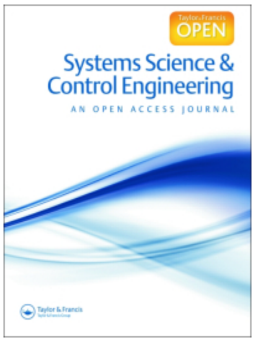


Taylor & Francis
Taylor & Francis Group

Systems Science & Control Engineering
An Open Access Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tssc20

# Statistical inference for partially linear errors-in-variables panel data models with fixed effects

**Bangqiang He , Minxiu Yu & Jinming Zhou**

**To cite this article:** Bangqiang He , Minxiu Yu & Jinming Zhou (2021) Statistical inference for partially linear errors-in-variables panel data models with fixed effects, Systems Science & Control Engineering, 9:1, 1-10, DOI: 10.1080/21642583.2020.1856212

**To link to this article:** https://doi.org/10.1080/21642583.2020.1856212



Published online: 21 Dec 2020.

SYSTEMS SCIENCE & CONTROL ENGINEERING: AN OPEN ACCESS JOURNAL
2021, VOL. 9, NO. 1, 1–10
https://doi.org/10.1080/21642583.2020.1856212

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

# Statistical inference for partially linear errors-in-variables panel data models with fixed effects

Bangqiang He, Minxiu Yu and Jinming Zhou

School of Mathematics and Finance, Anhui Polytechnic University, Wuhu, People's Republic of China

**ABSTRACT**

In this paper, 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 asymptotic 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. We conduct simulation studies to demonstrate the finite sample performance of our proposed method and we also present an illustrative empirical application.

**ARTICLE HISTORY**
Received 1 July 2020
Accepted 23 November 2020

**KEYWORDS**
Panel data; partially linear model; fixed effect; errors-in-variables

## 1. Introduction

Panel data records information on each individual unit over time, the rich information contained in panel data allows researchers to estimate complex models and answer questions that may not be possible using time series or cross-sectional data alone. Panel data analysis has received a lot of attention during the last two decades due to applications in many disciplines, such as economics, finance, biology, engineering and social sciences. Baltagi (2005) and Hsiao (2003) provided excellent overviews of statistical inference and econometric analysis of parametric panel data models. Semiparametric regression models reduce the high risk of misspecification relative to a fully parametric model and avoid some serious drawbacks of purely nonparametric methods such as the curse of dimensionality, difficulty of interpretation, and lack of extrapolation capability. In the last two decades, various forms of semiparametric regression models have been developed. These include the varying-coefficient model (e.g. Zhao & Lin, 2019), the partially linear regression model (e.g. Engle et al., 1986), the varying-coefficient partially linear model (e.g. J. Fan & Huang, 2005). The partially linear panel data models with fixed effects is a useful tool for econometric analysis; see, e.g. Henderson et al. (2008), Hu (2014), and Li et al. (2011).

In this paper, we consider the following partially linear panel data models with fixed effects (e.g. Su & Ullah, 2006):

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

where $Y_{it}$ is the response, $X_{it}$ and $U_{it}$ are of dimensions $p \times 1$ and $q \times 1$, respectively, $\beta = (\beta_1, \dots, \beta_p)^{\tau}$ is a vector of $p$-dimensional unknown parameters, and the superscript $\tau$ denotes the transpose of a vector or matrix. $m(U_{it})$ is a unknown function and $\alpha_i$ is the unobserved individual effects, $\varepsilon_{it}$ is the random model error. Here, we assume $\varepsilon_{it}$ to be i.i.d. with zero mean and finite variance $\sigma^2 > 0$. We allow $\alpha_i$ to be correlated with $X_{it}$, and $U_{it}$ with an unknown correlation structure. Hence, model (1) is a fixed effects model.

Measurement error data are often encountered in many fields, including economics, biomedical sciences and epidemiology. Simply ignoring measurement errors (errors-in-variables), will result in biased estimators. Handing the measurement errors in covariates is generally a challenge for statistical analysis. For the past two decades, regression analysis with measurement errors had much progress. A detailed study can be found in the research of Fuller (1987), J. Fan and Truong (1993), Liang et al. (1999), G. L. Fan et al. (2013) and so on. A statistical analysis of model (1) with additive measurement errors, however, still seems to be missing. The objective of the present paper is to fill this gap.

**CONTACT** Minxiu Yu yuminxiu01@126.com

Specifically, we consider the following partially linear errors-in-variables panel data models with fixed effects

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

where the covariate variable $X_{it}$ is measured with additive error and is not directly observable. Instead, we observe $Z_{it}$ with where the measurement errors $\eta_{it}$ are independent and identically distributed, independent of $(X_{it}, U_{it}, \varepsilon_{it})$, we assume that $\mathrm{Cov}(\eta_{it}) = \Sigma_\eta$ with $\Sigma_\eta$ assumed to be known, as in the papers of Zhu and Cui (2003), You and Chen (2006), and other. When $\Sigma_\eta$ is unknown, we can estimate it by repeatedly measuring $Z_{it}$; see Liang et al. (1999) for details.

The individual effects are often viewed as nuisance parameters in the fixed effects panel data regression model. Because of the diverging number of nuisance parameters, the estimation of the parametric and nonparametric components in the partially linear panel data models with fixed effects is no longer straightforward. Great efforts have been made to develop suitable methods for estimations in the last few years. For the model (1), Su and Ullah (2006) adapted a local linear dummy variable approach to remove the unknown fixed effects. In this paper, our aim is to extend the results in Zhou et al. (2010) for fixed effects partially linear regression models with errors in variables to partially linear error-in-variables panel data models with fixed effects. we make statistical inference for the parametric $\beta$ in partially linear panel data models with fixed effects. Following the estimation procedure proposed by J. Fan et al. (2005), the profile least-squared estimations of $\beta$ can be obtained. Based on this, a normal-based confidence region for the parametric is constructed. However, because of the attenuation the previously proposed estimators, the profile least squares estimator, are no longer consistent. We propose a modified profile least squares estimator. The resulting estimator is shown to be consistent and asymptotically normal. Consistently estimating the error variance is also considered.

The layout of the remainder of this paper is as follows. In Section 2, we present the estimators of the parametric component, the nonparametric component, as well as the error variance. Assumption conditions and the main result are given in Section 3. The simulated example is provided in Section 4. A real-data example is given in Section 5. All the mathematical proofs of the asymptotic results are given in Appendix.

## 2. Model and methodology

To introduce our estimation, we rewrite model (1) in a matrix format yields

$$Y = X\beta + m(U) + B\alpha + \varepsilon, \tag{3}$$

where $B$ is a $nT \times n$ matrix. For our application, the matrix is in a specific form, $B = I_n \otimes i_T$ with $\otimes$ the Kronecker product, $I_n$ denotes the $n \times n$ identity matrix, and $i_T$ denotes the $n \times 1$ vector of ones. There are many approaches to estimate the parameters $\{\beta_j, j = 1, \ldots, p\}$ and the functions $m(\cdot)$. The main idea is from the profile least squares (PLS) approach proposed by J. Fan and Huang (2005): suppose that we have a random sample $\{(U_{it}, X_{it1}, \ldots, X_{itp}, Y_{it}),\ i = 1, \ldots, n,\ t = 1, \ldots, T\}$ from model (3). Let $\theta = (\alpha^\tau, \beta^\tau)^\tau$. Given $\theta$, one can apply a local linear regression technique to estimate the nonparametric component $m(\cdot)$ in (3). For $U_{it}$ in a small neighbourhood of $u$, one can approximated $m(U_{it})$ locally by a linear function as below

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

where $m'(u) = \partial m(u)/\partial u$.

This leads to the following weighted local least squares problem: find $a$ and $b$ to minimize

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{(Y_{it} - X_{it}^{\tau}\beta - \alpha_i) - [a + b(U_{it} - u)]\right\}^2 \times K_h(U_{it} - u), \tag{4}$$

where $a$ and $b$ are q-dimensional column vectors, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and $h$ is a sequence of positive numbers tending to zero, called bandwidth. For notational convenience, let

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

$$Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1T} \\ \vdots \\ Y_{nT} \end{pmatrix},$$

and $W_u = \mathrm{diag}(K_h(U_{11} - u), \ldots, K_h(U_{1T} - u), \ldots, K_h(U_{nT} - u))$. Then the solution of problem (4) is given by

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

In (3), $m(U_{it})$ is replaced with $\hat{a}$. Then we have

$$\hat{m}(U_{it}) = (1,0)(D_{U_{it}}^\tau W_{U_{it}} D_{U_{it}})^{-1} D_{U_{it}}^\tau W_{U_{it}} (Y - X\beta - B\alpha). \quad (5)$$

Now we consider a way of removing the unknown fixed effects motivated by a least squares dummy variable model in parametric panel data analysis, for which we solve the following optimization problem:

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

Supposing that $N = nT, \widetilde{X} = (I_N - S)X, \widetilde{Y} = (I_N - S)Y, \widetilde{B} = (I_N - S)B$,

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

we have

$$\widetilde{Y} = \widetilde{X}\beta + \widetilde{B}\alpha + \varepsilon.$$

By the least squares method and a slight complex computation, we have

$$\widehat{\beta}^* = (\widetilde{X}^\tau \widetilde{X} - \widetilde{X}^\tau \widetilde{M}_{\widetilde{B}} \widetilde{X})^{-1} \widetilde{X}^\tau (I - \widetilde{M}_{\widetilde{B}}) \widetilde{Y},$$
$$\widetilde{\alpha} = (\widetilde{B}^\tau \widetilde{B})^{-1} \widetilde{B}^\tau (\widetilde{Y} - \widetilde{X}\beta), \quad (6)$$

where $\widetilde{M}_{\widetilde{B}} = \widetilde{B}(\widetilde{B}^\tau \widetilde{B})^{-1} \widetilde{B}^\tau$ is a projection matrix of $\widetilde{B}$.

However, $X_{it}$'s can not be observed in our case and we just have $Z_{it}$. If we ignore the measurement error and replace $X_{it}$ with $Z_{it}$ in (6), (6) can be used to show that the resulting estimate is inconsistent. It is well known that in linear regression or partially linear regression, inconsistency caused by the measurement error can be overcome by applying the so-called 'correction for attenuation', see Fuller (1987) and Liang et al. (1999) for more details. In the context of partially linear regression, we introduce the modified least squares method to estimate $\beta$ as follows:

$$\widehat{\beta} = \{\widetilde{Z}^\tau (I - \widetilde{M}_{\widetilde{B}}) \widetilde{Z} - n(T-1)\Sigma_\eta\}^{-1} \widetilde{Z}^\tau (I - \widetilde{M}_{\widetilde{B}}) \widetilde{Y}. \quad (7)$$

Moreover, note that $E(m_j(U_{it})) = E[Y_{it} - X_{it}^\tau \beta - \alpha_i \mid U_{it}] = E[Y_{it} - Z_{it}^\tau \beta - \alpha_i \mid U_{it}]$ and results of J. Fan et al. (2005) suggest that the profile least squares estimator of the nonparametric components $m = [m(U_{11}), \ldots, m(U_{1T}), \ldots, m(U_{nT})]^\tau$ can be achieved with the following equation

$$\begin{bmatrix} I_N & \bar{M}_B \\ S & I_N \end{bmatrix} \begin{bmatrix} B\alpha \\ m \end{bmatrix} = \begin{bmatrix} \bar{M}_B \\ S \end{bmatrix} (Y - Z\widehat{\beta}),$$

where $\bar{M}_B = B(I_N - i_N i_N^\tau / N)(B^\tau B)^{-1} B^\tau$. According to the results of Opsomer and Ruppert (1997), the estimate of $m$ has the following form

$$\widehat{m} = \{I - (I - S\bar{M}_B)^{-1}(I - S)\}(Y - Z\widehat{\beta}). \quad (8)$$

Sometimes, it is also necessary to estimate the error variance $\sigma^2 = E(\varepsilon_{it}^2)$ for such tasks as the construction of confidence regions, model-based tests, model selection procedures, single-to-noise ratio determination and so on. From $E[Y_{it} - X_{it}^\tau \beta - B_i^\tau \alpha_i - m(U_{it})]^2 = \sigma^2$ and $E[Y_{it} - Z_{it}^\tau \beta - B_i^\tau \alpha_i - m(U_{it})]^2 = \sigma^2 + \beta^\tau \Sigma_\eta \beta$, we define an estimator of $\sigma^2$ as

$$\widehat{\sigma} = \frac{T}{N(T-1)} (Y - Z\widehat{\beta} - \widehat{m})^\tau (I - M_B)(Y - Z\widehat{\beta} - \widehat{m}) - \widehat{\beta}^\tau \Sigma_\eta \widehat{\beta}, \quad (9)$$

with $M_B = B(B^\tau B)^{-1} B$.

## 3. Main results

In this section, we will be establishing the asymptotic properties of the proposed estimators from the previous section. Before formulating the main results, we first give the following assumptions:

(A1) The random vector $U_{it}$ has a continuous density function $f(\cdot)$ with a bounded support $\mathcal{U}$. $0 < \inf_{u \in \mathcal{U}} f(\cdot) \leq \sup_{u \in \mathcal{U}} f(\cdot) < \infty$.

(A2) Let $\Phi(u) = E(X_{it} \mid U_{it} = u)$. The functions $\Phi(\cdot)$ and $m(\cdot)$ have bounded second partial derivatives on $\mathcal{U}$.

(A3) $(\alpha_i, X_{it}, U_{it}, \varepsilon_{it})$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ are i.i.d. There exists some $\delta > 2$ such that $E\|X_{it}\|^{2+\delta} < \infty$ and $E|\varepsilon_{it}|^{2+\delta} < \infty$, where $\|a\| = (\sum_{i=1} a_i^2)^{1/2}$ is the usual Euclidean norm of a vector $a$.

(A4) The kernel $K(v)$ is a symmetric probability density function with a continuous derivative on its compact support $\mathcal{U}$.

(A5) $E(Y_{it} \mid X_i, U_i, \alpha_i) = E(Y_{it} \mid X_{it}, U_{it}, \alpha_i) = X_{it}^\tau \beta + m(U_{it}) + \alpha_i$.

(A6) $E|\breve{X}_{it}|^{2+\delta} < \infty$, $\Sigma = E[\breve{X}_{it}\breve{X}_{it}^{\tau}]$ is non-singular, where $\breve{X}_{it} = X_{it} - E(X_{it} \mid U_{it})$.

(A7) The bandwidth $h$ satisfies $h \to 0$, $Nh^8 \to 0$ and $Nh^2/(\log N)^2 \to \infty$ as $N \to \infty$.

**Theorem 3.1:** *Suppose that Assumptions* (A1)–(A7) *hold. For model* (3), *if $\beta$ is the true value of the parameter, then the proposed estimator $\widehat{\beta}$ of $\beta$ is asymptotically normal, namely*

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{T}{N(T-1)}\Sigma^{-1}\Sigma_1\Sigma^{-1}\right), \tag{10}$$

*where* $\Sigma_1 = [E(\varepsilon - \eta^{\tau}\beta)^2\Sigma + \sigma^2\Sigma_\eta + E[(\eta\eta^{\tau} - \Sigma_\eta)\beta]^{\oplus 2}]$ *and* $\mathbf{A}^{\oplus 2}$ *means* $\mathbf{A}\mathbf{A}^{\tau}$.

Further, $\widehat{\Sigma}^{-1}\widehat{\Sigma}_1\widehat{\Sigma}^{-1}$ is a consistent estimator of $\Sigma^{-1}\Sigma_1\Sigma^{-1}$ where $\widehat{\Sigma} = \frac{T}{N(T-1)}\widetilde{Z}^{\tau}(I - \widetilde{M}_{\widetilde{B}})\widetilde{Z} - \Sigma_\eta$ and $\widehat{\Sigma}_1 = \frac{1}{N}[\widetilde{Z}\mathrm{diag}(\widetilde{Y}_{11} - \widetilde{Z}_{11}^{\tau}\widehat{\beta}, \ldots, \widetilde{Y}_{1T} - \widetilde{Z}_{1T}^{\tau}\widehat{\beta}, \ldots, \widetilde{Y}_{nT} - \widetilde{Z}_{nT}^{\tau}\widehat{\beta}) + I_N(\Sigma_\eta\widehat{\beta})^{\tau}]^{\oplus 2}$.

**Theorem 3.2:** *Suppose that Assumptions* (A1)–(A7) *hold. Then the risk of the profile least squares $\widehat{m}$ is bounded as follows*:

$$\begin{aligned} &MSE\{\widehat{m} \mid U_{11}, \ldots, U_{1T} \cdots, U_{nT}\} \\ &\leq \frac{T}{(\sqrt{T}-1)^2}\left\{\frac{\mu_2^2 h^4}{4}E[\widehat{m}''(u)] + \frac{(\sigma^2 + \beta^{\tau}\Sigma_\eta\beta)\nu_0|\mathcal{U}|}{Nh}\right\} \\ &\quad + o_p\left(h^4 + \frac{1}{Nh}\right), \end{aligned}$$

*where* $\mu_i = \int u^i K(u)\,\mathrm{d}u$, $\nu_i = \int u^i K^2(u)\,\mathrm{d}u$, $\widehat{m} = [\widehat{m}(U_{11}), \ldots, \widehat{m}(U_{1T}), \ldots, \widehat{m}(U_{nT})]^{\tau}$, $MSE\{\widehat{m} \mid U_{11}, \ldots, U_{1T}, \ldots, U_{nT}\} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} E[\widehat{m}(U_{it}) - m(U_{it}) \mid U_{11}, \ldots, U_{1T}, \ldots, U_{nT}]^2$.

**Theorem 3.3:** *Suppose that Assumptions* (A1)–(A7) *hold. Then it holds that*

$$\sqrt{N}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, \Theta), \tag{11}$$

*where* $\Theta = E(\varepsilon - \eta^{\tau}\beta)^4 + (3 - T)/(T-1)(\sigma^2 + \beta^{\tau}\Sigma_\eta\beta)^2$. *Define* $\widehat{\Psi} = (\widehat{\psi}_{11}, \ldots, \widehat{\psi}_{1T}, \ldots, \widehat{\psi}_{nT})^{\tau} = (I - M_B)(Y - Z\widehat{\beta} - M)$ *and*

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

*then $\widehat{\Theta}$ is a consistent estimator of $\Theta$.*

## 4. Simulation studies

In this section, we carry out some simulation experiments to study the finite sample performance of the estimators $\widehat{\beta}$, $\widehat{m}$ and $\widehat{\sigma}^2$, which are defined in Section 2. Firstly, we consider the following partially linear errors-in-variables panel data models with fixed effects:

$$\begin{aligned} &Y_{it} = X_{it}^{\tau}\beta + m(U_{it}) + \alpha_i + \varepsilon_{it}, \\ &Z_{it} = X_{it} + \eta_{it}, \\ &\quad i = 1, \ldots, n;\ t = 1, \ldots, T, \end{aligned} \tag{12}$$

where $\beta = (\beta_1, \beta_2)^{\tau} = (1, \sqrt{2})^{\tau}/\sqrt{3}$, $m(U_{it}) = \cos(2\pi U_{it})$, $U_{it} \sim U(0,1)$, $X_{it} = (X_{it1}, X_{it2})^{\tau} \overset{i.i.d}{\sim} N((1,1)^{\tau}, \mathrm{diag}(4,4))$, $\alpha_i = \rho\bar{X}_i + w_i$ with $\rho = 0.5$, 1 and $w_i \sim N(0,1)$ for $i = 1, 2, \ldots, n$, and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it1}$. We use $\rho$ to control the correlation between $\alpha_i$ and $\bar{X}_i$. The measurement error $\eta_{it} \sim N(0, \Sigma_\eta)$ where we take $\Sigma_\eta = 0.2^2 I_2$ and $0.4^2 I_2$ to represent different levels of measurement error.

In our simulations, we took the sample sizes $(n, T) = (50, 4), (50, 6), (100, 4)$ and $(100, 6)$, respectively, and we choose the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \leq 1\}$. The 'leave-one-subject-out' cross-validation bandwidth $CV(h)$ is obtained by minimizing

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

where $\hat{\beta}_{[i]}$ and $\hat{m}_{[i]}(U_{it})$ are estimators of $\beta$ and $m(U_{it})$, respectively, which are computed with all of the measurements but not the $i$th subject. We consider 500 replications with the nominal level $1 - \alpha = 0.95$. We calculate the sample means and standard deviations (SD) of the proposed estimators for the parametric components $\beta_1$, $\beta_2$, and error variance $\sigma^2$ in Section 2. In order to compare the proposed modified PLS estimation method for the parametric component and estimator of the error variance with the existing method, we also calculate the sample means and SD of the modified PLS estimators and naive estimators (neglecting the measurement errors). The naive estimators for $(\beta_1, \beta_2)^{\tau}$ and $\sigma^2$ are defined as

$$\breve{\beta} = \left\{\widetilde{Z}^{\tau}(I - \widetilde{M}_{\widetilde{B}})\widetilde{Z}\right\}^{-1}\widetilde{Z}^{\tau}(I - \widetilde{M}_{\widetilde{B}})\widetilde{Y},$$

and

$$\breve{\sigma} = \frac{T}{N(T-1)}(Y - Z\breve{\beta} - \breve{m})^{\tau}(I - M_B)(Y - Z\breve{\beta} - \breve{m}),$$

where

$$\breve{m} = \left\{I - (I - S\bar{M}_B)^{-1}(I - S)\right\}(Y - Z\breve{\beta}).$$

Simulation results are summarized in Tables 1 and 2. From Tables 1 and 2, we make the following observations:

**Table 1.** Sample means and SD of modified PLS estimators and naive estimators when $\rho = 0.5$.

| $\Sigma_\eta$ | $(n,T)$ | | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}^2$ | $\check{\beta}_1$ | $\check{\beta}_2$ | $\check{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| $0.2^2 I_2$ | (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 |

**Table 2.** Sample means and SD of modified PLS estimators and naive estimators when $\rho = 1$.

| $\Sigma_\eta$ | $(n,T)$ | | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}^2$ | $\check{\beta}_1$ | $\check{\beta}_2$ | $\check{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| $0.2^2 I_2$ | (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 |

(1) There is a clear difference between the SD of the naive estimators and those of the proposed estimators. We think this difference is caused by the measurement errors.
(2) The modified profile least squares estimators for the parametric component and the estimator of the error variance are asymptotically unbiased and have smaller SD than those of the naive estimators.
(3) When $\rho$ increases, the proposed estimator behave better than those of the naive estimators. The SD becomes a little bit bigger.

From Figure 1, we can see that the modified profile least squares estimator of the nonparametric component outperforms the naive profile least squares estimator. The latter is biased.

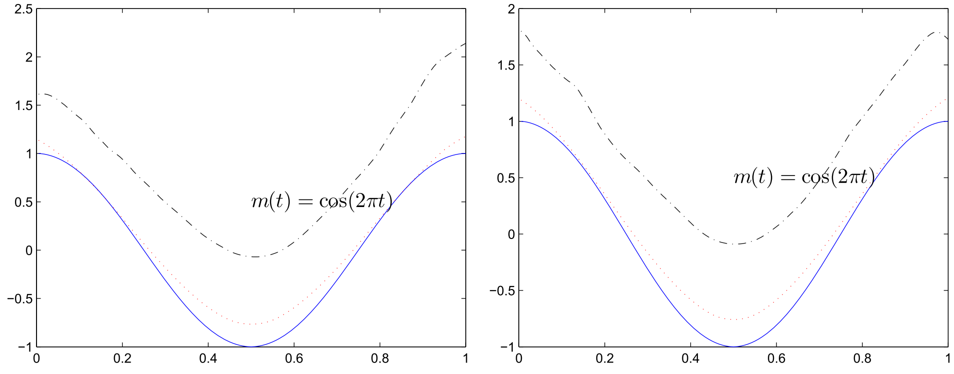


**Figure 1.** The estimators of nonparametric component $m(\cdot)$ with $(n,T) = (50,4)$ and $\Sigma_\eta = 0.2$ (left panel) and $\Sigma_\eta = 0.4$ (right panel). $\cos(2\pi t)$ (solid curve), the proposed estimator (dotted curve) and the naive estimator (dot-dashed curve).

## 5. Real-data analysis

To examine the performance of the proposed method, we analyse a climate data set from the UK met office web site: http://www.metoffice.gov.uk/climate/uk/. The dataset contains mean maximum temperature (in Celsius degrees), mean minimum temperatures, total rainfall (in millimetres), total raindays $\geq 1.0$ mm, total sunshine duration (in hours) and days of air frost (in days) from 37 stations covering UK. The objectives of the study are to describe the common trend in the mean maximum temperature series and the relationship of the mean maximum temperatures with total rainfall and total sunshine during the period of January 2006 to December 2015. Data from 16 stations are selected according to data availability. This dataset has been studied by D. G. Li et al. (2011).

Let $Y_{it}$ be the log-transformed mean maximum temperature in $t$th month in station $i$, $X_{1it}$ be the log-transformed total sunshine duration in $t$th month in station $i$, because of the measuring mechanism, $X_{1it}$ can only observe the surrogate variables $Z_{1it}$. $X_{2it}$ be the log-transformed total rainfall in $t$th month in station $i$, $X_{3it}$ be the log-transformed total raindays $\geq 1.0$ mm in $t$th month in station $i$, respectively. We consider the following model:

$$\begin{aligned} Y_{it} &= X_{1it}\beta_1 + X_{2it}\beta_2 + X_{3it}\beta_3 + m(U_{it}) + \alpha_i + \varepsilon_{it}, \\ Z_{1it} &= X_{1it} + \eta_{it}, \\ & i = 1, \ldots, n;\ t = 1, \ldots, T, \end{aligned}$$

where $U_{it} = t/T$, $\alpha_i = \rho\bar{X}_i$ for $i = 1, 2, \ldots, n$, and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it1}$. The measurement error $\eta_{it} \sim N(0, \Sigma_\eta)$ where we take $\Sigma_\eta = 0.2^2 I_2$. A natural question is whether the coefficient of total sunshine duration, the total rainfall and the total raindays $\geq 1.0$ mm are statistically significant. To answer this question, the proposed PLR tests are employed. As a result, The PLR tests statistic are $\lambda_n(\beta_1) = 80.12$, $\lambda_n(\beta_2) = 0.8490$ and $\lambda_n(\beta_3) = 5.1266$, respectively. Which provides stark evidence that the variables $X_{1it}$ and $X_{3it}$ are significant for the mean maximum temperature at the 0.1 significant level, and which indicates that the coefficient of the variables $X_{2it}$ is zero. We apply the modified least squares method in Section 2 to models (5.1), and obtain the estimators of parameter $\beta = (\beta_1, \beta_3)$ as $\hat{\beta} = (0.6861, 0.2365)$ with $SD = 0.040$. which shows that the total raindays $\geq 1.0$ mm has no significant effect on the mean maximum temperature, but total sunshine duration is highly positively associated with the mean maximum temperature. In other words, longer sunshine during tends to result in higher maximum temperatures. In addition, the estimators of $m(\cdot)$ are shown in Figure 2.

From Figure 2, we can see that from the beginning of 2006 to the end of 2008, there is an upward trend in the monthly mean maximum temperatures. Thereafter, there is a slight decrease from the beginning of 2009 to the end of 2011. Then from the beginning of 2013 to the end of 2015, there is an increase in the maximum temperatures. From Figure 2, we also can see that the modified profile least squares estimator of the nonparametric component outperforms the naive local profile least squares estimator. The latter has a greater volatility.

## 6. Conclusion

In this paper, we have studied the estimation of a partially linear errors-in-variables panel data models with fixed effects. When measurement errors are ignored, the usual profile least squares lead to biased estimators of the parametric and nonparametric components. To deal with this problem, we have proposed a modified profile least squares estimator for the parametric and nonparametric components by correcting the attenuation. We have showed that they were consistent and asymptotically normal.

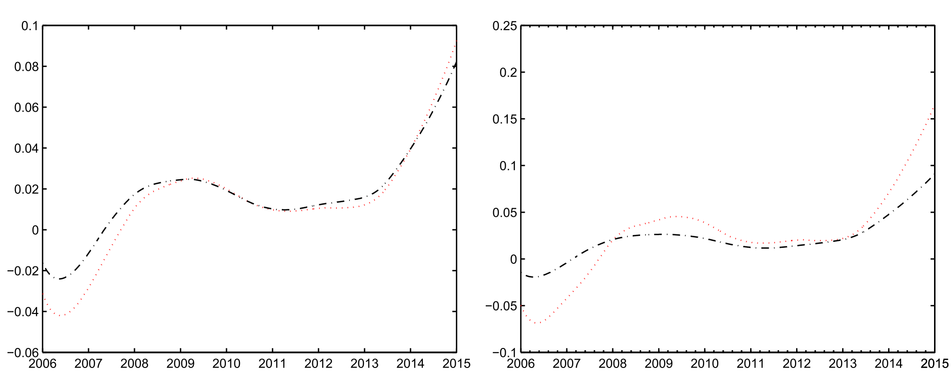


**Figure 2.** The estimators of nonparametric component $m(\cdot)$ with $\Sigma_\eta = 0.2$ (left panel) and $\Sigma_\eta = 0.4$ (right panel). The proposed estimator (dot-dashed curve) and the naive estimator (dotted curve).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research is supported by the National Social Science Fund of China [grant number 18BTJ034].

## References


Baltagi, B. H. (2005). *Econometric analysis of panel data* (2nd ed.). Wiley.

Engle, R. F., Granger, W. J., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, *80*, 310–319. https://doi.org/10.1080/01621459.1986.10478274

Fan, J., & Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*(6), 1031–1057. https://doi.org/10.3150/bj/1137421639

Fan, G. L., Liang, H. Y., & Wang, J. F. (2013). Statistical inference for partially time-varying coefficient errors-in-variables models. *Journal of Statistical Planning and Inference*, *143*, 505–519. https://doi.org/10.1016/j.jspi.2012.08.017

Fan, J., Peng, H., & Huang, T. (2005). Semilinear high-dimensional model for normalization of microarrays data: A theoretical analysis and partial consistency. *Journal of the American Statistical Association*, *100*, 781–813. https://doi.org/10.1198/016214504000001781

Fan, J., & Truong, Y. K. (1993). Nonparametric regression with errors-in-variables. *Annals of Statistics*, *21*, 1900–1925. https://doi.org/10.1214/aos/1176349402

Fuller, W. A. (1987). *Measurement error models*. Wiley.

He, B. Q., Hong, X. J., & Fan, G. L. (2020). Penalized empirical likelihood for partially linear errors-in-variables panel data models with fixed effects. *Statistical Papers*, *61*, 2351–2361. https://doi.org/10.1007/s00362-018-1049-2

Henderson, D. J., Carroll, R. J., & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, *144*, 257–275. https://doi.org/10.1016/j.jeconom.2008.01.005

Hsiao, C. (2003). *Analysis of panel data* (2nd ed.). Cambridge University Press.

Hu, X. M. (2014). Estimation in a semi-varying coefficient model for panel data with fixed effects. *Journal of Systems Science and Complexity*, *27*, 594–604. https://doi.org/10.1007/s11424-014-2263-1

Li, D. G., Chen, J., & Gao, J. T. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal*, *14*(3), 387–408. https://doi.org/10.1111/j.1368-423X.2011.00350.x

Liang, H., Härdle, W., & Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, *27*, 1519–1535. https://doi.org/10.1214/aos/1017939140

Opsomer, J. D., & Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, *25*, 186–211. https://doi.org/10.1214/aos/1034276626

Su, L. J., & Ullah, A. (2006). Profile likelihood estimation of partially linear panel data models with models with fixed effects. *Economics Letters*, *92*, 75–81. https://doi.org/10.1016/j.econlet.2006.01.019

You, J. H., & Chen, G. M. (2006). Estimation in a statistical inference in a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis*, *97*, 324–341. https://doi.org/10.1016/j.jmva.2005.03.002

Zhao, Y. Y., & Lin, J. G. (2019). Estimation and test of jump discontinuities in varying coefficient models with empirical applications. *Computational Statistics & Data Analysis*, *139*, 145–163. https://doi.org/10.1016/j.csda.2019.05.003

Zhou, H. B., You, J. H., & Zhou, B. (2010). Statistical inference for fixed-effects partially linear regression models with errors in variables. *Statistical Papers*, *51*, 629–650. https://doi.org/10.1007/s00362-008-0150-3

Zhu, L. X., & Cui, H. J. (2003). A semiparametric regression model with errors in variables. *Scandinavian Journal of Statistics*, *30*, 429–442. https://doi.org/10.1111/sjos.2003.30.issue-2


## Appendix. Proofs of the main results

In order to establish this theorem, we need to show some useful lemmas. For the convenience and simplicity, let $\vartheta_k = \int u^k K(u)\,\mathrm{d}t$, $c_N = \{\log(1/h)/(Nh)\}^{1/2} + h^2$ and set $\Phi(U) = E(\mathbf{X} \mid U)$.

**Lemma A.1:** *Suppose that Assumptions* (A1)–(A5) *hold. Then*

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

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

***Proof:*** Note that

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

Each element of the above matrix is in the form of a kernel regression. Similar to the proof of Lemma A.2 in J. Fan and Huang (2005), we can derive the desired result. ∎

**Lemma A.2:** *Suppose that Assumptions* (A1)–(A7) *hold, we have*

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

*where* $\Sigma = E\{[X - E(X \mid U)]^\tau [X - E(X \mid U)]\}$.

***Proof:*** Similar to the proof of Lemma A.3 in He et al. (2020). ∎

**Lemma A.3:** *Under the conditions of Theorem* 3.1, *we have*

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

***Proof:*** By Lemma A.1 and a similar proof to that of Lemma A.2, we have

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

Defining $\Omega = (\Omega_{11}, \ldots, \Omega_{1T}, \ldots, \Omega_{nT}) = (I - \widetilde{M}_{\widetilde{B}})(I-S)m$, we have

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

and

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

We now deal with the first term. By Assumption (A6), it is easy to see that

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

Hence, Lemma A.3 holds. ■

**Lemma A.4:** *Under the conditions of Theorem* 3.1, *we have*

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

***Proof:*** By Lemma A.1, we have $S\varepsilon = O_p(c_N)$. Similar to the proof of Lemma A.3 and under Assumption (A7), we have $\frac{1}{\sqrt{N}}\widetilde{Z}^\tau(I - \widetilde{M}_{\widetilde{B}})S\varepsilon = O(\sqrt{N}c_N^2) = o_p(1)$. Therefore, by Lemma A.2 and the central limit theorem, we have

$$\begin{aligned}
\frac{1}{\sqrt{N}}\widetilde{Z}^\tau(I-\widetilde{M}_{\widetilde{B}})(I-S)\varepsilon &= \frac{1}{\sqrt{N}}\widetilde{Z}^\tau(I-\widetilde{M}_{\widetilde{B}})\varepsilon + o_p(1) \\
&\xrightarrow{p} N\left(0, \frac{T-1}{T}\sigma^2(\Sigma + \Sigma_\eta)\right).
\end{aligned}$$

■

**Lemma A.5:** *Under the conditions of Theorem* 3.1, *we have*

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

*where* $\lambda(A)$ *denotes the eigenvalue of matrix* $A$.

***Proof:*** Because $\bar{M} = B(I_N - i_N i_N^\tau/N)(B^\tau B)^{-1}B^\tau = \frac{1}{T}I_n \oplus (i_T i_T^\tau) - \frac{1}{N}i_N i_N^\tau$ and $Si_N = i_N$, it is easy to show that

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

For the term $K_2$, we have

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

By Lemma A.1, we have

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

which holds uniformly for $i, j = 1, \ldots, n$. Hence

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

or

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

Therefore,

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

It is obvious that the eigenvalues of $K_1$ satisfy $0 \leq \lambda(K_1) \leq \frac{1}{T}$. Similar to the proof of Lemma A.5 in J. Fan et al. (2005), we have

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

■

***Proof of Theorem 3.1.:*** Since $\widetilde{Y} = \widetilde{B}\alpha + \widetilde{X}\beta + (I-S)m + \widetilde{\varepsilon}$ and $(I - \widetilde{M}_{\widetilde{B}})\widetilde{B} = 0$, by the definition of $\widehat{\beta}$, we have

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

By Lemmas A.2–A.4, we obtain

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

Therefore, applying Lemma A.1 and the fact $(A + aB)^{-1} \to A^{-1}$ as $a \to 0$, we have

$$\begin{aligned}
\sqrt{N}(\widehat{\beta} - \beta) &= \left[\frac{T-1}{T}\Sigma\right]^{-1} \\
&\quad \cdot [\widetilde{Z}^\tau(I-\widetilde{M}_{\widetilde{B}})(\varepsilon - \eta\beta) + n(T-1)\Sigma_\eta] + o_p(1) \\
&= \left[\frac{T-1}{T}\Sigma\right]^{-1} \cdot [(X - E(X \mid U))(I-\widetilde{M}_{\widetilde{B}})(\varepsilon - \eta\beta) \\
&\quad + \eta^\tau(I-\widetilde{M}_{\widetilde{B}})\varepsilon + \{n(T-1)\Sigma_\eta - \eta^\tau(I-\widetilde{M}_{\widetilde{B}})\}\beta] \\
&\quad + o_p(1)
\end{aligned}$$

$$= \left[\frac{T-1}{T}\Sigma\right]^{-1} \cdot J + o_p(1).$$

Note that

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

Therefore,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Cov}\left(\frac{1}{\sqrt{N}}J\right) &= \frac{T-1}{T}\{E(\varepsilon - \eta\beta)^2\Sigma + \sigma^2\Sigma_\eta \\ &+ E[(\eta\eta^\tau - \Sigma_\eta)\beta]^{\oplus 2}\} = \frac{T-1}{T}\Sigma_1.\end{aligned}$$

Invoking the Slutsky theorem and the central limit theorem, we obtain the desired result. The proof of $\widehat{\Sigma}^{-1}\widehat{\Sigma}_1\widehat{\Sigma}^{-1}$ being a consistent estimator of $\Sigma^{-1}\Sigma_1\Sigma^{-1}$ is straight forward. We here omit the details. ■

***Proof of Theorem 3.2.:*** According to the definition of $\widehat{m}$, we have

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

Because $\sum_{i=1}^n \alpha_i = 0$, it is easy to show that $(I - S\bar{M}_B)^{-1}S(I - \bar{M}_B)B\alpha_i = 0$. Then we have

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

and

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

Similar to the proof of Theorem 3 in J. Fan et al. (2005), we obtain

$$L_1 \le \frac{T}{(\sqrt{T}-1)^2}\frac{\mu_2^2 h^4}{4}E[m''(u)] + o_p(h^4) \tag{A1}$$

and

$$\begin{aligned}L_2 &\le \frac{T}{(\sqrt{T}-1)^2}\frac{1}{N}E\|S(I - \bar{M}_B)\varepsilon\|^2 \\ &= \frac{T}{(\sqrt{T}-1)^2}\frac{\sigma^2 + \beta^\tau\Sigma_\eta\beta}{N}\,\mathrm{tr}[S(I - \bar{M}_B)S] \\ &\le \frac{T}{(\sqrt{T}-1)^2}\frac{\sigma^2 + \beta^\tau\Sigma_\eta\beta}{N}\,\mathrm{tr}(SS^\tau).\end{aligned} \tag{A2}$$

By the law of large numbers, we have

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

which, together with (A1) and (A2), Theorem 3.2 holds. ■

***Proof of Theorem 3.3.:*** By (9) and $(I - M_B)B = 0$, $\widehat{\sigma}^2$ can be decomposed as

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

Therefore,

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

and

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

Because $\varepsilon_{it}$ and $\eta_{it}$ are i.i.d, we have

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

Then, by some simple calculation, we have

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

Therefore,

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

By the same argument as proving Theorem 3 of Zhou et al. (2010), we can show that $\Pi_i = o_p(N^{-\frac{1}{2}}),\ \ i = 2, 3$. Equation (11) holds.

We now show the consistency of $\widehat{\Theta}$. To simplify the notation, we write

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

and $\zeta_{it} = \varepsilon_{it} - \eta_{it}\widehat{\beta} - \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})$. Then

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

For $J_1$, we have

$$\begin{aligned}J_1 &= \frac{1}{T}\sum_{t=1}^{T} E\left[\varepsilon_{it} - \eta_{it}\widehat{\beta} - \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})\right]^4 + o_p(1) \\ &= E(\varepsilon_{it} - \eta_{it}\widehat{\beta})^4 + \frac{1}{T^2}\left[\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})\right]^4 \\ &\quad + \frac{6}{T^3}\sum_{t=1}^{T} E(\varepsilon_{it} - \eta_{it}\widehat{\beta})^2\left[\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})\right]^2 \\ &\quad - \frac{4}{T^2}\sum_{t=1}^{T} E(\varepsilon_{it} - \eta_{it}\widehat{\beta})^3\left[\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})\right] \\ &\quad - \frac{4}{T^2}\left[\sum_{t=1}^{T}(\varepsilon_{it} - \eta_{it}\widehat{\beta})\right]^4 + o_p(1) \\ &= \left(1 + \frac{6}{T^2} - \frac{3}{T^3} - \frac{4}{T}\right) E(\varepsilon_{it} - \eta_{it}\widehat{\beta})^4 \\ &\quad + \frac{3 - T}{T - 1}\left(6 - \frac{9}{T}\right)(\sigma^2 + \beta^{\tau}\Sigma_{\eta}\beta).\end{aligned}$$

Moreover, according to Hölder inequality and the fact that $\frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}\Xi_{it}^{4} = o_p(1)$, we can show that $J_s = o_p(1)$ for $s = 2, 3, 4$ and 5. Thus, the consistency result of $\Theta$ follows. ■