Statistics and Probability Letters 123 (2017) 128–138

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Block empirical likelihood for partially linear panel data models with fixed effects

Bang-Qiang He [a], Xing-Jian Hong [b,*], Guo-Liang Fan [a]

[a] School of Mathematics and Physics, Anhui Polytechnic University, Wuhu 241000, China

[b] School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou 310018, China

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form 20 November 2016
Accepted 21 November 2016
Available online 19 December 2016

MSC:
62G15
62G20

Keywords:
Block empirical likelihood
Partially linear model
Panel data
Fixed effect

ABSTRACT

In this article, we consider a partially linear panel data models with fixed effects. In order to accommodate the within-group correlation, we apply the block empirical likelihood procedure to partially linear panel data models with fixed effects, and prove a nonparametric version of Wilks' theorem which can be used to construct the confidence region for the parametric. By the block empirical likelihood ratio function, the maximum empirical likelihood estimator of the parameter is defined and the asymptotic normality is shown. A simulation study and a real data application are undertaken to assess the finite sample performance of our proposed method.

1. Introduction

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. The double-index panel data models enable researchers to estimate complex models and extract information which may be difficult to obtain by applying purely cross section or time series models. There exists a rich literature on parametric linear and nonlinear panel data models. For an overview of statistical inference and econometric analysis of parametric panel data models, we refer to the books by Baltagi (2005) and Hsiao (2003). To avoid imposing the strong restrictions assumed in the parametric panel data models, some nonparametric methods have been used in both panel data model estimation and specification testing (e.g., Hjellvik et al., 2004; Henderson et al., 2008; Cai and Li, 2008). While the nonparametric approach is useful in exploring hidden structures and reducing modeling biases, they can be too flexible to draw concise conclusions, and face the *curse of dimensionality* due to a large number of covariates. To overcome these shortcomings, we use semiparametric approaches which are the compromises between the general nonparametric modeling and fully parametric specification.

Consider the following partially linear panel data models with fixed effects:

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

where Y_{it} is the response, $(X_{it}^{\tau}, U_{it}) \in R^p \times R$ 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, $g(U_{it}) = (g_1(U_{it}), \ldots, g_q(U_{it}))^{\tau}$ is

* Corresponding author.
E-mail address: myemailh@126.com (X.-J. Hong).

http://dx.doi.org/10.1016/j.spl.2016.11.021
0167-7152/

a q-dimensional unknown functions and υ_i is the unobserved individual effects. Denote by $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ the random error vector of the ith subject and $\{\varepsilon_i, i = 1, \ldots, n\}$ are mutually independent with $E(\varepsilon_i|X_{it}, U_{it}) = 0$.

Model (1.1) is called a fixed effects model if υ_i is correlated with X_{it} and (or) U_{it} with an unknown correlation structure. Model (1.1) is called a random effects model if υ_i is uncorrelated with X_{it} and U_{it}. The fixed effects specification has the advantage of robustness compared to the random effects specification (e.g., Baltagi, 2005; Horowitz and Lee, 2004). However, the analysis of the fixed effects panel data is more challenging because of increasing number of parameters with the sample size, yielding the famous Neyman and Scott (1948) problem. In this paper, we are concerned with the fixed effects case.

Obviously, model (1.1) includes many usual parametric, nonparametric and semiparametric regression models. For example, when $\upsilon_i = 0$, model (1.1) reduces to the partially linear panel data model. Many researchers have explored the partially linear panel data model (e.g., Roy, 1997; Li and Ullah, 1998; You and Zhou, 2006a). Roy (1997) has used the partially linear panel data model to study the calorie and income relationship for two years panel data of rural south India. When $\upsilon_i = 0$ and $g = 0$, the model becomes the well-known parametric panel data regression model, which has been widely applied in economics (cf., Ahn and Schmidt, 2000; Hsiao, 2003). When $\upsilon_i = 0$ and $\beta = 0$, the model reduces to the panel data nonparametric model, which has been investigated by Ruckstuhl et al. (2000).

For partially linear panel data models, Li and Ullah (1998) constructed a feasible semiparametric generalized least squares estimator for the coefficient of the linear component and derived the asymptotic normality of the proposed estimator. For the model (1.1), Su and Ullah (2006) adapted a local linear dummy variable approach to remove the unknown fixed effects. In this paper, we make statistical inference for the parametric β in partially linear panel data models with fixed effects. Following the estimation procedure proposed by Li and Ullah (1998) and Su and Ullah (2006), the least-squared estimations of β can be obtained. Based on this, a normal-based confidence region for the parametric is constructed. But such a construction is inconvenient because it involves estimating complex asymptotic covariance of the estimators. To end this, we recommend using the empirical likelihood (EL) method to construct the confidence regions for β. Based on the EL method, we can construct immediately an approximate confidence region for the regression parameter. One motivation is that empirical likelihood inference does not involve the asymptotic covariance of the estimators, which is rather complex structure for the partially linear panel data models with fixed effects. Another motivation is that the confidence region based on EL approach does not impose prior constraints on the region shape, and the shape and orientation of confidence regions are determined completely by the data. Therefore, The EL method has been used by many authors, such as Shi and Lau (2000), Wang and Jing (1999), You and Zhou (2006b), Fan et al. (2012) and so on.

The usual empirical likelihood method cannot be applied, however, to partially linear panel data models with fixed effects due to correlation within groups. To accommodate the within-group correlation, we apply the block empirical likelihood procedure proposed by You et al. (2006) to model (1.1), establish a block empirical log-likelihood ratio for the parametric component, and derive a nonparametric version of Wilks' theorem which can be used to construct the block empirical likelihood confidence region with asymptotically correct coverage probability for the parametric component.

The rest of this paper is organized as follows. Section 2 introduces the methodology and empirical log-likelihood ratio function for β. Assumptions and the main result are given in Section 3. Some simulation studies and a real-data example are conducted in Section 4. The proofs of the main results are relegated to Section 5.

2. The model and methodology

To introduce our estimation, we assume that model holds with the restriction $\sum_{i=1}^{n} \upsilon_i = 0$. Let $\upsilon = (\upsilon_2, \ldots, \upsilon_n)^\tau$ and $\upsilon_0 = (-\sum_{i=2}^{n} \upsilon_i, \upsilon_2, \ldots, \upsilon_n)^\tau$. We rewrite model (1.1) in a matrix format yields

$$Y = X\beta + g(U) + M\upsilon + \varepsilon, \tag{2.1}$$

where $M = [-i_{n-1}\, I_{n-1}]^\tau \otimes i_T$ is an $nT \times (n-1)$ matrix, I_n denotes the $n \times n$ identity matrix, and i_n denotes the $n \times 1$ vector of ones. There are many approaches to estimating the parameters $\{\beta_j, j = 1, \ldots, p\}$ and the functions $\{g_i(\cdot), i = 1, \ldots, q\}$. The main idea is from the profile least squares approach proposed by 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 (2.1). Let $\theta = (\upsilon^\tau, \beta^\tau)^\tau$. Given θ, one can apply a local linear regression technique to estimate the nonparametric component $\{g_j(\cdot), j = 1, \ldots, q\}$ in (2.1). For U_{it} in a small neighborhood of u, one can approximate $g_j(U_{it})$ locally by a linear function as below

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

where $g_j'(u) = \partial g_j(u)/\partial u$. This leads to the following weighted local least-squares problem: find $\{(a_j, b_j), j = 1, \ldots, q\}$ to minimize

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(Y_{it} - X_{it}^\tau\beta - \upsilon_i\right) - \sum_{j=1}^{q}\left[a_j + b_j(U_{it} - u)\right]\right\}^2 K_h(U_{it} - u), \tag{2.2}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and h is a sequence of positive numbers tending to zero, called bandwidth. Simple calculation yields that

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

where

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

and $W_u = \mathrm{diag}(K_h(U_{11}-u), \ldots, K_h(U_{1T}-u), \ldots, K_h(U_{nT}-u))$.

Then we can estimate $g(u)$ by

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

Let $N = nT$, $\Omega = (I_q, 0_{q\times q})(D_u^{\tau} W_u D_u)^{-1} D_u^{\tau} W_u$, and the $[(i-1)T+t]$th element of Ω be $\omega_{it}(u)$. Then according to Chiou and Müller (1999), we have

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

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

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:

$$\hat{\theta} = \arg\min_{\theta}\left[Y - X\beta - M\upsilon - \Omega(Y - X\beta - M\upsilon)\right]^{\tau}[Y - X\beta - M\upsilon - \Omega(Y - X\beta - M\upsilon)]. \tag{2.3}$$

Supposing that $\widetilde{X} = (I_{nT} - \Omega)X$, $\widetilde{Y} = (I_{nT} - \Omega)Y$, $\widetilde{M} = (I_{nT} - \Omega)M$, we have $\widetilde{\upsilon} = (\widetilde{M}^{\tau}\widetilde{M})^{-1}\widetilde{M}^{\tau}(\widetilde{Y} - \widetilde{X}\beta)$. Let $\widetilde{H} = I_{nT} - \widetilde{M}(\widetilde{M}^{\tau}\widetilde{M})^{-1}\widetilde{M}^{\tau}$, we can obtain $\widetilde{H}\widetilde{M}\upsilon = 0$. Hence, the fixed effects term $M\upsilon$ 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:

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

Similar to Owen (1990) and Shi and Lau (2000), we can treat $\{\hat{r}_{it}(\beta),\; i = 1, \ldots, n;\; t = 1, \ldots, T\}$ as a random sieve approximation of the sequence of random errors $\{\varepsilon_{it},\; i = 1, \ldots, n;\; t = 1, \ldots, T\}$. In order to deal with the correlation within groups, we use the block empirical likelihood procedure proposed by You et al. (2006). Similar to You et al. (2006), the profile empirical likelihood for β is defined as

$$\mathcal{L}_n(\beta) = \sup\left\{\prod_{i=1}^{n} np_i \,\middle|\, \sum_{i=1}^{n} p_i \sum_{t=1}^{T} \widetilde{X}_{it} e_{it}^{\tau}\widetilde{H}(\widetilde{Y}_{it} - \widetilde{X}_{it}^{\tau}\beta) = 0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\right\}. \tag{2.4}$$

For a given β, a unique maximum exists, provided that 0 is inside the convex hull of the points $\psi_i(\beta) = \sum_{t=1}^{T}\widetilde{X}_{it}e_{it}^{\tau}\widetilde{H}(\widetilde{Y}_{it} - \widetilde{X}_{it}^{\tau}\beta)$, $i = 1, \ldots, n$. The maximum of (2.4) may be found via the method of Lagrange multiplier. The optimal value for p_i satisfying (2.4) may be shown to $p_i = \frac{1}{n[1+\lambda^{\tau}\psi_i(\beta)]}$. We define the corresponding profile block empirical log-likelihood ratio as

$$\mathcal{LR}_n(\beta) = -\sum_{i=1}^{n}\log\{1 + \lambda^{\tau}\psi_i(\beta)\}, \tag{2.5}$$

where $\lambda(\beta)$ is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\beta)}{1 + \lambda^{\tau}\psi_i(\beta)} = 0. \tag{2.6}$$

We show in the next section that if β is the true parameter vector, $\mathcal{LR}_n(\beta)$ is asymptotically chi-square distributed. Maximizing $\{\mathcal{LR}_n(\beta)\}$ can obtain the maximum block empirical likelihood estimator (MBELE) $\widehat{\beta}$ of β. According to Qin and Lawless (1994), $\widehat{\beta}$ is also equal to the solution of the estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^{\tau}e_{it}^{\tau}\widetilde{H}(\widetilde{Y}_{it} - \widetilde{X}_{it}^{\tau}\beta) = 0. \tag{2.7}$$

The solution of the estimating equation (2.7) is the usual profile least squares estimator. Therefore, the MBELE of β is identical to the profile least squares estimator, which is given by

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

3. The asymptotic result

In this section, we will show that if β is the true parameter vector, $\mathcal{LR}_n(\beta)$ is asymptotically χ^2-distributed. Before formulating the main results, we first give the following some assumptions.

(A1) The random vector U_{it} has a continuous density function $f(\cdot)$ with a compact support $\mathcal{U}$ on R. $0 < inf_{u\in\mathcal{U}} f(u) \le \sup_{u\in\mathcal{U}} f(u) < \infty$.
(A2) Let $m(u) = E(X_{it}|U_{it} = u)$. The functions $m(u)$ and $g(u)$ are twice continuously differentiable on $\mathcal{U}$.
(A3) $(\upsilon_i, X_i, U_i, \varepsilon_i), i = 1, \ldots, n$, are i.i.d. There exists some $\delta > 2$ such that $E\|X_{it}\|^{2+\delta} < \infty$ and $E|\varepsilon_i|^{2+\delta} < \infty$, where $\|a\| = (\sum_{i=1} a_i^2)^{1/2}$ is the usual Euclidean norm of a vector a.
(A4) The kernel $K(\upsilon)$ is a symmetric probability density function with a continuous derivative on its compact support $\mathcal{U}$.
(A5) $E|\breve{X}_{it}|^{2+\delta} < \infty$, $\Psi = \sum_{t=1}^{T}[\breve{X}_{it}(\breve{X}_{it} - \sum_{s=1}^{T}\frac{\breve{X}_{is}}{T})^\tau]$ is positive definite, where $\breve{X}_{it} = X_{it} - E(X_{it}|U_{it})$.
(A6) $E(Y_{it}|X_i, U_i, \upsilon_i) = E(Y_{it}|X_{it}, U_{it}, \upsilon_{it}) = X_{it}\beta + g(U_{it}) + \upsilon_i$.
(A7) The bandwidth h satisfies $h \to 0$, $nh^8 \to 0$ and $nh^2/(logn)^2 \to \infty$ as $n \to \infty$.
(A8) As $k \to \infty$, $P\Big[0 \in ch\big\{\sum_{t=1}^{T}\widetilde{X}_{1t}e_{it}^\tau\widetilde{H}(\widetilde{Y}_{1t} - \widetilde{X}_{1t}^\tau\beta), \ldots, \sum_{t=1}^{T}\widetilde{X}_{kt}e_{it}^\tau\widetilde{H}(\widetilde{Y}_{kt} - \widetilde{X}_{kt}^\tau\beta)\big\}\Big] \to 1$ where "ch" denotes the convex hull of a set in R^p.

Remark 3.1. Assumptions (A1)–(A8) while look a bit lengthy, are actually quite mild and can be easily satisfied. (A1)–(A6) can be founded in Su and Ullah (2006). (A7) is assumed in Hu (2014). (A8) has been used by many authors (e.g., Shi and Lau, 2000; You et al., 2006).

Theorem 3.1. *Suppose that* (A1)–(A8) *hold. For model* (2.1), *if* β_0 *is the true value of the parameter, then* $\mathcal{LR}_n(\beta_0) \xrightarrow{d} \chi_p^2$, *as* $n \to \infty$, *where* χ_p^2 *is a standard chi-square random variable with p degrees of freedom and* $\xrightarrow{d}$ *stands for convergence in distribution.*

As a consequence of the theorem, confidence regions for the parameter β_0 can be constructed. More precisely, for any $0 \le \alpha < 1$, let c_α be the $1-\alpha$ quantile of χ^2 distribution such that $P(\chi_p^2 > c_\alpha) \le 1 - \alpha$. Then $\ell_{EL}(\alpha) = \{\beta \in R^p : \mathcal{LR}_n(\beta_0) \le c_\alpha\}$ constitutes a confidence region for β_0 with asymptotic coverage probability $1-\alpha$ because the event that β_0 belongs to $\ell_{EL}(\alpha)$ is equivalent to the event that $\mathcal{LR}_n(\beta_0) \le c_\alpha$.

Maximizing $\{-\mathcal{LR}_n(\beta_0)\}$ can obtain the maximum empirical likelihood estimator (MELE) $\widehat{\beta}$ of β_0. Write

$$\widehat{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}\left(\widetilde{X}_{it} - \sum_{s=1}^{T}\frac{\widetilde{X}_{is}}{T}\right)^\tau. \tag{3.1}$$

If the matrix $\widehat{\Psi}$ is invertible, with the similar proof Theorem 1 in Qin and Lawless (1994) and Theorem 2 in Xue and Zhu (2008), then MELE $\widehat{\beta}$ of β_0 can be represented as

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

Then the asymptotic normality of $\widehat{\beta}$ is stated in the following theorem.

Theorem 3.2. *Suppose that* (A1)–(A8) *hold, if* β_0 *is the true value of the parameter, then we have* $\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$, *where* $V = \Psi^{-1}\Theta\Psi^{-1}$, $\Theta = \sum_{t=1}^{T}\sum_{s=1}^{T}E\Big\{\Big[\breve{X}_{it}(\breve{X}_{is} - \sum_{l=1}^{T}\frac{\breve{X}_{il}}{T})^\tau\Big](\varepsilon_{it}\varepsilon_{is})\Big\}$.

To construct confidence regions, we can obtain $\widehat{\Psi}^{-1}\widehat{\Theta}\widehat{\Psi}^{-1}$ as the estimator of $\Psi^{-1}\Theta\Psi^{-1}$, where $\widehat{\Psi}$ is defined in (3.1), $\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{X}_{it}(\widetilde{X}_{it} - \sum_{s=1}^{T}\frac{\widetilde{X}_{is}}{T})^\tau\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{is}$ and $\widehat{\varepsilon}_{it} = \widetilde{Y}_{it} - \widetilde{X}_{it}^\tau\widehat{\beta} - \widehat{\upsilon}_i$. By Theorem 3.2, we obtain that

$$[\widehat{\Psi}^{-1}\widehat{\Theta}\widehat{\Psi}^{-1}]^{-1/2}\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, I_p), \tag{3.2}$$

or

$$(\widehat{\beta} - \beta_0)^\tau n[\widehat{\Psi}^{-1}\widehat{\Theta}\widehat{\Psi}^{-1}]^{-1}(\widehat{\beta} - \beta_0) \xrightarrow{d} \chi_p^2. \tag{3.3}$$

Thus, a confidence region for β_0 can be obtained from (3.2) and (3.3).

Table 1
CP of confidence regions for (β_1, β_2) based on the BEL and the AN with nominal confidence level 0.95.

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

Note: "CP_{BE} and CP_{AN}" denote the CP based on the BEL and the AN, respectively.

Table 2
CP and AL of the confidence intervals based on the BEL and the AN with nominal confidence level 0.95.

Para	ρ	(n, T)	*CE*	*CN*	*AE*	*AN*	ρ	(n, T)	*CE*	*CN*	*AE*	*AN*
β_1	0.5	(50,4)	0.935	0.895	0.1246	0.2309	1	(50,4)	0.931	0.770	0.1279	0.2541
		(50,6)	0.942	0.903	0.1146	0.1851		(50,6)	0.939	0.798	0.1049	0.2018
		(80,4)	0.950	0.921	0.1078	0.1806		(80,4)	0.941	0.807	0.0992	0.1931
		(80,6)	0.957	0.928	0.0899	0.1418		(80,6)	0.949	0.865	0.0816	0.1518
β_2		(50,4)	0.938	0.670	0.1376	0.2363		(50,4)	0.924	0.654	0.1532	0.2541
		(50,6)	0.941	0.780	0.1231	0.1853		(50,6)	0.937	0.712	0.1324	0.2034
		(80,4)	0.946	0.823	0.1152	0.1803		(80,4)	0.941	0.810	0.1267	0.1900
		(80,6)	0.954	0.895	0.0956	0.1415		(80,6)	0.947	0.853	0.1087	0.1512

Note: "*CE* and *CN*" denote the CP based on the BEL and the AN, respectively; "*AE* and *AN*" denote the AL based on the BEL and the AN, respectively.

4. Numerical studies

4.1. Simulation study

In this section, we carry out some simulation experiments to illustrate the finite sample performance of the proposed block empirical likelihood (BEL) method over the asymptotically normal (AN).

Firstly, we consider the following partially linear panel data models with fixed effects:

$$Y_{it} = X_{it}^{\tau}\beta + g(U_{it}) + \upsilon_i + \varepsilon_{it}, \quad i = 1, \dots, n, \; t = 1, \dots, T, \tag{4.1}$$

where $\beta = (\beta_1, \beta_2)^{\tau} = (1, \sqrt{2})^{\tau}/\sqrt{3}, g(U_{it}) = \sin(2\pi U_{it}), U_{it} \sim U(0, 1),\ \upsilon_i = \rho\bar{U}_i + w_i$ with $\rho = 0.5,\ 1$ and $w_i \sim N(0, 1)$ for $i = 2, 3, \dots, n$, and $\upsilon_i = -\sum_{i=2}^{n} \upsilon_i, \bar{U}_i = \frac{1}{T}\sum_{t=1}^{T} U_{it}$. We use ρ to control the correlation between υ_i and $\bar{U}_i$.

In our simulations, we took the sample sizes $(n, T) = (50, 4), (50, 6), (80, 4)$ *and* $(80, 6)$, respectively, and we choose the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \le 1\}$. The "leave-one-subject-out" cross-validation bandwidth $CV(h)$ is obtained by minimizing

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

where $\hat{\beta}_{[i]}$ and $\hat{g}_{[i]}(U_{it})$ are estimators of β and $g(U_{it})$ which are computed with all of the measurements but not the ith subject. We consider 500 replications with the nominal level $1 - \alpha = 0.95$. The numerical results are reported in Tables 1–3, where Table 1 reports some representative coverage probabilities (CP) of confidence regions for (β_1, β_2), Table 2 reports the CP and average lengths (AL) of confidence intervals for β_1 and β_2, and Table 3 reports the average biases and standard deviations (SD) of the proposed MBELE for the parametric components β_1 and β_2.

From Tables 1 and 2, we see that the method based on the BEL performs better than the AN method since the BEL method gives higher CP and shorter AL of confidence intervals than the AN method. Also, it is obvious to see that the CP of the confidence regions/intervals tend to increase and the AP decrease as the sample size (n, T) become larger. It is also interesting to note that the CP of the confidence intervals tend to decrease and the AL increase as ρ gets larger. In addition, it can be seen from Table 3 that the proposed estimators of the parametric components are asymptotically unbiased and have small SD. Further, it is reasonable that the average biases and SD decrease as the sample size increases.

4.2. Application to CD4 data

We now apply the proposed procedure to the CD4 data from the Multi-Center AIDS Cohort study. The dataset contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during the follow-up period between 1984 and 1991. Details about the related design, methods and medical implications of the Multi-Center AIDS Cohort study have been described by Kaslow et al. (1987). This dataset has been studied by many authors, such as Xue and Zhu (2007) and Zhou and Lin (2014). Although all the individuals were scheduled to have clinical visits semi-annually for taking the measurement of CD4 percentage, due to the various reasons, some individuals missed scheduled visits, which resulted in unequal numbers of measurements and different measurement times across individuals. We select

Table 3
The finite sample average biases and SD of the proposed MBELE for the parametric components β_1 and β_2.

(n, T)	β_1				β_2			
	$\rho = 0.5$		$\rho = 1$		$\rho = 0.5$		$\rho = 1$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
(50,4)	0.0019	0.0536	0.0021	0.0837	0.0026	0.0609	−0.0021	0.0571
(50,6)	−0.0013	0.0520	−0.0016	0.0523	0.0014	0.0531	0.0023	0.0528
(80,4)	0.0007	0.0461	0.0009	0.0471	0.0010	0.0452	0.0014	0.0508
(80,6)	0.0001	0.0404	0.0006	0.0426	0.0006	0.0418	0.0008	0.0438

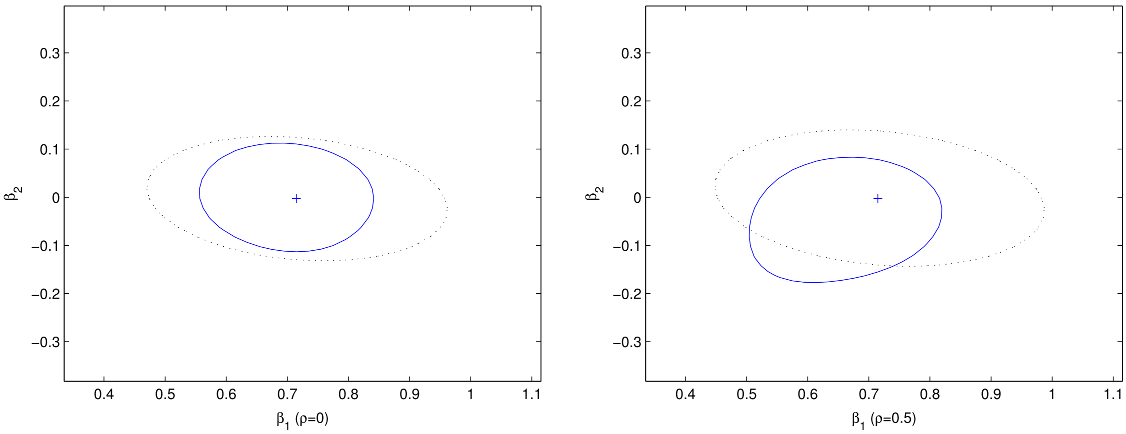

Fig. 1. Application to CD4 data. The 95% confidence regions based on the AN (dotted curve) and BEL (solid curve) for $\beta = (\beta_1, \beta_2)$.

158 individuals in which there have the first six measurements. Hence, we can obtain equal numbers of measurements for each individuals.

Let t_{ij} be the time for years of the tth measurement of the ith individual after HIV infection, Y_{it} be the ith individual's CD4 cell percentage at time t_{ij}. We take two covariates for the study: X_{1it}, the ith individual's preCD4 percentage measured at time t_{ij}; X_{2it}, the individual's smoking status, which takes binary values 1 or 0, according to whether an individual is a smoker or nonsmoker. We consider the following model:

$$Y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + g(U_{it}) + \upsilon_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \tag{4.2}$$

where $U_{it} = t_{ij}/T$ and υ_i is a state-specific effect that may include time, religion, race and education. $\upsilon_i = \rho\bar{U}_i + w_i$ with $\bar{U}_i = \frac{1}{T}\sum_{t=1}^{T} U_{it}$ and $w_i \sim N(0, 1)$. β_1 and β_2 describe the effects for pre-infection CD4 percentage and cigarette smoking. Using the BEL and AN methods, we obtained $1-\alpha = 0.95$ confidence regions for (β_1, β_2) that are presented in Fig. 1. These figures show that BEL gives smaller confidence regions than the AN does, and the BEL confidence regions become deviation from the center as $\rho = 0$ gets $\rho = 0.5$.

5. Proofs of the main result

For the convenience and simplicity, let $\vartheta_k = \int u^k K(u)dt$ and $P = (I_{nT} - \Omega)^\tau (I_{nT} - \Omega)$.

Lemma 5.1. *Suppose that conditions* (A1)–(A8) *hold. Then we have*

$$E\left|g(U_{it}) - \sum_{k=1}^{n}\sum_{l=1}^{T}\omega_{kl}(U_{it})g(U_{kl})\right|^2 = O(h^4). \tag{5.1}$$

Proof. Write

$$\omega_{it}(u) = \frac{\frac{1}{N}K_h(U_{it}-u)[A_{N2}(u) - (U_{it}-u)A_{N1}(u)]}{A_{N0}(u)A_{N2}(u) - A_{N1}^2(u)} = \frac{V_{Nit}(u)}{V_N(u)}, \tag{5.2}$$

where $V_{Nit}(u) = \frac{1}{N}K_h(U_{it}-u)[A_{N2}(u) - (U_{it}-u)A_{N1}(u)]$, $V_N(u) = A_{N0}(u)A_{N2}(u) - A_{N1}^2(u)$. Let $u = U_{kl}$, $k = 1, \ldots, n$; $l = 1, \ldots, T$. Note that $\sum_{k=1}^{n}\sum_{l=1}^{T} V_{Nit}(u)(U_{it} - u) = 0$ and $\sum_{k=1}^{n}\sum_{l=1}^{T} V_{Nit}(u) = V_N(u)$. Write $G(U_{it}, u) = g(U_{it}) - g(u)$

$+ g'(u)(U_{it} - u)$. A simple calculation yields

$$g(U_{it}) - \sum_{k=1}^{n}\sum_{l=1}^{T}\omega_{kl}(U_{it})g(U_{kl}) = \frac{\sum_{k=1}^{n}\sum_{l=1}^{T}V_{Nit}(u)G(U_{it}, u)}{V_N(u)}. \tag{5.3}$$

Let $E_{U_{it}}(\chi_n)$ be the conditional expectation of χ_n given U_{it}, and denote $\chi_n = O_r(a_n)$, if $E(|\chi_n|^r) = O(a_n^r)$. Using the Cauchy–Schwarz inequality, we can obtain

$$O_r(a_n)O_r(b_n) = O_{r/2}(a_nb_n), \tag{5.4}$$

$$\chi_n = E_{U_{it}}[\chi_n] + O_4\big((E|\chi_n - E_{U_{it}}[\chi_n]|^4)^{1/4}\big). \tag{5.5}$$

By condition (A1), we know that $\sup_u f(u) < \infty$. Then $\exists L > 0$, for any y and u, $|f(y) - f(u)| \leqslant L|y-u|$. For $A_{Ns}(u)$, $s = 0, 1, 2$, we can obtain that

$$\begin{aligned} E_{U_{it}}[A_{Ns}(u)] &= E\left[\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{T}(U_{it}-u)^sK_h(U_{it}-u)\right] \\ &= \int(U_{it}-u)^sK_h(U_{it}-u)f(U_{it})dU_{it} = h^s\int u^sK(u)du[f(u)+O(h)] \\ &= h^sf(u)\vartheta_s[1+O(h)]. \end{aligned} \tag{5.6}$$

By Theorem 1 in Yang (2001), it is easy to obtain

$$\begin{aligned} &E\Big\{|A_{Ns}(u) - E_{U_{it}}[A_{Ns}(u)]|^4\Big\} \\ &\leq \frac{c}{N^4}Nh^{4s-3}\int|u^sK(u)|^4du[Ef(u)+Lh] + \frac{c}{N^4}E\Big\{Nh^{2s-3}\int u^{2s}K^2(u)du[Ef(u)+Lh]\Big\} \\ &= O(N^{-3}h^{4s-3}) + O(N^{-2}h^{2s-2}). \end{aligned} \tag{5.7}$$

Combining (5.5) and (5.7), we obtain

$$A_{Ns}(u) = E_{U_{it}}[A_{Ns}(u)] + O_4(h^s(Nh)^{-1/2}) = h^sf(u)\vartheta_s[1 + O_4(h + (Nh)^{-1/2})], \quad s = 0, 1, 2. \tag{5.8}$$

From (5.4) and (5.8), we can derive

$$V_N(u) = A_{N0}(u)A_{N2}(u) - A_{N1}^2(u) = h^2f^2(u)\vartheta_s[1 + O_2(h + (Nh)^{-1/2})]. \tag{5.9}$$

Let $V_{Nh}(u) = \frac{1}{h^2}V_N(u)$, $V(u) = f^2(u)\vartheta_s$. Since $\mathcal{U}$ is a compact set, using (5.6) and Dvoretzky inequality, and invoking the standard method of dealing with the kernel density estimation, we can derive, for arbitrary $\varepsilon > 0$,

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

Because the series $\sum_{N=1}^{\infty}\exp(-\gamma Nh^2) < \infty$, for any $\gamma > 0$, by hypothesis, an application of the Borel–Cantelli lemma proves that

$$\sup_{u\in\mathcal{U}}|V_{Nh}(u) - V(u)| \to 0, \quad \text{a.s.} \tag{5.10}$$

From Condition (A1), we know that $\inf_{u\in\mathcal{U}}f(u) \geqslant c_0 > 0$, Therefore, when n is large enough, we have

$$\inf_{u\in\mathcal{U}}|V_{Nh}(u)| \geqslant \inf_{u\in\mathcal{U}}|V(u)| - \sup_{u\in\mathcal{U}}|V_{Nh}(u) - V(u)| > \vartheta_2c_0^2 > 0, \quad \text{a.s.} \tag{5.11}$$

A simple calculation yields

$$\begin{aligned} \sum_{k=1}^{n}\sum_{l=1}^{T}V_{Nit}(u)G(U_{it}, u) = &\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{T}G(U_{it}, u)K_h(U_{it}-u)A_{N2}(u) \\ &-\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{T}G(U_{it}, u)(U_{it}-u)K_h(U_{it}-u)A_{N1}(u). \end{aligned} \tag{5.12}$$

Similar to the proof of (5.8), for $s = 0, 1$, we have

$$\frac{1}{Nh^{2+s}}\sum_{k=1}^{n}\sum_{l=1}^{T}G(U_{it},u)(U_{it}-u)^sK_h(U_{it}-u)$$
$$= h^{-(2+s)}E_{U_{it}}[G(U_{it},u)(U_{it}-u)^sK_h(U_{it}-u)] + O_4((Nh)^{-1/2}) = d_{Ns} + O_4((Nh)^{-1/2}), \quad (5.13)$$
$$|d_{Ns}| \le c\int |u|^{s+2}K(u)du[f(u)+O(h)] = O(1).$$

Combining (5.8), (5.12) and (5.13), we obtain

$$\sum_{k=1}^{n}\sum_{l=1}^{T}V_{Nit}(u)G(U_{it},u) = h^4 f(u)\vartheta_2 d_{N0} + O_2(h+(Nh)^{-1/2}). \quad (5.14)$$

This together with Eq. (5.9), we conclude that

$$E\left|\frac{\sum_{k=1}^{n}\sum_{l=1}^{T}V_{Nit}(u)G(U_{it},u)}{V_N(u)}\right|^2 \le ch^{-4}E\left[\sum_{k=1}^{n}\sum_{l=1}^{T}V_{Nit}(u)G(U_{it},u)\right]^2 = O(h^4). \quad (5.15)$$

Hence, the prove of Lemma 5.1 is completed.

Lemma 5.2. *Suppose that conditions* (A1)–(A8) *hold. Then*

(a) $\omega_{it}(u) = o(1)$. (5.16)

(b) $(M^\tau PM)^{-1} = (M^\tau M)^{-1} + O(\zeta_n) = T^{-1}I_n + O(\zeta_n)$ (5.17)

for sufficiently large n. *where* $\zeta_n = \frac{\sqrt{lnn}}{nh}i_n i_n^\tau$.

Proof. (a) Similar to the proof of (5.14), we have

$$E[V_{Nit}(U_{it})] = N^{-1}h^4 f(u)\vartheta_2[1+O_2(h+(Nh)^{-1/2})]. \quad (5.18)$$

Combining (5.9), we can prove (5.16).

(b) $M^\tau PM = M^\tau(I_{nT}-\Omega)^\tau(I_{nT}-\Omega)M = M^\tau M - M^\tau\Omega M - M^\tau\Omega^\tau M + M^\tau\Omega^\tau\Omega M = \Delta_{11}-\Delta_{12}-\Delta_{13}+\Delta_{14}$. By (a), we can show that the (i,j)th element of Δ_{1l} is $[\Delta_{1l}]_{ij} = o(1)$, $l = 2, 3, 4$. Similar to the proof of Lemma A.2 in Su and Ullah (2006). We can prove $(M^\tau PM)^{-1} = (M^\tau M)^{-1} + O(\zeta_n) = T^{-1}I_n + O(\zeta_n)$.

Lemma 5.3. *Suppose that conditions* (A1)–(A8) *hold. Then*

(a) $n^{-1}X^\tau PX \xrightarrow{p} \sum_{t=1}^{T}\left[(X_{it}-E(X_{it}|U_{it}))(X_{it}-E(X_{it}|U_{it}))^\tau\right]$.

(b) $n^{-1}X^\tau PM(M^\tau M)^{-1}M^\tau PX \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[(X_{it}-E(X_{it}|U_{it}))(X_{is}-E(X_{is}|U_{is}))^\tau\right]$.

(c) $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^\tau e_{it}^\tau\widetilde{H}\widetilde{X}_{it} \xrightarrow{p} \Psi$.

Proof. (a), (b) are the direct result of the Lemma A.3 in Su and Ullah (2006).

(c) By (a), (b) and Lemma 5.2,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^\tau e_{it}^\tau\widetilde{H}\widetilde{X}_{it} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^\tau e_{it}^\tau(I_{nT}-\widetilde{M}(\widetilde{M}^\tau\widetilde{M})^{-1}\widetilde{M}^\tau)\widetilde{X}_{it}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^\tau\widetilde{X}_{it} - \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{X}_{it}^\tau\widetilde{M}(\widetilde{M}^\tau\widetilde{M})^{-1}\widetilde{M}^\tau\widetilde{X}_{it}$$
$$= n^{-1}X^\tau PX - n^{-1}X^\tau PM(M^\tau M)^{-1}M^\tau PX \xrightarrow{p} \Psi.$$

Lemma 5.4. *Suppose that conditions* (A1)–(A8) *hold. Then*

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

$$\text{(b) } n^{-1/2}X^\tau PM(M^\tau M)^{-1}M^\tau P\varepsilon \xrightarrow{p} n^{-1/2}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}(X_{it} - E(X_{it}|U_{it}))\varepsilon_{is} + o_P(1).$$

Proof. (a), (b) are the direct result of the Lemma A.6 in Su and Ullah (2006).

Lemma 5.5. *Under the conditions of Theorem 3.1, we have*

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

Proof. By the definition of $\psi_i(\beta)$, we have

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

By Lemma 5.3, it is easy to show $E(I_{1n}) = 0$, $Var(I_{1n}) = \Theta$, and I_{1n} satisfies the conditions of the *Cramér–Wold* device and the Lindeberg condition. Using the central limits theorem, we have $I_{1n} \xrightarrow{d} N(0, \Theta)$. Similar to the proof of lemma A.5 in Su and Ullah (2006), we have $\frac{1}{\sqrt{n}}\widetilde{X}^\tau(I - \widetilde{H})\widetilde{g}(U) = o_p(1)$. Then it is easy to get that

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

Therefore, the proof is completed.

Lemma 5.6. *Under the conditions of Theorem 3.1, Let* $B_{it} = \widetilde{X}_{it}e_{it}^\tau H$ *we have*

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

Proof.

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

By the law of large numbers, we can derive that $\Lambda_1 \xrightarrow{d} \Theta$. By Lemma 5.4 and Cauchy–Schwarz inequality, it can prove $\Lambda_j \xrightarrow{d} \Theta,\ j = 2, 3, 4$, Lemma 5.6 is proved.

Lemma 5.7. *Under the conditions of Theorem 3.1, we have*

$$\text{(i) } \max_{1\le i\le n}\|\psi_i(\beta)\| = o_p(n^{1/2}), \tag{5.19}$$

$$\text{(ii) } \lambda(\beta) = O_p(n^{-1/2}). \tag{5.20}$$

Proof. Similar to the arguments given by Owen (1988), from Eq. (2.6) and Lemmas 5.5 and 5.6, we can easily verify Eq. (5.20). We now prove (5.19).

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

For Π_1, note that $E[\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{X}_{it}e_{it}^{\tau}\widetilde{H}(I_{nT}-S)v_{it}\widetilde{X}_{is}e_{is}^{\tau}\widetilde{H}(I_{nT}-S)v_{is}]=\Theta_{s,s}+o(1)<\infty$, where $\Theta_{s,s}$ denotes the sth diagonal element of Θ. B_{it} is independent and identically distributed across the i index for each fixed t. By lemma 3 in Owen (1990), we obtain $\Pi_1=o_P(n^{1/2})$. Now see the second term Π_2, by the Markov inequality, we obtain

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

Therefore, $\Pi_2=o_p(n^{1/2})$ and $\max_{1\le i\le n}\|\psi_i(\beta)\|=o_p(n^{1/2})$.

Proof of Theorem 3.1. By (2.6), we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\beta)}{1+\lambda^{\tau}\psi_i(\beta)}=\frac{1}{n}\sum_{i=1}^{n}\psi_i(\beta_0)-\frac{1}{n}\sum_{i=1}^{n}\psi_i(\beta)\psi_i^{\tau}(\beta)\lambda+\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\beta)[\lambda^{\tau}\psi_i(\beta)]^2}{1+\lambda^{\tau}\psi_i(\beta)}=0. \tag{5.21}$$

Applying the Taylor expansion, from (2.5), we obtain that

$$\mathcal{LR}_n(\beta)=2\sum_{i=1}^{n}\left\{\lambda^{\tau}\psi_i(\beta)-\frac{[\lambda^{\tau}\psi_i(\beta)]^2}{2}\right\}+o_p(1). \tag{5.22}$$

In view of Lemmas 5.5–5.7, we have

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\psi_i(\beta)[\lambda^{\tau}\psi_i(\beta_0)]^2}{1+\lambda^{\tau}\psi_i(\beta)}\right\| &\le \frac{1}{n}\sum_{i=1}^{n}\frac{\|\psi_i(\beta)\|^3\|\lambda\|^2}{|1+\lambda^{\tau}\psi_i(\beta)|} \\ &\le \|\lambda\|^2\max_{1\le i\le n}\|\psi_i(\beta)\|\frac{1}{n}\sum_{i=1}^{n}\|\psi_i(\beta)\|^2=O_p(n^{-1})o_p(n^{1/2})O_p(1)=o_p(n^{-1/2}),\end{aligned}$$

which, together with (5.21), yields that $\sum_{i=1}^{n}[\lambda^{\tau}\psi_i(\beta)]^2=\sum_{i=1}^{n}\lambda^{\tau}\psi_i(\beta)+o_p(1)$ and

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

Then, by (5.22), we have

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

which, combining with Lemmas 5.5 and 5.6, proves Theorem 3.1.

Proof of Theorem 3.2. Following the similar arguments as were used in the proof of Theorem 2 in Xue and Zhu (2008), we have

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

By Lemma 5.3, we can prove $\widehat{\Psi}\xrightarrow{P}\Psi$ by the law of large numbers. Together with Lemma 5.4 and the Slutsky Theorem, we complete the proof of Theorem 2.

Acknowledgments

The authors are grateful to the Editor and 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 Natural Science Foundation of China (11401006); the Financial engineering research and development center of Anhui Polytechnic University (JRGCKF201504). Anhui province higher education promotion plan scientific research general project (TSKJ2016B09).

References

Ahn, S.C., Schmidt, P., 2000. Estimation of linear panel data models using GMM. In: Mátyás, L. (Ed.), Generalized Method of Moments Estimation. Cambridge University Press, Cambridge.
Baltagi, B.H., 2005. Econometric Analysis of Panel Data, second ed. Wiley, New York.
Cai, Z., Li, Q., 2008. Nonparametric estimation and varying coefficient dynamic panel data models. Econometric Theory 24, 1321–1342.
Chiou, J.M., Müller, H.G., 1999. Nonparametric quasi-likelihood. Ann. Statist. 27, 36–64.
Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11 (6), 1031–1057.
Fan, G.L., Liang, H.Y., Huang, Z.S., 2012. Empirical likelihood for partially time-varying coefficient models with dependent observations. J. Nonparametr. Stat. 24, 71–84.
Henderson, D.j., Carroll, R.J., Li, Q., 2008. Nonparametric estimation and testing of fixed effects panel data models. J. Econometrics 144, 257–275.
Hjellvik, V., Chen, R., Tjøstheim, D., 2004. Nonparametric estimation and testing in panels of intercorrelated time series. J. Time Series Anal. 25, 831–872.
Horowitz, J.L., Lee, S., 2004. Semiparametric estimation of a panel data proportional hazards model with fixed effects. J. Econometrics 119, 155–198.
Hsiao, C., 2003. Analysis of Panel Data, second ed. Cambridge University Press.
Hu, X.M., 2014. Estimation in a semi-varying coefficient model for panel data with fixed effects. J. Syst. Sci. Complex. 27, 594–604.
Kaslow, R.A., Ostrow, D.G., Detels, J.P., Polk, B.F., Rinaldo, C.R., 1987. The multicenter AIDS cohort study: rational, oranization and selected characteristics of the prticipants. Am. J. Epidemiol. 126, 310–318.
Li, Q., Ullah, A., 1998. Estimating partially linear panel data models with one-way error components. Econometrica 17 (2), 145–166.
Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A.B., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90–120.
Qin, J., Lawless, B., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22, 300–325.
Roy, N., 1997. Nonparametric and semiparametric analysis of panel data models: An application to Calorie-Income relation for rural south India (Ph.D. thesis), University of California, Riverside.
Ruckstuhl, A.F., Welsh, A.H., Carroll, R.J., 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statist. Sinica 10, 51–71.
Shi, J., Lau, T.S., 2000. Empirical likelihood for partially linear models. J. Multivariate Anal. 72, 132–148.
Su, L.J., Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with models with fixed effects. Econom. Lett. 92, 75–81.
Wang, Q.H., Jing, B.Y., 1999. Empirical likelihood for partial linear models with fixed designs. Statist. Probab. Lett. 41, 425–433.
Xue, L.G., Zhu, L.X., 2007. Empirical likelihood semiparametric regression analysis for longitudinal data. Biometrika 94 (4), 921–937.
Xue, L.G., Zhu, L.X., 2008. Empirical likelihood-based inference in a partially linear model for longitudinal data. Sci. China Ser. A 51 (1), 115–130.
Yang, S.C., 2001. Moment inequalities for the partial sums of random variables. Sci. China Ser. A 44 (1), 1–6.
You, J.H., Chen, G.M., Zhou, Y., 2006. Block empirical likeliood for longitudinal partially linear regression models. Canad. J. Statist. 34 (1), 79–96.
You, J.H., Zhou, Y., 2006a. Statistical inference in a panel data semiparametric regression model with serially correlated errors. J. Multivariate Anal. 97, 844–873.
You, J.H., Zhou, Y., 2006b. Empirical likeliood for semiparametric varying-coefficient partially linear regression models. Statist. Probab. Lett. 76, 412–422.
Zhou, X.C., Lin, J.G., 2014. Empirical likeliood for varying-coefficient semiparametric mixed-effects errors-in-variables models with longitudinal data. Stat. Methods Appl. 23, 51–69.