The Econometrics Journal

Econometrics Journal (2011), volume **14**, pp. 387–408.
doi: 10.1111/j.1368-423X.2011.00350.x

Non-parametric time-varying coefficient panel data models with fixed effects

DEGUI LI[†,‡], JIA CHEN[†,‡,§] AND JITI GAO[†,‡]

[†]*School of Economics, The University of Adelaide, Adelaide SA 5005, Australia.*
E-mail: degui.li08@gmail.com

[‡]*Department of Econometrics and Business Statistics, Monash University Caulfield Campus, VIC 3145, Australia.*
E-mail: jiti.gao@monash.edu

[§]*School of Mathematics and Physics, The University of Queensland, Brisbane QLD 4072, Australia.*
E-mail: jiachen1981@gmail.com

First version received: March 2010; final version accepted: April 2011

Summary This paper is concerned with developing a non-parametric time-varying coefficient model with fixed effects to characterize non-stationarity and trending phenomenon in a non-linear panel data model. We develop two methods to estimate the trend function and the coefficient function without taking the first difference to eliminate the fixed effects. The first one eliminates the fixed effects by taking cross-sectional averages, and then uses a non-parametric local linear method to estimate both the trend and coefficient functions. The asymptotic theory for this approach reveals that although the estimates of both the trend function and the coefficient function are consistent, the estimate of the coefficient function has a rate of convergence of $(Th)^{-1/2}$, which is slower than $(NTh)^{-1/2}$ as the rate of convergence for the estimate of the trend function. To estimate the coefficient function more efficiently, we propose a pooled local linear dummy variable approach. This is motivated by a least squares dummy variable method proposed in parametric panel data analysis. This method removes the fixed effects by deducting a smoothed version of cross-time average from each individual. It estimates both the trend and coefficient functions with a rate of convergence of $(NTh)^{-1/2}$. The asymptotic distributions of both of the estimates are established when T tends to infinity and N is fixed or both T and N tend to infinity. Both the simulation results and real data analysis are provided to illustrate the finite sample behaviour of the proposed estimation methods.

Keywords: *Fixed effects*, *Local linear estimation*, *Non-stationarity*, *Panel data*, *Time-varying coefficient function*.

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 and biology. 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 non-linear panel data models (see Baltagi, 1995, Arellano, 2003, and Hsiao, 2003). However, it is known that the parametric panel data models may be misspecified, and estimators obtained from misspecified models are often inconsistent. To address such issues, some non-parametric methods have been used in both panel data model estimation and specification testing (see, e.g., Ullah and Roy, 1998, Lin and Ying, 2001, Fan and Li, 2004, Hjellvik et al., 2004, Cai and Li, 2008, Henderson et al., 2008, Mammen et al., 2009, and Zhang et al., 2009).

Meanwhile, trending econometric modelling of non-stationary processes has also gained a great deal of attention in recent years. For example, it is generally believed that the increase in carbon dioxide emissions through the twentieth century has caused global warming problem and it is important to model the trend of the global temperature. Some existing literature, such as Gao and Hawthorne (2006), revealed that the parametric linear trend does not approximate well the behaviour of global temperature data. Hence, non-parametric modelling of the trending phenomenon has since attracted interest. One of the key features of the non-parametric trending model is that it allows for the data to 'speak for themselves' with regard to choosing the form of the trend. For the recent development in non-parametric and semiparametric trending modelling of time series or panel data, see Gao and Hawthorne (2006), Cai (2007), Gao (2007), Robinson (2008), Atak et al. (2011) and the references therein.

While there is a rich literature on parametric and non-parametric time-varying coefficient time series models (Robinson, 1989, Phillips, 2001, and Cai, 2007), as far as we know, little work has been done in the panel data case. The recent work by Robinson (2011) may be among the first to introduce a trending time-varying model for the panel data case where cross-sectional dependence is incorporated. In both theory and applications, various explanatory variables are of significant interest when modelling the trend of a panel data. Thus, it may be more informative and useful to add such explanatory variables into a time-varying panel data model when modelling the trend of a panel data.

This paper thus proposes using a non-parametric trending time-varying coefficient panel data model of the form

$$\begin{aligned} Y_{it} &= f_t + \sum_{j=1}^{d} \beta_{t,j} X_{it,j} + \alpha_i + e_{it} \\ &= f_t + X_{it}^{\top}\beta_t + \alpha_i + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \end{aligned} \tag{1.1}$$

where $X_{it} = (X_{it,1}, \ldots, X_{it,d})^{\top}$, $\beta_t = (\beta_{t,1}, \ldots, \beta_{t,d})^{\top}$, all β_t and f_t are unknown functions, $\{\alpha_i\}$ reflects unobserved individual effect, and $\{e_{it}\}$ is stationary and weakly dependent for each i and independent of $\{X_{it}\}$ and $\{\alpha_i\}$, T is the time series length and N is the cross section size.

Model (1.1) is called a fixed effects model when $\{\alpha_i\}$ is allowed to be correlated with $\{X_{it}\}$ with an unknown correlation structure. Model (1.1) is called a random-effects model when $\{\alpha_i\}$ is uncorrelated with $\{X_{it}\}$. For the purpose of identification, we assume that

$$\sum_{i=1}^{N} \alpha_i = 0, \tag{1.2}$$

throughout the paper.

Model (1.1) includes many interesting non-parametric panel data models. For example, when $\{\beta_t\}$ does not vary over time and reduces to a vector of constants, model (1.1) becomes a semiparametric trending panel data model with fixed effects. When $\beta_t \equiv \overline{\mathbf{0}}_d$ ($\overline{\mathbf{0}}_d$ is a

d-dimensional null vector), model (1.1) reduces to a non-parametric trending panel data model as discussed in Robinson (2011), which allows for cross-sectional dependence for $\{e_{it}\}$.

The aim of this paper is to construct consistent estimates for the time trend f_t and time-varying coefficient vector β_t before we establish asymptotic properties for the estimates. As in Robinson (1989, 2011) and Cai (2007), we suppose that the time trend function f_t and the coefficient vector β_t satisfy

$$f_t = f\left(\frac{t}{T}\right) \quad \text{and} \quad \beta_{t,j} = \beta_j\left(\frac{t}{T}\right), \quad t = 1, \ldots, T, \tag{1.3}$$

where $f(\cdot)$ and $\beta_j(\cdot)$ are unknown smooth functions.

In this paper, we consider two classes of local linear estimates. As $\sum_{i=1}^{N}\alpha_i = 0$, the first method eliminates the fixed effects by taking cross-sectional averages, and we call it the averaged local linear method. We establish asymptotic distributions for the resulting estimates of $f(\cdot)$ and $\beta(\cdot)$ under mild conditions. The asymptotic results reveals that as both T and N tend to infinity, the rate of convergence for the estimate of the coefficient function $\beta(\cdot)$ is $O_P((Th)^{-1/2})$ while the rate of convergence for the estimate of the trend function $f(\cdot)$ is $O_P((NTh)^{-1/2})$. To improve the rate of convergence for the coefficient function, a local linear dummy variable approach is proposed. This is motivated by a least squares dummy variable method proposed in parametric panel data analysis (see, e.g. chapter 3 of Hsiao, 2003). This method removes the fixed effects by deducting a smoothed version of cross-time average from each individual. As a consequence, it is shown that the rate of convergence of $O_P[(NTh)^{-1/2}]$ for both of the estimates is achievable. The simulation study in Section 3 confirms that the local linear dummy variable estimate of $\beta(\cdot)$ outperforms the averaged local linear estimate.

The rest of this paper is organized as follows. Two classes of local linear estimates as well as their asymptotic distributions are given in Section 2. The simulated example is provided in Section 3. A real data example is given in Section 4. Section 5 summarizes some conclusions and discusses future research. All the mathematical proofs of the asymptotic results are given in Appendix.

2. NON-PARAMETRIC ESTIMATION METHOD AND ASYMPTOTIC THEORY

In this section, we introduce two classes of local linear estimates and establish the asymptotic distributions of the proposed estimates. In Section 2.1, we consider the averaged local linear estimation method. Section 2.2 discusses the local linear dummy variable approach.

2.1. *Averaged local linear estimation*

To introduce the estimation method, we introduce some notation. Define

$$Y_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} Y_{it}, \quad X_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} X_{it} \quad \text{and} \quad e_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} e_{it}.$$

By taking averages over i and using $\sum_{i=1}^{N}\alpha_i = 0$, we have

$$Y_{\cdot t} = f_t + X_{\cdot t}^{\top}\beta_t + e_{\cdot t}, \quad t = 1, \ldots, T, \tag{2.1}$$

where the individual effects α_is are eliminated.

Letting $Y = (Y_{\cdot 1}, \ldots, Y_{\cdot T})^\top$, $f = (f_1, \ldots, f_T)^\top$, $B(X, \beta) = (X_{\cdot 1}^\top \beta_1, \ldots, X_{\cdot T}^\top \beta_T)^\top$ and $e = (e_{\cdot 1}, \ldots, e_{\cdot T})^\top$, model (2.1) can then be rewritten as the following vector form:

$$Y = f + B(X, \beta) + e. \tag{2.2}$$

We then adopt the conventional local linear approach (see, e.g. Fan and Gijbels, 1996) to estimate

$$\beta_*(\cdot) = (f(\cdot), \beta_1(\cdot), \ldots, \beta_d(\cdot))^\top .$$

For given $0 < \tau < 1$, define

$$M(\tau) = \begin{pmatrix} 1 & X_{\cdot 1}^\top & \frac{1-\tau T}{Th} & \frac{1-\tau T}{Th} X_{\cdot 1}^\top \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{\cdot T}^\top & \frac{T-\tau T}{Th} & \frac{T-\tau T}{Th} X_{\cdot T}^\top \end{pmatrix}$$

and

$$W(\tau) = \operatorname{diag}\left[K\left(\frac{1-\tau T}{Th}\right), \ldots, K\left(\frac{T-\tau T}{Th}\right)\right].$$

Assuming that $\beta_*(\cdot)$ has continuous derivatives of up to the second order, by Taylor expansion we have

$$\beta_*\left(\frac{t}{T}\right) = \beta_*(\tau) + \beta_*'(\tau)\left(\frac{t}{T} - \tau\right) + O\left[\left(\frac{t}{T} - \tau\right)^2\right], \tag{2.3}$$

where $0 < \tau < 1$ and $\beta_*'(\cdot)$ is the derivative of $\beta_*(\cdot)$. Based on the local linear approximation in (2.3), $(\beta_*^\top(\tau), h[\beta'_*(\tau)]^\top)^\top$ can then be estimated by solving the optimization problem:

$$\arg \min_{a \in \mathbf{R}^{d+1}, b \in \mathbf{R}^{d+1}} [Y - M(\tau)(a^\top, b^\top)^\top]^\top W(\tau) [Y - M(\tau)(a^\top, b^\top)^\top]. \tag{2.4}$$

Following the standard argument, the local linear estimator of $\beta_*(\tau)$ is

$$\widehat{\beta}_*(\tau) = [I_{d+1},\ \mathbf{0}_{d+1}][M^\top(\tau) W(\tau) M(\tau)]^{-1} M^\top(\tau) W(\tau) Y, \tag{2.5}$$

where I_{d+1} is a $(d+1) \times (d+1)$ identity matrix and $\mathbf{0}_{d+1}$ is a $(d+1) \times (d+1)$ null matrix.

To establish asymptotic results, we need to introduce the following regularity conditions. Here and in the sequel, define $\mu_j = \int u^j K(u)\, du$ and $\nu_j = \int u^j K^2(u)\, du$ for $j = 0, 1, 2$.

ASSUMPTION 2.1. *The probability kernel function $K(\cdot)$ is symmetric and Lipschitz continuous with a compact support* $[-1, 1]$.

ASSUMPTION 2.2 (a) $\{(X_i, e_i), i \geq 1\}$ *is a sequence of independent and identically distributed (i.i.d.) variables, where $X_i = (X_{it}, t \geq 1)$ and $e_i = (e_{it}, t \geq 1)$. Furthermore, for each $i \geq 1$, $\{(X_{it}, e_{it}), t \geq 1\}$ is stationary and α–mixing with mixing coefficient α_k satisfying $\alpha_k = O(k^{-\tau})$, where $\tau > \frac{(\delta+2)}{\delta}$ for some $\delta > 0$ involved in (b) below.* (*b*) *$E(X_{it}) = \overline{\mathbf{0}}_d$ and there exists a positive definite matrix $\Sigma_X := E(X_{it} X_{it}^\top)$. Furthermore, $E(\|X_{it}\|^{2(2+\delta)}) < \infty$*, where $\|\cdot\|$ is the L_2-distance. (*c*) *The error process $\{e_{it}\}$ is independent of $\{X_{it}\}$ with $E[e_{it}] = 0$ and $\sigma_e^2 = E[{e_{it}}^2]$*

$< \infty$. *Furthermore,* $E[|e_{it}|^{2+\delta}] < \infty$ *and* $\sum_{t=-\infty}^{\infty} c_X(t)c_e(t)$ *is positive definite, where* $c_X(t) = E(X_{is}X_{i,s+t}^\top)$ *and* $c_e(t) = E(e_{is}e_{i,s+t})$.

ASSUMPTION 2.3. *The trend function* $f(\cdot)$ *and the coefficient function* $\beta(\cdot)$ *have continuous derivatives of up to the second order.*

ASSUMPTION 2.4. *The bandwidth* h *satisfies that* $h \to 0$ *and* $Th \to \infty$ *as* $T \to \infty$.

REMARK 2.1 (a) The conditions on the kernel function $K(\cdot)$ in Assumption 2.1 are imposed for brevity of our proofs and they can be weakened. For example, the compact support assumption can be removed if we impose certain restriction on the tail of the kernel function. In Assumption 2.2 (a), we assume that $\{(X_i, e_i), i \geq 1\}$ is cross-sectionally independent and each time series component is α-mixing. Such assumptions are reasonable and verifiable and cover many linear and non-linear time series models (see, e.g. Fan and Yao, 2003, Gao, 2007, and Li and Racine, 2007). Assumption 2.2 (c) imposes the homoscedasticity assumption on $\{e_{it}\}$. In fact, the independence between $\{e_{it}\}$ and $\{X_{it}\}$ may be weakened through allowing $e_{it} = \sigma(X_{it}, t)\epsilon_{it}$, where $\sigma(x, t)$ is a positive function and Lipschitz-continuous in t, and $\{\epsilon_{it}\}$ satisfies Assumption 2.2 (c) with $E[\epsilon_{it}] = 0$ and $E[\epsilon_{it}^2] = 1$. (b) Both Assumptions 2.3 and 2.4 are mild common conditions on the smoothness of the functions and the bandwidth involved in the local linear fitting for the fixed-design case. Since both the stationarity and mixing condition are only required to impose on the parametric regressors $\{X_{it}\}$, and the non-parametric kernel estimation is only involved in the fixed-design case, there is no need to impose any kind of relationship between the mixing coefficient α_k and the bandwidth h in Assumption 2.4.

Define

$$\Delta_X^* = \begin{pmatrix} 1 & \bar{\mathbf{0}}_d^\top \\ \bar{\mathbf{0}}_d & \Sigma_X \end{pmatrix}, \quad \Lambda_X = \nu_0 \begin{bmatrix} \sigma_e^2 + 2\sum_{t=1}^{\infty} c_e(t) & \bar{\mathbf{0}}_d^\top \\ \bar{\mathbf{0}}_d & \sum_{t=-\infty}^{\infty} c_X(t)c_e(t) \end{bmatrix}.$$

We state an asymptotic distribution for $\widehat{\beta}_*(\tau)$ in the following theorem.

THEOREM 2.1. *Consider models (1.1) and (1.2). Suppose that Assumptions 2.1–2.4 are satisfied. Then, as* $T \to \infty$,

$$D_{NT}[\widehat{\beta}_*(\tau) - \beta_*(\tau) - b(\tau)h^2 + o_P(h^2)] \xrightarrow{d} N\left(\bar{\mathbf{0}}_{d+1}, \Delta_X^{*-1}\Lambda_X\Delta_X^{*-1}\right) \tag{2.6}$$

for given $0 < \tau < 1$, *where* $D_{NT} = \mathrm{diag}(\sqrt{NTh}, \sqrt{Th}I_d)$, $b(\tau) = \frac{1}{2}\mu_2\beta_*''(\tau)$ and $\beta_{\prime *}'(\cdot) = [f''(\cdot), \beta_{\prime 1}'(\cdot), \ldots, \beta''{}_d(\cdot)]^\top$. *In particular,*

$$\sqrt{NTh}[\widehat{f}(\tau) - f(\tau) - b_f(\tau)h^2 + o_P(h^2)] \xrightarrow{d} N\left\{0, \nu_0\left[\sigma_e^2 + 2\sum_{t=1}^{\infty} c_e(t)\right]\right\} \tag{2.7}$$

and

$$\sqrt{Th}[\widehat{\beta}(\tau) - \beta(\tau) - b_\beta(\tau)h^2 + o_P(h^2)] \xrightarrow{d} N\left\{\bar{\mathbf{0}}_d, \Sigma_X^{-1}\nu_0\left[\sum_{t=-\infty}^{\infty} c_X(t)c_e(t)\right]\Sigma_X^{-1}\right\}, \tag{2.8}$$

where $b_f(\tau) = \frac{1}{2}\mu_2 f''(\tau)$, $b_\beta(\tau) = \frac{1}{2}\mu_2\beta''(\tau)$ *and* $\beta''(\cdot) = [\beta_{\prime 1}'(\cdot), \ldots, \beta_{\prime d}'(\cdot)]^\top$.

REMARK 2.2 (a) The asymptotic distribution (2.6) is established by letting $T \to \infty$. This implies that two cases are included: (i) T tends to infinity and N is fixed and (ii) both T and N tend to infinity. Note that Assumption 2.4 is stronger than the condition $h \to 0$ and $TNh \to \infty$ as either $T \to \infty$ or both $T \to \infty$ and $N \to \infty$. Such a condition is reasonable and sufficient because the panel data series is being treated as a pooled time series with a sample size of NT in the construction of our estimation method. (b) The above theorem complements some existing results (see, e.g. Robinson, 1989, 2011, and Cai, 2007). Theorem 2.1 considers only the interior point in the interval (0, 1) and the case of the boundary points can be dealt with analogously (see, e.g. Theorem 4 in Cai, 2007). It can be seen from (2.7) and (2.8) that the rate of convergence of $\widehat{\beta}(\tau)$ is slower than that of $\widehat{f}(\tau)$, which is confirmed by the simulation study in Section 3.

2.2. *Local linear dummy variable approach*

As shown in Theorem 2.1, the rate of convergence of the averaged local linear estimate of $\beta(\cdot)$ is $(Th)^{-1/2}$. To get an estimate that has faster rate of convergence, we next propose a local linear dummy variable approach. Recently, Su and Ullah (2006) considered a profile likelihood dummy variable approach in partially linear models with fixed effects, and Sun et al. (2009) discussed a local linear dummy variable method in varying coefficient models with fixed effects.

Note that (1.1) can be rewritten as

$$\widetilde{Y} = \widetilde{f} + \widetilde{B}(X, \beta) + \widetilde{D}\alpha + \widetilde{e}, \tag{2.9}$$

where

$$\begin{aligned}
&\widetilde{Y} = \left(Y_1^\top, \ldots, Y_N^\top\right)^\top, \quad Y_i = (Y_{i1}, \ldots, Y_{iT})^\top, \\
&\widetilde{e} = \left(e_1^\top, \ldots, e_N^\top\right)^\top, \quad e_i = (e_{i1}, \ldots, e_{iT})^\top, \\
&\widetilde{f} = \overline{I}_N \otimes (f_1, \ldots, f_T)^\top = \overline{I}_N \otimes f, \\
&\widetilde{B}(X, \beta) = \left(X_{11}^\top \beta_1, \ldots, X_{1T}^\top \beta_T, X_{21}^\top \beta_1, \ldots, X_{NT}^\top \beta_T\right)^\top, \\
&\alpha = (\alpha_1, \ldots, \alpha_N)^\top, \quad \widetilde{D} = I_N \otimes \overline{I}_T,
\end{aligned}$$

in which $\overline{I}_k$ is a k-dimensional vector of ones and f is as defined in Section 2.1.

As $\sum_{i=1}^N \alpha_i = 0$, (2.9) can be further rewritten as

$$\widetilde{Y} = \widetilde{f} + \widetilde{B}(X, \beta) + \widetilde{D}^* \alpha^* + \widetilde{e}, \tag{2.10}$$

where $\alpha^* = (\alpha_2, \ldots, \alpha_N)^\top$ and $\widetilde{D}^* = (-\overline{I}_{N-1},\ I_{N-1})^\top \otimes \overline{I}_T$. Based on the Taylor expansion in (2.3), we have

$$\widetilde{f} + \widetilde{B}(X, \beta) \approx \widetilde{M}(\tau)\left\{\beta_*^\top(\tau),\ h[\beta_*'(\tau)]^\top\right\}^\top,$$

where $\beta_*(\cdot) = [f(\cdot), \beta_1(\cdot), \ldots, \beta_d(\cdot)]^\top$ and $\widetilde{M}^\top(\tau) = [M_1^\top(\tau), \ldots, M_N^\top(\tau)]$ with

$$M_i(\tau) = \begin{pmatrix} 1 & X_{i1}^\top & \dfrac{1 - \tau T}{Th} & \dfrac{1 - \tau T}{Th} X_{i1}^\top \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{iT}^\top & \dfrac{T - \tau T}{Th} & \dfrac{T - \tau T}{Th} X_{iT}^\top \end{pmatrix}.$$

The algorithm for the local linear dummy variable method is described as follows.

Define $\widetilde{W}(\tau) = I_N \otimes W(\tau)$. For given $0 < \tau < 1$, minimize

$$\left\{\widetilde{Y} - \widetilde{M}(\tau)\left[\beta_*^\top(\tau),\ h(\beta_*'(\tau))^\top\right]^\top - \widetilde{D}^*\alpha^*\right\}^\top \widetilde{W}(\tau)\left\{\widetilde{Y} - \widetilde{M}(\tau)\left[\beta_*^\top(\tau),\ h(\beta_*'(\tau))^\top\right]^\top - \widetilde{D}^*\alpha^*\right\} \tag{2.11}$$

with respect to $\{\beta_*^\top(\tau), h[\beta_*{}'(\tau)]^\top\}^\top$ and α^*.

Taking derivative of (2.11) with respect to α^* and setting the result to zero, we obtain

$$\widehat{\alpha}^* := \widehat{\alpha}^*(\tau) = \left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\left\{\widetilde{Y} - \widetilde{M}(\tau)\left[\beta_*^\top(\tau),\ h\left(\beta_*'(\tau)\right)^\top\right]^\top\right\}.$$

Replacing α^* in (2.11) by $\widehat{\alpha}^*$, we obtain the concentrated weighted least squares:

$$\left\{\widetilde{Y} - \widetilde{M}(\tau)\left[\beta_*^\top(\tau),\ h\left(\beta_*'(\tau)\right)^\top\right]^\top\right\}^\top \widetilde{W}^*(\tau)\left\{\widetilde{Y} - \widetilde{M}(\tau)\left[\beta_*^\top(\tau),\ h\left(\beta_*'(\tau)\right)^\top\right]^\top\right\}, \tag{2.12}$$

where $\widetilde{W}^*(\tau) = \widetilde{\mathcal{K}}^\top(\tau)\widetilde{W}(\tau)\widetilde{\mathcal{K}}(\tau)$ and $\widetilde{\mathcal{K}}(\tau) = I_{NT} - \widetilde{D}^*\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)$. Observe that for any τ, $\widetilde{\mathcal{K}}(\tau)\widetilde{D}^*\alpha^* = 0$. Hence, the fixed effects term $\widetilde{D}^*\alpha^*$ is eliminated in (2.12).

Minimizing (2.12) with respect to $\{\beta_*^\top(\tau), h[\beta_*{}'(\tau)]^\top\}^\top$, we obtain the estimate of $\beta_*(\tau)$ as

$$\widetilde{\beta}_*(\tau) = \left[I_{d+1},\ \mathbf{0}_{d+1}\right]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{Y}. \tag{2.13}$$

Then $\widetilde{\beta}_*(\tau)$ is called the local linear dummy variable estimator of $\beta_*(\tau)$ and its asymptotic distribution is given in the following theorem.

THEOREM 2.2. *Consider models (1.1) and (1.2). Suppose that Assumptions 2.1–2.4 are satisfied. For given* $0 < \tau < 1$*, as* $T \to \infty$,

$$\sqrt{NTh}\left[\widetilde{\beta}_*(\tau) - \beta_*(\tau) - b(\tau)h^2 + o_P(h^2)\right] \xrightarrow{d} N\left(\overline{\mathbf{0}}_{d+1}, \Delta_X^{*-1}\Lambda_X\Delta_X^{*-1}\right). \tag{2.14}$$

In particular,

$$\sqrt{NTh}\left[\widetilde{f}(\tau) - f(\tau) - b_f(\tau)h^2 + o_P(h^2)\right] \xrightarrow{d} N\left\{0,\ \nu_0\left[\sigma_e^2 + 2\sum_{t=1}^{\infty} c_e(t)\right]\right\} \tag{2.15}$$

and

$$\sqrt{NTh}\left[\widetilde{\beta}(\tau) - \beta(\tau) - b_\beta(\tau)h^2 + o_P(h^2)\right] \xrightarrow{d} N\left\{\overline{\mathbf{0}}_d,\ \Sigma_X^{-1}\nu_0\left[\sum_{t=-\infty}^{\infty} c_X(t)c_e(t)\right]\Sigma_X^{-1}\right\}. \tag{2.16}$$

REMARK 2.3. As in Theorem 2.1, Theorem 2.2 covers the case where N is either fixed or N goes to infinity. Moreover, the local linear dummy variable estimates of both $f(\cdot)$ and $\beta(\cdot)$ have a rate of convergence of $(NTh)^{-1/2}$. This implies that the local linear dummy variable estimate of $\beta(\cdot)$ is asymptotically more efficient than the averaged local linear estimate of $\beta(\cdot)$. In addition, as the averaged local linear method, the individual fixed effects are eliminated in the estimation procedure without taking the first difference, and the fixed effects do not affect the asymptotic distributions of the two estimates.

In Section 3 below, we provide a simulated example to compare the small sample behaviour of the proposed local linear estimation methods with that of the ordinary local linear approach ignoring the fixed effects in non-parametric random effects panel data models.

3. SIMULATION

Throughout this section, we use a quadratic kernel function $K(u) = \frac{3}{4}(1 - x^2)I(|x| \leq 1)$. In the averaged local linear estimation method, a conventional leave-one-out cross validation method is used to the averaged model (2.1) to select an optimal bandwidth. In contrast, in the local linear dummy variable method, a leave-one-unit-out cross validation method, which was proposed by Sun et al. (2009), is applied to select an optimal bandwidth, as the conventional leave-one-out cross validation method fails to provide satisfactory results due to the existence of the fixed effects. See Sun et al. (2009) for more details on this leave-one-unit-out method.

Consider a trending time-varying coefficient model of the form

$$Y_{it} = f(t/T) + \beta(t/T)X_{it} + \alpha_i + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{3.1}$$

where $f(u) = u^2 + u + 1$, $\beta(u) = \sin(\pi u)$, $\{X_{it}\}$ is generated by the AR (1) process

$$X_{it} = \frac{1}{2}X_{i,t-1} + x_{it}, \ t \geq 1, \quad X_{i,0} = 0,$$

$\{x_{it}, t \geq 1\}$ is a sequence of independent and identically distributed $N(0, 1)$ random variables, $\{x_{it}, t \geq 1\}$ is independent of $\{x_{jt}, t \geq 1\}$ for $i \neq j$,

$$\alpha_i = \theta_0 X_{i\cdot} + u_i, \quad i = 1, \ldots, N-1, \quad \alpha_N = -\sum_{i=1}^{N-1} \alpha_i, \tag{3.2}$$

$X_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, $\theta_0 = 0, 1, 2$, $\{u_i, i \geq 1\}$ is a sequence of independent and identically distributed $N(0, 1)$ random variables, $\{e_{it}\}$ is an AR(1) process generated by

$$e_{it} = \frac{1}{2}e_{i,t-1} + \eta_{it},$$

$\{\eta_{it}, t \geq 1\}$ is a sequence of independent and identically distributed $N(0, 1)$ random variables, $\{\eta_{it}, t \geq 1\}$ is independent of $\{\eta_{jt}, t \geq 1\}$ for $i \neq j$, and $\{\eta_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$ is independent of $\{X_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$.

Model (3.1) is similar to the simulated example in Sun et al. (2009) and they studied the case of random design varying coefficient panel data models. It is easy to check that (3.1) becomes the random effects case when $\theta_0 = 0$ in (3.2). Otherwise ($\theta_0 = 1, 2$), model (3.1) is a fixed effects panel data model. We next compare three classes of local linear estimation methods: the averaged local linear estimates (ALLE), local linear dummy variable estimates (LLDVE) and the ordinary local linear estimates (OLLE) ignoring the fixed effects. We compare the average mean squared

Table 1. AMSE for estimators of $f(\cdot)$ ($\theta_0 = 0$).

$N\backslash T$		10	20	30
10	ALLE	0.1008 (0.1345)	0.0482 (0.0458)	0.0359 (0.0300)
	LLDVE	0.0669 (0.0652)	0.0490 (0.0362)	0.0338 (0.0267)
	OLLE	0.0722 (0.0732)	0.0523 (0.0387)	0.0349 (0.0282)
20	ALLE	0.0523 (0.0587)	0.0264 (0.0267)	0.0241 (0.0181)
	LLDVE	0.0357 (0.0290)	0.0253 (0.0189)	0.0272 (0.0168)
	OLLE	0.0378 (0.0305)	0.0262 (0.0195)	0.0283 (0.0173)
30	ALLE	0.0389 (0.0478)	0.0221 (0.0197)	0.0133 (0.0108)
	LLDVE	0.0244 (0.0185)	0.0230 (0.0139)	0.0142 (0.0096)
	OLLE	0.0251 (0.0188)	0.0240 (0.0146)	0.0145 (0.0099)

Table 2. AMSE for estimators of $f(\cdot)$ ($\theta_0 = 1$).

$N\backslash T$		10	20	30
10	ALLE	0.1073 (0.1313)	0.0478 (0.0435)	0.0350 (0.0274)
	LLDVE	0.0750 (0.0630)	0.0440 (0.0349)	0.0322 (0.0229)
	OLLE	0.0928 (0.0821)	0.0489 (0.0386)	0.0334 (0.0239)
20	ALLE	0.0551 (0.0777)	0.0386 (0.0370)	0.0263 (0.0210)
	LLDVE	0.0360 (0.0294)	0.0337 (0.0199)	0.0210 (0.0149)
	OLLE	0.0412 (0.0331)	0.0354 (0.0213)	0.0217 (0.0153)
30	ALLE	0.0362 (0.0372)	0.0222 (0.0211)	0.0224 (0.0192)
	LLDVE	0.0261 (0.0200)	0.0178 (0.0127)	0.0136 (0.0095)
	OLLE	0.0292 (0.0226)	0.0184 (0.0130)	0.0138 (0.0096)

errors (AMSE) of the three estimators. The AMSE of an estimator $\widehat{f}$ of f is defined as

$$\mathrm{AMSE}(\widehat{f}) = \frac{1}{R}\sum_{r=1}^{R}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{f}_{(r)}(t/T) - f(t/T)\right]^2\right\},$$

where $\widehat{f}_{(r)}(\cdot)$ denotes the estimate of $f(\cdot)$ in the rth replication and R is the number of replications which is chosen as $R = 500$ in our simulation. Similarly, the AMSE of an estimator of β is

$$\mathrm{AMSE}(\widehat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{\beta}_{(r)}(t/T) - \beta(t/T)\right]^2\right\}.$$

The simulation results for estimates of $f(\cdot)$ and $\beta(\cdot)$ with different values of N and T ($N, T =$ 10, 20, 30) and with both random effects ($\theta_0 = 0$) and fixed effects ($\theta_0 = 1$, 2) are summarized in Tables 1–6. Tables 1–3 contain the AMSEs and their standard deviations (in parentheses) of the three estimates of $f(\cdot)$ for the three cases of $\theta_0 = 0$, 1 and 2, and those of $\beta(\cdot)$ are listed in Tables 4–6.

From Tables 1–6, we can see that the performances of all the three estimates of the trend function $f(\cdot)$ are satisfactory even when both N and T are small. And all of the three estimates of $f(\cdot)$ improve as either T or N increases. However, when we observe the simulation results for

Table 3. AMSE for estimators of $f(\cdot)$ ($\theta_0 = 2$).

$N \backslash T$		10	20	30
10	ALLE	0.0991 (0.1124)	0.0505 (0.0465)	0.0367 (0.0310)
	LLDVE	0.0703 (0.0573)	0.0654 (0.0379)	0.0335 (0.0289)
	OLLE	0.1100 (0.1559)	0.0819 (0.0590)	0.0358 (0.0308)
20	ALLE	0.0529 (0.0610)	0.0290 (0.0261)	0.0274 (0.0228)
	LLDVE	0.0361 (0.0282)	0.0342 (0.0209)	0.0168 (0.0123)
	OLLE	0.0481 (0.0402)	0.0383 (0.0236)	0.0179 (0.0141)
30	ALLE	0.0417 (0.0589)	0.0190 (0.0164)	0.0133 (0.0113)
	LLDVE	0.0262 (0.0201)	0.0179 (0.0123)	0.0136 (0.0090)
	OLLE	0.0353 (0.0305)	0.0197 (0.0135)	0.0145 (0.0094)

Table 4. AMSE for estimators of $\beta(\cdot)$ ($\theta_0 = 0$).

$N \backslash T$		10	20	30
10	ALLE	0.7644 (1.2696)	0.3171 (0.3169)	0.2474 (0.1827)
	LLDVE	0.0852 (0.0394)	0.0292 (0.0212)	0.0276 (0.0183)
	OLLE	0.1520 (0.1425)	0.0841 (0.0775)	0.0629 (0.0577)
20	ALLE	0.8655 (1.1818)	0.3220 (0.3303)	0.3006 (0.3583)
	LLDVE	0.0280 (0.0190)	0.0169 (0.0122)	0.0139 (0.0079)
	OLLE	0.0784 (0.0783)	0.0467 (0.0382)	0.0508 (0.0463)
30	ALLE	1.0126 (1.3883)	0.3972 (0.3986)	0.2111 (0.1966)
	LLDVE	0.0195 (0.0133)	0.0137 (0.0072)	0.0081 (0.0050)
	OLLE	0.0511 (0.0450)	0.0426 (0.0358)	0.0248 (0.0214)

Table 5. AMSE for estimators of $\beta(\cdot)$ ($\theta_0 = 1$).

$N \backslash T$		10	20	30
10	ALLE	0.7638 (1.0143)	0.3235 (0.3252)	0.2457 (0.1701)
	LLDVE	0.0543 (0.0428)	0.0376 (0.0244)	0.0287 (0.0188)
	OLLE	0.2029 (0.2430)	0.0942 (0.0965)	0.0616 (0.0649)
20	ALLE	0.9463 (1.3754)	0.5140 (0.6736)	0.2929 (0.3756)
	LLDVE	0.0291 (0.0203)	0.0189 (0.0107)	0.0123 (0.0081)
	OLLE	0.1237 (0.1125)	0.0790 (0.0588)	0.0388 (0.0337)
30	ALLE	0.8382 (0.9442)	0.3929 (0.4229)	0.3202 (0.3308)
	LLDVE	0.0198 (0.0130)	0.0116 (0.0074)	0.0083 (0.0054)
	OLLE	0.1044 (0.0906)	0.0442 (0.0407)	0.0271 (0.0220)

estimates of $\beta(\cdot)$, we find that when keep T fixed, the performance of the average local linear estimate of the coefficient function $\beta(\cdot)$ does not necessarily improve as N increases. But as T increases, its performance improves significantly. In contrast, the local linear dummy variable estimate and the ordinary local linear estimate perform better and better when either N or T increases. This confirms the asymptotic theory given in Section 2: the rate of convergence of

Table 6. AMSE for estimators of $\beta(\cdot)$ ($\theta_0 = 2$).

$N \backslash T$		10	20	30
10	ALLE	0.8282 (1.3952)	0.3184 (0.3061)	0.1806 (0.1707)
	LLDVE	0.0579 (0.0467)	0.0378 (0.0233)	0.0361 (0.0206)
	OLLE	0.4129 (0.4169)	0.2416 (0.2895)	0.0717 (0.0686)
20	ALLE	0.8852 (1.3935)	0.3420 (0.3886)	0.3376 (0.3589)
	LLDVE	0.0276 (0.0191)	0.0188 (0.0108)	0.0164 (0.0100)
	OLLE	0.3279 (0.2615)	0.1535 (0.1214)	0.0548 (0.0572)
30	ALLE	0.9106 (1.3650)	0.3009 (0.3011)	0.2032 (0.1814)
	LLDVE	0.0202 (0.0138)	0.0112 (0.0072)	0.0086 (0.0057)
	OLLE	0.2981 (0.2168)	0.1012 (0.0846)	0.0525 (0.0417)

$(Th)^{-1/2}$ for the averaged local linear estimate of $\beta(\cdot)$ is slower than the rate of convergence of $(NTh)^{-1/2}$ for the local linear dummy variable estimate. However, the averaged local linear estimate and the local linear dummy variable estimate of $f(\cdot)$ have the same rate of convergence of $(NTh)^{-1/2}$.

The simulation results also reveal that in the random effects case ($\theta_0 = 0$), the performances of the local linear dummy variable estimate and the ordinary local linear estimator are comparable with the local linear dummy variable estimate performing slightly better than the ordinary local linear estimate. As θ_0 increases, the performance of the ordinary local linear estimate becomes worse, while the performance of the dummy variable local linear dummy variable estimate is not influenced by the increase of θ_0 and remains quite stable and satisfactory.

4. REAL DATA ANALYSIS

We analyse a climate data set from the UK met office web site: *http://www.metoffice.gov.uk/climate/uk/stationdata/*. The data set contains monthly mean maximum temperatures (in Celsius degrees), mean minimum temperatures (in Celsius degrees), days of air frost (in days), total rainfall (in millimetres) and total sunshine duration (in hours) from 37 stations covering UK. Here we study the common trend in the mean maximum temperature series and the relationship of the mean maximum temperatures with total rainfall and total sunshine duration during the period of January/1999 to December/2008. Data from 16 stations are selected according to data availability (records start at different time for different stations and data for some part of the period Jan/1999–Dec/2008 are missing at some stations).

Denote by Y_{it} the log-transformed mean maximum temperature in t-th month in station i, and $X_{it,1}$ and $X_{it,2}$ the log-transformed total rainfall and total sunshine duration, respectively. Let $X_{it} = (X_{it,1}, X_{it,2})^\top$. As there exists seasonality in both Y and X, we first remove the seasonality from the observations of Y and X. We also remove the trends in X, as we assume in the paper that X is stationary. Let Y^*_{it} be the seasonally adjusted values of Y_{it} and X^*_{it} be the seasonally adjusted and detrended values of X_{it}. We then assume the following model:

$$Y^*_{it} = f(t/T) + \beta_1(t/T)X^*_{it,1} + \beta_2(t/T)X^*_{it,2} + \alpha_i + e_{it}, \quad 1 \le i \le N, 1 \le t \le T \qquad (4.1)$$

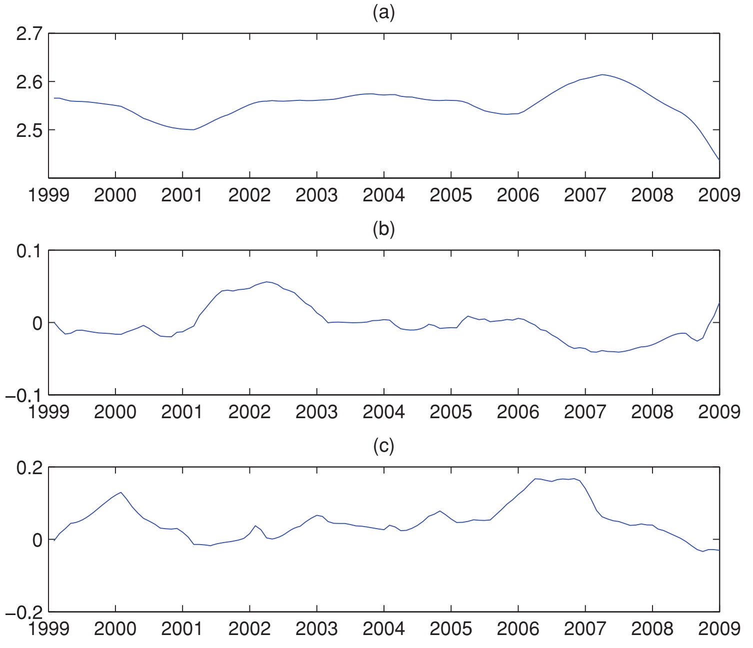

Figure 1. The local linear dummy variable estimates of (a) $f(\cdot)$, (b) $\beta_1(\cdot)$, (c) $\beta_2(\cdot)$.

with α_i being the station-specific effects and $i = 16$ and $T = 120$. Our purpose is to estimate $f(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ to get the common trend in the mean maximum temperature series from the 16 stations and to see the relationship between the mean maximum temperatures with the total rainfall and total sunshine duration. As evidenced by the simulation results in Section 3, the local linear dummy variable method generally outperforms the averaged local linear method. Hence, we use the local linear dummy variable method to estimate $f(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ in (4.1). The estimated curves are given in Figure 1(a)–(c).

The estimated trend curve in Figure 1(a) shows that from the beginning of 1999 to the end of 2000, there is a slight decrease in the monthly mean maximum temperatures. Thereafter, there is an overall upward trend from the beginning of 2001 to the end of 2006. Then from the beginning of 2007 to the end of 2008, there is a drop in the maximum temperatures. On the other hand, the estimated coefficient functions in Figure 1(b) and (c) do exhibit certain variance over time. And the coefficient function for the total sunshine duration ($\beta_2(\cdot)$) is generally above zero, which indicates that there is an overall positive relationship between the mean maximum temperatures and the total sunshine duration. In other words, longer sunshine duration tends to result in higher maximum temperatures.

5. CONCLUSION AND DISCUSSION

We have considered a non-parametric time-varying coefficient panel data model with fixed effects. Two classes of non-parametric estimates have been proposed and studied. The first one is based on an averaged local linear estimation method while the second one relies on a local linear dummy variable approach. Asymptotic distributions of the proposed estimates have been established with the second estimation method providing a faster rate of convergence than the first one. Both the simulation results and real data analysis have been provided to illustrate the asymptotic theory and support the finite-sample performance of the proposed estimates.

The asymptotic distribution of the averaged local linear and dummy variable estimators for the case of $E(X_{it}) \neq \overline{\mathbf{0}}_d$ can be obtained as corollaries of Theorems 2.1 and 2.2. Note that (1.1) can be rewritten as

$$Y_{it} = f_t + X_{it}^{\top}\beta_t + \alpha_i + e_{it} = f_t^* + X_{it}^{*\top}\beta_t + \alpha_i + e_{it}, \tag{5.1}$$

where $f_t^* = f_t + E(X_{it}^{\top})\beta_t$ and $X_{it}^* = X_{it} - E(X_{it})$. By Theorems 2.1 and 2.2, we can establish the asymptotic distributions of the averaged local linear estimator and dummy variable estimator of $(f^*(\tau), \beta^{\top}(\tau))$. Then by the Cramér-Wold device, we can further obtain asymptotic distributions for the two estimators of $(f(\tau), \beta^{\top}(\tau))$ by noting that

$$\left[f^*(\tau), \beta^{\top}(\tau)\right]^{\top} = \begin{bmatrix} 1 & E(X_{it}^{\top}) \\ \overline{\mathbf{0}}_d & I_d \end{bmatrix} \left[f(\tau), \beta^{\top}(\tau)\right]^{\top}.$$

There are some limitations in this paper. The first one is the assumption on cross-sectional independence. One future topic is to extend the discussion of Robinson (2011) to model (1.1) under cross-sectional dependence. The second one is that there is no endogeneity between $\{e_{it}\}$ and $\{X_{it}\}$. Another future topic is to accommodate such endogeneity in a general model.

ACKNOWLEDGMENTS

The authors would like to thank the Co-Editor Professor Jianqing Fan and two referees for their valuable comments, which substantially improved an earlier version of this paper. Thanks also go to Professor Peter Phillips for his constructive comments when the authors discussed the paper with him during his visit to Adelaide in February 2010. This project was supported by the Australian Research Council Discovery Grant Program under Grant Number: DP0879088.

REFERENCES

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

Atak, A., O. Linton and Z. Xiao (2011). A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. Forthcoming in *Journal of Econometrics*.

Baltagi, B. H. (1995). *Econometrics Analysis of Panel Data*. New York: John Wiley.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics 136*, 163–88.

Cai, Z. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory 24*, 1321–342.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, J. and R. Li (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association 99*, 710–23.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. London: Chapman and Hall/CRC.

Gao, J. and K. Hawthorne (2006). Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal 9*, 332–55.

Henderson, D., R. J., Carroll and Q. Li (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics 144*, 257–75.

Hjellvik, V., R. Chen and D. Tjøstheim (2004). Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis 25*, 831–72.

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Li, Q. and J. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

Lin, D. and Z. Ying (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association 96*, 103–26.

Mammen, E., B., Støve and D. Tjøstheim (2009). Nonparametric additive models for panels of time series. *Econometric Theory 25*, 442–81.

Peligrad, M. and S. Utev (1997). Central limit theorems for linear processes. *Annals of Probability 25*, 443–56.

Phillips, P. C. B. (2001). Trending time series and macroeconomic activity: some present and future challengers. *Journal of Econometrics 100*, 21–27.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: a Comparative Approach*. Cambridge: The MIT Press.

Robinson, P. (1989). Nonparametric estimation of time-varying parameters. In P. Hackl (Eds.), *Statistical Analysis and Forecasting of Economic Structural Change*, 164–253. Berlin: Springer.

Robinson, P. (2011). Nonparametric trending regression with cross-sectional dependence. Forthcoming in *Journal of Econometrics*.

Su, L. and A. Ullah (2006). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters 92*, 75–81.

Sun, Y., R. J. Carroll and D. Li (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics 25*, 101–29.

Ullah, A. and N. Roy (1998). Nonparametric and semiparametric econometrics of panel data. In A. Ullah and D. E. A. Giles (Eds.), *Handbook of Applied Economics and Statistics*, 579–604. New York: Marcel Dekker.

Zhang, W., J. Fan and Y. Sun (2009). A semiparametric model for cluster data. *Annals of Statistics 37*, 2377–408.

APPENDIX A: PROOFS OF THE MAIN RESULTS

PROPOSITION A.1. *Suppose that Assumptions 2.1, 2.2 (a), (b) and 2.4 are satisfied. Then, as $T, N \to \infty$ simultaneously,*

$$P_{NT}^{\top} M^{\top}(\tau) W(\tau) M(\tau) P_{NT} - \Lambda_\mu \otimes \Delta_X^* = o_P(1), \tag{A.1}$$

where $P_{NT} = \mathrm{diag}(\frac{1}{\sqrt{Th}}, \sqrt{\frac{N}{Th}} I_d, \frac{1}{\sqrt{Th}}, \sqrt{\frac{N}{Th}} I_d)$, $\Lambda_\mu = \mathrm{diag}(\mu_0, \mu_2)$ *and* Δ_X^* *is as defined in Section 2.*

Proof: Observe that

$$
\begin{aligned}
&P_{NT}^{\top} M^{\top}(\tau) W(\tau) M(\tau) P_{NT} \\
&\quad = \begin{bmatrix} \sum_{t=1}^{T} X_{\cdot t}^{*} X_{\cdot t}^{*\top} K\left(\frac{t-\tau T}{Th}\right) & \sum_{t=1}^{T} X_{\cdot t}^{*} X_{\cdot t}^{*\top} \left(\frac{t-\tau T}{Th}\right) K\left(\frac{t-\tau T}{Th}\right) \\ \sum_{t=1}^{T} X_{\cdot t}^{*} X_{\cdot t}^{*\top} \left(\frac{t-\tau T}{Th}\right) K\left(\frac{t-\tau T}{Th}\right) & \sum_{t=1}^{T} X_{\cdot t}^{*} X_{\cdot t}^{*\top} \left(\frac{t-\tau T}{Th}\right)^{2} K\left(\frac{t-\tau T}{Th}\right) \end{bmatrix},
\end{aligned}
$$

where $X_{\cdot t}^{*} = P_{NT}^{*}(1, X_{\cdot t}^{\top})^{\top}$ and $P_{NT}^{*} = \mathrm{diag}(\frac{1}{\sqrt{Th}}, \sqrt{\frac{N}{Th}} I_d)$.

We need only to prove that

$$
\sum_{t=1}^{T} X_{\cdot t}^{*} X_{\cdot t}^{*\top} K\left(\frac{t-\tau T}{Th}\right) = \mu_0 \Delta_X^{*} + o_P(1), \tag{A.2}
$$

since the proofs for the other components in $P_{NT}^{\top} M^{\top}(\tau) W(\tau) M(\tau) P_{NT}$ are analogous.

For simplicity, define $Q_t = X_{\cdot t}^{*} X_{\cdot t}^{*\top}$, $K_t = K(\frac{t-\tau T}{Th})$ and $X_{it}^{*} = P_{NT}^{*}(1, X_{it}^{\top})^{\top}$. By Assumption 2.2 (a) (b), we have

$$
\frac{Th}{N^2} E\left[\left(\sum_{i=1}^{N} X_{it}^{*}\right)\left(\sum_{i=1}^{N} X_{it}^{*}\right)^{\top}\right] = \begin{pmatrix} 1 & \bar{\mathbf{0}}_d^{\top} \\ \bar{\mathbf{0}}_d & \Sigma_X \end{pmatrix} = \Delta_X^{*}.
$$

By Assumption 2.1 and the above equation, it is easy to show that as $T, N \to \infty$,

$$
\begin{aligned}
\sum_{t=1}^{T} E\left[Q_t K_t\right] &= \frac{1}{N^2} \sum_{t=1}^{T} K_t \left(E\left[\left(\sum_{i=1}^{N} X_{it}^{*}\right)\left(\sum_{i=1}^{N} X_{it}^{*}\right)^{\top}\right]\right) \\
&= \frac{1}{Th} \sum_{t=1}^{T} K_t \left(\frac{Th}{N^2} E\left[\left(\sum_{i=1}^{N} X_{it}^{*}\right)\left(\sum_{i=1}^{N} X_{it}^{*}\right)^{\top}\right]\right) \\
&= \mu_0 \Delta_X^{*} \left(1 + O\left(\frac{1}{Th}\right)\right).
\end{aligned} \tag{A.3}
$$

We then consider the variance of $\sum_{t=1}^{T} Q_t K_t$. Note that

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{t=1}^{T} Q_t K_t\right) &= \sum_{t=1}^{T} \mathrm{Var}\left(Q_t K_t\right) + \sum_{t=1}^{T} \sum_{s \neq t} \mathrm{Cov}\left(Q_t K_t, Q_s K_s\right) \\
&=: \Pi_{NT}(1) + \Pi_{NT}(2).
\end{aligned}
$$

It follows from Assumption 2.2 (a) (b) that

$$
\sup_{1 \le t \le T} E\left[\left\| X_{\cdot t}^{*} \right\|^{2(2+\gamma)}\right] \le C(Th)^{-(2+\gamma)}, \tag{A.4}
$$

for some $0 \le \gamma \le \delta$ and some constant $C > 0$, where δ is as defined in Assumption 2.2, and we have used an existing inequality of the form $E[||\sum_{i=1}^{N} X_{it}||^{2(2+\gamma)}] \le N^{2+\gamma} E[||X_{1t}||^{2(2+\delta)}]$.

Furthermore, by Assumption 2.1 and (A.4) with $\gamma = 0$, we have

$$\Pi_{NT}(1) = \sum_{t=1}^{T} K_t^2 \mathrm{Var}(Q_t) = \sum_{t=1}^{T} K_t^2 \mathrm{Var}(X_{\cdot t}^* X_{\cdot t}^{*\top}) \leq \frac{C}{T^2h^2} \sum_{t=1}^{T} K_t^2 = O\left(\frac{1}{Th}\right). \tag{A.5}$$

Meanwhile, for $\Pi_{NT}(2)$, notice that

$$\Pi_{NT}(2) = \sum_{t=1}^{T} \sum_{0<|s-t|\leq q_T} \mathrm{Cov}(Q_t K_t, Q_s K_s) + \sum_{t=1}^{T} \sum_{|s-t|>q_T} \mathrm{Cov}(Q_t K_t, Q_s K_s) = \Pi_{NT}(3) + \Pi_{NT}(4), \tag{A.6}$$

where $q_T \to \infty$ and $q_T = o(Th)$.

As $\{X_i,\, i \geq 1\}$ is cross-sectional independent and for each i, and $\{X_{it},\, t \geq 1\}$ is stationary and α-mixing, both $\{X_{\cdot t}\}$ and $\{Q_t\}$ are still stationary and α-mixing with the same mixing coefficient α_k. As a consequence, we are able to employ some existing results for α-mixing processes.

By Assumption 2.1, 2.2 (a) and (b), (A.4) and Lemma A.1 of Gao (2007), we have

$$\Pi_{NT}(4) \leq C \sum_{t=1}^{T} K_t \sum_{|s-t|>q_T} \mathrm{Cov}\,(Q_t, Q_s) \leq \frac{C}{T^2h^2} \sum_{t=1}^{T} K_t \sum_{k>q_T} \alpha_k^{\delta/(2+\delta)} = o\left(\frac{1}{Th}\right), \tag{A.7}$$

as $q_T \to \infty$ and $\alpha_k = O(k^{-\tau})$ for $\tau > (\delta + 2)/\delta$.

Since $K(\cdot)$ is Lipschitz continuous by Assumption 2.1, we have

$$|K_t - K_s| \leq C\frac{q_T}{Th} \quad \text{when } |t - s| \leq q_T. \tag{A.8}$$

Standard calculation gives

$$\begin{aligned}\Pi_{NT}(3) &= \sum_{t=1}^{T} K_t^2 \sum_{0<|s-t|\leq q_T} \mathrm{Cov}\,(Q_t, Q_s) \\ &\quad - \sum_{t=1}^{T} K_t \sum_{0<|s-t|\leq q_T} (K_t - K_s)\mathrm{Cov}\,(Q_t, Q_s) \\ &=: \Pi_{NT}(5) + \Pi_{NT}(6).\end{aligned} \tag{A.9}$$

By (A.8) and the fact that $\sum_{t=2}^{\infty} \|\mathrm{Cov}(Q_1, Q_t)\| = O\left[(Th)^{-2}\right]$, we have

$$\|\Pi_{NT}(6)\| \leq \frac{Cq_T}{T^3h^3} \sum_{t=1}^{T} K_t = o\left(\frac{1}{Th}\right). \tag{A.10}$$

Furthermore, similarly to the proof of (A.5), we can show that

$$\|\Pi_{NT}(5)\| \leq \frac{C}{T^2h^2} \sum_{t=1}^{T} K_t^2 = O\left(\frac{1}{Th}\right). \tag{A.11}$$

In view of (A.5)–(A.7) and (A.9)–(A.11), we have as $Th \to \infty$

$$\mathrm{Var}\left(\sum_{t=1}^{T} Q_t K_t\right) = O\left(\frac{1}{Th}\right). \tag{A.12}$$

By (A.3) and (A.12), we have shown that (A.2) holds. The proof of Proposition A.1 is thus completed. □

PROPOSITION A.2. *Suppose that Assumptions 2.1, 2.2 (a) (b), 2.3 and 2.4 are satisfied. Then, as $T, N \to \infty$,*

$$I_{NT}(1) = \frac{1}{2}\mu_2 \beta_*''(\tau)h^2 + o_P(h^2). \tag{A.13}$$

Proof: By Assumption 2.3 and Taylor expansion, we have

$$\beta_*(t/T) \approx \beta_*(\tau) + \beta_*'(\tau)(t/T - \tau) + \frac{1}{2}\beta_*''(\tau)(t/T - \tau)^2.$$

Equation (A.13) then follows immediately from the definition of $I_{NT}(1)$, the above equation and Proposition A.1. □

PROPOSITION A.3. *Suppose that Assumptions 2.1, 2.2 and 2.4 are satisfied. Then, as $T, N \to \infty$,*

$$D_{NT} I_{NT}(2) \xrightarrow{d} N\left(\overline{\mathbf{0}}_{d+1}, \Delta_X^{*-1}\Lambda_X \Delta_X^{*-1}\right), \tag{A.14}$$

where D_{NT} and Λ_X are defined as in Theorem 2.1.

Proof: By Assumption 2.2 (a) (c), it is easy to check that

$$E\left[I_{NT}(2)\right] = \overline{\mathbf{0}}_{d+1}. \tag{A.15}$$

We then calculate the variance of $D_{NT}I_{NT}(2)$. Noticing that

$$\begin{aligned} N^2 E\left(X_{\cdot t} X_{\cdot t}^{\top} e_{\cdot t}^2\right) &= \frac{N^2}{N^4} E\left(\sum_{i=1}^{N} X_{it} \sum_{j=1}^{N} X_{jt}^{\top} \sum_{k=1}^{N} e_{kt} \sum_{l=1}^{N} e_{lt}\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} E(X_{it}X_{it}^{\top})E(e_{kt}^2) \\ &= \Sigma_X \sigma_e^2 \end{aligned}$$

and

$$N^2 \mathrm{Cov}\left(X_{\cdot t}e_{\cdot t}, X_{\cdot s}e_{\cdot s}\right) = E\left(X_{it}X_{is}^{\top}\right) E\left(e_{it}e_{is}\right) = c_X(|s-t|)c_e(|s-t|),$$

we have as $T, N \to \infty$,

$$\frac{N^2}{T}\mathrm{Var}\left(\sum_{t=1}^{T} X_{\cdot t} e_{\cdot t}\right) \longrightarrow \sum_{t=-\infty}^{\infty} c_X(t)c_e(t). \tag{A.16}$$

By (A.16), the fact that $\frac{1}{Th}\sum_{t=1}^{T} K_t^2 = \nu_0 + o(1)$ and following the proof of (A.12), we have

$$\frac{N^2}{Th}\mathrm{Var}\left(\sum_{t=1}^{T} K_t X_{\cdot t} e_{\cdot t}\right) = \nu_0\left[\sum_{t=-\infty}^{\infty} c_X(t)c_e(t)\right] + o(1). \tag{A.17}$$

Analogously, we have

$$\frac{N}{Th}\mathrm{Var}\left(\sum_{t=1}^{T} K_t e_{\cdot t}\right) = \nu_0\left[\sigma_e^2 + 2\sum_{t=1}^{\infty} c_e(t)\right] + o(1) \tag{A.18}$$

and

$$\frac{N^{3/2}}{Th} E\left(\sum_{t=1}^{T} K_t^2 X_{\cdot t} e_{\cdot t}^2\right) = \overline{\mathbf{0}}_d. \tag{A.19}$$

Furthermore, by (A.17)–(A.19) and Proposition A.1 and noting that

$$D_{NT} I_{NT}(2) = D_{NT}\,[I_{d+1},\ \mathbf{0}_{d+1}]\, P_{NT}\left[P_{NT}^{\top} M^{\top}(\tau) W(\tau) M(\tau) P_{NT}\right]^{-1} P_{NT}^{\top} M^{\top}(\tau) W(\tau) e,$$

we have

$$\mathrm{Var}\,[D_{NT} I_{NT}(2)] = \Delta_X^{*-1}\Lambda_X\Delta_X^{*-1} + o(1). \tag{A.20}$$

By Assumption 2.2, (A.15), (A.20) and Theorem 2.2 of Peligrad and Utev (1997), we have shown that (A.14) holds. □

Proof of Theorem 2.1: We only consider the case when T and N tend to infinity simultaneously. The proof for the case when T tends to infinity and N is fixed is similar and we therefore omit the details here.

Note that

$$\begin{aligned}\widehat{\beta}_*(\tau) &= [I_{d+1},\ \mathbf{0}_{d+1}]\left[M^{\top}(\tau)W(\tau)M(\tau)\right]^{-1} M^{\top}(\tau)W(\tau)Y \\ &= [I_{d+1},\ \mathbf{0}_{d+1}]\left[M^{\top}(\tau)W(\tau)M(\tau)\right]^{-1} M^{\top}(\tau)W(\tau)\,(B(X,\beta_*) + e)\end{aligned} \tag{A.21}$$

and

$$\begin{aligned}\widehat{\beta}_*(\tau) - \beta_*(\tau) =\ & \{[I_{d+1},\ \mathbf{0}_{d+1}][M^{\top}(\tau)W(\tau)M(\tau)]^{-1}M^{\top}(\tau)W(\tau)[f + B(X,\beta)] - \beta_*(\tau)\} \\ & + [I_{d+1},\ \mathbf{0}_{d+1}][M^{\top}(\tau)W(\tau)M(\tau)]^{-1}M^{\top}(\tau)W(\tau)e \\ =:\ & I_{NT}(1) + I_{NT}(2).\end{aligned} \tag{A.22}$$

Note that (2.7) and (2.8) can be derived from (2.6). The detailed proof of (2.6), follows from Propositions A.1–A.3. □

PROPOSITION A.4. *Suppose that Assumption 2.1, 2.2 (a), (b) and 2.4 are satisfied. Then, as $T, N \to \infty$ simultaneously,*

$$\frac{1}{NTh}\widetilde{M}^{\top}(\tau)\widetilde{W}^{*}(\tau)\widetilde{M}(\tau) - \Lambda_\mu \otimes \Delta_X^{*} = o_P(1), \tag{A.23}$$

where Λ_μ is defined as in Proposition A.1.

Proof: By the definition of $\widetilde{W}^{*}(\tau)$, we have

$$\widetilde{M}^{\top}(\tau)\widetilde{W}^{*}(\tau)\widetilde{M}(\tau) = \widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{M}(\tau) - \widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{D}^{*}\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^{*}\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{M}(\tau). \tag{A.24}$$

Following the proof of Proposition A.1, we have

$$\frac{1}{NTh}\widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{M}(\tau) = \Lambda_\mu \otimes \Delta_X^{*} + o_P(1). \tag{A.25}$$

In view of (A.24), we need only to prove

$$\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^* \left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1} \widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{M}(\tau) = o_P(NTh). \tag{A.26}$$

To do so, we first consider the term $\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*$. Define $Z_T = \sum_{t=1}^T K(\frac{t-T\tau}{Th})$. By the definition of $\widetilde{D}^*$ and standard calculation, we have

$$\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^* = \begin{pmatrix} 2Z_T & Z_T & \cdots & Z_T \\ \vdots & \vdots & & \vdots \\ Z_T & Z_T & \cdots & 2Z_T \end{pmatrix} = \begin{pmatrix} Z_T & Z_T & \cdots & Z_T \\ \vdots & \vdots & & \vdots \\ Z_T & Z_T & \cdots & Z_T \end{pmatrix} + \operatorname{diag}(Z_T, \ldots, Z_T).$$

Letting $A = \operatorname{diag}(Z_T, \ldots, Z_T)$, $B = (\sqrt{Z_T}, \ldots, \sqrt{Z_T})^\top$, $C = 1$ and $D = (\sqrt{Z_T}, \ldots, \sqrt{Z_T})$, then $\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^* = A + BCD$. Noticing that

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}, \tag{A.27}$$

which can be found in Poirier (1995), we have

$$\left(\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right)^{-1} = \begin{pmatrix} \frac{1}{Z_T} - \frac{1}{NZ_T} & -\frac{1}{NZ_T} & \cdots & -\frac{1}{NZ_T} \\ -\frac{1}{NZ_T} & \frac{1}{Z_T} - \frac{1}{NZ_T} & \cdots & -\frac{1}{NZ_T} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{NZ_T} & -\frac{1}{NZ_T} & \cdots & \frac{1}{Z_T} - \frac{1}{NZ_T} \end{pmatrix}. \tag{A.28}$$

Define $\overline{Z}_T(i) = \sum_{t=1}^T K(\frac{t-T\tau}{Th})X_{it}$ and $\widetilde{Z}_T(i) = \sum_{t=1}^T(\frac{t-T\tau}{Th})K(\frac{t-T\tau}{Th})X_{it}$. By standard arguments, we have

$$\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^* = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \overline{Z}_T(2) - \overline{Z}_T(1) & \overline{Z}_T(3) - \overline{Z}_T(1) & \cdots & \overline{Z}_T(N) - \overline{Z}_T(1) \\ 0 & 0 & \cdots & 0 \\ \widetilde{Z}_T(2) - \widetilde{Z}_T(1) & \widetilde{Z}_T(3) - \widetilde{Z}_T(1) & \cdots & \widetilde{Z}_T(N) - \widetilde{Z}_T(1) \end{bmatrix}.$$

which, together with (A.28), implies

$$\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^* \left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1} \widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{M}(\tau) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{NT} & 0 & C_{NT} \\ 0 & 0 & 0 & 0 \\ 0 & D_{NT} & 0 & B_{NT} \end{pmatrix}, \tag{A.29}$$

where

$$A_{NT} = \sum_{k=2}^{N} A_{NT}(k)[\overline{Z}_T(k) - \overline{Z}_T(1)]^\top,$$
$$B_{NT} = \sum_{k=2}^{N} B_{NT}(k)[\widetilde{Z}_T(k) - \widetilde{Z}_T(1)]^\top,$$
$$C_{NT} = \sum_{k=2}^{N} A_{NT}(k)[\widetilde{Z}_T(k) - \widetilde{Z}_T(1)]^\top,$$
$$D_{NT} = \sum_{k=2}^{N} B_{NT}(k)[\overline{Z}_T(k) - \overline{Z}_T(1)]^\top,$$

$$A_{NT}(k) = \overline{Z}_T(k)\frac{N-1}{NZ_T} - \frac{1}{NZ_T}\sum_{j=1,\neq k}^{N} \overline{Z}_T(j) = \frac{\overline{Z}_T(k)}{Z_T} - \frac{1}{NZ_T}\sum_{i=1}^{N}\overline{Z}_T(i) \text{ and}$$
$$B_{NT}(k) = \widetilde{Z}_T(k)\frac{N-1}{NZ_T} - \frac{1}{NZ_T}\sum_{j=1,\neq k}^{N} \widetilde{Z}_T(j) = \frac{\widetilde{Z}_T(k)}{Z_T} - \frac{1}{NZ_T}\sum_{i=1}^{N}\widetilde{Z}_T(i).$$

By the weak consistency of the non-parametric estimator in the fixed design case, we have for each $i \geq 1$,

$$\frac{1}{Th}Z_T = \mu_0 + o_P(1), \quad \frac{1}{Th}\overline{Z}_T(i) = o_P(1), \quad \frac{1}{Th}\widetilde{Z}_T(i) = o_P(1). \tag{A.30}$$

By (A.29) and (A.30), we have shown that (A.26) holds. The proof of Proposition A.4 is completed. □

PROPOSITION A.5. *Suppose that Assumptions 2.1, 2.2 (a) (b), 2.3 and 2.4 are satisfied. Then, as $T, N \to \infty$ simultaneously,*

$$\Xi_{NT}(1) = \frac{1}{2}\mu_2\beta_*''(\tau)h^2 + o_P(h^2) = b(\tau)h^2 + o_P(1). \tag{A.31}$$

Proof: By Assumption 2.3, Taylor expansion of $\beta_*(\cdot)$ at τ and Proposition A.4, the proof of (A.31) follows. □

PROPOSITION A.6. *Suppose that Assumptions 2.1, 2.2 and 2.4 are satisfied. Then, we have*

$$\sqrt{NTh}\,\Xi_{NT}(3) \xrightarrow{d} N\left(\overline{\mathbf{0}}_{d+1}, \Delta_X^{*-1}\Lambda_X\Delta_X^{*-1}\right) \tag{A.32}$$

as $T, N \to \infty$ simultaneously.

Proof: By Propositions A.4 and A.5, to prove (A.32), it suffices to prove

$$\frac{1}{\sqrt{NTh}}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{e} \xrightarrow{d} N\left(\overline{\mathbf{0}}_{2d+2}, \Lambda_\nu \otimes \Lambda_X\right), \tag{A.33}$$

where $\Lambda_\nu = \mathrm{diag}(\nu_0, \nu_2)$.

By the definition of $\widetilde{W}^*(\tau)$, we have

$$\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{e} = \widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{e} - \widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^*\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{e}. \tag{A.34}$$

By (A.34), it is enough to show that

$$\frac{1}{\sqrt{NTh}}\widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{e}\xrightarrow{d} N\left(\overline{\mathbf{0}}_{2d+2},\,\Lambda_{\nu}\otimes\Lambda_X\right). \tag{A.35}$$

and

$$\widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{D}^{*}\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^{*}\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{e}=o_P(\sqrt{NTh}). \tag{A.36}$$

Define

$$L_j\left(\frac{t-T\tau}{Th}\right)=\left(\frac{t-T\tau}{Th}\right)^{j}K\left(\frac{t-T\tau}{Th}\right)$$

for $j=0,1$.

To prove (A.35), we need only to prove

$$\frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}L_j\left(\frac{t-T\tau}{Th}\right)\begin{pmatrix}1\\X_{it}\end{pmatrix}e_{it}\xrightarrow{d} N\left(\overline{\mathbf{0}}_{d+1},\,\nu_{2j}\Lambda_X\right). \tag{A.37}$$

By Assumption 2.2, we have

$$\frac{1}{NTh}\mathrm{Var}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}L_j\left(\frac{t-T\tau}{Th}\right)\begin{pmatrix}1\\X_{it}\end{pmatrix}e_{it}\right]=\nu_{2j}\Lambda_X+o(1). \tag{A.38}$$

Denote $\{\xi_j,\,j=1,\ldots,NT\}=\{(X_{11},e_{11}),\ldots,(X_{1T},e_{1T}),(X_{21},e_{21}),\ldots,(X_{NT},e_{NT})\}$, then $\{\xi\}$ is stationary and mixng with α-mixing coefficient

$$\widehat{\alpha}_k=\begin{cases}\alpha_k \text{ or } 0, & k<T,\\ 0, & k\geq T.\end{cases}$$

By (A.38) and Theorem 2.2 of Peligrad and Utev (1997), equation (A.37) holds.

We then turn to the proof of (A.36). Let $e_T(i)=\sum_{t=1}^{T}K(\frac{t-T\tau}{Th})e_{it}$. Following the proof of (A.29), we have

$$\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{e}=\left[e_T(2)-e_T(1),\ldots,e_T(N)-e_T(1)\right]^{\top}. \tag{A.39}$$

By (A.28), (A.29) and (A.39), we have

$$\widetilde{M}^{\top}(\tau)\widetilde{W}(\tau)\widetilde{D}^{*}\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^{*}\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{e}=\left(0,U_{NT}^{\top},0,V_{NT}^{\top}\right)^{\top}, \tag{A.40}$$

where

$$U_{NT}=\sum_{k=2}^{N}A_{NT}(k)[e_T(k)-e_T(1)]=\sum_{k=1}^{N}A_{NT}(k)e_T(k)$$

and

$$V_{NT}=\sum_{k=2}^{N}B_{NT}(k)[e_T(k)-e_T(1)]=\sum_{k=1}^{N}B_{NT}(k)e_T(k),$$

in which $A_{NT}(k)$ and $B_{NT}(k)$ are defined in the proof of Proposition A.4.

By Assumption 2.2 (a) (c), we have

$$E\left(U_{NT}^2\right)=o(NTh)\ \text{ and }\ E\left(V_{NT}^2\right)=o(NTh),$$

which implies

$$U_{NT} = o_P\left(\sqrt{NTh}\right) \text{ and } V_{NT} = o_P\left(\sqrt{NTh}\right). \tag{A.41}$$

In view of (A.40) and (A.41), (A.36) holds. The proof of Proposition A.6 is therefore completed. □

Proof of Theorem 2.2: Note that

$$\begin{aligned}
\widetilde{\beta}_*(\tau) - \beta_*(\tau) =\ & [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{Y} - \beta_*(\tau)\\
=\ & \left\{[I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\left[\widetilde{f} + \widetilde{B}(X,\beta)\right] - \beta_*(\tau)\right\}\\
& + [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{D}^*\alpha^*\\
& + [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{e}\\
=:\ & \Xi_{NT}(1) + \Xi_{NT}(2) + \Xi_{NT}(3).
\end{aligned} \tag{A.42}$$

By the definition of $\widetilde{W}^*(\tau)$, we have

$$\widetilde{W}^*(\tau) = \widetilde{\mathcal{K}}^\top(\tau)\widetilde{W}(\tau)\widetilde{\mathcal{K}}(\tau) = \widetilde{W}(\tau) - \widetilde{W}(\tau)\widetilde{D}^*\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau),$$

which implies that

$$\begin{aligned}
\Xi_{NT}(2) &= [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^*\alpha^*\\
&\quad - [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^*\left[\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\right]^{-1}\widetilde{D}^{*\top}\widetilde{W}(\tau)\widetilde{D}^*\alpha^*\\
&= [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^*\alpha^*\\
&\quad - [I_{d+1},\ \mathbf{0}_{d+1}]\left[\widetilde{M}^\top(\tau)\widetilde{W}^*(\tau)\widetilde{M}(\tau)\right]^{-1}\widetilde{M}^\top(\tau)\widetilde{W}(\tau)\widetilde{D}^*\alpha^*\\
&\equiv \overline{\mathbf{0}}_{d+1}.
\end{aligned}$$

Therefore, in order to establish the asymptotic distribution (2.14) in Theorem 2.2, we need only to consider $\Xi_{NT}(1)$ and $\Xi_{NT}(3)$. Propositions A.4–A.6 establish the asymptotic properties of $\Xi_{NT}(1)$ and $\Xi_{NT}(3)$, which lead to (2.14). And (2.14) implies (2.15) and (2.16). Theorem 2.2 holds for both N being fixed and N tending to infinity, but we only give the proof for the case of N tending to infinity simultaneously with T. The proof for the case of fixed N is similar. □