合 $\{1, 2, \dots, T\}$ 成 $2q_n + 1$ 子集伴随着大-块的大小 $r = r_T$ 和小-块的大小 $s = s_T$ . 集合 $q = q_T = \lfloor \frac{T}{r_T + s_T} \rfloor$ 和定义随机变量如下,对于 $0 \leq j \leq q$ ,

$$\pi_{ij} = \sum_{t=k_m}^{k_m+r-1} \tilde{B}_{it}, \quad \xi_{ij} = \sum_{t=l_m}^{l_m+s-1} \tilde{B}_{it}, \quad \text{and} \quad \zeta_{iq} = \sum_{t=q(r+s)+1}^n \tilde{B}_{it}.$$

其中  $k_m = (m-1)(r+s) + 1$ ,  $l_m = (m-1)(r+s) + r + 1$ ,  $m = 1, \dots, q$ . 因此

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{B}_{it} = \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^n \sum_{j=0}^q \pi_{ij} + \sum_{i=1}^n \sum_{j=0}^q \xi_{ij} + \sum_{i=1}^n \zeta_{iq} \right\} = \frac{1}{\sqrt{N}} \{S_{n,1} + S_{n,2} + S_{n,3}\}.$$

为了证明式子(5-22), 这足以表明下面的结论: 当  $N \rightarrow \infty$ ,

$$\frac{1}{N} E(S_{n,2})^2 \rightarrow 0, \quad \frac{1}{N} E(S_{n,3})^2 \rightarrow 0, \quad (5-23)$$

$$\left| E \exp(itS_{n,1}) - \prod_{m=1}^q E \exp(it\pi_{ij}) \right| \rightarrow 0, \quad (5-24)$$

$$\text{Var}[(N)^{-1/2} S_{n,1}] \rightarrow a^\tau \Delta a, \quad (5-25)$$

$$\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^q E[\pi_{ij}^2 I\{|\pi_{ij}| > \epsilon \sqrt{n}\}] \rightarrow 0, \quad \forall \epsilon > 0. \quad (5-26)$$

关系式(5-23) 暗示  $S_{n,2}/\sqrt{N}$  和  $S_{n,3}/\sqrt{N}$  在概率上是渐近可忽略的; 式(5-24) 表明和项  $\pi_j$  在  $S_{n,1}/\sqrt{N}$  是渐近独立, 并且式子(5-25) 和(5-26) 是标准的Lindeberg-Feller 条件对于  $S_{n,1}/\sqrt{N}$  为独立设置的渐近正态的。

首先, 我们证明式(5-23). 为这个目的, 我们选取大块的大小为  $r_T$  且  $r_T = [T^{1/t}]$  和小块大小为  $s_T = [T^{1/(1+t)}]$ , 其中  $t$  被指定在条件(A4) 和  $[x]$  表示  $x$  的整数部分. 因此, 它可以很容易地显示条件 (A4) :

$$\frac{s_T}{r_T} \rightarrow 0, \quad \frac{r_T}{n} \rightarrow 0, \quad \text{and } q_T \alpha(s_T) \leq CT^{-1/(t+1)t} \rightarrow 0. \quad (5-27)$$

观察

$$\frac{1}{N} E(S_{n,2})^2 = \sum_{j=0}^q \text{Var}(\xi_{ij}) + 2 \sum_{0 \leq j_1 < j_2 \leq q} \text{Cov}(\xi_{ij_1}, \xi_{ij_2}) \equiv \Pi_1 + \Pi_2. \quad (5-28)$$

从平稳性和引理5.3 有

$$\Pi_1 = q_n \text{Var}(\xi_{ij}) = q_n \text{Var}\left(\sum_{j=1}^{s_T} \tilde{B}_{ij}\right) = q_T s_T [a^\tau \Delta a + o(1)]. \quad (5-29)$$

下一步考虑项  $\Pi_2$ . 令  $\tilde{r}_j = j(r_n + s_n)$ , 因此  $\tilde{r}_j - \tilde{r}_i \geq r_T$  对于所有的  $j_2 > j_1$ , 我们因此有

$$|\Pi_2| \leq \sum_{0 \leq j_1 < j_2 \leq q} \sum_{j_1=1}^{s_T} \sum_{j_2=1}^{s_T} |\text{Cov}(\tilde{B}_{\tilde{r}_j+r_T+j_1}, \tilde{B}_{\tilde{r}_j+r_T+j_2})| \leq \sum_{j_1=1}^{T-r_n} \sum_{j_2=j_1+r_T}^T |\text{Cov}(\tilde{B}_{j_1}, \tilde{B}_{j_2})|.$$

通过平稳性和引理5.3, 我们有

$$|\Pi_2| \leq 2T \sum_{j=r_T+1}^n |\text{Cov}(\tilde{B}_{i1}, \tilde{B}_{ij})| = o(T). \quad (5-30)$$

因此, 通过式子(5-27)-(5-30), 我们得到

$$\frac{1}{N} E(S_{n,2})^2 = O\left(\frac{q_T s_T}{T}\right) + o(1) = o(1). \quad (5-31)$$

关于  $\frac{1}{N} E(S_{n,3})^2$ , 从条件(A4) 和(A7). 由此得出结论

$$\begin{aligned} \frac{1}{N} E(S_{n,3})^2 &= \frac{1}{N} \sum_{i=1}^n \sum_{s=q(r+s)+1}^T \text{Var}(\tilde{B}_i) + \frac{1}{N} \sum_{i=1}^n \sum_{q(r+s)+1 \leq j_1 < j_2 \leq T} \text{cov}(\tilde{B}_{ij_1}, \tilde{B}_{ij_2}) \\ &\leq c_1 \frac{T - q(r+s)}{T} + c_2 \sum_{q(r+s)+1 \leq j_1 < j_2 \leq T} \text{cov}(\tilde{B}_{ij_1}, \tilde{B}_{ij_2}) = o(1). \end{aligned} \quad (5-32)$$

合并(5-31)和(5-32), 我们建立(5-23)。为了建立(5-24), 通过补充的引理A.1, 我们得到

$$\left| E \exp\left(it \sum_{j=1}^{q_T} \pi_{ij}\right) - \prod_{j=1}^{q_T} E[\exp(it\pi_{ij})] \right| \leq 16 \frac{T}{r_T} \alpha(s_T),$$

那将趋向于0作为  $T \rightarrow \infty$  利用(5-27)。

关于式子(5-25), 通过稳定性, 式子(5-27) 引理5.3, 很容易看出

$$\sum_{j=0}^q E(\pi_{ij}^2) = q_T E(\pi_{i1}^2) = q_T r_T \frac{1}{r_T} \text{Var}\left(\sum_{j=0}^{r_T} \tilde{B}_{ij}\right) \rightarrow a^T \Sigma a.$$

它仍然需要建立(5-26). 为这个目的, 通过利用引理5.3, 我们得到

$$\begin{aligned} E[\pi_{ij}^2 I\{|\pi_{ij}| > \epsilon\sqrt{T}\}] &\leq cT^{-\frac{\delta}{2}} E\left|\sum_{t=k_m}^{k_m+r-1} \tilde{B}_{it}\right|^{2+\delta} \\ &\leq cT^{-\frac{\delta}{2}} \left\{ r^{1+\frac{\delta}{2}} \max_{1 \leq t \leq T} [E(\tilde{B}_{it}^4)]^{\frac{2+\delta}{4}} + r^{1+\epsilon_1} [E(\tilde{B}_{it}^4)]^{\frac{2+\delta}{4}} \right\} \\ &\leq cT^{-\frac{\delta}{2}} \{c_1 r^{1+\frac{\delta}{2}} + c_2 r^{1+\epsilon_1}\}. \end{aligned}$$

因此, 利用  $r$  的定义, 我们得到

$$\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^q E[\pi_{ij}^2 I\{|\pi_{ij}| > \epsilon\sqrt{T}\}] \leq cT^{-1-\frac{\delta}{2}} \{c_1 r^{1+\frac{\delta}{2}} + c_2 r^{1+\epsilon_1}\} = o_p(1).$$

因此,  $I_{1n} \xrightarrow{d} N(0, \Delta)$  成立。相似Hu (2014) 的引理(4.4)的证明, 我们得到  $\frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T \tilde{Z}_{it} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M] = o_p(1)$ 。那么很容易得到  $I_{2n} = o_p(1)$ 。因此, 引理5.4的证明完成。

引理 5.5. 在定理5.1条件下, 我们得到

$$\frac{1}{\sqrt{N}} \sum_{i=1}^n \eta_i(\beta) \eta_i^T(\beta) \xrightarrow{p} \Delta.$$

证明. 我们同样用引理5.4的注记, 且表示  $R_{it} = \tilde{Z}_{it} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M]$ , 因此

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^n \eta_i(\beta) \eta_i^T(\beta) &= \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T B_{it} B_{it}^T + \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T R_{it} R_{it}^T \\ &+ \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T B_{it} R_{it}^T + \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{t=1}^T R_{it} B_{it}^T = \Xi_1 + \Xi_2 + \Xi_3 + \Xi_4. \end{aligned}$$

我们能相似引理5.4得到  $\Xi_1 \xrightarrow{p} \Delta$ , 通过大数定律和  $\Xi_v \xrightarrow{p} 0, v = 2, 3, 4$ . 因此, 引理5.5被完成。

引理 5.6. 在定理5.1条件下, 我们得到

$$(i) \max_{1 \leq i \leq n} \|\eta_i(\beta)\| = o_p(N^{1/2}), \quad (5-33)$$

$$(ii) \lambda(\beta) = O_p(N^{-1/2}). \quad (5-34)$$

证明. 相似于Owen (1998)给的结论, 从式子(5-9)和引理5.4和引理5.5, 我们能容易证明式子(5-34)。我们现在证明式子(5-33)。

$$\begin{aligned} \max_{1 \leq i \leq n} \|\eta_i(\beta)\| &\leq \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{Z}_{it} e_{it}^T \Lambda (I_{nT} - S) v_{it} \right\| + \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{Z}_{it} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M] \right\| \\ &= M_1 + M_2. \end{aligned}$$

对于  $M_1$ , 注意  $E[\sum_{t=1}^T \sum_{s=1}^T \tilde{Z}_{it} e_{it}^T \Lambda (I_{nT} - S) v_{it} \tilde{Z}_{is} e_{is}^T \Lambda (I_{nT} - S) v_{is}] = \Delta_{s,s} + o(1) < \infty$ , 其中  $\Delta_{s,s}$  表示第  $s$  个对角元素  $\Delta$ ,  $B_{it}$  是独立同分布的在截面第  $i$  个体为每一个固定时间。通过Owen (1990)的引理3, 我们得到  $M_1 = o_p(N^{1/2})$ 。现在看第二项  $M_2$ , 运用Markov 不等式, 我们得到

$$\begin{aligned} P(M_2 > N^{1/2}) &\leq \frac{1}{N} \sum_{i=1}^n E \left\| \sum_{t=1}^T \tilde{Z}_{it} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M] \right\|^2 \\ &= \frac{1}{N} \sum_{i=1}^n E \left\| \sum_{t=1}^T \sum_{s=1}^T \tilde{Z}_{it} \tilde{Z}_{is} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M] e_{is}^T \Lambda [X_{is} \alpha(u_{is}) - s_{is}^T M] \right\|^2 \\ &\leq \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \left\{ E \left[ \tilde{Z}_{it} e_{it}^T \Lambda [X_{it} \alpha(u_{it}) - s_{it}^T M]^2 \right] \right\}^{1/2} \left\{ E \left[ \tilde{Z}_{is} e_{is}^T \Lambda [X_{is} \alpha(u_{is}) - s_{is}^T M]^2 \right] \right\}^{1/2} \\ &= o_p(1) \rightarrow 0. \end{aligned}$$

因此,  $M_2 = o_p(N^{1/2})$  和  $\max_{1 \leq i \leq n} \|\eta_i(\beta)\| = o_p(N^{1/2})$ 。

证明定理5.1. 应用泰勒展开式, 从(5-8), 我们能得到下式

$$\log \mathcal{L}_n(\beta) = 2 \sum_{i=1}^n \left\{ \lambda^\tau \eta_i(\beta) - \frac{[\lambda^\tau \eta_i(\beta)]^2}{2} \right\} + o_p(1). \quad (5-35)$$

应用泰勒展开式, 从(2-6), 我们能得到下式通过式子(5-9), 因此

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^\tau \eta_i(\beta)} \\ &= \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) - \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) \eta_i^\tau(\beta) \lambda + \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta) [\lambda^\tau \eta_i(\beta)]^2}{1 + \lambda^\tau \eta_i(\beta)}. \end{aligned} \quad (5-36)$$

鉴于引理5.4-5.6, 我们得到

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta) [\lambda^\tau \eta_i(\beta)]^2}{1 + \lambda^\tau \eta_i(\beta)} \right\| &\leq \frac{1}{n} \sum_{i=1}^n \frac{\|\eta_i(\beta)\|^3 \|\lambda\|^2}{|1 + \lambda^\tau \eta_i(\beta)|} \\ &\leq \|\lambda\|^2 \max_{1 \leq i \leq n} \|\eta_i(\beta)\| \frac{1}{n} \sum_{i=1}^n \|\eta_i(\beta)\|^2 = O_p(N^{-1}) o_p(N^{1/2}) O_p(1) = o_p(N^{-1/2}), \end{aligned}$$

那合并(5-36), 得到  $\sum_{i=1}^n [\lambda^\tau \eta_i(\beta)]^2 = \sum_{i=1}^n \lambda^\tau \eta_i(\beta) + o_p(1)$  和

$$\lambda = \left[ \sum_{i=1}^n \eta_i(\beta) \eta_i^\tau(\beta) \right]^{-1} \sum_{i=1}^n \eta_i(\beta) + o_p(N^{-1/2}).$$

因此, 通过式子(5-35), 我们得到

$$\log \mathcal{L}_n(\beta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\beta) \right)^\tau \left( \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) \eta_i^\tau(\beta) \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\beta) \right) + o_p(1),$$

那合并引理5.4和5.5, 证明了定理5.1。

证明定理5.2. 顺延着Hu 和Li (2011)的定理2的证明的相似结论, 我们能证明

$$\hat{\beta} - \beta = \hat{\Psi}^{-1} \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) + o_p(n^{-1/2}).$$

通过引理5.3, 我们能证明  $\hat{\Psi} \xrightarrow{P} \Psi$  通过利用大数定律, 合并引理5.4 和Slutsky 定理, 我们能够容易完成定理5.2的证明。

## 5.6 本章补充的引理

这节补充一些引理, 那被用来证明主要结论在上一节。令随机变量  $\{Z_i, i \geq 1\}$  是一组平稳 $\alpha$ -混合序列伴随混合系数是  $\{\alpha(m)\}$ 。

引理A.1. [Volkonskii 和Roazanov, 1959] 令  $U_1, \dots, U_n$  是 $\alpha$ -混合随机变量分别关于在 $\sigma$ -代数上的测量  $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_m}^{j_m}$ , 伴随着  $1 \leq i_1 < j_1 < \dots < j_m \leq$

$n$ ,  $i_{l+1} - j_l \geq w \geq 1$  和  $|U_j| \leq 1$  对于  $l, j = 1, 2, \dots, m$ . 因此

$$\left| E\left(\prod_{j=1}^m U_j\right) - \prod_{j=1}^m EU_j \right| \leq 16(m-1)\alpha_w,$$

其中  $\mathcal{F}_a^b = \sigma\{U_i, a < i \leq b\}$  表示  $\sigma$ -域来自于  $U_{a+1}, U_{a+2}, \dots, U_b$ ,  $\alpha_n$  是混合系数。

**引理A.2. [Hall 和 Heyde, 1980, 推论A.2, p.278]** 假定  $X$  and  $Y$  是随机变量, 以至  $E|X|^p < \infty, E|Y|^q < \infty$ , 其中  $p, q > 1, p^{-1} + q^{-1} < 1$ . 因此

$$|EXY - EXEY| \leq 8\|X\|_p\|Y\|_q \left\{ \sup_{A \in \sigma(X), B \in \sigma(Y)} |P(A \cap B) - P(A)P(B)| \right\}^{1-p^{-1}-q^{-1}}.$$

**引理A.3. [Yang, 2007, Theorem 2.2]** 令  $\lambda > 2, \mu > 0, EZ_i = 0$  和  $E|Z_i|^{\lambda+\mu} < \infty$ . 假定  $\alpha(n) = O(n^{-r})$  对于  $r > \lambda(\lambda + \mu)/(2\mu)$ . 因此, 对于任意指定的  $\varepsilon > 0$ , 存在实数  $C = C(r, \mu, \varepsilon, \lambda)$  以致

$$E \max_{1 \leq k \leq n} \left| \sum_{i=1}^k Z_i \right|^\lambda \leq C \left\{ n^\varepsilon \sum_{i=1}^n E|Z_i|^\lambda + \left( \sum_{i=1}^n \|Z_i\|_{\lambda+\mu}^2 \right)^{\lambda/2} \right\}.$$

**引理A.4. [Fan 和 Yao, 2003, Theorem 2.18(ii), p.73]** 假定  $P(|X_t| \leq b) = 1$ . 因此

(ii) 对于每一个  $q = 1, \dots, [n/2]$  和  $\varepsilon > 0$ ,

$$P(|S_n| > n\varepsilon) \leq 4 \exp\left(-\frac{\varepsilon^2 q}{8v^2(q)}\right) + 22\left(1 + \frac{4b}{\varepsilon}\right)^{\frac{1}{2}} q \alpha\left(\frac{n}{2q}\right),$$

其中  $v^2(q) = 2\sigma^2(q)/p^2 + b\varepsilon/2, p = n/(2q)$ , 和

$$\begin{aligned} \sigma^2(q) = \max_{0 \leq j \leq 2q-1} E \{ & ([jp] + 1 - jp)X_1 + X_2 + \dots + X_{[(j+1)p][jp]} \\ & + (jp + p - [jp + p])X_{[(j+1)p]-[jp]+1} \}^2. \end{aligned}$$





## 第6章 带固定效应面板数据部分线性时变系数模型的经验似然

### 6.1 引言

面板数据的使用在微观经济学和宏观经济学的实证研究中越来越受欢迎。面板数据集具有多种优势比纯截面或时间序列数据集，其中最重要的一点是，面板数据为研究人员提供了模拟横截面单元间异质性和随着时间的变化结构的变化的灵活方法。例如，经济机制可能会受到偏好改变、技术进步、制度变迁和经济转轨影响，随着时间的推移，这可能会导致经济变量之间的函数关系的变化。面板数据通常包括不同背景和不同抽样的个体单位的时间序列。不同的个体特征是它的异质性，其中大部分是不可观测的。尽管传统的面板数据模型经常假设均匀的斜率便于估计和推断，这样假设经常被实证研究所拒绝(例如: Lee, Pesaran 和Smith1997; Su和Chen, 2013)在面板数据模型的斜率异质性建模越来越多被关注。虽然个体异质性和结构变化可能共存，存在的面板数据模型最多只处理这两个重要特性之一。

对面板数据模型结构的变化研究可以分为两类，一类是考虑突然的变化，另一类是模拟平滑的变化。对于前一类方法，参见例如Bai(2010), Kim(2011)和Qian和Su(2014)。后一类方法主要动机来自时变(功能)系数模型或非参数回归模型的时间序列框架。例如，Li, Chen和Gao(2011), Fan和Yao(2003)研究的时变系数面板数据模型，并利用局部线性虚拟变量方法来估计函数的系数。在过去二十年中，时变系数模型已经获得了很多关注，因为它们在经济和金融方面有许多应用。例如，金融市场模型将个人股票的回报与市场指数或另一个股票的回报相关联，该系数通常被称为资本资产定价模型中的 $\beta$ 系数(例如，参见Tsay 2002)，研究表明 $\beta$ 系数可能随时间变化。

这一章考虑下面的带固定效应面板数据部分线性时变系数模型

$$Y_{it} = Z_{it}^T \beta + X_{it}^T \alpha(t/T) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (6-1)$$

其中 $Y_{it}$  是响应变量,  $(Z_{it}, X_{it}) \in R^p \times R^q$ ,  $\beta = (\beta_1, \dots, \beta_p)^\tau$  是一个 $p$ -维未知参数, 和上标 $\tau$  表示一个向量的转置。  $\alpha(t/T) = (\alpha_1(t/T), \dots, \alpha_q(t/T))^\tau$  是一个 $q$ -维向量的未知函数和 $\mu_i$  是不能观察的个体效应,  $v_{it}$  是随机误差。这儿, 我们假定 $\mu_i$  是独立同分布的均值为0 和有限的方差 $\sigma^2 > 0$ 。我们允许 $\mu_i$  是相关的与 $Z_{it}, X_{it}$  伴随一个的相关结构, 此时模型(6-1)被叫固定效应模型。如果 $\mu_i$  是与 $X_{it}$  和 (或)  $Z_{it}$  是不相关的, 模型(6-1)被叫随机效应模型。固定效应的样本稳定性要优于随机效应样本(Baltagi, 2005; Horowitz 和Lee, 2004 等)。分析带固定效应的面板数据具有更多的挑战。在此, 我们只考虑固定效应情况。

近来趋势时变系数模型得到大量的注意, 因为他们可以应用到经济, 金融, 和环境的评价中。这篇文章的动机是去分析英国的温度变化。我们都知道, 气象数据具有长期的趋向性。研究趋向性一个共同的方法利用半参数时变系数模型, 在这一章我们考虑用带固定效应面板数据部分线性时变系数模型去描述英国地区的近些年的温度变化。我们建立该模型的参数估计和时变函数的估计。对于本章, 我们跟第5章一样, 假定 $\{(X_{it}, Z_{it}, Y_{it}, v_{it}), t \geq 1\}$  是平稳 $\alpha$ -混合的对于每个 $i$ 。

剩下的这一章组成是这样的: 6.2节介绍方法和建立 $\beta$  经验对数似然比率函数, 假定和主要结论被给在6.3节, 一些模拟安排在6.4节, 主要结果的证明安排在6.5节。

## 6.2 模型和方法

假定模型拥有下面的约束条件 $\sum_{i=1}^n \mu_i = 0$ 。令 $\mu = (\mu_2, \dots, \mu_n)^\tau$  和 $\mu_0 = (-\sum_{i=2}^n \mu_i, \mu_2, \dots, \mu_n)^\tau$ 。我们重新写模型(6-1) 成矩阵的形式

$$Y = Z\beta + X\alpha(t/T) + H\mu + V, \quad (6-2)$$

其中 $H = [-i_{n-1} \quad I_{n-1}] \otimes i_T$  是一个 $nT \times (n-1)$  矩阵,  $I_n$  表示 $n \times n$  单位矩阵,  $i_n$  表示 $n \times 1$  个1的向量, 有许多方法可以估计参数 $\{\beta_j, j = 1, \dots, p\}$  和时变系数函数 $\{\alpha_i(\cdot), i = 1, \dots, q\}$ 。我们运用剖面似然方法: 假定我们有一个随机样本 $\{(Z_{it1}, \dots, Z_{itp}, X_{it1}, \dots, X_{itq}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T\}$  来自模型(6-1)。令 $\gamma = (\mu^\tau, \beta^\tau)^\tau$ 。指定 $\gamma$ , 我们能够运用一个局部线性回归技术估计变系数函数 $\{\alpha_j(\cdot), j = 1, \dots, q\}$  来自模型(6-1)。我们能估计 $\alpha_j(\cdot)$  局部线性函数

$$\alpha_j(\cdot) \approx \alpha_j(\omega) + \alpha_j'(\omega)(\cdot - \omega) \equiv a_j + b_j(\cdot - \omega), \quad j = 1, \dots, q,$$

其中  $0 < \omega < 1, \alpha'_j(\omega) = \partial \alpha_j(\tau) / \partial \omega$ 。这导致下面的权局部最小二乘问题：发现  $\{(a_j, b_j), j = 1, \dots, q\}$  最小化

$$\sum_{i=1}^n \sum_{t=1}^T \left\{ \left( Y_{it} - Z_{it}^T \beta - \mu_i \right) - \sum_{j=1}^q \left[ a_j + b_j(t/T - \omega) \right] X_{it} \right\}^2 K_{h_1}(t/T - \omega), \quad (6-3)$$

其中  $K_h(\cdot) = K(\cdot/h)/h$ ,  $K(\cdot)$  是一个核函数和  $h$  是一系列接近于0的正数, 叫窗宽。简单计算得到

$$(\hat{\alpha}_1(\cdot), \dots, \hat{\alpha}_q(\cdot), h\hat{a}'_1(\cdot), \dots, h\hat{a}'_q(\cdot))^T = (D_\omega W_\omega D_\omega)^{-1} D_\omega^T W_\omega (Y - Z\beta - D\mu),$$

其中

$$Z = \begin{pmatrix} Z_{11}^T \\ \vdots \\ Z_{1T}^T \\ \vdots \\ Z_{N1}^T \\ \vdots \\ Z_{nT}^T \end{pmatrix}, \quad X = \begin{pmatrix} X_{11}^T \\ \vdots \\ X_{1T}^T \\ \vdots \\ X_{N1}^T \\ \vdots \\ X_{nT}^T \end{pmatrix}, \quad D_\omega = \begin{pmatrix} X_{11}^T & \frac{1-\omega T}{Th} X_{11}^T \\ \vdots & \vdots \\ X_{1T}^T & \frac{T-\omega T}{Th} X_{1T}^T \\ \vdots & \vdots \\ X_{N1}^T & \frac{T-\omega T}{Th} X_{N1}^T \\ \vdots & \vdots \\ X_{nT}^T & \frac{T-\omega T}{Th} X_{nT}^T \end{pmatrix},$$

$Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{nT})^T$  和  $W_u = \text{diag}(K_h(\frac{1-\omega T}{T}), \dots, K_h(\frac{T-\omega T}{T}), \dots, K_h(\frac{T-\omega T}{T}))$ .

表示光滑矩阵  $S$  是

$$S = \begin{pmatrix} (X_{11}^T \ 0_q^T)(D_{\omega_1}^T W_{\omega_1} D_{\omega_1})^{-1} D_{\omega_1}^T W_{\omega_1} \\ \vdots \\ (X_{1T}^T \ 0_q^T)(D_{\omega_T}^T W_{\omega_T} D_{\omega_T})^{-1} D_{\omega_T}^T W_{\omega_T} \\ \vdots \\ (X_{N1}^T \ 0_q^T)(D_{\omega_1}^T W_{\omega_1} D_{\omega_1})^{-1} D_{\omega_1}^T W_{\omega_1} \\ \vdots \\ (X_{nT}^T \ 0_q^T)(D_{\omega_T}^T W_{\omega_T} D_{\omega_T})^{-1} D_{\omega_T}^T W_{\omega_T} \end{pmatrix} = \begin{pmatrix} s_{11} \\ \vdots \\ s_{1T} \\ \vdots \\ s_{N1} \\ \vdots \\ s_{nT} \end{pmatrix}. \quad (6-4)$$

一个估计实参数 $\gamma$ 可以通过最小化下式:

$$\hat{\gamma} = \arg \min_{\gamma} [Y - Z\beta - H\mu - S(Y - Z\beta - H\mu)]^r [Y - Z\beta - H\mu - S(Y - Z\beta - H\mu)], \quad (6-5)$$

令 $Z_i = (Z_{i1}^r, \dots, Z_{iT}^r)^r$ ,  $Y_i = (Y_{i1}, \dots, Y_{iT})^r$ ,  $\tilde{Z} = (I_{nT} - S)Z$ ,  $\tilde{H} = (I_{nT} - S)H$ ,  $\tilde{Y} = (I_{nT} - S)Y$ ,  $\tilde{\mu} = (\tilde{H}^r \tilde{H})^{-1} \tilde{H}^r (\tilde{Y} - \tilde{Z}\beta)$  和  $\Phi = I_{nT} - \tilde{H}(\tilde{H}^r \tilde{H})^{-1} \tilde{H}^r$ 。参数 $\gamma$ 的剖面似然的估计被指定通过下式:

我们引入下面的辅助随机向量:

$$\Xi_i(\beta) = \tilde{Z}_i \Phi (\tilde{Y}_i - \tilde{Z}_i \beta), \quad i = 1, \dots, n. \quad (6-6)$$

注意 $E(\Xi_i(\beta)) = 0$  如果 $\beta$  是实值参数。因此, 相似于Owen (1990), 我们定义经验对数似然比率如下

$$\log \mathcal{L}_n(\beta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) : \sum_{i=1}^n p_i \Xi_i(\beta) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \quad (6-7)$$

利用拉格朗日乘数方法, 我们能得到 $p_i = \frac{1}{n[1 + \lambda^r \Xi_i(\beta)]}$ , 和 $\log \mathcal{L}_n(\beta)$  能被表示如下

$$\log \mathcal{L}_n(\beta) = 2 \sum_{i=1}^n \log \{1 + \theta^r \Xi_i(\beta)\}, \quad (6-8)$$

其中 $\theta(\beta)$  被决定下式

$$\frac{1}{n} \sum_{i=1}^n \frac{\Xi_i(\beta)}{1 + \theta^r \Xi_i(\beta)} = 0. \quad (6-9)$$

### 6.3 假定和主要结论

在这一章, 我们将显示如果 $\beta$  是实参数向量,  $\log \mathcal{L}_n(\beta)$  是渐近卡方分布。在得到主要结论前, 我们给出下面的一些假定

(A1) 核函数 $K(v)$  是一个在连续可微的对称概率密度函数在紧支撑 $[-1, 1]$  上。

(A2)  $E(Z_{it}) = 0_q, E(X_{it}) = 0_p$ , 存在正实数矩阵 $\Sigma_Z = E(Z_{it} Z_{it}^r)$ ,  $\Sigma_X = E(X_{it} X_{it}^r)$ ,  $\Sigma_{XZ} = E(Z_{it} X_{it}^r)$  和  $\tilde{\Sigma} = \Sigma_Z - \Sigma_{XZ}^r \Sigma_X^{-1} \Sigma_{XZ}$ 。

(A3)  $\{\alpha_j(\cdot), j = 1, \dots, q\}$  是二级连续可微函数在 $[0, 1]$  上。

(A4) (i)  $\{X_{it}, Y_{it}, Z_{it}, v_{it}\}$  是独立同分布的对于截面第 $i$  个体为每一个固定时间 $t$ , 并且是严平稳的为 $t$  序列对于每个固定的 $i$ 。(ii) 对于固定的 $i$ , 过

程 $\{X_{it}, Y_{it}, Z_{it}, v_{it}\}$  是一个严平稳的 $\alpha$ -混合序列伴随这混合系数满足条件 $\alpha(k) = O(k^{-\kappa})$ ,  $\kappa = \frac{(2+\delta)(1+\delta)}{\delta}$  和 $\delta > 2$ 。

(A5) 存在一些 $\delta > 2$  以致 $E\|X_i\|^{2+\delta} < \infty$ ,  $E\|Z_i\|^{2+\delta} < \infty$ ,  $E\|v_i\|^{2+\delta} < \infty$ , 其中 $\|\cdot\|$  是这个 $L_2$ -距离。

(A6) 令 $N = nT$ , 窗框 $h$  满足 $Nh \rightarrow \infty$ ,  $Nh^6 \rightarrow 0$ ,  $(\log N)^{\kappa+1/2} N^{-((\kappa-\frac{1}{2})-\frac{1}{2\delta})} T^{\frac{\kappa+1}{\delta}} h^{\kappa-3} \rightarrow 0$ . 其中 $\kappa$  和 $\delta$  被定义在上面的A4(ii)。

(A7)  $E|\check{Z}_{it}|^{2+\delta} < \infty$ ,  $\Psi = \sum_{t=1}^T [\check{Z}_{it}(\check{Z}_{it} - \sum_{s=1}^T \frac{\check{Z}_{is}}{T})^\tau]$  是正的有限的, 其中 $\check{Z}_{it} = Z_{it} - \Sigma_X^{-1} \Sigma_{ZX} X_{it}$ 。

**定理 6.1.** 假定(A1)-(A7) 条件满足. 对于模型(5-1), 如果 $\beta_0$  是正的实值参数, 则有 $\log \mathcal{L}_n(\beta) \xrightarrow{d} \chi_p^2$ , 作为 $n \rightarrow \infty$ , 其中 $\chi_p^2$  是一个标准自由度为 $p$  卡方随机变量和 $\xrightarrow{d}$  代表依分布收敛。

作为定理的一个结果, 参数的 $\beta_0$  的置信区间能被建立。更准确的说, 对于任何 $0 \leq \alpha < 1$ , 令 $c_\alpha$  是 $\chi^2$  分布的 $1 - \alpha$  分位数以至 $P(\chi_p^2 > c_\alpha) \leq 1 - \alpha$ 。因此 $\ell_{EL}(\alpha) = \{\beta \in R^p : \log \mathcal{L}_n(\beta) \leq c_\alpha\}$  建立 $\beta$  一个置信区间是渐近 $1 - \alpha$  收敛的, 因为 $\beta$  属于 $\ell_{EL}(\alpha)$  等价于 $\log \mathcal{L}_n(\beta) \leq c_\alpha$ 。

### 6.4 Monte Carlo 模拟研究

在这节, 我们将做有限样本的模拟来阐明我们建议的经验似然方法优于渐近正态方法。我们考虑下面的带固定效应面板数据时变系数模型:

$$Y_{it} = Z_{it}^T \beta + X_{it} \alpha(t/T) + \mu_i + v_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (6-10)$$

其中 $\beta = (\beta_1, \beta_2)^T = (1, 2)^T$ ,  $\alpha(t/T) = \sin(2\pi t/T)$ ,  $\mu_i = \rho \bar{X}_i + w_i$  伴随着 $\rho = 1, 2$  和 $w_i \sim N(0, 1)$  对于 $i = 2, 3, \dots, n$ , 和 $\mu_i = -\sum_{i=2}^n \mu_i$ ,  $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}$ . 我们用 $\rho$  去控制 $\mu_i$  和 $\bar{X}_i$  的相关关系.  $X_{it}$  和 $Z_{it}$  是来自下面的AR(1) 模型,

$$\begin{aligned} Z_{it,1} &= 0.6Z_{it-1,1} + e_{i,1}, & Z_{it,2} &= 0.8Z_{it-1,2} + e_{i,2}, & X_{it} &= 0.6X_{it-1} + u_{i,1}, \\ v_{it} &= 0.5v_{it-1} + \varepsilon_i, & \varepsilon_i &\stackrel{i.i.d}{\sim} N(0, 1), \end{aligned}$$

其中 $u_i = (u_{i,1}, u_{i,2})^T \stackrel{i.i.d}{\sim} N((0, 0)^T, \text{diag}(1, 1))$  和 $e_i = (e_{i,1}, e_{i,2})^T \stackrel{i.i.d}{\sim} N((0, 0)^T, \text{diag}(1, 1))$ . 容易证明 $\{X_{it}, Z_{it}, v_{it}\}$  是平稳和 $\alpha$ -混合的。

现在我们做模拟，我们选取的核为  $K(u) = \frac{15}{16}(1 - u^2)^2 I\{|u| \leq 1\}$ . 采取去掉第  $i$  个观测主题的交错核实验验证(cross validation, CV) 方法来获得窗宽  $CV(h)$  被用最小化下式

$$CV(h) = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T (Y_{it} - Z_{it}^T \hat{\beta}_i - X_{it}^T \hat{\alpha}_i(t/T))^2,$$

其中  $(\hat{\beta}_1, \hat{\beta}_2)^T$  和  $\hat{\alpha}_i(t/T)$  分别是  $(\beta_1, \beta_2)^T$  和  $\alpha(t/T)$  的估计，那计算所有的测量但除掉第  $i$  个观测主题.

我们考虑两种方法计算个体  $\beta_i (i = 1, 2)$  的置信区间的覆盖率(CPs): 经验似然(EL)和正态逼近方法(NA). 样本大小  $(n, T)$  我们分别选取  $(30, 6), (30, 10), (50, 6)$  和  $(50, 10)$ , 我们将做 500 次重复在置信水平分别为  $1 - \alpha = 0.90$  和  $0.95$ , 一些有代表性的  $(\beta_1, \beta_2)$  的置信区间的覆盖率被报道在表 6.1, 从表 6.1, 我们可以看出经验似然方法表现很好. 因为我们还能看出覆盖率随着样本容量  $(n, T)$  的变大而增大. 同样我们能发现置信区间的覆盖率变小随着  $\rho$  变大. 我们还可以看出经验似然方法优于渐近正态分布方法因为经验似然方法给出更高的覆盖率和更小的置信区间比渐近正态方法.

表 6.1 CP 表示  $(\beta_1, \beta_2)$  的置信区间依据于经验似然(EL)和正态逼近方法(NA)在置信水平是  $1 - \alpha = 0.90$  和  $0.95$  的覆盖率

$\rho$	$1 - \alpha$	$(n, T)$	$CP_{EL}$	$CP_{NA}$	$\rho$	$1 - \alpha$	$(n, T)$	$CP_{EL}$	$CP_{NA}$
0	0.90	(30,6)	0.8621	0.738	1	0.90	(30,6)	0.8242	0.8040
		(30,10)	0.8762	0.770			(30,10)	0.866	0.8333
		(50,6)	0.8633	0.786			(50,6)	0.856	0.8410
		(50,10)	0.8920	0.816			(50,10)	0.872	0.8533
	0.95	(30,6)	0.9233	0.912		0.95	(30,6)	0.908	0.896
		(30,10)	0.9367	0.922			(30,10)	0.916	0.906
		(50,6)	0.9300	0.917			(50,6)	0.920	0.9176
		(50,10)	0.9467	0.936			(50,10)	0.9344	0.9268

Note: “ $CP_{EL}$  和  $CP_{NA}$ ” 表示 CP 分别依据于经验似然和正态逼近方法.

## 6.5 主要结果的证明

为了表示的方面，我们简化符号。令  $\varepsilon_N = \{(Nh)^{-1} \log N\}^{1/2}$ ,  $c_N =$

$\varepsilon_N + h^2$ ,  $v_k = \int t^k K(t) dt$ ,  $v_k = \int t^k K^2(t) dt$  和  $R = (I_{nT} - S)^\tau (I_{nT} - S)$ 。  
 注意  $S = (s_{11}, \dots, s_{1T}, \dots, s_{nT})^\tau$ 。表示一个典型的元素  $s(u)$  通过  $s_{it}(u) = (X_{it} \ 0)(D_u^\tau W_u D_u)^{-1} D_{it} K_h(U_{it} - u)$ 。

**引理 6.1.** 假设条件(A1)-(A7) 都满足. 因此我们有

$$D_\omega^\tau W_\omega D_\omega = \Lambda \otimes \Sigma_X [1 + c_N], \quad (6-11)$$

其中  $\Lambda = \begin{pmatrix} v_0 & 0 \\ 0 & v_2 \end{pmatrix}$  和  $\Sigma_X = E(X_{it} X_{it}^\tau)$ 。

**证明.** 观察下式

$$D_\omega^\tau W_\omega D_\omega = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^\tau K_h\left(\frac{t-\omega T}{Th}\right) & \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right) K_h\left(\frac{t-\omega T}{T}\right) \\ \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right) K_h\left(\frac{t-\omega T}{T}\right) & \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right)^2 K_h\left(\frac{t-\omega T}{T}\right) \end{pmatrix} \quad (6-12)$$

利用条件(A2), 我们得到

$$\begin{aligned} E\left[\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right)^j K_h\left(\frac{t-\omega T}{T}\right)\right] &= E\left[X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right)^j K_h\left(\frac{t-\omega T}{T}\right)\right] \\ &= \int \Sigma_X \cdot \left(\frac{t-\omega T}{Th}\right)^j K_h\left(\frac{t-\omega T}{T}\right) dU_{it} = [\Sigma_X + O(h)] \int u^j K(u) du [1 + O(h)] \\ &= \Sigma_X v_j [1 + O_P(h^2)], \end{aligned} \quad (6-13)$$

我们由何 (2016) 的引理5.1, 可以得到

$$\sup \left| \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \left\{ X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right)^j K_h\left(\frac{t-\omega T}{T}\right) - E\left[X_{it} X_{it}^\tau \left(\frac{t-\omega T}{Th}\right)^j K_h\left(\frac{t-\omega T}{T}\right)\right] \right\} \right| = O_P(e_N), \quad (6-14)$$

合并(6-13) 和(6-14), 我们得到这个引理6.1成立。

**引理 6.2.** 在假设条件(A1)-(A6) 条件下. 我们有

$$Z^\tau (I - S)^\tau (I - S) Z = (\Sigma_Z - \Sigma_{XZ}^\tau \Sigma_X^{-1} \Sigma_{XZ}) [1 + c_N],$$

证明. 观察下式

$$D_{\omega}^{\tau} W_{\omega} Z = \begin{pmatrix} \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} Z_{it}^{\tau} K_h\left(\frac{t-\omega T}{Th}\right) \\ \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} Z_{it}^{\tau}\left(\frac{t-\omega T}{Th}\right) K_h\left(\frac{t-\omega T}{T}\right) \end{pmatrix}$$

因为  $E(X_{it} Z_{it}^{\tau}) = \Sigma_{XZ}$ , 顺从引理6.1, 我们得到

$$\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} Z_{it}^{\tau} K_h\left(\frac{t-\omega T}{Th}\right) = \Sigma_{XZ} v_0 [1 + O_P(h^2)],$$

和

$$\frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} Z_{it}^{\tau}\left(\frac{t-\omega T}{Th}\right) K_h\left(\frac{t-\omega T}{T}\right) = \Sigma_{XZ} v_0 \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} [1 + O_P(h^2)],$$

合并引理6.1, 我们得到

$$(X_{it}^{\tau} 0)(D_{\omega}^{\tau} W_{\omega} D_{\omega})^{-1} D_{\omega}^{\tau} W_{\omega} Z = \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T X_{it} \Sigma_X^{-1} \Sigma_{XZ} [1 + O_P(h^2)]$$

因此

$$Z^{\tau}(I - S)^{\tau}(I - S)Z = (Z_{it} - \Sigma_{XZ}^{\tau} \Sigma_X^{-1} X_{it})(Z_{it} - \Sigma_{XZ}^{\tau} \Sigma_X^{-1} X_{it})^{\tau} [1 + O_P(h^2)],$$

得到

$$Z^{\tau}(I - S)^{\tau}(I - S)Z = (\Sigma_Z - \Sigma_{XZ}^{\tau} \Sigma_X^{-1} \Sigma_{XZ}) [1 + O_P(h^2)],$$

**引理 6.3.** 假设条件(A1)-(A6) 拥有. 因此我们能得到

$$(H^{\tau}(I - S)^{\tau}(I - S)H)^{-1} = (H^{\tau}H)^{-1} + O(\zeta_n) = T^{-1}I_n + O(\zeta_n).$$

对于充分大的  $n$ , 其中  $\zeta_n = \frac{\sqrt{\ln n}}{nh} i_n i_n^{\tau}$ .

证明.  $H^{\tau}RH = H^{\tau}(I_{nT} - S)^{\tau}(I_{nT} - S)H = H^{\tau}H - H^{\tau}SH - H^{\tau}S^{\tau}H + H^{\tau}S^{\tau}SH = \Omega_{11} - \Omega_{12} - \Omega_{13} + \Omega_{14}$ . 利用  $s_{it}(u) = N^{-1}(X_{it} \ 0)[N^{-1}D_{\omega}^{\tau}W_{\omega}D_{\omega}]^{-1}D_{\omega}K_h\left(\frac{t-\omega T}{Th}\right) = N^{-1}K_h\left(\frac{t-\omega T}{Th}\right)f^{-1}(\cdot)\{1 + O_P(c_n)\} = O\left(\frac{1}{Nh}\right)$ . 我们能显示  $\Omega_{1l}$  的第  $(i, j)$  个元素是  $[\Omega_{1l}]_{ij} = O\left(\frac{1}{Nh}\right)$ ,  $l = 2, 3, 4$ . 相似于 Su 和 Ullah (2006) 引理A.2 的证明. 我们容易证明  $(H^{\tau}RH)^{-1} = (H^{\tau}H)^{-1} + O(\zeta_n) = T^{-1}I_n + O(\zeta_n)$ .

**引理 6.4.** 假设条件(A1)-(A8) 都拥有.

$$\frac{1}{N} \tilde{Z} \Phi \tilde{Z} \rightarrow \frac{T-1}{T} \tilde{\Sigma}. \quad (6-15)$$



证明. 从引理6.3, 我们得到

$$\tilde{Z}\tilde{Z} = Z^r(I-S)^r(I-S)Z = nT(\Sigma_Z - \Sigma_{XZ}^r\Sigma_X^{-1}\Sigma_{XZ})[1 + O_P(h^2)], \quad (6-16)$$

通过Su 和Ullah (2006) 引理3, 我们得到

$$\tilde{Z}\tilde{H}(\tilde{H}^r\tilde{H})^{-1}\tilde{H}^r\tilde{Z} = Z^r(I-S)^r(I-S)Z = n(\Sigma_Z - \Sigma_{XZ}^r\Sigma_X^{-1}\Sigma_{XZ})[1 + O_P(h^2)], \quad (6-17)$$

通过Slutsky 定理和, 我们得到

$$\frac{1}{N}\tilde{Z}\Phi\tilde{Z} = \tilde{Z}\tilde{Z} + \tilde{Z}\tilde{H}(\tilde{H}^r\tilde{H})^{-1}\tilde{H}^r\tilde{Z} \rightarrow \frac{T-1}{T}\tilde{\Sigma}.$$

**引理 6.5.** 假设条件(A1)-(A8) 都拥有. 令 $e_{it}$  是向量 $nT \times 1$ 第 $it$  元素是1, 其余元素都是0. 因此

- (a)  $Z_{it}e_{it}^r(I-S)^r(I-S)v_{it} = (Z_{it} - \Sigma_X^{-1}\Sigma_{XZ}X_{it})v_{it} + o_P(1).$
- (b)  $Z_{it}e_{it}^rRH(H^rH)^{-1}H^rRv_{it} = \frac{1}{T}(Z_{it} - \Sigma_X^{-1}\Sigma_{XZ}X_{it})v_{it} + o_P(1).$
- (c) 令 $B_{it} = \tilde{Z}_{it}e_{it}^r\Phi(I_{nT} - S)v_{it}$ , 因此  $Var(\frac{1}{\sqrt{N}}\sum_{i=1}^n\sum_{s=1}^T B_{it}) \xrightarrow{d} \Delta$

**引理 6.6.** 在定理6.1条件下, 我们得到

$$(i) \max_{1 \leq i \leq n} \|\Xi_i(\beta)\| = o_p(N^{1/2}), \quad (6-18)$$

$$(ii) \lambda(\beta) = O_p(N^{-1/2}). \quad (6-19)$$

证明. 相似于Owen (1998)给的结论, 从式子(6-9) 和引理6.4 和引理6.5, 我们能容易证明式子(6-19). 我们现在证明式子(6-18).

$$\begin{aligned} \max_{1 \leq i \leq n} \|\Xi_i(\beta)\| &\leq \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{Z}_{it}e_{it}^r\Phi(I_{nT} - S)v_{it} \right\| \\ &\quad + \max_{1 \leq i \leq n} \left\| \sum_{t=1}^T \tilde{Z}_{it}e_{it}^r\Phi[X_{it}\alpha(t/T) - s_{it}^rM] \right\| \\ &= R_1 + R_2. \end{aligned}$$

对于 $R_1$ , 注意 $E[\sum_{t=1}^T \sum_{s=1}^T \tilde{Z}_{it}e_{it}^r\Phi(I_{nT} - S)v_{it}\tilde{Z}_{is}e_{is}^r\Phi(I_{nT} - S)v_{is}] = \Delta_{s,s} + o(1) < \infty$ , 其中 $\Delta_{s,s}$  表示第 $s$ 个对角元素 $\Delta$ .  $B_{it}$  是独立同分布的在截面第 $i$  个体为每一个固定时间. 通过Owen (1990)的引理3, 我们得到 $R_1 = o_p(N^{1/2})$ . 现在看第二

项 $R_2$ , 运用Markov 不等式, 我们得到

$$\begin{aligned}
 P(R_2 > N^{1/2}) &\leq \frac{1}{N} \sum_{i=1}^n E \left\| \sum_{t=1}^T \tilde{Z}_{it} e_{it}^\tau \Phi[X_{it}\alpha(t/T) - s_{it}^\tau M] \right\|^2 \\
 &= \frac{1}{N} \sum_{i=1}^n E \left\| \sum_{t=1}^T \sum_{s=1}^T \tilde{Z}_{it} \tilde{Z}_{is} e_{it}^\tau \Phi[X_{it}\alpha(t/T) - s_{it}^\tau M] e_{is}^\tau \Phi[X_{is}\alpha(t/T) - s_{is}^\tau M] \right\| \\
 &\leq \frac{1}{N} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \left\{ E[\tilde{Z}_{it} e_{it}^\tau \Phi[X_{it}\alpha(t/T) - s_{it}^\tau M]^2] \right\}^{1/2} \left\{ E[\tilde{Z}_{is} e_{is}^\tau \Phi[X_{is}\alpha(t/T) - s_{is}^\tau M]^2] \right\}^{1/2} \\
 &= o_p(1) \rightarrow 0.
 \end{aligned}$$

因此,  $R_2 = o_p(N^{1/2})$  and  $\max_{1 \leq i \leq n} \|\Xi_i(\beta)\| = o_p(N^{1/2})$ .

证明定理6.1. 应用泰勒展开式, 从(6-8), 我们能得到下式

$$\log \mathcal{L}_n(\beta) = 2 \sum_{i=1}^n \left\{ \lambda^\tau \Xi_i(\beta) - \frac{[\lambda^\tau \Xi_i(\beta)]^2}{2} \right\} + o_p(1). \quad (6-20)$$

应用泰勒展开式, 从(6-6), 我们能得到下式

$$\begin{aligned}
 0 &= \frac{1}{n} \sum_{i=1}^n \frac{\Xi_i(\beta)}{1 + \lambda^\tau \Xi_i(\beta)} \\
 &= \frac{1}{n} \sum_{i=1}^n \Xi_i(\beta_0) - \frac{1}{n} \sum_{i=1}^n \Xi_i(\beta) \Xi_i^\tau(\beta) \lambda + \frac{1}{n} \sum_{i=1}^n \frac{\Xi_i(\beta) [\lambda^\tau \Xi_i(\beta)]^2}{1 + \lambda^\tau \Xi_i(\beta)}. \quad (6-21)
 \end{aligned}$$

鉴于引理6.4-6.6, 我们得到

$$\begin{aligned}
 \left\| \frac{1}{n} \sum_{i=1}^n \frac{\Xi_i(\beta) [\lambda^\tau \Xi_i(\beta_0)]^2}{1 + \lambda^\tau \Xi_i(\beta)} \right\| &\leq \frac{1}{n} \sum_{i=1}^n \frac{\|\Xi_i(\beta)\|^3 \|\lambda\|^2}{|1 + \lambda^\tau \Xi_i(\beta)|} \\
 &\leq \|\lambda\|^2 \max_{1 \leq i \leq n} \|\Xi_i(\beta)\| \frac{1}{n} \sum_{i=1}^n \|\Xi_i(\beta)\|^2 = O_p(N^{-1}) o_p(N^{1/2}) O_p(1) = o_p(N^{-1/2}),
 \end{aligned}$$

那合并(6-20), 得到  $\sum_{i=1}^n [\lambda^\tau \Xi_i(\beta)]^2 = \sum_{i=1}^n \lambda^\tau \Xi_i(\beta) + o_p(1)$  和

$$\lambda = \left[ \sum_{i=1}^n \Xi_i(\beta) \Xi_i^\tau(\beta) \right]^{-1} \sum_{i=1}^n \Xi_i(\beta_0) + o_p(N^{-1/2}).$$

因此, 通过式子(6-21), 我们得到

$$\log \mathcal{L}_n(\beta) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \Xi_i(\beta) \right)^\tau \left( \frac{1}{n} \sum_{i=1}^n \Xi_i(\beta) \Xi_i^\tau(\beta) \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \Xi_i(\beta) \right) + o_p(1),$$

那合并引理6.4和6.5, 证明了定理6.1。

## 第7章 结论与展望

回顾整篇论文, 反思所做之研究, 深刻地体会到科学研究的艰难、崇高和神圣及由此产生的人生意义和幸福, 也坚定了自己从事科学研究的决心和信心。在信息高度畅通的今天, 理论和技术都在不断发展。为此, 只有使自己的知识不断更新, 才能跟上时代前进的步伐。

半参数面板数据模型兼顾了参数模型与非参数模型的优点, 一方面, 函数形式更加灵活多样; 另一方面, 克服了“维数灾祸”问题。因此, 关于半参数面板数据模型的估计问题是现今最前沿、最热门的研究领域之一。含个体效应面板数据模型和动态面板数据模型的参数估计问题是理论计量经济学中的一个研究热点, 尤其含固定效应面板数据模型的参数估计问题不仅是研究热点还是一个难点。此外, 考虑模型的变量选择是现代统计中的热门研究课题之一, 而变量选择对高维数据的处理起着至关重要的作用。有效的变量选择通过剔除多余的变量能够给出最优的预测变量。从而得到最简洁的模型。为此, 通过构建半参数惩罚函数, 研究了半参数部分线性面板数据模型中参数部分的变量选择问题。下面给出本文的总结: 这篇文章主要是将经验似然方法应用到带固定效应面板数据半参数回归模型。其主要思想是应用经验似然方法提出一个似然比统计量, 在给定的假设条件下, 证明了这个统计量的渐近性质, 由此可以构造回归模型中未知参数的置信域。具体如下:

第一, 我们研究了带固定效应部分线性面板数据模型的参数估计的经验似然, 并构造了未知参数的置信域。证明了调整的经验对数似然比函数在真参数点时是渐近卡方分布的。同时, 模拟研究表明经验似然方法优于正态逼近方法。

第二, 考虑带固定效应面板数据部分线性误差变量模型的统计推断, 对于协变量为附加误差情况下, 我们提出了一个修改的剖面最小二乘估计参数和非参数部分, 参数部分的渐近正态性和非参数部分的收敛速度被建立。一致估计的误差方差也被估计。此外, 我们介绍的剖面似然比检验并证明它在原假设条件下是一个渐进的卡方分布。我们进行模拟研究, 证明了我们所提出的方法更优于不考虑误差变量情况下的估计。

第三, 对于带固定效应面板数据的高维部分线性误差变量模型, 当误差

## 参考文献

- [1] 柴根象, 洪圣岩. 半参数回归模型[M]. 安徽教育出版社, 1995.
- [2] 王启华. 经验似然统计推断方法综述[J]. 数学进展, 2004,32(2):141-151.
- [3] 王启华, 史宁中, 耿直. 现代统计研究基础[M]. 北京: 科学出版社, 2010.
- [4] 李高荣, 冯三营, 薛留根. 纵向数据单指标模型中参数的经验似然置信域[J]. 应用概率统计, 2010,26 (4) : 190-206.
- [5] Ahn, S.C., & Schmidt, P. Estimation of linear panel data models using GMM, in: L. Mátyás (Ed.), *Generalized Method of Moments Estimation*[M]. Cambridge: Cambridge University Press, 2000.
- [6] Amemiya Y., Fuller W. A. Estimation for the multivariate errors-in-variables model with estimated error covariance matrix[J]. *Annals of Statistics*, 1984, 12:497-509.
- [7] Antoniadis, A. Wavelets in statistics: A review[J]. *Journal Italian Statistics Association*, 1997, 6:97-144.
- [8] Antoniadis, A., Fan, J. Regularization of wavelets approximations[J]. *Journal of the American Statistical Association*, 2001, 96:939-967.
- [9] Arellano, M. *Panel Data Econometrics*[M]. New York: Oxford University Press, 2003.
- [10] Auestad, B., & Tjøstheim, D. Identification of nonlinear time series: First order characterization and order determination[J]. *Biometrika*, 1990, 77: 669-687.
- [11] Bai, J. Common breaks in means and variances for panel data[J]. *Journal of Econometrics*, 2010, 157, 78-92.
- [12] Bai Y., Zhu Z., Fung W. K. Weighted empirical likelihood for generalized linear models with longitudinal data[J]. *Journal of Statistics Planning and Inference*, 2010, 140:3446-3456.
- [13] Baltagi B. H. *Econometric Analysis of Panel Data*[M]. 2nd ed. New York: Wiley, 2005.
- [14] Baran S. A consistent estimator for linear models with dependent observations[J]. *Communications in Statistics — Theory and Methods*, 2004,33(10):2469-2486.
- [15] Cai Z.W., Fan J. Q., Li R.Z. Efficient estimation and inference for varying coefficient model[J]. *Journal of the American Statistical Association*, 2000a, 95(451):888 - 902.

方差为已知和未知情形下。我们建议一个修正的回归参数和最大经验似然比率回归参数。同时,依据惩罚经验似然方法,这个模型的参数估计和变量选择被调查,这个建议的惩罚经验似然被证明具有神特征,同时,我们建议的惩罚经验似然比率统计在原假设下具有渐近卡方分布,其结果可以用来构造未知参数的置信域。模拟和实证结果用来评估经验似然方法的性能。

第四,研究带固定效应面板数据的半变系数模型的经验似然,考虑协变量是 $\alpha$ -混合下情形下带固定效应的面板数据的半变系数模型,由于 $\alpha$ -混合的广泛实用性及应用背景,且在诸多条件下是最弱的。我们提出了在考虑 $\alpha$ -混合下带固定效应面板数据的半变系数模型模型下的经验似然,得到未知参数的经验对数似然比统计量,获得了Wilks定理的非参数形式。证明了建议的对数似然比统计量是渐近卡方分布的,同时也获得未知参数的最大似然估计量和证明在合适条件是渐近正态的。模拟结果用来评估经验似然方法的性能。

最后,我们研究带固定效应面板数据部分线性时变系数模型的经验似然,同样考虑在协变量是 $\alpha$ -混合下情形下,未知参数的经验对数似然比统计量,证明了建议的对数似然比统计量是渐近卡方分布的,我们也获得参数的最大似然估计量和证明在合适条件是渐近正态的。同样证明了调整的经验对数似然比函数是渐近卡方分布的。由此可以构造了未知参数的置信域。最后,通过Monte Carlo仿真实验验证建议经验似然方法比正态逼近方法更好。

本文在理论和方法上较为详细地剖析了部分带固定效应面板数据半参数模型的推断以及高维面板数据的变量选择问题,取得了一些成果,但关于半参数面板数据模型的研究尚不完善,还有一些问题有待进一步研究,主要包括以下几个方面:虽然我们成功地将经验似然方法应用到带固定效应面板数据半参数几种回归模型,但仍有许多问题等待我们进一步探讨和研究。比如带固定效应面板数据半参数含误差变量的变系数模型,带固定效应面板数据单指标部分线性模型。同样考虑协变量含误差变量时的情况,带固定效应面板数据单指标部分线性模型。更深入考虑变量是高维时,能不能用惩罚经验似然进行变量选择后,构造了未知参数的置信域。另外,还有许多其他的带固定效应面板数据半参数回归模型下的统计推断等待我们进一步研究,为进一步完善和发展现有的成果,我们还有大量的研究工作要做。

- 
- [16] Cai Z.W., Fan J. Q., Yao Q.W. Functional-coefficient regression models for nonlinear times series[J]. *Journal of the American Statistical Association*, 2000b, 95(451):941 - 956.
- [17] Cai Z., Li Q. Nonparametric estimation and varying coefficient dynamic panel data models[J]. *Econometric Theory*, 2008,24: 1321-1342.
- [18] Carroll R.J., Ruppert, D., Stefanski, L.A. *Measurement error in nonlinear models*[M]. London: Chapman & Hall, 1995.
- [19] Chen, J., Gao, J. T., & Li, D. G. Estimation in single-index panel data models with heterogeneous link functions[J]. *Econometric Reviews*, 2013,32(8): 928-955.
- [20] Chen M. H., Ren Z., and Hu S. H. Strong consistency of a class of estimators in partial linear model[J]. *Acta Mathematica Sinica*, 1998,41:429-439.
- [21] Chen H. Convergence rates for parametric components in a partial linear model[J]. *Annals of Statistics*, 1988,16:136-146.
- [22] Chen S. X. Empirical likelihood confidence intervals for linear regression model coefficients. *Journal of Multivariate Analysis*, 1994,49:24-40.
- [23] Chiou, J. M., & Müller, H. G. Nonparametric quasi-likelihood[J]. *Annals of Statistics*, 1999,27: 36-64.
- [24] Cui H. and Li R. On parameter estimation for semi-linear errors-in-variables models[J]. *Journal of Multivariate Analysis*, 1998,64:1-24.
- [25] Cui H. J., Chen S. X. Empirical likelihood confidence region for parameter in the errors-in-variables models[J]. *Journal of Multivariate Analysis*, 2003,84:101-115.
- [26] Cui H. and Kong E. Empirical likelihood confidence region for parameters in semilinear errors-in-variables models[J]. *Scandinavian Journal of Statistics*, 2006,33:153-168.
- [27] Chen X. and Cui H. J. Empirical likelihood inference for partial linear models under martingale difference[J]. *Statistics and Probability Letters*, 2008,78:2895-2901.
- [28] Engle, R. F., Granger, W. J., Rice, J. & Weiss, A. Semiparametric estimates of the relation between weather and electricity sales[J]. *Journal of the American Statistical Association*, 1986,80: 310-319.
- [29] Fazekas I. and Kukush A. G. Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms[J]. *Computers and Mathematics with Applications*, 1997,34(10):23-39.

- [30] Fan G. L. and Liang H. Y. Empirical likelihood inference for semiparametric model with linear process errors[J]. *Journal of the Korean Statistical Society*, 2010,39:55–65.
- [31] Fan, G. L., Liang, H. Y., & Huang, Z. S. Empirical likelihood for partially time-varying coefficient models with dependent observations[J]. *Journal of Nonparametric Statistics*, 2012, 24: 71-84.
- [32] Fan, G. L., Liang, H. Y., & Wang, J. F. Statistical inference for partially time-varying coefficient errors-in-variables models[J]. *Journal of Statistics Planning and Inference*, 2013,143: 505-519.
- [33] Fan, J., Peng, H., & Huang, T. Semilinear high-dimensional model for normalization of microarrays data: a theoretical analysis and partial consistency[J]. *Journal of the American Statistical Association*, 2005, 100: 781-813.
- [34] Fan, J., & Huang, T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models[J]. *Bernoulli*, 2005,11(6): 1031-1057.
- [35] Fan, J., & Zhang, W. Statistical estimation in varying coefficient models[J]. *Annals of Statistics*, 1999,27: 1491-1518.
- [36] Fan, J., & Yao, Q. W. *Nonlinear times series: nonparametric and parametric methods*[M]. New York: Springer, 2003. .
- [37] Fan J. Q., Truong Y. Nonparametric regression with errors in variables[J]. *The Annals of Statistics*, 1993,21(4):1900-1925.
- [38] Fan J. Q., Li R Z. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. *Journal of the American Statistical Association*, 2001,96: 1348-1360.
- [39] Fan J. and Yao Q. Efficient estimation of conditional variance functions in stochastic regression[J]. *Biometrika*, 1998,85:645–660.
- [40] Fan J. and Gijbels I. Variable bandwidth and local linear regression smoothers[J]. *The Annals of Statistics*, 1992,20:2008–2036.
- [41] Fan J. and Gijbels I. *Local Polynomial Modeling and Its Applications*[M]. London: Chapman and Hall, 1996.
- [42] Fan J. Q., Truong Y. Nonparametric regression with errors in variables[J]. *The Annals of Statistics*, 1993, 21(4):1900-1925.
- [43] Fan J. and Zhang J. T. Two-step estimation of functional linear models with applications to longitudinal data[J]. *Journal of the Royal Statistical Society. Series B*, 2000,62:303–322.

- [44] Frank I E. and Friedman J.H. A statistical view of some chemometrics regression tools[J]. *Technometrics*,1993, 35:109–148.
- [45] Fuller, W. A. *Measurement error models*[M]. New York: Wiley, 1987.
- [46] Gao J. T. and Anh V. V. Semiparametric regression under long-range dependent errors[J]. *Journal of Statistics Planning and Inference*, 1999,80:37–57.
- [47] Gao J. T., Chen X. R., and Zhao L. C. Asymptotic normality of a class of estimators in partial linear models[J]. *Acta Mathematica Sinica*, 1994,37:256–268.
- [48] Glesser, L. J. Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models[J]. *Contemporary Mathematics*, 1990, 112:99-114.
- [49] Hamilton S. A. and Truong Y. K. Local linear estimation in partly linear models[J]. *Journal of Multivariate Analysis*, 1997,60:1–19.
- [50] Hastie T J,Tibshirani R J. Varying-coefficient models[J].*Journal of the Royal Statistical Society Series B*,1993,55:757-796.
- [51] Hsiao C. *Analysis of Panel Data*,2nd ed [M]. Cambridge, Cambridge University Press, 2003.
- [52] Härdle W., Hall, P. Ichimura, H. Optimal smoothing in single-index models[J]. *Annals of Statistics*, 1993,21:157–178.
- [53] Härdle W., Liang H., and Gao J. *Partially Linear Models*[M]. Heidelberg: Physica-Verlag, 2000.
- [54] Hall P. and Heyde C. C. *Martingale Limit Theory and Its Applications*[M]. New York: Academic Press, 1980.
- [55] Heckman N. Spline smoothing in a partially linear model[J]. *Journal of the Royal Statistical Society Series B*,1986,48:244-248.
- [56] He, B.Q., Hong, X.J.,& Fan, G.L. Empirical likelihood for semi-varying coefficient models for panel data with fixed effects[J].*Journal of the Korean Statistical Society*. 2016,45(3):395-408.
- [57] He, B.Q., Hong, X.J.,& Fan, G.L. Block empirical likelihood for partially linear panel data models with fixed effects[J]. *Statistics and Probability Letters*. 2017,123(4):128-138.
- [58] Henderson D.j., Carroll R.J., Li Q. Nonparametric estimation and testing of fixed effects panel data models[J]. *Journal of Econometrics*, 2008,144:257-275.
- [59] Henderson, D.j.,Ullah, A. A nonparametric random effects estimator[J]. *Economics Letters*, 2005,88: 403-407.



- [60] Horowitz, J. L., & Lee, S. Semiparametric estimation of a panel data proportional hazards model with fixed effects[J]. *Journal of Econometrics*, 2004, 119:155-198.
- [61] Hjellvik, V., Chen, R., & Tjøstheim, D. Nonparametric estimation and testing in panels of intercorrelated time series[J]. *Journal of Time Series Analysis*, 2004, 25:831-872.
- [62] Huang Z. S. Empirical likelihood for varying-coefficient single-index model with right-censored data[J]. *Metrika*, 2012, 75:55-71.
- [63] Huang Z. S., Zhang R. Q. Efficient empirical-likelihood-based inferences for the single-index model[J]. *Journal of Multivariate Analysis*, 2011, 102(5):937-947.
- [64] Hu, X. M. Estimation in a semi-varying coefficient model for panel data with fixed effects[J]. *Journal of the System Science and Complexity*, 2014, 27:594-604.
- [65] Hu, Y., & Li, T. T. Empirical likelihood for semiparametric varying coefficient partially linear models with longitudinal data[J]. *Statistics and Probability Letters*, 2011, 80: 111-121.
- [66] Ichimura, H. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models[J]. *Journal of Econometrics*, 1993, 58: 71-120.
- [67] Kaslow, R. A., Ostrow, D.G., Detels, J.P., Polk, B.F., Rinaldo, C.R. The multicenter AIDS cohort study: rationale, organization and selected characteristics of the participants[J]. *American Journal of Epidemiology*, 1987, 126:310-318.
- [68] Kim, D. Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence[J]. *Journal of Econometrics*, 2011, 164: 310-330.
- [69] Leng, C., Tang, C. Y. Penalized empirical likelihood and growing dimensional general estimating equations[J]. *Biometrika*, 2012, 99(3):703-716.
- [70] Liang H. Y. and Jing B. Y. Asymptotic normality in partial linear models based on dependent errors[J]. *Journal of Statistics Planning and Inference*, 2009, 139:1357-1371.
- [71] Liang H., Harder W., and Carroll R. J. Estimation in a semiparametric partially linear errors-in-variables model[J]. *Annals of Statistics*, 1999, 27:1519-1535.
- [72] Li D.G., Chen J., Gao J. T. Non-parametric time-varying coefficient panel data models with fixed effects[J]. *Econometrics Journal*. 2011, (14):387-408.
- [73] Li D.G., Chen J., Gao J. T. Estimation in Partially Linear Single-Index Panel Data Models With Fixed Effects[J]. *Journal of Business and Economic Statistics*, 2013, 31(3):315-330.

- [74] Li G. R. and Xue L. G. Empirical likelihood confidence region for the parameter in a partially linear errors-in-variables model[J]. *Communication in Statistics- Theory and Methods*, 2008,37:1552–1564.
- [75] Li G. R., Tian P., and Xue L. G. Generalized empirical likelihood inference in semiparametric regression model for longitudinal data[J]. *Acta Mathematica Sinica (Engl. Ser.)*, 2008,24(12):2029–2040.
- [76] Li, Q., & Ullah, A. Estimating partially linear panel data models with one-way error components[J]. *Econometrica*, 1998,17(2): 145-166.
- [77] Liebscher E. Strong convergence of sums of  $\alpha$ -mixing random variables with applications to density estimation[J]. *Stochastic Processes Appl*, 1996,65(1):69–80.
- [78] Liu J. X. and Chen X. R. Consistency of ls estimator in simple linear cv regression models[J]. *Acta Mathematica Scientia. Ser. B Engl. Ed*, 2005,25:50–58.
- [79] Masry, E., & Tjøstheim, D. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality[J]. *Econometric Theory*, 1995,11: 258-289.
- [80] Neyman, J., & Scott, E. Consistent estimates based on partially consistent observations[J]. *Econometrica*, 1948,16: 1-32.
- [81] Opsomer, J. D., & Ruppert, D. Fitting a bivariate additive model by local polynomial regression[J]. *Annals of Statistics*, 1997,25: 186-211.
- [82] Owen A. B. Empirical likelihood ratio confidence intervals for a single functional[J]. *Biometrika*, 1988,75:237–249.
- [83] Owen A. B. *Empirical Likelihood*[M]. New York: Chapman and Hall, 2001.
- [84] Owen A. B. Empirical likelihood ratio confidence regions[J]. *Annals of Statistics*, 1990,18:90–120.
- [85] Powell, J. L., Stock, J. H., Stoker, T. M. Semiparametric Estimation of Index Coefficients[J]. *Econometrica*, 1989, 57:1403-1430.
- [86] Qin G. Y., Bai Y., Zhu Z. Y. Robust empirical likelihood inference for generalized partial linear models with longitudinal data[J]. *Journal of Multivariate. Analysis*. 2012, 105(1):32-44.
- [87] Qin, J., Lawless, B. Empirical likelihood and general estimating equations[J]. *Annals of Statistics*, 1994,22: 300-325.

- [88] Qin G. Y., Zhu Z. Y., Fung W K. Robust estimation of covariance parameters in partial linear model for longitudinal data[J]. *Journal of Statistics Planning and Inference*, 2009, 139:558-570.
- [89] Robinson, P.M. Nonparametric Trending Regression with Cross-sectional Dependence[J]. *Journal of Econometrics*, 2012, 169:4-14.
- [90] Roy, N. Nonparametric and semiparametric analysis of panel data models: An application to Calorie-Income relation for rural south India[D]. Riverside; University of California. 1997.
- [91] Rodriguez-Poo, J. M., Soberon, A. Nonparametric estimation of fixed effects panel data varying coefficient models[J]. *The Econometrics Journal*, 2014, 133(1):107-138.
- [92] Severini T, Staniswlis J. Quasi-likelihood estimation in semiparametric models[J]. *Journal of the Royal Statistical Sociation Series*, 1994, 89:501 - 511.
- [93] Shi, J., & Lau, T. S. Empirical likelihood for partially linear models[J]. *Journal of Multivariate Analysis*, 2000,72: 132-148.
- [94] Speckman P. Kernel smoothing in partial linear models[J]. *Journal of the Royal Statistical Sociation Series B*, 1988,50:413-436.
- [95] Su, L. and Chen, Q. Testing Homogeneity in Panel Data Models with Interactive Fixed Effects[J]. *Econometric Theory*, 2013, 29: 1079-1135.
- [96] Su L. J. Ullah A. Profile likelihood estimation of partially linear panel data models with models with fixed effects[J]. *Econometrics Letters*, 2006,92:75-81.
- [97] Sun X. Q., You J. H., and Chen G. M. et al. Convergence rates of estimators in partial linear regression models with  $ma(\infty)$  error process[J]. *Communications in Statistics - Theory and Methods*, 2002,31:2251-2273.
- [98] Tang, C.Y., Leng, C. Penalized high-dimensional empirical likelihood[J]. *Biometrika*, 2010, 97:905-920.
- [99] Tibshiani, R. Regression shrinkage and selection via the Lasso[J]. *Journal of the Royal Statistical Society. Series B*, 1996,58(1):267-288.
- [100] Tsay, R.S. *Analysis of Financial Time Series*[M]. New York: Wiley, 2002.
- [101] Tosteson, T., Stefanski, L.A., Schafer, D.W. A measurement error model for binary and ordinal regression[J]. *Statistics in Medicine*, 1989,8:1139-1147.
- [102] Volkonskii, V. A., & Rozanov, Y. A. Some limit theorems for random functions[J]. *Theory of Probability and its Applications*, 1959,4: 178-197.

- [103] Wang, N., Carroll, R.J., Liang, K.Y. Quasi-likelihood and variance functions in measurement error models with replicates[J]. *Biometrics*, 1996, 52: 423-432.
- [104] Wang Q. H. and Li G. Empirical likelihood semiparametric regression analysis under random censorship[J]. *Journal of Multivariate Analysis*, 2002, 83: 469-486.
- [105] Wang Q. H. and Jing B. Y. Empirical likelihood for partial linear models with fixed designs[J]. *Statistics and Probability Letters*, 1999, 41: 425-433.
- [106] Wang Q. H. and Jing B. Y. Empirical likelihood for partial linear models[J]. *Annals of the Institute of Statistical Mathematics*, 2003, 55: 585-595.
- [107] Wilks S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses[J]. *Annals of the Institute of Statistical Mathematics*, 1938, 9: 60-62.
- [108] Wong H., Liu F., and Chen M. et al. Empirical likelihood based diagnostics for heteroscedasticity in partially linear errors-in-variables models[J]. *Journal of Statistics Planning and Inference*, 2009, 139: 916-929.
- [109] Wooldridge J. M. *Econometric analysis of cross section and panel data*[M]. Cambridge: MIT Press, 2002.
- [110] Xue L. G. and Zhu L. X. Empirical likelihood for single-index models[J]. *Journal of Multivariate Analysis*, 2006, 97: 1295-1312.
- [111] Xue L. G. and Zhu L. X. Empirical likelihood-based inference in a partially linear model for longitudinal data[J]. *Science in China Series A: Mathematics*, 2008, 51(1): 115-130.
- [112] Xue L. G. and Zhu L. X. Empirical likelihood semiparametric regression analysis for longitudinal data[J]. *Biometrika*, 2007, 94(4): 921-937.
- [113] Yang, S. C. Moment inequalities for the partial sums of random variables[J]. *Science in China (Series A)*, 2001, 44(1): 1-6.
- [114] Yang, S. C. Maximal moment inequality for partial sums of strong mixing sequences and application[J]. *Acta Mathematica Sinica, English Series*, 2007, 23: 1013-1024.
- [115] You, J. H., & Zhou, Y. Statistical inference in a panel data semiparametric regression model with serially correlated errors[J]. *Journal of Multivariate Analysis*, 2006, 97: 844-873.
- [116] You, J. H., Chen, G. M., & Zhou, Y. Block empirical likelihood for longitudinal partially linear regression models[J]. *The Canadian Journal of Statistics*, 2006, 34(1): 79-96.
- [117] You, J. H., & Zhou, Y. Empirical likelihood for semiparametric varying-coefficient partially linear regression models[J]. *Statistics and Probability Letters*, 2006, 76: 412-422.

- 
- [118] You, J. H., Zhou, X., & Zhou, Y. Series estimation in partially linear in-slide regression models[J]. *Scandinavian Journal of Statistics*, 2011,38:89-107.
- [119] You J. H., Chen G. M., and Zhou Y. Statistical inference of partially linear regression models with heteroscedastic errors[J]. *Journal of Multivariate Analysis*, 2007,98:1539–1557.
- [120] You J. H. and Chen G. M. Semiparametric generalized least squares estimation in partially linear regression models with correlated errors[J]. *Journal of Statistics Planning and Inference*, 2007,137:117–132.
- [121] Zeger S. L. and Diggle P. J. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters[J]. *Biometrics*, 1994,50:689–699.
- [122] Zhou, H. B., You, J. H., Zhou, B. Statistical inference for fixed-effects partially linear regression models with errors in variables[J]. *Statistical Papers*, 2010,51:629-650.
- [123] Zhou, X. C., & Lin, J.G. Empirical likelihood for varying-coefficient semiparametric mixed-effects errors-in-variables models with longitudinal data[J]. *Statistical Methods & Applications*, 2014,23:51-69.
- [124] Zhu L. X. and Xue L. G. Empirical likelihood confidence regions in a partially linear single-index model[J]. *Journal of the Royal Statistical Society. Series B*, 2006,68:549–570.
- [125] Zhu L. X. and Cui H. A semi-parametric regression model with errors in variables[J]. *Scandinavian Journal of Statistics*, 2003,30:429–442.
- [126] Zhu, L. P., You, J. H., & Xu, Q. F. Statistical inference for single-index panel data models[J]. *Scandinavian Journal of Statistics*, 2014,41:830-843.
- [127] Zhang, R. Q., & Huang, Z. S. Statistical inference on parametric part for partially linear single-index model[J]. *Science in China series A: mathematics*, 2009,52(10):2227-2242.
- [128] Zhang R. Q., Huang Z. S., Lv Y. Z. Statistical inference for the index parameter in single-index models[J]. *Journal of Multivariate Analysis*, 2010, 101(4): 1026-1041.

## 攻读博士学位期间发表和接受的学术论文与研究成果

### 已发表和接受的论文:

- [1] He Bang-Qiang, Hong Xing-Jian , Fan Guo-Liang. Empirical likelihood for semi-varying coefficient models for panel data with fixed effects [J]. Journal of the Korean Statistical Society, 2016, 45(3):395-408. (SCI)
- [2] He Bang-Qiang, Hong Xing-Jian , Fan Guo-Liang. Block empirical likelihood for partially linear panel data models with fixed effects[J]. Statistics and Probability Letters, 2017,123(4):128-138. (SCI)
- [3] He Bang-Qiang, Ni Wenqing,Zhou Jin-Ming. Monotone empirical bayes test for the parameter of pareto distribution under random censorship[J]. Journal of Interdisciplinary Mathematics, 2016, 19 (4) : 709-720. (EI)
- [4] He Bang-Qiang, Hong Xing-Jian , Fan Guo-Liang. Statistical inference for partially linear errors-in-variables panel data models with fixed effects[J]. Econometric Reviews. (Accepted) (SSCI)
- [5] 何帮强, 洪兴建. 基尼系数计算与分解方法综述[J]. 统计与决策, 2016, 458(14): 13-17. (CSSCI)
- [6] 何帮强, 胡小文. 利率市场化下严格通胀制与灵活通胀制的政策效应分析[J]. 浙江工商大学学报, 2017, 142 (1) : 66-73. (CSSCI)

### 研究成果:

- [1] 主持: 2016年安徽省高等教育提升计划科学研究一般项目 (TSKJ2016B09): 半参数面板数据模型的经验似然估计及其应用。
- [2] 参与:2015年浙江省自然科学基金项目(Y15G030009): 居民收入增长与经济发展同步的保障机制及对策研究。

## 致谢

逾尺的札记和研究纪录凝聚成这么薄薄的一本，高兴和欣慰之余，不禁感慨系之。回头望去，从坚定信念要读博士，到复习考博之路，到艰辛的读博生涯，往事仿佛就在昨天。一位位支持的家人、师长、朋友、同学、同事从脑海里不断的涌现，不觉泪水盈眶，心中全是感激的信号，顿时觉得自己无比富有，无比的幸福！这几年的学习和生活经历已经成为我一生中重要的一页和不可多得的宝贵财富，它也会令我对即将迈入的新生活充满信心；回顾在浙江工商日日夜夜，有太多的东西值得我去珍惜、回味和总结，有太多的人值得我去留恋和表达感激之情。

首先我要感谢导师洪兴建教授，洪老师渊博的专业知识，严谨的治学态度，精益求精的工作作风，诲人不倦的高尚师德，严以律己和宽以待人的崇高风范，朴实无华和平易近人的人格魅力对我影响深远。不仅使我掌握了基本的研究方法，还使我明白了许多待人接物与为人处事的道理。这些都为我树立了一辈子的学习典范，深刻地影响着我未来的工作和生活。在攻读学位期间，从专业课程的学习、文献的阅读、论文的选题、框架结构的确定、论文的修改和定稿，无不倾注着洪老师的心血和汗水，衷心感谢洪老师数年来的精心指导、关心和帮助！除了感激之外，我还有丝许的惭愧，因为我与导师所要求的还相差甚远。希望在以后的人生道路上继续得到他的指导和帮助。

我也深深感谢浙江工商大学统计学院各位领导和老师们给予我的帮助和指导，特别感谢李金昌教授、苏为华教授、陈振龙教授、蔡光辉教授、陈钰芬教授、程开明教授、徐霏婷教授等，教授们精彩的授课，使我的专业理论和学术水平得到极大拓展与提高，为论文的完成打下了坚实的基础。还要感谢我的同学倪文清、陈锦其、陈龙等同学，他们的友谊和关怀无处不在、无微不至，伴我度过这愉快的研究生活。另外还要感谢范国良、周金明，徐红霞等同事，正是他们的无私的帮助，使我能够使用Matlab软件和论文的写作。

另外，我得感谢我的家人。从考博到整个博士学习阶段，父母始终给我精神上的鼓励；爱人夏菊红无怨无悔地承担了全部家务，并给我精神上和事业上的支持，为我的学习和研究工作付出了很多；女儿何硕怡小朋友的快乐成长也是我学习和科研的动力。

## 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含本人为获得浙江工商大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

签名： 何帮强 日期：2017年5月18日

## 关于论文使用授权的说明

本学位论文作者完全了解浙江工商大学有关保留、使用学位论文的规定：浙江工商大学有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅，可以将学位论文的全部或部分内 容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并且本人电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

签名： 何帮强 导师签名： 何帮强  
日期：2017年5月18日