Statistical Papers
https://doi.org/10.1007/s00362-018-1049-2

REGULAR ARTICLE

Penalized empirical likelihood for partially linear errors-in-variables panel data models with fixed effects

Bang-Qiang He[1] · Xing-Jian Hong[2] · Guo-Liang Fan[1]

Received: 20 August 2017 / Revised: 10 September 2018

Abstract
For the partially linear errors-in-variables panel data models with fixed effects, we, in this paper, study asymptotic distributions of a corrected empirical log-likelihood ratio and maximum empirical likelihood estimator of the regression parameter. In addition, we propose penalized empirical likelihood (PEL) and variable selection procedure for the parameter with diverging numbers of parameters. By using an appropriate penalty function, we show that PEL estimators have the oracle property. Also, the PEL ratio for the vector of regression coefficients is defined and its limiting distribution is asymptotically chi-square under the null hypothesis. Moreover, empirical log-likelihood ratio for the nonparametric part is also investigated. Monte Carlo simulations are conducted to illustrate the finite sample performance of the proposed estimators.

Keywords Panel data · Penalized empirical likelihood · Partially linear model · Fixed effect · Errors-in-variables

1 Introduction

The analysis of panel data is the subject of one of the most active and innovative bodies of literature in econometrics. Panel data sets have various advantages over that of pure time-series or cross-sectional data sets, among which the most important one is perhaps that the panel data provide researchers a flexible way to model both heterogeneity among cross-sectional units and possible structural changes over time. Arellano (2003), Baltagi (2005) and Hsiao (2003) provided excellent overviews of statistical inference and econometric analysis of parametric panel data models. However, a misspecified parametric panel data model may result in misleading inference. Therefore, econometricians and statisticians have developed some flexible nonpara-

✉ Xing-Jian Hong
myemailh@126.com

1 School of Mathematics and Physics, Anhui Polytechnic University, Wuhu 241000, China

2 School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou 310018, China

metric and semi-parametric panel data models. For example, Su and Ullah (2007) proposed a class of two-step estimators for nonparametric panel data with random effects. Cai and Li (2008) studied dynamic nonparametric panel data models. Henderson et al. (2008) considered nonparametric panel data model with fixed effects. Rodriguez-Poo and Soberon (2014) considered varying coefficient fixed effects panel data models, established direct semiparametric estimations. Chen et al. (2013) studied partially linear single-index panel data models with fixed effects, proposed a dummy variable method to remove fixed effects and established a semi-parametric minimum average variance estimation procedure. Baltagi and Li (2002) discussed partially linear panel data models with fixed effects, developed the series estimation procedure and the profile likelihood estimation technique. Hu (2014) proposed the profile likelihood procedure to estimate semi-varying coefficient model for panel data with fixed effects. The partially linear panel data models with fixed effects are widely used in econometric analysis; see, e.g., Henderson et al. (2008), Horowitz and Lee (2004), Hu (2017) and Li et al. (2011).

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

$$Y_{it} = X_{it}^{\tau}\beta + g(Z_{it}) + \mu_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{1.1}$$

where Y_{it} is the response, $(X_{it}, Z_{it}) \in R^p \times R$ are strictly exogenous variables, $\beta = (\beta_1, \ldots, \beta_p)^{\tau}$ is a vector of p-dimensional unknown parameters, and the superscript τ denotes the transpose of a vector or matrix. $g(Z_{it})$ is a unknown functions and μ_i is the unobserved individual effects, ε_{it} is the random model error. Here, we assume ε_{it} to be i.i.d. with zero mean and finite variance $\sigma^2 > 0$. We allow μ_i to be correlated with X_{it}, and Z_{it} with an unknown correlation structure. Hence, model (1.1) is a fixed effects model.

It is well known that in many fields, such as engineering, economics, biology, biomedical sciences and epidemiology, observations are measured with error. For example, urinary sodium chloride level (Wang et al. 1996) and serum cholesterol level (Carroll et al. 1995) are often subjects to measurement errors. Simply ignoring measurement errors (errors-in-variables), known as the naive method, would result in biased estimators. Handing the measurement errors in covariates is generally a challenge for statistical analysis. For the past two decades, errors-in-variables can be handled by means of corrected score function (Nakamura 1990), corrected likelihood method (Hanfelt and Liang 1997), the instrumental variables estimation approach (Schennach 2007) and so on.

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

$$\begin{cases} Y_{it} = X_{it}^{\tau}\beta + g(Z_{it}) + \mu_i + \varepsilon_{it}, \\ W_{it} = X_{it} + \nu_{it}, \end{cases} \quad i = 1, \ldots, n; \; t = 1, \ldots, T. \tag{1.2}$$

where the covariate variables X_{it} are measured with additive error and are not directly observable. Instead, X_{it} are observed $W_{it} = X_{it} + \nu_{it}$, where the measurement errors ν_{it} are independent and identically distributed, independent of $(X_{it}, Z_{it}, \varepsilon_{it})$, with mean zero and covariance matrix Σ_ν. We will assume that Σ_ν is known, as in the papers of Zhu and Cui (2003) and You and Chen (2006) and other. When Σ_ν is unknown, we can estimate it by repeatedly measuring W_{it}; see Liang et al. (1999) and Fan et al. (2013) for details.

It is well-known that high-dimensional data analysis arises frequently in many contemporary statistical studies. The emergence of high-dimensional data, such as the gene expression values in microarray, brings challenges to many traditional statistical methods and theory. One important aspect of the high-dimensional data under the regression setting is that the number of covariates is diverging. When dimensionality diverges, variable selection through regularization has proven to be effective. As argued in Hastie et al. (2009) and Fan and Lv (2008), penalized likelihood can properly adjust the bias-variance trade-off so that the performance improvement can be achieved; Various powerful penalization methods have been developed for variable selection. Fan and Li (2001) proposed a unified approach via nonconcave penalized least squares to automatically and simultaneously select variables. Li and Liang (2008) developed the nonconcave penalized quasilikelihood method for variable selection in semiparametric regression model. Recently, a new and efficient variable selection approach, PEL introduced for the first time by Tang and Leng (2010), was applied to analyze mean vector in multivariate analysis and regression coefficients in linear models with diverging number of parameters. As demonstrated in Tang and Leng (2010), the PEL has merits in both efficiency and adaptivity stemming from a nonparametric likelihood method. Also, the PEL method possesses the same merit of the empirical likelihood (EL) which only uses the data to determine the shape and orientation of confidence regions and without estimating the complex covariance. As far as we know, there are a few papers related to the PEL approach, such as Ren and Zhang (2011) proposed the PEL approach for variable selection in moment restriction models; Leng and Tang (2012) applied the PEL approach to parametric estimation and variable selection for general estimating equations; Wang and Xiang (2017) studied PEL inference for sparse additive hazards regression with a diverging number of covariates.

It is worth pointing out that there is no result available in the literature when the number of covariates is diverging in partially linear errors-in-variables panel data models with fixed effects. In this paper, our aim is to extend the results in Fan et al. (2016) for high-dimensional partially linear varying coefficient model with measurement errors to partially linear error-in-variables panel data models with fixed effects. Our contribution can be summarized as follows. Following the estimation procedure proposed by Fan et al. (2005), we first adapt a local linear dummy variable approach to remove the unknown fixed effects. Moreover, we utilize the EL method to construct confidence regions of unknown parameter and establish asymptotic normality of maximum empirical likelihood (MEL) estimator of the parameter. At last, for building sparse models, we propose an estimating equation-based PEL, a unified framework for variable selection in optimally combining estimating equations. More specifically, this method has the oracle property. Moreover, PEL ratio statistic shows the Wilks' phenomenon, facilitating hypothesis testing and constructing confidence regions.

The layout of the remainder of this paper is as follows. In Sect. 2, we construct corrected-attenuation EL ratio and test statistic as well as define the MEL and PEL estimators of the parameter and give their asymptotic properties. Moreover, empirical log-likelihood ratio for the nonparametric part is also investigated. Finally, we briefly introduce computational algorithm. The simulated example is provided in Sect. 3. Section 4 summarizes some conclusions and discusses future research. Assumption conditions and the proofs of the asymptotic results are given in Appendix.

2 Methodology and main results

2.1 Modified empirical likelihood

We give vector and matrix notations in the following. Let $Y = (Y_1^\tau, \ldots, Y_n^\tau)^\tau$, $X = (X_1^\tau, \ldots, X_n^\tau)^\tau$, $Z = (Z_1^\tau, \ldots, Z_n^\tau)^\tau$, $\mu_0 = (\mu_1^\tau, \ldots, \mu_n^\tau)^\tau$ and $\varepsilon = (\varepsilon_1^\tau, \ldots, \varepsilon_n^\tau)^\tau$ be $nT \times 1$ vectors, where $Y_i = (Y_{i1}, \ldots, Y_{iT})^\tau$, $X_i = (X_{i1}, \ldots, X_{iT})^\tau$, $Z_i = (Z_{i1}, \ldots, Z_{iT})^\tau$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\tau$, and $D_0 = I_n \otimes i_T$ with $\otimes$ the Kronecker product, I_n denotes the $n \times n$ identity matrix, and i_n denotes the $n \times 1$ vector of ones. We rewrite model (1.1) in a matrix format which yields

$$Y = X\beta + g(Z) + D_0\mu_0 + \varepsilon. \tag{2.1}$$

For the identification purpose, we impose the restriction $\sum_{i=1}^n \mu_i = 0$. Letting $D = [-i_{n-1} \quad I_{n-1}] \otimes i_T$ and $\mu = (\mu_2, \ldots, \mu_n)^\tau$, model (2.1) can then be rewritten as

$$Y = X\beta + g(Z) + D\mu + \varepsilon. \tag{2.2}$$

Let $G_{it}(z, h) = (1, (Z_{it} - z)/h)^\tau$, $K_h(z) = K(\cdot/h)/h$ with a kernel function $K(\cdot)$ and a bandwidth h. The diagonal matrices

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

and $\zeta = (\mu^\tau, \beta^\tau)^\tau$. Given ζ, we can estimate the functions $g(\cdot) = (g(z), \{hg'(z)\}^\tau)^\tau$ by

$$\sum_{i=1}^n \sum_{t=1}^T \left\{ \left(Y_{it} - X_{it}^\tau \beta - \mu_i \right) - \left[g(z) + hg'(z)(Z_{it} - z) \right] \right\}^2 K_h(Z_{it} - z). \tag{2.3}$$

Let $G(z, h) = [G_1^\tau(z, h), \ldots, G_n^\tau(z, H)]^\tau$, $G_i(z, h) = (G_{i1}(z, h), \ldots, G_{iT}(z, h))^\tau$, $g'(z) = \partial g(z)/\partial z$ and $z = z_{it}$ is in a neighborhood of Z_{it}. Then the solution of problem (2.3) is given by

$$g(\cdot) = [G^\tau(z,h)W_H(u)G(z,h)]^{-1}G^\tau(z,h)W_H(u)(Y - X\beta - D\mu)$$
$$= S(z,H)(Y - X\beta - D\mu).$$

In particular, the estimator for $g(z)$ is given by

$$\hat{g}(z) = s(z,h)(Y - X\beta - D\mu). \tag{2.4}$$

where $s(z,h) = (1\ \ 0)S(z,h)$.

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:

$$\hat{\zeta} = \arg\min_{\zeta} \left[Y - X\beta - D\mu - S(Y - X\beta - D\mu)\right]^\tau [Y - X\beta - D\mu - S(Y - X\beta - D\mu)], \tag{2.5}$$

where the smoothing matrix S is

$$S = \begin{pmatrix} (1\ \ 0)[G^\tau(Z_{11},h)W_{11}(z,h)G(Z_{11},h)]^{-1}G^\tau(Z_{11},h)W_{11}(z,h) \\ \vdots \\ (1\ \ 0)[G^\tau(Z_{1T},h)W_{1T}(z,h)G(Z_{1T},h)]^{-1}G^\tau(Z_{11},h)W_{1T}(z,h) \\ \vdots \\ (1\ \ 0)[G^\tau(Z_{nT},h)W_{nT}(z,h)G(Z_{nT},h)]^{-1}G^\tau(Z_{nT},h)W_{nT}(z,h) \end{pmatrix}$$
$$= \begin{pmatrix} S_{11} \\ \vdots \\ S_{1T} \\ \vdots \\ S_{nT} \end{pmatrix}.$$

Supposing that $\widetilde{X} = (I_{nT} - S)X$, $\widetilde{Y} = (I_{nT} - S)Y$, $\widetilde{D} = (I_{nT} - S)D$, we have $\widetilde{\mu} = (\widetilde{D}^\tau \widetilde{D})^{-1}\widetilde{D}^\tau(\widetilde{Y} - \widetilde{X}\beta)$. Let $H = I_{nT} - \widetilde{D}(\widetilde{D}^\tau\widetilde{D})^{-1}\widetilde{D}^\tau$, we can obtain $H\widetilde{D}\mu = 0$. Hence, the fixed effects term $D\mu$ is eliminated in (2.3). Let e_{it} be the $nT \times 1$ vector with its $\{(i-1)T + t\}$th element being 1 and others 0. We state the approximate residuals as the following:

$$\Lambda_i(\beta) = \sum_{t=1}^{T} \widetilde{X}_{it}^\tau H(\widetilde{Y}_{it} - \widetilde{X}_{it}\beta), \quad i = 1, \ldots, n. \tag{2.6}$$

However, X_{it}'s can not be observed in our case and we just have W_{it}. If we ignore the measurement error and replace X_{it} with W_{it} in (2.4) directly, (2.4) 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 Liang et al. (1999) and Hu et al. (2009) for more details.

$$\Gamma_i(\beta) = \sum_{t=1}^{T} \widetilde{W}_{it}^{\tau} H(\widetilde{Y}_{it} - \widetilde{W}_{it}\beta) - (T-1)\Sigma_\nu \beta, \quad i = 1, \ldots, n. \tag{2.7}$$

Note that $E(\Gamma_i(\beta)) = 0$, if β is the true parameter. Therefore, similar to Owen (1990), we define a corrected-attenuation empirical likelihood (CAEL) ratio of β as.

$$R_n(\beta) = -\max\left\{\sum_{i=1}^{n} \ln(np_i)|p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\Gamma_i(\beta) = 0\right\}. \tag{2.8}$$

With the assumption that 0 is inside the convex hull of the point $(\Gamma_1(\beta), \ldots, \Gamma_n(\beta))$, a unique value for $R_n(\beta)$ exists. By the Lagrange multiplier method, one can obtain that

$$R_n(\beta) = \sum_{i=1}^{n} \ln\{1 + \gamma^\tau \Gamma_i(\beta)\}, \tag{2.9}$$

where γ is determined by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\Gamma_i(\beta)}{1 + \gamma^\tau \Gamma_i(\beta)} = 0. \tag{2.10}$$

Theorem 2.1 *Suppose that the conditions of* (B1)–(B7) *in the Appendix hold. Further assume that* $E(\varepsilon^3|X, Z) = 0$ *almost surely or* $k \geq 8$. *If* β_0 *is the true value of the parameter vector and* $\stackrel{d}{\to}$ *stands for convergence in distribution,* $p^{3+2/(k-2)}/n \to 0$ *as* $n \to \infty$, *then* $(2p)^{-1/2}(2R_n(\beta_0) - p) \stackrel{d}{\to} N(0, 1)$.

Define $\hat{\beta}_{ME} = \arg\min_\beta R_n(\beta)$, *which is the MEL estimator of the parameter* β.

Theorem 2.2 *Under the conditions of Theorem 2.1, we have*

$$\sqrt{n} A_n \Omega^{-1/2}(\widehat{\beta}_{ME} - \beta_0) \stackrel{d}{\to} N(0, \Delta).$$

A_n *represents a projection of the diverging dimensional vector to a fixed dimension* s, *and* A_n *is a* $s \times p$ *matrix such that* $A_n A_n^\tau \to \Delta, \Delta$ *is a* $s \times s$ *nonnegative symmetric matrix with fixed* s, *and* $\Omega = \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$, $\Sigma_1 = (T-1)\big\{E(\varepsilon_{11} - \nu_{11}\beta_0)^2\Sigma_2 + \sigma^2\Sigma_\nu + E[(\nu_{11}\nu_{11}^\tau - \Sigma_\nu)\beta_0]^2\big\}$, $\Sigma_2 = E\{[X_{11} - E(X_{11}|Z_{11})]^\tau[X_{11} - E(X_{11}|Z_{11})]\}$ *and* $\Sigma_0 = (T-1)\Sigma_2$.

Further, $\widehat{\Sigma}_2^{-1}\widehat{\Sigma}_1\widehat{\Sigma}_2^{-1}$ *is a consistent estimator of* $\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}$ *where* $\widehat{\Sigma}_2 = \frac{1}{n}\widetilde{W}^\tau H \widetilde{W} - (T-1)\Sigma_\nu$ *and* $\widehat{\Sigma}_1 = \{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}[\widetilde{W}_{it}^\tau H(\widetilde{Y}_{it} - \widetilde{W}_{it}\widehat{\beta})] + (T-1)(\Sigma_\nu\widehat{\beta})\}^{\oplus 2}$ *and* $\mathbf{A}^{\oplus 2}$ *means* $\mathbf{A}\mathbf{A}^\tau$. *By Theorem 2.2, we obtain that*

$$(\widehat{\beta}_{ME} - \beta_0)^\tau n[\widehat{\Sigma}_2^{-1}\widehat{\Sigma}_1\widehat{\Sigma}_2^{-1}]^{-1}(\widehat{\beta}_{ME} - \beta_0) \stackrel{d}{\to} \chi_s^2.$$

2.2 Penalized empirical likelihood for variable selection

We use the PEL by combining the profile likelihood method and the smoothly clipped absolute deviation (SCAD) penalized approach. The SCAD penalty is defined in terms of its first derivative.

We define the penalized empirical likelihood (PEL) as follows,

$$\mathcal{L}_n(\beta) = R_n(\beta) + n\sum_{j=1}^{p} p_\lambda(|\beta_j|), \tag{2.11}$$

where $p_\lambda(\cdot)$ is a penalty function with tuning parameter λ. See Fan and Li (2001) for example of this function. In this paper, we use the smoothly clipped absolute deviation penalty, whose first derivative satisfies

$$p'_\lambda(\theta) = \theta\left\{I(\theta \le \lambda) + \frac{(a\lambda-\theta)_+}{(a-1)\lambda}I(\theta > \lambda)\right\},\ \ \theta > 0,\ \lambda > 0, \tag{2.12}$$

for some $a > 2$ and $p'_\lambda(0) = 0$. Following Fan and Li (2001), we set $a = 3.7$ in our work.

Maximizing the PEL function (2.8) is equivalent to minimizing

$$\mathcal{L}_n(\beta) = \sum_{i=1}^{n}\ln\{1+\gamma^\tau\Gamma_i(\beta)\} + n\sum_{j=1}^{p} p_\lambda(|\beta_j|), \tag{2.13}$$

Let $\mathcal{B} = \{j : \beta_{0j}\}$ be the set of nonzero components of the true parameter vector β_0 and its cardinality $|\mathcal{B}| = d$ where d is allowed grow as $n \to \infty$. Without loss of generality, one can partition the parameter vector as $\beta = (\beta_1^\tau, \beta_2^\tau)^\tau$ where $\beta_1 \in \mathbb{R}^d$ and $\beta_2 \in \mathbb{R}^{p-d}$. Hence, the true parameter $\beta_0 = (\beta_{10}^\tau, 0^\tau)^\tau$ and we write $\widehat{\beta} = (\widehat{\beta}_1^\tau, \widehat{\beta}_2^\tau)^\tau$ called PEL estimator which is the minimizer of (2.13). The matrix Ω can be decomposed as a block matrix according to the arrangement of β_0 as $\Omega = \begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix}$.

Theorem 2.3 *Suppose that Assumptions* (B1)–(B9) *hold. If* $p^5/n \to 0$, *then with probability tending to 1, the PEL estimator* $\widehat{\beta}$ *satisfies*

(a) (*Sparity*): $\widehat{\beta}_2 = 0$;
(b) (*Asymptotic normality*): $n^{1/2}W_n\Omega_p^{-1/2}(\widehat{\beta}_1 - \beta_{10}) \xrightarrow{d} N(0, G)$, *where* $W_n \in R^{q\times d}$ *such that* $W_nW_n^\tau \to G$ *for* $G \in R^{q\times q}$ *with fixed* q *and* $\Omega_p = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$.

A remarkable advantage of PEL lies in testing hypotheses and constructing confidence regions for β. *To understand this more clearly, we consider the problem of testing linear hypothesis*:

$$H_0 : L_n\beta_{10} = 0 \quad vs \quad H_1 : L_n\beta_{10} \ne 0$$

where L_n is $q \times s$ matrix such that $L_n L_n^\tau = I_q$ for a fixed and finite q. A nonparametric profile likelihood ratio statistic is constructed as

$$\widetilde{\mathcal{L}}_n(\hat{\beta}) = -\left\{\mathcal{L}_n(\hat{\beta}) - \min_{\beta, L_n\beta_{10}=0} \mathcal{L}_n(\beta)\right\}.$$

We summarize the property of the test statistic in the following theorem.

Theorem 2.4 *Under the conditions of Theorem 2.3. Then under the null hypothesis H_0, we have*

$$2\widetilde{\mathcal{L}}_n(\hat{\beta}) \xrightarrow{d} \chi_q^2, \quad as\ n \to \infty.$$

As a consequence of the theorem, confidence regions for the parameter β can be constructed. More precisely, for any $0 \le \alpha < 1$, let c_α be such that $P(\chi_q^2 > c_\alpha) \le 1-\alpha$. Then $\ell_{PEL}(\alpha) = \{\beta \in R^p : \widetilde{\mathcal{L}}_n(\beta) \le c_\alpha\}$ constitutes a confidence region for β with asymptotic coverage α because the event that β belongs to $\ell_{PEL}(\alpha)$ is equivalent to the event that $\widetilde{\mathcal{L}}_n(\beta) \le c_\alpha$.

2.3 Empirical likelihood for the nonparametric part

For model (2.2), we solve the the following optimization problem:

$$\hat{\mu} = \arg\min_{\mu} \left[Y - X\beta - g(Z) - D\mu\right]^\tau [Y - X\beta - g(Z) - D\mu],$$

we have $\mu = (D^\tau D)^{-1} D^\tau [Y - X\beta - g(Z)]$. For given β and z, an auxiliary random vector for nonparametric part can be stated as

$$\Xi_i\{g(z)\} = \sum_{t=1}^{T} (I_{it} - Q)[Y_{it} - X_{it}\beta - g(z)]K_h(Z_{it} - z), \quad i = 1, \ldots, n.$$

where $Q = D(D^\tau D)^{-1} D^\tau$. Note that $E[\Xi_i\{g(z)\}] = 0$ if $g(z)$ is the true parameter. Thus we can define an empirical log-likelihood ratio statistic for $g(z)$ by using the methodology in Owen (1988). We introduce an adjusted auxiliary random vector for $g(z)$ as follows.

$$\hat{\Xi}_i\{g(z)\} = \sum_{t=1}^{T} (I_{it} - Q)[Y_{it} - X_{it}\hat{\beta} - g(z) - \{\hat{g}(Z_{it}) - \hat{g}(z)\}]K_h(Z_{it} - z),$$
$$i = 1, \ldots, n.$$

By the adjustment in $\hat{\Xi}_i\{g(z)\}$, we not only correct the bias, but also avoid undersmoothing the function $g(z)$, as proved in the Appendix. The adjusted empirical log-likelihood ratio for $g(z)$ can be defined as

$$\mathcal{Q}_n(g(z)) = -\max\left\{\sum_{i=1}^{n}\ln(n\tilde{p}_i)|\tilde{p}_i \geq 0, \sum_{i=1}^{n}\tilde{p}_i = 1, \sum_{i=1}^{n}\tilde{p}_i\,\hat{\Xi}_i\{g(z)\} = 0\right\}.$$

By the Lagrange multiplier method, one can obtain that

$$\mathcal{Q}_n(g(z)) = \sum_{i=1}^{n}\ln\{1 + \phi^\tau\,\hat{\Xi}_i\{g(z)\}\},$$

where ϕ is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\Xi}_i\{g(z)\}}{1 + \phi^\tau\,\hat{\Xi}_i\{g(z)\}} = 0.$$

Theorem 2.5 *Suppose that the conditions of* (B1)–(B9) *in the Appendix hold. For a given* $z \in \mathcal{Z}$, *if* $g(z)$ *is the true value of the parameter, then*

$$2\mathcal{Q}_n(g(z)) \xrightarrow{d} \chi_1^2.$$

2.4 Computational algorithm

This section employs the local quadratic approximation algorithm to obtain the minimizer of PEL ratio defined by (2.13). Specifically, for each $j = 1, \ldots, p$, $[p_\lambda(|\beta_j|)]'$ can be locally approximated by the quadratic function defined as $[p_\lambda(|\beta_j|)]' = p'_\lambda(|\beta_j|)sgn(\beta_j) \approx \{p'_\lambda(|\beta_{j0}|/|\beta_{j0}|)\}\beta_j$ at an initial value β_{j0} of β_j is not close to 0; otherwise, we set $\hat{\beta}_j = 0$. In other words, in a neighborhood of a given nonzero β_{j0}, we assume that $p_\lambda(|\beta_j|) \approx p_\lambda(|\beta_{j0}|) + \frac{1}{2}\{p'_\lambda(|\beta_{j0}|/|\beta_{j0}|\}(\beta_j^2 - \beta_{j0}^2)$. We then make use of algorithm (see Owen, 2001) to obtain the minimum through nonlinear optimization. The procedure is repeated until convergence.

We apply the following Bayesian information criterion (BIC) to select the tuning parameter λ, which is defined by

$$BIC(\lambda) = -2\mathcal{L}_n(\beta_\lambda) + \ln(n)df_\lambda,$$

where df_λ is the number of nonzero estimated parameters. Then the optimal tuning parameter is the minimizer of the BIC.

3 Simulation studies

In this section, we carry out some simulation to study the finite sample performance of our proposed method. Throughout this section, we choose the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \leq 1\}$ and use the “leave-one-subject-out” cross-validation bandwidth method to select the optimal handwidth h_{opt}.

Table 1 Comparison of coverage probability for MEL, PEL, NMEL and NPEL

Σ_v	$1-\alpha$	p	(n, T)	MEL	PEL	NMEL	NPEL
$0.2I_2$	0.90	10	(50,4)	0.818	0.827	0.810	0.834
		10	(50,6)	0.852	0.863	0.878	0.879
		15	(100,6)	0.865	0.885	0.880	0.895
	0.95	10	(50,4)	0.924	0.930	0.918	0.936
		10	(50,6)	0.932	0.939	0.926	0.940
		15	(100,6)	0.937	0.946	0.929	0.947
$0.4I_2$	0.90	10	(50,4)	0.795	0.807	0.801	0.818
		10	(50,6)	0.836	0.846	0.821	0.836
		15	(100,6)	0.862	0.851	0.826	0.857
	0.95	10	(50,4)	0.914	0.923	0.912	0.921
		10	(50,6)	0.925	0.932	0.920	0.928
		15	(100,6)	0.928	0.937	0.923	0.939

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

$$\begin{cases} Y_{it} = X_{it}^{\tau}\beta + g(Z_{it}) + \mu_i + \varepsilon_{it}, \\ W_{it} = X_{it} + \nu_{it} \end{cases} \quad i = 1, \ldots, n;\ t = 1, \ldots, T. \tag{3.1}$$

where $\beta = (3, 1.5, 0, 0, 2, 0, \ldots, 0)^{\tau}$, $g(Z_{it}) = \cos(2\pi Z_{it})$, $Z_{it} \sim U(0, 1)$, $\mu_i = \frac{1}{2}\bar{Z}_i + w_i$ and $w_i \sim N(0, 0.1^2)$ for $i = 1, 2, \ldots, n$, and $\bar{Z}_i = \frac{1}{T}\sum_{t=1}^{T} Z_{it}$. The measurement error $\nu_{it} \sim N(0, \Sigma_\nu)$ where we take $\Sigma_\nu = 0.2^2 I_{10}$ and $0.4^2 I_{10}$ to represent different levels of measurement error. The covariate X_{it} is a p-dimensional normal distribution random vector with mean zero and covariance matrix $\text{cov}(X_{it}, X_{jt}) = 0.5^{|i-j|}$.

In our simulations, we take p as the integer part of $10(6n)^{1/5.1} - 20$ and the sample sizes $(n, T) = (50, 4)$, $(50, 6)$ and $(100, 6)$, respectively. In order to show the performance of the proposed methods, we compare MEL and PEL estimators with the native maximum empirical likelihood (NMEL) and native penalized empirical likelihood (NPEL) estimators that the neglecting the measurement errors with a direct replacement of X by W in our proposed estimators. In each case the number of simulated realizations is 500.

Seen from Table 1, when the nominal level is 0.9 and 0.95, shows the coverage probability of confidence region for the whole β constructed by MEL and PEL method, respectively. From the results, we can see that the PEL confidence region has slightly higher coverage probability than the NEL confidence region, and the coverage probability tends to the nominal level as the sample size increases.

From Table 2, we can see the average model errors (ME) and the standard deviations (SD) of the β_1 that is nonzero components of β. based on PEL and MEL estimators decreases as the sample size increases and the PEL estimator gives the smallest ME

Table 2 ME and SD of β_1 for MEL, PEL, NPEL and NMEL estimators

Σ_v	p	(n, T)	PEL		MEL		NPEL		NMEL	
			ME	SD	ME	SD	ME	SD	ME	SD
$0.2I_2$	10	(50,4)	0.288	0.177	0.780	0.284	0.344	0.285	0.980	0.241
	10	(50,6)	0.287	0.173	0.778	0.389	0.232	0.270	0.847	0.258
	15	(100,6)	0.185	0.275	0.790	0.506	0.246	0.296	0.819	0.328
	10	(50,4)	0.314	0.162	0.771	0.236	0.232	0.205	0.794	0.211
	10	(50,6)	0.282	0.161	0.758	0.279	0.225	0.197	0.771	0.259
	15	(100,6)	0.207	0.250	0.799	0.225	0.218	0.145	0.810	0.237
$0.4I_2$	10	(50,4)	0.295	0.207	0.986	0.328	0.334	0.268	0.847	0.376
	10	(50,6)	0.236	0.266	0.942	0.336	0.326	0.206	0.829	0.477
	15	(100,6)	0.272	0.281	0.926	0.457	0.245	0.190	0.973	0.338
	10	(50,4)	0.164	0.173	0.832	0.271	0.293	0.268	0.875	0.296
	10	(50,6)	0.155	0.142	0.820	0.260	0.192	0.260	0.815	0.239
	15	(100,6)	0.102	0.164	0.883	0.287	0.196	0.170	0.816	0.395

Table 3 Simulation results for variable selection selection based on the PEL and NPEL methods

p	(n, T)	PEL ($\Sigma_v = 0.2$)		EL($\Sigma_v = 0.2$)		NPEL($\Sigma_v = 0.4$)		NMEL ($\Sigma_v = 0.4$)	
		C	I	C	I	C	I	C	I
10	(50,4)	5.98	0	5.79	0	5.44	0	5.24	0.02
10	(50,6)	6.20	0	6.17	0	5.93	0	5.54	0
15	(100,6)	13.13	0	12.82	0	12.66	0	12.32	0
10	(50,4)	6.85	0	6.77	0	6.34	0	6.28	0
10	(50,6)	6.13	0	6.21	0	5.96	0	5.39	0
15	(100,6)	13.53	0	13.22	0	13.45	0	13.33	0

and SD among the estimators based on PEL, MEL, NPEL and NMEL methods for all settings. The ME is defined as $ME(\hat{\beta}_1) = (\hat{\beta}_1 - \beta_1)^\tau E(X^\tau X)(\hat{\beta}_1 - \beta_1)$.

Table 3 summaries the variable selection results, where important variable have large effects. The column labeled "C" gives average number of correct zeros and column labeled "I" gives the average number of incorrect zeros. From Table 3, it can be seen that variable selection method based on the PEL select all three true predictors and the average number of correct zeros are close to $p - 5$ in all settings. Further the smaller measurement errors lead to better performance. It can also be seen that the PEL approach perform better than the NPEL method for all settings. These findings imply that the model selection result based on the PEL approach effectively reduces model complexity and the selected model is very close to the true model in terms of nonzero coefficients.

From Fig. 1, We see that the method based on the EL performs slighter better than the NA method since the EL method gives shorter confidence intervals than the NA method which is shown in Theorem 4 in Xue and Zhu (2008). Besides, interestingly,

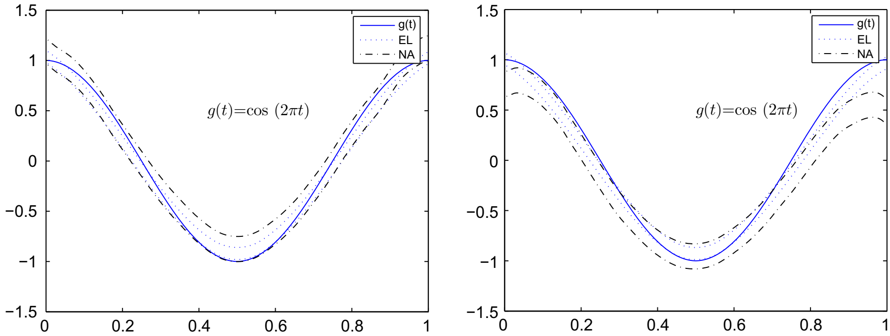

Fig. 1 95% confidence intervals for $g(z)$ for $\Sigma_v = 0.2$ (left panel) and $\Sigma_v = 0.4$ (right panel) based on EL (dotted curve) and NA (dot-dashed curve). The solid curve is the estimated cure of $g(z)$

seen from Fig. 1, $\Sigma_v = 0.2$ gives shorter confidence intervals and narrower confidence bands than $\Sigma_v = 0.4$ for $g(z)$. This shows the empirical likelihood ratio generally works well.

4 Conclusion remarks

The partially linear panel data models with fixed effects has received a lot of attention. But there have been few studies about partially linear errors-in-variables panel data models with fixed effects. We apply empirical likelihood both for parameter and nonparametric parts. Moreover, we propose PEL and variable selection procedure for the parameter with diverging numbers of parameters. By using an appropriate penalty function, we show that PEL estimators has 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. Still, more work is needed to extend the method to more complex settings, including errors-in-function, cross-sectional dependence and spatial panel data model. The results presented in this paper provide the foundation for additional work in these directions.

Acknowledgements The authors are grateful to two anonymous referees for providing detailed lists of comments and suggestions which greatly improved the presentation of the paper. This research is supported by the National Social Science Fund of China (18BTJ034).

Appendix: Proofs of the main results

We use Frobenius norm of a matrix A, defined as $||A|| = \{tr(A^\tau A)\}^{1/2}$. Before we give the details of the proofs, we present some regularity conditions.

(B1) The random vector Z_{it} has a continuous density function $f(\cdot)$ with a bounded support $\mathcal{Z}$. $0 < inf_{z\in\mathcal{Z}} f(\cdot) \le sup_{z\in\mathcal{Z}} f(\cdot) < \infty$.

(B2) The functions $E(X_{it}|Z_{it}=z)$ and $g(\cdot)$ have two bounded and continuous derivatives on $\mathcal{Z}$.

(B3) The kernel $K(v)$ is a symmetric probability density function with a continuous derivative on its compact support $\mathcal{Z}$.

(B4) $(\mu_i, W_{it}, Z_{it}, \varepsilon_{it})$, $i=1,\ldots,n$, $t=1,\ldots,T$ are i.i.d. $E(\varepsilon|W,Z,\mu)=0$ almost surely. Furthermore, for some integer $k\geq 4$, $E(||W\varepsilon||^k)\leq\infty$, $E(||W||^k)\leq\infty$, $E(|\varepsilon|^k)\leq\infty$.

(B5) $E|\breve{X}_{it}|^{2+\delta}<\infty$, $\Sigma=E[\breve{X}_{it}\breve{X}_{it}^{\tau}]$ is non-singular, where $\breve{X}_{it}=X_{it}-E(X_{it}|Z_{it})$.

(B6) The bandwidth h satisfies $h\to 0$, $Nh^8\to 0$ and $Nh^2/(logN)^2\to\infty$ as $n\to\infty$.

(B7) Σ_1 and Σ_2 are positive definite matrices with all eigenvalues being uniformly bounded away from zero and infinity.

(B8) Let $\varpi_1=\sum_{t=1}^{T}\frac{T-1}{T}(X_{it}-E(X_{it}|U_{it}))(\varepsilon_{it}-\nu_{it}\beta_0)$, $\varpi_2=\sum_{t=1}^{T}\frac{T-1}{T}\nu_{it}\varepsilon_{it}$, $\varpi_3=\sum_{t=1}^{T}\frac{T-1}{T}(\nu_{it}\nu_{it}^{\tau}-\Sigma_\nu)\beta_0$ and ϖ_{sj}, $j=1,\ldots,p$ be the j-th component of ϖ_s. For k of condition (B4), there is a positive constant c such that as $n\to\infty$, $E(||\varpi_s/\sqrt{p}||^k)\leq c$, $s=1,2,3$.

(B9) The $p_\lambda(\cdot)$ satisfy $\underset{j\in\mathcal{B}}{max}\, p'_\lambda(|\beta_{j0}|)=o((np)^{-1/2})$ and $\underset{j\in\mathcal{B}}{max}\, p''_\lambda(|\beta_{j0}|)=o(p^{-1/2})$.

Note that the obove conditions are assumed to hold uniformly in $z\in\mathcal{Z}$. Conditions (B1)–(B9) while look a bit lengthy, are actually quite mild and can be easily satisfied. (B1)–(B2) are standard in the literature on local linear/polynomial estimation. B5 implies $E(\varepsilon_{it}|X_i,Z_i,\mu_i)=E(\varepsilon_{it}|X_{it},Z_{it},\mu_{it})=0$. (B1)–(B5) can be founded in Su and Ullah 2006. (B6) and (B7) have been used in Zhou et al. (2010).

For the convenience and simplicity, let $\vartheta_k=\int z^k K(z)dz$, $c_N=\{log(1/h)/(Nh)\}^{1/2}+h^2$ and $\widetilde{M}_{\widetilde{D}}=\widetilde{D}(\widetilde{D}^{\tau}\widetilde{D})^{-1}\widetilde{D}^{\tau}$

$$a_n=\max_{1\leq j\leq p}\{p'_\lambda(\beta_{j0})|,\beta_{j0}\neq 0\},\quad b_n=\max_{1\leq j\leq p}\{p''_\lambda(\beta_{j0})|,\beta_{j0}\neq 0\},$$
$$B_n=\{\beta:||\beta-\beta_0||\geq d_n\},\quad d_n=n^{-1/3-\delta}+a_n,\quad 0<\delta<1/6$$

Lemma A.1 *Suppose that Assumptions* (B1)–(B6) *hold. Then*

$$G^{\tau}(z,h)W_h(z)G(z,h)=Nf(z)\times\begin{pmatrix}1&0\\0&v_2\end{pmatrix}\{1+O_p(c_N)\},$$
$$G^{\tau}(z,h)W_h(z)X=Nf(z)E(X|Z)\times(1,0)^{\tau}\{1+O_p(c_N)\},$$

Proof Note that

$$G^{\tau}(z,h)W_h(z)G(z,h)$$
$$=\begin{pmatrix}\sum_{i=1}^{n}\sum_{t=1}^{T}K_h(Z_{it}-z) & \sum_{i=1}^{n}\sum_{t=1}^{T}(\frac{Z_{it}-z}{h})K_h(Z_{it}-z)\\ \sum_{i=1}^{n}\sum_{t=1}^{T}(\frac{Z_{it}-z}{h})K_h(Z_{it}-z) & \sum_{i=1}^{n}\sum_{t=1}^{T}(\frac{Z_{it}-z}{h})^2K_h(Z_{it}-z)\end{pmatrix}.$$

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

Lemma A.2 *Suppose that Assumptions* (B1)–(B6) *hold, we have*

$$E|g(Z_{it}) - \sum_{k=1}^{n}\sum_{l=1}^{T} S_{kl}g(Z_{kl})|^2 = O(h^4). \tag{A.1}$$

Proof Similar to the proof of Lemma 5.1 in He et al. (2017). □

Lemma A.3 *Suppose that Assumptions* (B1)–(B6) *hold, we have*

$$\frac{1}{N}\widetilde{W}^{\tau}H\widetilde{W} \xrightarrow{d} \frac{T-1}{T}(\Sigma_2 + \Sigma_{\nu}),$$

where $\Sigma_2 = E\{[X_{11} - E(X_{11}|Z_{11})]^{\tau}[X_{11} - E(X_{11}|Z_{11})]\}$.

Proof By Lemma A.1, we can obtain

$$[I_q \ \ 0_q^{\tau}]^{-1}(G^{\tau}(z,h)W_h(z)G(z,h))^{-1}G^{\tau}(z,h)W_h(z)X = E(X|Z) + O_p(c_N).$$

Then we have

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

and

$$\widetilde{W} = \widetilde{X} + \nu + O_p(c_N) \triangleq A + O_p(c_N).$$

By the law of large numbers, we have

$$\frac{1}{N}\widetilde{W}^{\tau}\widetilde{W} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{[X_{it} - E(X_{it}|Z_{it})]^{\tau}[X_{it} - E(X_{it}|Z_{it})] + \nu_{it}^{\tau}\nu_{it}\right\} \\ + O_p(c_N) \xrightarrow{p} \Sigma_2 + \Sigma_{\nu}. \tag{A.2}$$

Hence, to prove the lemma, we consider the limit of $N^{-1}\widetilde{W}^{\tau}\widetilde{M}_{\widetilde{D}}\widetilde{W}$. It is easy to show that $N^{-1}\widetilde{W}^{\tau}\widetilde{M}_{\widetilde{D}}\widetilde{W} = N^{-1}A^{\tau}\widetilde{M}_{\widetilde{D}}A + O_p(c_N)$. Let $(\widetilde{M}_{\widetilde{D}})_{e_{kl}e_{it}} \triangleq m_{e_{kl}e_{it}}$ and $(A)_{it} \triangleq a_{it} = \widetilde{W}_{it}$, where $e_{kl} = (k-1)T + l$. Then

$$\frac{1}{N}A^{\tau}\widetilde{M}_{\widetilde{D}}A = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{k=1}^{n}\sum_{l=1}^{T} a_{kl}m_{e_{kl}e_{it}}a_{it} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} a_{it}m_{e_{it}e_{it}}a_{it} \\ + \frac{1}{N}\sum_{e_{kl}\neq e_{it}}\sum_{k=1}^{n}\sum_{l=1}^{T} a_{kl}m_{e_{kl}e_{it}}a_{it} \triangleq I_1 + I_2.$$

For the term I_2, we have

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

Note that $(X_{11}, Z_{11}), \ldots, (X_{nT}, Z_{nT})$ are i.i.d. and $E(a_{it}|Z_{it}) = 0$, when $e_{kl} \neq e_{rs}$ and $e_{it} \neq e_{uv}$, we have

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

Using the same argument and $m_{e_{kl}e_{it}} = m_{e_{it}e_{kl}}$, we have

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

By Conditions (B3), we obtain

$$EI_2^2 \leq \frac{2c}{N^2} \sum_{e_{kl} \neq e_{it}} E(m_{e_{kl}e_{it}})^2 \leq \frac{1}{N^2} tr(\widetilde{M}^2) \leq \frac{2c}{N},$$

where c is a constant. Hence

$$EI_2 = o_p(1). \tag{A.3}$$

Note that I_1 can be decomposed as

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

By the definition of S, it is easy to show that

$$S = (S_{11}, \ldots, S_{1T}, S_{21}, \ldots, S_{nT})^{\tau} [I + diag\{O_p(c_n)\}],$$

where

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

Let D_1 is the first column vector of D, thus we have

$$D_1^\tau(I-S^\tau)(I-S)D_1 = \left\{T - 2\sum_{t=1}^{T}\left[\sum_{l=1}^{T}\frac{K_h(Z_{1l}-Z_{1t})}{Nf(Z_{1t})}\right] + \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sum_{l=1}^{T}\frac{K_h(Z_{1l}-Z_{it})}{Nf(Z_{it})}\right]^2\right\}\{1+O_p(c_N)\}.$$

Because

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

and

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

we have

$$D_1^\tau(I-S^\tau)(I-S)D_1 = T\left[1+O_p\left(\frac{1}{Nh}\right)\right].$$

Consider the projection matrix, for $i = 1, \ldots, T$, we obtain

$$\begin{aligned}(\widetilde{M}_{\widetilde{D}_1})_{ii} &= (I-S)D_1[D_1^\tau(I-S^\tau)(I-S)D_1]^{-1}D_1^\tau(I-S^\tau)\\ &= \frac{1}{T}\left[1+O_p\left(\frac{1}{Nh}\right)\right]\left[1-\sum_{l=1}^{T}\frac{K_h(Z_{1l}-Z_{it})}{Nf(Z_{it})}\{1+O_p(c_N)\}\right]^2\\ &= \frac{1}{T}+O_p\left(\frac{1}{Nh}\right).\end{aligned}$$

Because $\widetilde{D}_1$ is the first column vector of $\widetilde{D}$. It is easy to show that $\widetilde{M}_{\widetilde{D}}\widetilde{M}_{\widetilde{D}_1} = \widetilde{M}_{\widetilde{D}_1}\widetilde{M}_{\widetilde{D}} = \widetilde{M}_{\widetilde{D}_1}$. Hence, $\widetilde{M}_{\widetilde{D}} - \widetilde{M}_{\widetilde{D}_1}$ is also a projection matrix. Thus $\widetilde{M}_{\widetilde{D}} - \widetilde{M}_{\widetilde{D}_1} = (\widetilde{M}_{\widetilde{D}} - \widetilde{M}_{\widetilde{D}_1})^2 \geq 0$. We obtain $(\widetilde{M}_{\widetilde{D}})_{ii} \geq (\widetilde{M}_{\widetilde{D}_1})_{ii} = \frac{1}{T} + O_p(\frac{1}{Nh}),\ i = 1, \ldots, T$. By a similar argument, we can show that $(\widetilde{M}_{\widetilde{D}})_{ii} \geq \frac{1}{T} + O_p(\frac{1}{Nh}), i = T+1, \ldots, N$. Thus, we have

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

then, it is easy to show that

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

Hence, we have

$$\Pi_2 = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} m_{e_{it}e_{it}} E(a_{it}a_{it}) = \frac{1}{T}(\Sigma + \Sigma_\eta) + O_p\left(\frac{1}{Nh}\right). \tag{A.4}$$

By (A.2), it is easy to show that

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

By the law of large numbers, Π_1 is bounded as

$$\begin{aligned}\Pi_1 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T}(m_{e_{it}e_{it}} - T^{-1})(a_{it}a_{it} - E(a_{it}a_{it}))\\ &\quad + \frac{1}{NT}\sum_{i=1}^{n}\sum_{t=1}^{T}(a_{it}a_{it} - E(a_{it}a_{it}))\\ &\le \frac{1}{N}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(m_{e_{it}e_{it}} - T^{-1})^2\right]^{1/2} + o_p(1) = o_p(1).\end{aligned} \tag{A.5}$$

By (A.4) and (A.5), we have

$$I_1 = \frac{1}{T}(\Sigma_2 + \Sigma_\eta) + o_p(1). \tag{A.6}$$

By (A.2), (A.3) and (A.6), the lemma holds. □

Lemma A.4 *Under the conditions of Theorem 2.1, if β is the true value of the parameter, we have*

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \nu_{it} H\widetilde{g}_{it} = o_p(N^{1/2}), \tag{A.7}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \varepsilon_{it} H\widetilde{g}_{it} = o_p(N^{1/2}), \tag{A.8}$$

Proof Since the proof of (A.8) is similar of (A.7), we prove only (A.7)here. Let $\zeta_N = N^{1/2}/log(N)$,

$$\begin{aligned}P\left(|\sum_{i=1}^{n}\sum_{t=1}^{T}\nu_{it}H\widetilde{g}_{it}| > \zeta_N\right) &\le P\left(|\sum_{i=1}^{n}\sum_{t=1}^{T}\nu_{it}H\widetilde{g}_{it}| > \zeta_N, max_{i,t}|\widetilde{g}_{it}| \le ch^4\right)\\ &\quad + P(max_{i,t}|\widetilde{g}_{it}| \ge ch^4)\end{aligned} \tag{A.9}$$

The second term is $o_p(1)$ by Lemma A.2. For the first term, let R_{it} be the event that $|\widetilde{g}_{it}| \leq ch^4$. Then

$$
\begin{aligned}
&P\left(|\sum_{i=1}^{n}\sum_{t=1}^{T}\nu_{it}H\widetilde{g}_{it}| > \zeta_N, \{I(R_{it})=1, \forall i, t\}\right)\\
&\leq \zeta_N^{-2}\sum_{i=1}^{n}\sum_{t=1}^{T}E[\nu_{it}H\widetilde{g}_{it}\{I(R_{it})=1\}]^2\\
&\quad + \zeta_N^{-2}\sum_{i\neq k}^{n}\sum_{t\neq s}^{T}E[\nu_{it}H\widetilde{g}_{it}\nu_{ks}\Omega\widetilde{g}_{ks}\{I(R_{ks})=1\}].
\end{aligned}
$$

Since $\widetilde{g}_{it}\{I(R_{it})=1\} \leq ch^4$ is independent of ν_{it}, the first term is $O\{N\zeta_N^{-2}c^2h^8\} = o(1)$. The second term is easily seen to equal zero. □

Lemma A.5 *Under the conditions of Theorem 2.1, if β_0 is the true value of the parameter, we have*

$$
\max_{1\leq i\leq n} ||\Gamma_i(\beta_0)|| = o_p(\sqrt{n/p}), \tag{A.10}
$$

$$
\frac{\{N^{-1/2}\sum_{i=1}^{n}\Gamma_i^{\tau}(\beta_0)\}\Sigma_1^{-1}\{N^{-1/2}\sum_{i=1}^{n}\Gamma_i(\beta_0)\}-p}{\sqrt{2p}} \xrightarrow{d} N(0,1). \tag{A.11}
$$

Proof From the definition of $\Gamma_i(\beta)$ by (2.7), and a simple calculation, yields

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\Gamma_i(\beta_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\big\{&\widetilde{X}_{it}H\widetilde{g}_{it} - \widetilde{X}_{it}H\widetilde{\nu}_{it}\beta_0 + \widetilde{X}_{it}H\widetilde{\varepsilon}_{it}\\
&+\widetilde{\nu}_{it}H\widetilde{g}_{it} - \widetilde{\nu}_{it}H\widetilde{\nu}_{it}^{\tau}\beta_0 + \widetilde{\nu}_{it}H\widetilde{\varepsilon}_{it} + \Sigma_\nu\beta_0\big\}.
\end{aligned}
$$

By Lemma A.1, we have $S\varepsilon = O_p(c_N)$. Similar to the proof of Lemma A.3 and under Assumption (B7), we have $\frac{1}{\sqrt{N}}\widetilde{X}^{\tau}HS\varepsilon = O(\sqrt{N}c_N^2) = o_p(1)$. Therefore

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}H\widetilde{\varepsilon}_{it} = \frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{T-1}{T}(X_{it}-E(X_{it}|Z_{it}))\varepsilon_{it} \tag{A.12}
$$

Similar to the proofs of (A.12), we can derive that

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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{\nu}_{it}H\widetilde{\varepsilon}_{it}=\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{T-1}{T}\nu_{it}\varepsilon_{it},$$

which combining with Lemma A.4, it is easy to obtain

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\Gamma_i(\beta_0)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{T-1}{T}\big\{[(X_{it}-E(X_{it}|U_{it}))(\varepsilon_{it}-\nu_{it}\beta_0)]+\nu_{it}\varepsilon_{it}\\
&\quad+(\Sigma_\nu-\nu_{it}\nu_{it}^{\tau})\beta_0]\big\}+o_p(1)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{n}G_i(\beta_0)+o_p(1).
\end{aligned}\tag{A.13}$$

Therefore, we have

$$\begin{aligned}
\Gamma_i(\beta_0)&=\frac{T-1}{T}\sum_{t=1}^{T}\big\{[(X_{it}-E(X_{it}|U_{it}))(\varepsilon_{it}-\nu_{it}\beta_0)]+\nu_{it}\varepsilon_{it}\\
&\quad+(\Sigma_\nu-\nu_{it}\nu_{it}^{\tau})\beta_0]\big\}+o_p(1)\\
&=\sum_{t=1}^{T}\left\{\frac{T-1}{T}(X_{it}-E(X_{it}|U_{it}))(\varepsilon_{it}-\nu_{it}\beta_0)+\frac{T-1}{T}\nu_{it}\varepsilon_{it}\right.\\
&\quad\left.+\frac{T-1}{T}(\nu_{it}\nu_{it}^{\tau}-\Sigma_\nu)\beta_0\right\}+o_p(1)=\varpi_1+\varpi_2+\varpi_3+o_p(1).
\end{aligned}\tag{A.14}$$

Let $\varpi_s^*=\max_{1\le i\le n}||\varpi_{si}||$, $s=1,2,3$, and $\{\varpi_{si},i=1,\ldots,n\}$ is a sequence of independent random variables with common distribution. for any $\varepsilon\ge 0$, then

$$\begin{aligned}
P\{\varpi_1^*\ge(p)^{1/2}n^{1/k}\epsilon\}&\le\sum_{i=1}^{n}P\{||\varpi_{1i}||\ge(p)^{1/2}n^{1/k}\epsilon\}\\
&\le\frac{1}{np^{k/2}\epsilon^k}\sum_{i=1}^{n}E||\varpi_{1i}||^k\\
&\le\frac{1}{\epsilon^k}E||\varpi_{11}/p^{1/2}||^k.
\end{aligned}$$

From conditions (B4) and (B7) and Cauchy-Schwarz inequality yields that $\varpi_1^*=o_p(\sqrt{p}n^{1/k})$. By the condition $p=o(n^{(k-2)/(2k)}$ in Theorem 2.1, it is easy to check that $\varpi_1^*=o_p(\sqrt{n/p}n^{2-k/(2k)}p)=o_p(\sqrt{n/p})$. Similar to the proof, we obtain $\varpi_2^*=o_p(\sqrt{n/p})$ and $\varpi_3^*=o_p(\sqrt{n/p})$. Then $max_{1\le i\le n}||\Gamma_i(\beta_0)||=o_p(\sqrt{n/p})$.

By applying the martingale central limit theorem as give in Hall and Heyde (1980) and (A.13), it is easy to obtain (A.11). The proof of Lemma A.5 is thus completed. □

Lemma A.6 *Under the conditions of Theorem 2.1. Denote* $D_n = \{\beta : ||\beta - \beta_0|| \leq ca_n\}$ *Then* $||\gamma(\beta)|| = O_p(a_n)$*, for* $\beta \in D_n$.

Proof For $\beta \in D_n$, let $\gamma(\beta) = \rho\theta$, where $\rho \geq 0$, $\theta \in R^p$, and $||\theta|| = 1$. Set

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

From (2.7), we can obtain

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

Then

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

Observe that

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

Let $\mathcal{X}_{it} = \widetilde{W}_{it}H\widetilde{W}_{it}$ According to Condition (B.7) and Minkowski inequality, we have

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

Then we obtain that

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

which combining with (A.11)

$$\Gamma^*(\beta) = o_p(n/p) \tag{A.15}$$

For $\bar{\Gamma}(\beta)$, it is easy to see that

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

Similar to the proofs of (A.10) in Fan et al. (2016), we obtain

$$\theta^\tau \bar{\Gamma}(\beta) = O_p(a_n) \tag{A.16}$$

Therefore, it follows from (A.15) and (A.16), we have $max_{1\le i\le n}||\Gamma_i(\beta)|| = n^{-1}|\sum_{i=1}^{n}$.

From (2.7), similar to the proof (A.11) in Fan et al. (2016) and Lemma B.4 in Li et al. (2012), we can derive $tr[(J(\beta_0)-\Sigma_1)^2] = O_p(p^2(c_n^4+1/n))$ which means that all the eigenvalues of $J(\beta_0)$ converge to those of Σ_1 at the rate of $O_p(p^2(c_n^4+1/n))$. Therefore, by Lemma A.2, (2.7), (A.16), together with Condition (B7), we have

$$\begin{aligned}
J(\beta) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\big\{\widetilde{W}_{it}H(\widetilde{Y}_{it}-\widetilde{W}_{it}\beta_0)-(T-1)\Sigma_\nu\beta_0-\widetilde{W}_{it}H\widetilde{W}_{it}^\tau(\beta-\beta_0)\\
&\quad+(T-1)\Sigma_\nu(\beta-\beta_0)\big\}^{\oplus 2}\\
&= J(\beta_0)+\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\big\{\widetilde{W}_{it}H\widetilde{W}_{it}^\tau(\beta-\beta_0)+(T-1)\Sigma_\nu(\beta-\beta_0)\big\}^{\oplus 2}\\
&\quad-\frac{2}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\big\{\widetilde{W}_{it}H(\widetilde{Y}_{it}-\widetilde{W}_{it}\beta_0)-(T-1)\Sigma_\nu\beta_0\big\}\big\{\widetilde{W}_{it}H\widetilde{W}_{it}^\tau(\beta-\beta_0)\\
&\quad+(T-1)\Sigma_\nu(\beta-\beta_0)\big\}\\
&= \Sigma_1+O_p(p^2(c_n^4+1/n)).
\end{aligned} \tag{A.17}$$

we can obtain $\theta^\tau J(\beta)\theta = \theta^\tau\Sigma_1\theta \xrightarrow{P} c$. Therefore, we obtain $\rho \le c|\theta^\tau\bar{\Gamma}(\beta)| = O_p(a_n)$, then $||\gamma(\beta)|| = O_p(a_n)$. □

Lemma A.7 *Under the conditions of Theorem 2.1. as $n\to\infty$, with probability tending to 1, $R_n(\beta)$ has a minimum in D_n.*

Proof For $\beta\in D_n$,

$$H_{1n}(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}\frac{\Gamma_i(\beta)}{1+\gamma^\tau\Gamma_i(\beta)} = 0$$

According to Lemma A.6, we have $\gamma^\tau\Gamma_i(\beta) = o_p(1)$. Apply Taylor expansion to $H_{1n}(\beta,\gamma)$, we obtain $\bar{\Gamma}(\beta)-J(\beta)\gamma+\delta_n = 0$, where $\bar{\Gamma}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta)$, $\delta_n = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta)(\gamma^\tau\Gamma_i(\beta))^2/[1+\zeta_i]^3$ for some $|\zeta_i| \le |\gamma^\tau\Gamma_i(\beta)|$. We have

$\gamma = J(\beta)^{-1}\bar{\Gamma}(\beta) + J(\beta)^{-1}\delta_n$. Substituting γ into (2.6), it is easy to see that

$$2R_n(\beta) = n\bar{\Gamma}(\beta)^\tau J(\beta)^{-1}\bar{\Gamma}(\beta) - n\delta_n^\tau J(\beta)^{-1}\delta_n + \frac{2}{3}\sum_{i=1}^{n}(\gamma^\tau\Gamma_i(\beta))^3(1+\zeta_i)^{-4} \tag{A.18}$$

For $\beta \in \partial D_n$, where ∂D_n denotes the boundary of D_n, we write $\beta = \beta_0 + ca_n\phi$ where ϕ is a unit vector, we have a decomposition as $2R_n(\beta) = \Pi_0 + \Pi_1 + \Pi_2$, where $\Pi_0 = n\bar{\Gamma}(\beta_0)^\tau\Sigma_1^{-1}\bar{\Gamma}(\beta_0)$, $\Pi_1 = n(\bar{\Gamma}(\beta) - \bar{\Gamma}(\beta_0))^\tau J(\beta)^{-1}(\bar{\Gamma}(\beta) - \bar{\Gamma}(\beta_0))$, $\Pi_2 = n[\bar{\Gamma}(\beta_0)^\tau(J(\beta)^{-1} - \Sigma_1^{-1})\bar{\Gamma}(\beta_0) + 2\bar{\Gamma}(\beta_0)^\tau J(\beta)^{-1}(\bar{\Gamma}(\beta) - \bar{\Gamma}(\beta_0)] - n\delta_n^\tau J(\beta)^{-1}\delta_n + \frac{2}{3}\sum_{i=1}^{n}(\gamma^\tau\Gamma_i(\beta))^3(1+\zeta_i)^{-4}$ As $n \to \infty$, we see that

$$\begin{aligned}\Pi_1 &= n\left\{\left[\frac{1}{n}\widetilde{W}^\tau H\widetilde{W} - (T-1)\Sigma_\nu\right](\beta-\beta_0)\right\}^\tau J(\beta)^{-1}\left\{\left[\frac{1}{n}\widetilde{W}^\tau H\widetilde{W}\right.\right.\\ &\quad \left.\left. - (T-1)\Sigma_\nu\right](\beta-\beta_0)\right\}\\ &= c^2na_n^2\phi^\tau\Sigma_0\Sigma_1^{-1}\Sigma_0\phi\{1+o_p(1)\} = O_p(na_n^2),\end{aligned}$$

$\Pi_2/\Pi_1 \xrightarrow{P} 0$ and $2R_n(\beta_0) - \Pi_0 = o_p(1)$. This implies that for any c given, as $n \to \infty$, $Pr\{2[R_n(\beta) - R_n(\beta_0)] \geq c\} \to 1$. In addition, note that for n large,

$$\begin{aligned}\mathcal{L}_n(\beta) - \mathcal{L}_n(\beta_0) &= R_n(\beta) - R_n(\beta_0) + n\sum_{j=1}^{p}\{p_\lambda(|\beta_j|) - p_\lambda(|\beta_{j0}|)\\ &\geq R_n(\beta) - R_n(\beta_0) + n\sum_{j\in\mathcal{B}}\{p_\lambda(|\beta_j|) - p_\lambda(|\beta_{j0}|) \geq R_n(\beta) - R_n(\beta_0),\end{aligned}$$

where the last inequality holds due to Conditions (B9) and the unbiased property of the SCAD penalty so that $j \in \mathcal{B}$, $p_\lambda(|\beta_j|) = p_\lambda(|\beta_{j0}|)$ when n is large. Hence, $Pr\{\mathcal{L}_n(\beta) \geq \mathcal{L}_n(\beta_0)\} \to 1$ for $\beta \in \partial D_n$, which establishes Lemma A.7. □

Proof of Theorem 2.1 Let $U_i = \gamma^\tau\Gamma_i(\beta_0)$. Apply Taylor expansion to (2.10), we have

$$0 = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta_0)\left(1 - U_i + \frac{U_i^2}{1+U_i}\right) = \bar{\Gamma}(\beta_0) - J(\beta_0)\gamma + \delta_n, \tag{A.19}$$

where $\delta_n = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta_0)U_i^2 - \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta_0)\frac{U_i^3}{1+U_i}$.

From (A.11) and Lemma A.6, we have

$$\max_{1\leq i\leq n}|U_i| \leq ||\gamma(\beta)|| \max_{1\leq i\leq n}||\Gamma_i(\beta_0)|| = O_p(p/n^{1/2-1/r}).$$

Similar to the proof of (A.19) in Li et al. (2012), we can get $||\delta_n|| = o_p(p^{5/2}n^{-1}(n^{-1/2} + c_n^2)) + o_p(p^2n^{-1}c_n)$. From (A.19), we obtain that $\gamma = J(\beta_0)^{-1}\bar{\Gamma}(\beta_0) + J(\beta_0)^{-1}\delta_n$. Taylor expansion implies $\ln(1+U_i) = U_i - U_i^2/2 + U_i^3/3(1+\varsigma_i)^4$, for some ς_i

such that $|\varsigma_i| \le |U_i|$. Therefore, combining (A.16) and some elementary calculation, we have

$$\begin{aligned} 2R_n(\beta_0) &= 2\sum_{i=1}^n \ln\{1+U_i\} = n\bar{\Gamma}^\tau(\beta_0)J(\beta_0)^{-1}\bar{\Gamma}(\beta_0) - n\delta_n^\tau J(\beta_0)^{-1}\delta_n \\ &\quad + \frac{2}{3}\mathcal{R}_n\{1+o_p(1)\} \\ &= n\bar{\Gamma}^\tau(\beta_0)\Sigma_1^{-1}\bar{\Gamma}(\beta_0) + n\bar{\Gamma}^\tau(\beta_0)(J(\beta_0)^{-1} - \Sigma_1^{-1})\bar{\Gamma}(\beta_0) - n\delta_n^\tau J(\beta_0)^{-1}\delta_n^\tau \\ &\quad + \frac{2}{3}\mathcal{R}_n\{1+o_p(1)\} \end{aligned} \tag{A.20}$$

where $\mathcal{R}_n = \sum_{i=1}^n[\gamma^\tau\Gamma_i(\beta_0)]^3$, By using the proving method of (A.22) and Lemma B.6 in Li et al. (2012), we can easily derive $n\delta_n^\tau J(\beta_0)^{-1}\delta_n = o_p(\sqrt{p})$ and $n\bar{\Gamma}^\tau(\beta_0)(J(\beta_0)^{-1} - \Sigma_1^{-1})\bar{\Gamma}(\beta_0) = o_p(\sqrt{p})$. The proof of Theorem 2.1 is concluded from the above results together with (A.11). □

Proof of Theorem 2.2 Let $H_{1n}(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^n \frac{\Gamma_i(\beta)}{1+\gamma^\tau\Gamma_i(\beta)}$ and $H_{2n}(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^n \frac{\Gamma_i(\beta)}{1+\gamma^\tau\Gamma_i(\beta)}(\frac{\partial\Gamma_i(\beta)}{\partial\beta}^\tau)^\tau\gamma$. Note that $\hat{\beta}$ and $\hat{\gamma}$ satisfy $H_{1n}(\hat{\beta},\hat{\gamma}) = 0$ and $H_{2n}(\hat{\beta},\hat{\gamma}) = 0$ Let $\varphi = (\beta^\tau,\gamma^\tau)^\tau$, $\varphi_0 = (\beta_0^\tau, 0)^\tau$ and $\hat{\varphi}_0 = (\hat{\beta}_0^\tau,\hat{\gamma}_0^\tau)^\tau$. Then by

$$0 = H_{jn}(\hat{\beta},\hat{\gamma}) = H_{jn}(\beta_0^\tau,0) + \frac{\partial\Gamma_i(\beta,0)}{\partial\beta}(\hat{\beta}-\beta_0) + \frac{\partial\Gamma_i(\beta,0)}{\partial\gamma}(\hat{\gamma}-0) + \delta_{jn},$$

where δ_{jn} with $\delta_{jn} = \frac{1}{2}(\hat{\varphi}_0-\varphi_0)^\tau H''_{jn}(\varphi)(\hat{\varphi}_0-\varphi_0)$ for $j=1,2$. Here $H''_{jn}(\varphi)$ denotes the Hessian matric of $H_{jn}(\varphi)$. Then

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\beta}-\beta_0 \end{pmatrix} = \begin{bmatrix} \frac{\partial H_{1n}(\beta,\gamma)}{\partial\gamma} & \frac{\partial H_{1n}(\beta,\gamma)}{\partial\beta} \\ \frac{\partial H_{2n}(\beta,\gamma)}{\partial\gamma} & \frac{\partial H_{2n}(\beta,\gamma)}{\partial\beta} \end{bmatrix}^{-1}_{(\beta_0,0)} \begin{pmatrix} H_{1n}(\beta_0,0)+\delta_{1n} \\ H_{2n}(\beta_0,0)+\delta_{2n} \end{pmatrix},$$

from Lemma A.3, we have

$$n^{-1}\sum_{i=1}^n \frac{\partial\Gamma_i(\beta,0)}{\partial\beta} = \frac{1}{n}\widetilde{W}^\tau H\widetilde{W} - (T-1)\Sigma_\nu \xrightarrow{d} (T-1)\Sigma_2\{1+o_p(1)\} = \Sigma_0\{1+o_p(1)\}. \tag{A.21}$$

Note that $||\hat{\gamma}(\beta)|| = O_p(a_n)$ by Lemma A.6 and $||\hat{\beta}-\beta_0|| = O_p(a_n)$ by Lemma A.7. Then using the Cauchy-Schwarz inequality, we find

$$||\delta_{1n}||^2 = \sum_{k=1}^p \delta_{1n,k} \le c\sum_{k=1}^p ||\hat{\varphi}_0-\varphi_0||^4||H''_{jn}(\varphi)||^2 = O_p(a_n^4p^3). \tag{A.22}$$

and Condition (B9) yields that

$$||\sqrt{N}\Sigma_0^{-1}\delta_{1n}||^2 \le N\Sigma_0^{-2}||\delta_{1n}|| = O_p(Na_n^4p^3) = o_p(1).$$

and combining with $H_{1n}(\beta_0, 0) = n^{-1}\sum_{i=1}^{n}\Gamma_i(\beta_0)$, we have

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \Sigma_0^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta_0) + o_p(1)$$
$$= \Sigma_0^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{T-1}{T}\big\{[(X_{it} - E(X_{it}|U_{it}))(\varepsilon_{it} - \nu_{it}\beta_0)] + \nu_{it}\varepsilon_{it}$$
$$+ (\Sigma_\nu - \nu^\tau\nu_{11})\beta_0]\big\} + o_p(1) = \Sigma_0^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\beta_0) + o_p(1).$$

Note that

$$Cov(\Gamma_i(\beta_0)) = (T-1)\big\{E[(X_{11} - E(X_{11}|U_{11}))(\varepsilon_{11} - \nu_{11}\beta_0)]^2 + E[\nu_{11}^\tau\varepsilon_{11}]^2$$
$$+ E[(\Sigma_\nu - \nu^\tau\nu_{11})\beta_0]^2\big\}.$$

Therefore,

$$\lim_{n\to\infty} Cov(\Gamma_i(\beta_0)) = (T-1)\big\{E(\varepsilon_{11} - \nu_{11}\beta_0)^2\Sigma + \sigma^2\Sigma_\nu$$
$$+ E[(\nu_{11}\nu_{11}^\tau - \Sigma_\nu)\beta_0]^2\big\} = \Sigma_1.$$

Invoking the Slutsky theorem and the central limit theorem, we can prove Theorem 2.2. □

Proof of Theorem 2.3 From the Lemma A.7, we note that the minimizer of $\mathcal{L}_n(\beta)$ is in $\mathcal{D}_n$. Considering $\beta \in \mathcal{D}_n$, we have that for each of its components

$$\frac{1}{n}\frac{\partial\mathcal{L}_n(\beta)}{\partial\beta_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{\gamma\left[\sum_{t=1}^{T}\widetilde{W}_{it}H\widetilde{W}_{it} + (T-1)\Sigma_\nu\right]}{1+\gamma^\tau\Gamma_i(\beta)} + p'_\lambda(|\beta_j|)sign(\beta_j) = I_j + II_j, \tag{A.23}$$

First, $max_{j\in\mathcal{B}}|I_j| \le \gamma\Sigma_j(1+o_p(1)) = O_p(a_n)$, because $\gamma^\tau\Gamma_i(\beta) = o_p(1)$,where Σ_j denotes the jth column of Σ. as $\tau(n/p)^{1/2} \to \infty$. $Pr(max_{j\in\mathcal{B}}|I_j| > \tau/2) \to 0$. it can be seen that $p'_\lambda(|\beta_j|)sign(\beta_j)$ dominates the sign of $\frac{\partial\mathcal{L}_n(\beta)}{\partial\beta_i}$ asymptotically for all $j \notin \mathcal{B}$, as $n \to \infty$, for any $j \notin \mathcal{B}$, with probability tending to 1,

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

which implies that $\hat{\beta}_j = 0$ for all $j \notin \mathcal{B}^c$, with probability tending to 1. Thus part (a) of Theorem 2.3 follows.

Next, we establish part (b), Let Ψ_1 and Ψ_2 be matrices such that $\Psi_1\beta = \beta_1$ and $\Psi_2\beta = \beta_2$. As we have shown that as $n \to \infty$, $Pr(\hat{\beta}_2 = 0) \to 1$, thus by the Lagange multiplier method, finding the minimizer of $\mathcal{L}_n(\beta)$ is asymptotic equivalent to solve the minimization of the following objective function

$$\frac{1}{n}\sum_{i=1}^{n}\log\{1+\gamma^{\tau}\Gamma_i(\beta)\}+n\sum_{j=1}^{p}p_{\lambda}(|\beta_j|)+v^{\tau}\Psi_2\beta, \tag{A.24}$$

where v is $p-s$ dimensional column vector of an other Lagrange multiplier. Define

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

and $\tilde{Q}_{3n}(\beta,\gamma,v)=\Psi_2\beta$. where

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

The minimizer (β,γ,v) of (A.24) satisfies $\tilde{Q}_{in}(\beta,\gamma,v)=0,\ (i=1,2,3)$. Since $||\gamma||=O_p(a_n)$ for $\beta\in\mathcal{B}$, we can obtain that $||v||=O_p(a_n)$ from $\tilde{Q}_{2n}(\beta,\gamma,v)=0$, In order to expand $\tilde{Q}_{in}(\beta,\gamma,v)(i=1,2,3)$ around the value $(\beta_0,0,0)$, we first give the following partial derivatives,

$$\frac{\partial\tilde{Q}_{1n}(\beta_0,0,0)}{\partial\gamma}=-J(\beta_0),\ \frac{\partial\tilde{Q}_{1n}(\beta_0,0,0)}{\partial\beta}=\Sigma(\beta_0),\ \frac{\partial\tilde{Q}_{1n}(\beta_0,0,0)}{\partial v}=0,$$
$$\frac{\partial\tilde{Q}_{2n}(\beta_0,0,0)}{\partial\gamma}=\Sigma(\beta_0),\ \frac{\partial\tilde{Q}_{2n}(\beta_0,0,0)}{\partial\beta}=b'(\beta),\ \frac{\partial\tilde{Q}_{2n}(\beta_0,0,0)}{\partial v}=\Psi_2^{\tau},$$
$$\frac{\partial\tilde{Q}_{3n}(\beta_0,0,0)}{\partial\gamma}=0,\ \frac{\partial\tilde{Q}_{3n}(\beta_0,0,0)}{\partial\beta}=\Psi_2,\ \frac{\partial\tilde{Q}_{3n}(\beta_0,0,0)}{\partial v}=0.$$

Then by Taylor expansion, we immediately derive that

$$\begin{pmatrix}\tilde{Q}_{1n}(\beta_0,0,0)\\0\\0\end{pmatrix}=\begin{pmatrix}-\Sigma_1 & \Sigma_0 & 0\\ \Sigma_0^{\tau} & 0 & \Psi_2^{\tau}\\ 0 & \Psi_2 & 0\end{pmatrix}\begin{pmatrix}\tilde{\gamma}\\ \tilde{\beta}\\ \tilde{v}\end{pmatrix}+R_n, \tag{A.25}$$

where $\Sigma(\beta_0)=n^{-1}\sum_{i=1}^{n}\partial\Gamma_i(\beta_0)/\partial\beta$, $R_n=\sum_{l=1}^{5}R_n^{(l)}$, $R_n^{(1)}=(R_{1n}^{\tau 1},R_{2n}^{\tau 1},0)^{\tau}$, $R_{jn}^{\tau 1}\in R^p$ and the kth component of $R_{jn}^{\tau 1}$ for $j=1,2$ is given by

$$R_{jn,k}^{(1)}=\frac{1}{2}(\hat{\vartheta}-\vartheta_0)^{\tau}\frac{\partial^2\tilde{Q}_{jn,k}(\tilde{\vartheta})}{\partial\vartheta\partial\vartheta^{\tau}}(\hat{\vartheta}-\vartheta_0),$$

$\vartheta = (\beta, \gamma)^\tau$, $\tilde{\vartheta} = (\tilde{\beta}, \tilde{\gamma})^\tau$ satisfying $||\tilde{\vartheta} - \vartheta_0|| \le ||\hat{\vartheta} - \vartheta_0||$. $R_n^{(2)} = (0, b^\tau(\beta_0), 0)^\tau$, $R_n^{(3)} = [0, \{b'(\tilde{\vartheta})(\hat{\vartheta} - \vartheta_0)\}, 0]^\tau$, $R_n^{(4)} = [\{(J(\beta_0) - \Sigma_1))\hat{\gamma}\}^\tau + (\Sigma(\beta_0) - \Sigma_0)(\hat{\beta} - \beta)\}^\tau, 0, 0]^\tau$ and $R_n^{(5)} = [0, \{(\Sigma(\beta_0) - \Sigma_0)\hat{\gamma}\}^\tau, 0]^\tau$. Similar to the proof of (A.22), we can get $R_n^{(1)} = o_p(n^{-1/2})$. Given Condition (B8) and (B9), we see that $R_n^{(2)} = o_p(n^{-1/2})$ and $R_n^{(3)} = o_p(n^{-1/2})$. By (A.21) and (A.17) which together with Lemma A.6 yields that $R_n^{(4)} = o_p(n^{-1/2})$ and $R_n^{(5)} = o_p(n^{-1/2})$. Hence, we can get $R_n^{(k)} = o_p(n^{-1/2})$, $k = 1, \ldots, 5$.

Define $K_{11} = -\Sigma_1$, $K_{12} = [\Sigma_0, \ \ 0]$ and $K_{21} = K_{12}^\tau$,

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

and let $\kappa = (\beta^\tau, v^\tau)^\tau$. Then by inverted (A.25), we find

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\kappa} - \kappa_0 \end{pmatrix} = K^{-1}\left\{ \begin{pmatrix} -\tilde{Q}_{1n}(\beta_0, 0, 0) \\ 0 \end{pmatrix} + R_n \right\}, \tag{A.26}$$

As matrix K is partitioned into four blocks, it can be inverted blockwise as follows

$$K^{-1} = \begin{bmatrix} K_{11}^{-1} + K_{11}^{-1}K_{12}A^{-1}K_{21}K_{11}^{-1} & -K_{11}^{-1}K_{12}A^{-1} \\ -A^{-1}K_{21}K_{11}^{-1} & A^{-1} \end{bmatrix},$$

where $A = K_{22} - K_{21}K_{11}^{-1}K_{12} = \begin{bmatrix} \Omega^{-1} & \Psi_2^\tau \\ \Psi_2 & 0 \end{bmatrix}$ and Σ is defined in Theorem 2.2. Thus, we get

$$\hat{\kappa} - \kappa_0 = A^{-1}K_{21}K_{11}^{-1}\tilde{Q}_{1n}(\beta_0, 0, 0) + o_p(n^{-1/2}).$$

Matric A can also be inverted blockwise by using the analytic inversion formula,ie.,

$$A^{-1} = \begin{bmatrix} \Omega - \Omega\Psi_2^\tau(\Psi_2\Omega\Psi_2^\tau)^{-1}\Psi_2\Omega & \Omega\Psi_2^\tau(\Psi_2\Omega\Psi_2^\tau)^{-1} \\ (\Psi_2\Omega\Psi_2^\tau)^{-1}\Psi_2\Omega & -(\Psi_2\Omega\Psi_2^\tau)^{-1} \end{bmatrix},$$

Further, we have

$$\hat{\beta} - \beta_0 = [\Omega - \Omega\Psi_2^\tau(\Psi_2\Omega\Psi_2^\tau)^{-1}\Psi_2\Omega](\Sigma_0\Sigma_1\bar{\Gamma}(\beta_0) + o_p(n^{-1/2})).$$

It follows by an expansion of $\hat{\beta}_1$ that

$$\hat{\beta}_1 - \beta_0 = [\Psi_1\Omega - \Psi_1\Omega\Psi_2^\tau(\Psi_2\Omega\Psi_2^\tau)^{-1}\Psi_2\Omega](\Sigma_0\Sigma_1\bar{\Gamma}(\beta_0) + o_p(n^{-1/2})). \tag{A.27}$$

Then similar to the proof of Theorem 2.3 in Fan et al. (2016), we have $n^{1/2}W_n\Omega_p^{-1/2}$ $(\widehat{\beta}_1 - \beta_{10}) \xrightarrow{d} N(0, G)$, which completes the proof of Theorem 2.3. □

Proof of Theorem 2.4. Let $\hat{\beta}$ be the minimizer (2.13) and $U_i = \hat{\gamma}^\tau \Gamma_i(\beta)$. Taylor expansion gives

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

where $|\xi_i| \to |U_i|$ and $o_p(1)$ is due to the penalty function. From (A.26), we have $\hat{\gamma} = [\Sigma_1^{-1} + \Sigma_1^{-1}\Sigma_0\{\Omega - \Omega\Psi_2^\tau(\Psi_2\Omega\Psi_2^\tau)^{-1}\Psi_2\Omega\}\Sigma_0^\tau\Sigma_1^{-1}][\bar{\Gamma}(\beta_0) + o_p(n^{-1/2})]$.

Similar to Tang and Leng (2010), Substituting the expansion of $\hat{\gamma}$ and $\hat{\beta}$ given by (A.24) into U_i, we show that

$$2\mathcal{L}_n(\hat{\beta}) = n\bar{\Gamma}(\beta_0)^\tau\Psi_2^\tau(\Psi_2\Omega^{-1}\Psi_2^\tau)^{-1}\Psi_2\bar{\Gamma}(\beta_0) + o_p(1) \tag{A.28}$$

Under the null hypothesis, because $L_n L_n^\tau = I_q$, there exists $\tilde{\Psi}_2$ such that $\tilde{\Psi}_2\beta = 0$ and $\tilde{\Psi}_2\tilde{\Psi}_2^\tau = I_{p-d+q}$. Now by repeating the proof of Theorem 2.3, we establish that under the null hypothesis, the estimation of β can be obtained by minimizing (A.27), where Ψ_2 is replaced by $\tilde{\Psi}_2$, we can easily obtain that

$$2\mathcal{L}_n(\tilde{\beta}) = n\bar{\Gamma}(\beta_0)^\tau\tilde{\Psi}_2^\tau(\tilde{\Psi}_2\Omega^{-1}\tilde{\Psi}_2^\tau)^{-1}\tilde{\Psi}_2\bar{\Gamma}(\beta_0) + o_p(1).$$

Combining Eqs. (A.28), we have

$$\mathcal{L}_n = n\bar{\Gamma}(\beta_0)^\tau\Omega^{-1/2}(P_1 - P_2)\Omega^{-1/2}\bar{\Gamma}(\beta_0) + o_p(1).$$

where

$$P_1 = \Omega^{-1/2}\Psi_2^\tau(\Psi_2\Omega^{-1}\Psi_2^\tau)^{-1}\Psi_2\Omega^{-1/2},$$

and

$$P_2 = \Omega^{-1/2}\tilde{\Psi}_2^\tau(\tilde{\Psi}_2\Omega^{-1}\tilde{\Psi}_2^\tau)^{-1}\tilde{\Psi}_2\Omega^{-1/2},$$

are two idempotent matrices. As the rank of $P_1 - P_2$ is q, $P_1 - P_2$ can be written as $\Upsilon^\tau\Upsilon$, where Υ is $q \times p$ matrix such that $\Upsilon^\tau\Upsilon = I_q$, further, we see that

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

Then

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

and the proof of Theorem 2.4 is finished. □

Lemma A.8 *Under the conditions of Theorem 2.5. For a given z, if g(z) is the true value of the parameter, then*

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\Xi}_i\{g(z)\}-b(z)\xrightarrow{d}N(0,R). \tag{A.29}$$

$$\frac{1}{Nh}\sum_{i=1}^{n}\hat{\Xi}_i\{g(z)\}\hat{\Xi}_i^{\tau}\{g(z)\}\xrightarrow{P}R. \tag{A.30}$$

$$\max_{1\le i\le n}||\hat{\Xi}_i\{g(z)\}||=o_p(\sqrt{Nh}),\ \phi=O_p(N^{-1/2}). \tag{A.31}$$

where $b(z)=\left(\frac{N}{h}\right)^{1/2}\frac{T-1}{T}E[g(Z_{it})-g(z)]f(z)\int K(z)dz$ *and* $R=\sigma^2 f(z)\int K^2(z)dz$.

Proof Observe that

$$\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\Xi}_i\{g(z)\}-b(z)=S_1(z)+S_2(z)+S_3(z)$$

where

$$S_1(z)=\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{t=1}^{T}(I_{it}-Q)K_h(Z_{it}-z)\varepsilon_{it},$$

$$S_2(z)=\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{t=1}^{T}(I_{it}-Q)\{[g(Z_{it})-g(z)]K_h(Z_{it}-z)-(h/N)^{1/2}b(z)\},$$

$$S_3(z)=\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\sum_{t=1}^{T}(I_{it}-Q)X_{it}^{\tau}(\beta-\hat{\beta})K_h(Z_{it}-z).$$

It is not difficult to prove $E[S_1(z)]=0$ and $Var[S_1(z)]=R+o(1)$. $S_1(z)$ satisfies the conditions of the Cramer–Wold theorem and the Lindeberg condition. Therefore, we get

$$S_1(z)\xrightarrow{d}N(0,R). \tag{A.32}$$

We can also prove that

$$S_2(z)=o_p(1). \tag{A.33}$$

Theorems 2.2 and condition (B8) imply that $\beta-\hat{\beta}=O_p(N^{-1/2})$. Therefore, we get $S_3(z)=O_p(h^{1/2})$. This together with (A.32) and (A.33) proves (A.29).

Analogously to the proof of (A.30). We can verify (A.30) easily. As to (A.31), we find

$$\max_{1\le i\le n} ||\hat{\Xi}_i\{g(z)\}|| \le \max_{1\le i\le n} ||(I_{it}-Q)K_h(Z_{it}-z)\varepsilon_{it}|| \\ + \max_{1\le i\le n} ||(I_{it}-Q)[g(Z_{it})-g(z)]K_h(Z_{it}-z)|| \\ + \max_{1\le i\le n} ||(I_{it}-Q)X_{it}^{\tau}(\beta-\hat{\beta})K_h(Z_{it}-z)|| = J_1+J_2+J_3$$

From Markov inequality and conditions (B3) and (B4), one can obtain

$$P(J_1 \ge \sqrt{Nh}) \le (Nh)^{-s}\sum_{i=1}^{n} E[(I_{it}-Q)\varepsilon_{it}K_h(Z_{it}-z)]^{2s} \le C(Nh)^{1-s} \to 0$$

which implies that $J_1 = o_p(\sqrt{Nh})$. Using some arguments similar to those used in the proof of Lemma A.6, we can prove $J_2 = o_p(\sqrt{Nh})$ and $J_3 = o_p(\sqrt{Nh})$. Therefore we obtain that $max_{1\le i\le n} ||\hat{\Xi}_i\{g(z)\}|| = o_p(\sqrt{Nh})$.

Applying (A.30) and the proof in Owen (1990), one can derive that $\phi = O_p(N^{-1/2})$, which completes the proof of Lemma A.8. □

Proof of Theorem 2.5. Invoking some arguments similar to those used in the proof of can be proved Theorems 2.4, we can proof

$$2\mathcal{Q}_n(g(z)) = \left[\frac{1}{Nh}\sum_{i=1}^{n}\hat{\Xi}_i^2\{g(z)\}\right]^{-1}\left\{\frac{1}{\sqrt{Nh}}\sum_{i=1}^{n}\hat{\Xi}_i\{g(z)\} - b(z)\right\}^2$$

From Lemma A.8, we can prove that $2\mathcal{Q}_n(g(z)) \xrightarrow{d} \chi_1^2$. □

References

Arellano M (2003) Panel data econometrics. Oxford University Press, New York

Baltagi BH (2005) Econometric analysis of panel data, 2nd edn. Wiley, New York

Baltagi BH, Li D (2002) Series estimation of partially linear panel data models with fixed effects. Ann Econ Fin 3(1):103–116

Cai Z, Li Q (2008) Nonparametric estimation and varying coefficient dynamic panel data models. Econom Theory 24:1321–1342

Carroll RJ, Ruppert D, Stefanski LA (1995) Measurement error in nonlinear models. Chapman & Hall, London

Chen J, Gao J, Li D (2013) Estimation in partially linear single-index panel data models with fixed effects. J Bus Econ Stat 31(3):315–330

Fan J, Huang T (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11(6):1031–1057

Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96:1348–1360

Fan J, Lv J (2008) Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Ser B 70(5):849–911

Fan J, Peng H, Huang T (2005) Semilinear high-dimensional model for normalization of microarrays data: a theoretical analysis and partial consistency. J Am Stat Assoc 100:781–796
Fan GL, Liang HY, Wang JF (2013) Statistical inference for partially time-varying coefficient errors-in-variables models. J Stat Plan Inference 143:505–519
Fan GL, Liang HY, Shen Y (2016) Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors. J Multivar Anal 147:183–201
Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic Press, New York
Hanfelt JJ, Liang KY (1997) Approximate likelihood for generalized linear errors-in-variables models. J R Stat Soc Ser B 59(3):627–637
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York
He BQ, Hong XJ, Fan GL (2017) Block empirical likelihood for partially linear panel data models with fixed effects. Stat Probab Lett 123(4):128–138
Henderson DJ, Carroll RJ, Li Q (2008) Nonparametric estimation and testing of fixed effects panel data models. J Econom 144:257–275
Horowitz JL, Lee S (2004) Semiparametric estimation of a panel data proportional hazards model with fixed effects. J Econom 119:155–198
Hsiao C (2003) Analysis of panel data, 2nd edn. Cambridge University Press, New York
Hu XM (2014) Estimation in a semi-varying coefficient model for panel data with fixed effects. J Syst Sci Complex 27:594–604
Hu XM (2017) Semi-parametric inference for semi-varying coefficient panel data model with individual effects. J Multivar Anal 154:262–281
Hu XM, Wang Z, Zhao Z (2009) Empirical likelihood for semiparametric varying-coefficient partially linear errors-in-variables models. Stat Probab Lett 79(8):1044–1052
Leng C, Tang CY (2012) Penalized empirical likelihood and growing dimensional general estimating equations. Biometrika 99(3):703–716
Li R, Liang H (2008) Variable selection in semiparametric regression modeling. Ann Stat 36(1):261–286
Li DG, Chen J, Gao JT (2011) Non-parametric time-varying coefficient panel data models with fixed effects. Econometrics 14:387–408
Li G, Lin L, Zhu L (2012) Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. J Multiva Anal 105:85–111
Liang H, Härdle W, Carroll RJ (1999) Estimation in a semiparametric partially linear errors-in-variables model. Ann Stat 27:1519–1535
Nakamura T (1990) Corrected score function for errors-in-variables model: methodoology and application to generalized linear models. Biometrika 77(1):127–137
Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75:237–249
Owen AB (1990) Empirical likelihood ratio confidence regions. Ann Stat 18:90–120
Ren Y, Zhang X (2011) Variable selection using penalized empirical likelihood. Sci China Math 54(9):1829–1845
Rodriguez-Poo JM, Soberon A (2014) Direct semi-parametric estimation of fixed effects panel data varying coefficient models. Econom J 17(1):107–138
Schennach SM (2007) Instrumental variable estimation of nonlinear errors-in-variables models. Econometrics 75:201–239
Su LJ, Ullah A (2006) Profile likelihood estimation of partially linear panel data models with models with fixed effects. Econom Lett 92:75–81
Su L, Ullah A (2007) More efficient estimation of nonparametric panel data models with random effects. Econom Lett 96(3):375–380
Tang CY, Leng C (2010) Penalized high-dimensional empirical likelihood. Biometrika 97(4):905–920
Wang S, Xiang L (2017) Penalized empirical likelihood inference for sparse additive hazards regression with a diverging number of covariates. Stat Comput 27(5):1347–1364
Wang N, Carroll RJ, Liang KY (1996) Quasi-likelihood and variance functions in measurement error models with replicates. Biometrics 52:423–432
Xue LG, Zhu LX (2008) Empirical likelihood-based inference in a partially linear model for longitudinal data. Sci China Ser A 51(1):115–130
You JH, Chen GM (2006) Estimation in a statistical inference in a semiparametric varying-coefficient partially linear errors-in-variables model. J Multivar Anal 97:324–341

Zhou HB, You JH, Zhou B (2010) Statistical inference for fixed-effects partially linear regression models with errors in variables. Stat Papers 51:629–650

Zhu LX, Cui HJ (2003) A semiparametric regression model with errors in variables. Scand J Stat 30:429–442