MONASH University

Department of Econometrics and Business Statistics

Semiparametric Trending Panel Data Models with Cross-Sectional Dependence

Jia Chen, Jiti Gao and Degui Li

September 2011

Semiparametric Trending Panel Data Models
with Cross-Sectional Dependence

By Jia Chen, Jiti Gao and Degui Li[1]

The University of Adelaide and Monash University

Abstract

A semiparametric fixed effects model is introduced to describe the nonlinear trending phenomenon in panel data analysis and it allows for the cross–sectional dependence in both the regressors and the residuals. A pooled semiparametric profile likelihood dummy variable approach based on the first–stage local linear fitting is developed to estimate both the parameter vector and the nonparametric time trend function. As both the time series length T and the cross–sectional size N tend to infinity simultaneously, the resulting estimator of the parameter vector is asymptotically normal with a rate of convergence of $O_P\left(\frac{1}{\sqrt{NT}}\right)$. Meanwhile, the asymptotic distribution for the estimator of the nonparametric trend function is also established with a rate of convergence of $O_P\left(\frac{1}{\sqrt{NTh}}\right)$. Two simulated examples are provided to illustrate the finite sample performance of the proposed method. In addition, the proposed model and estimation method is applied to analyze a CPI data set as well as an input–output data set.

JEL classification: C13, C14, C23.

Keywords: Cross–sectional dependence, nonlinear time trend, panel data, profile likelihood, semiparametric regression.

Abbreviated title: Semiparametric trending regression

---

[1] *Corresponding Author*: Degui Li, Department of Econometrics and Business Statistics, Monash University, Caulfield Campus, Victoria 3145, Australia. Email: degui.li08@gmail.com

1. Introduction

Modeling time series with trend functions has attracted an increasing interest in recent years. Mainly due to the limitation and practical inapplicability of parametric trend functions, recent literature focuses on estimating time–varying coefficient trend functions using nonparametric estimation methods. Such studies include Robinson (1989) and Cai (2007). Phillips (2001) provides a review on the current development and future directions about modeling time series with trends. In the meantime, some other nonparametric and semi-parametric models are also developed to deal with time series with a trend function. Gao and Hawthorne (2006) propose using a semiparametric time series model to address the issue of whether the trend of a temperature series should be parametrically linear while allowing for the inclusion of some explanatory variables in a parametric component.

While there is a rich literature on parametric and nonparametric time–varying coefficient time series models, as far as we know, few work has been done in identifying and estimating the trend function in a panel data model. Atak, Linton and Xiao (2010) propose a semipara-metric panel data model to deal with the modeling of climate change in the United Kingdom. The authors consider using a model with a dummy variable in the parametric component while allowing for the time trend function to be nonparametrically estimated. More recently, Li, Chen and Gao (2010) extend the work of Cai (2007) in a trending time–varying coefficient time series model to a panel data time–varying coefficient model. In such existing studies, the residuals are assumed to be cross–sectionally independent. A recent work by Robinson (2010) may be among the first to introduce a nonparametric trending time–varying model for the panel data case under cross–sectional dependence.

In order to take into account existing information and contribution from a set of explanatory variables, this paper proposes extending the nonparametric model by Robinson (2008) to a semiparametric partially linear panel data model with cross–sectional dependence. In our discussion, both the residuals and explanatory variables are allowed to be cross–sectionally dependent. The model we consider in this paper is a semiparametric trending panel data model of the form

$$Y_{it} = X_{it}^\top \beta + f_t + \alpha_i + e_{it}, \tag{1.1}$$

$$X_{it} = g_t + x_i + v_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{1.2}$$

where β is a d–dimensional vector of unknown parameters, $f_t = f\left(\frac{t}{T}\right)$ and $g_t = g\left(\frac{t}{T}\right)$ are both time trend functions with $f(\cdot)$ and $g(\cdot)$ being unknown, both $\{e_{it}\}$ and $\{v_{it}\}$ are

independent and identically distributed (i.i.d.) across time but correlated among individuals. Note that $\{\alpha_i\}$ is allowed to be correlated with $\{X_{it}\}$ through some unknown structure, while $\{e_{it}\}$ is assumed to be independent of $\{v_{it}\}$. Throughout this paper, we impose the following restrictions on the fixed effects $\{\alpha_i\}$ and the individual effects $\{x_i\}$,

$$\sum_{i=1}^{N} \alpha_i = 0, \quad \text{and} \quad \sum_{i=1}^{N} x_i = \mathbf{0}_d, \tag{1.3}$$

where $\mathbf{0}_d$ is the d–dimensional null vector.

Models (1.1) and (1.2) cover and extend some existing models. When $\beta = 0$, model (1.1) reduces to the nonparametric model discussed in Robinson (2008). When $N = 1$, models (1.1) and (1.2) reduce to the models discussed in Gao and Hawthorne (2006). Meanwhile, model (1.2) allows for $\{X_{it}\}$ to have a trend function and thus be nonstationary. As a consequence, models (1.1) and (1.2) become more applicable in practice than some of the existing models discussed in Cai (2007), and Li, Chen and Gao (2010), in which $\{X_{it}\}$ is assumed to be stationary. Such practical situations include the modeling of the dependence between the share consumption, $\{Y_{it}\}$, on the total consumption, $\{X_{it}\}$, as well as the modeling of the dependence of the mean temperature series, $\{Y_{it}\}$, on the Southern Oscillation Index, $\{X_{it}\}$. Furthermore, we relax the cross–sectional independence assumption on both the regressors $\{X_{it}\}$ and the error process $\{e_{it}\}$. As pointed out in Chapter 10 of Hsiao (2003), this makes panel data models more practically applicable because there is no natural ordering for cross–sectional indices. As a result, appropriate modeling and estimation of cross–sectional correlatedness becomes difficult particularly when the dimension of cross–sectional observation N is large. To be able to study the asymptotic theory of our proposed estimation method in this paper, we will impose certain mild moment conditions on $\{e_{it}\}$ and $\{v_{it}\}$ as in (3.1)–(3.3) in Section 3. Different sets of cross–sectional dependence conditions are discussed in some existing literature, such as Pesaran (2006), and Su and Jin (2011).

The main objective of this paper is then to construct a consistent estimation method for the parameter vector β and the trending function $f(\cdot)$. Throughout the paper, both the time series length T and the cross sections size N are allowed to tend to infinity. A semiparametric dummy–variable based profile likelihood estimation method is developed to estimate both β and $f(\cdot)$ based on first–stage local linear fitting. The resulting estimator of β is shown to be asymptotically normal with a rate of convergence of $O_P\left(\frac{1}{\sqrt{NT}}\right)$. Meanwhile, an asymptotic distribution for the nonparametric estimator of the time trend function is also established with a rate of convergence of $O_P\left(\frac{1}{\sqrt{NTh}}\right)$. In addition, we also propose

a semiparametric estimator for the cross–sectional covariance matrix of $\{v_{it}, e_{it}\}$, which is useful in constructing the confidence intervals of the estimators of β and $f(\cdot)$.

The rest of the paper is organized as follows. A pooled semiparametric profile likelihood dummy variable method is proposed in Section 2 for estimating β and $f(\cdot)$. The asymptotic normality of the proposed estimators is established in Section 3. Some related discussions, such as estimation of some covariance matrices, the averaged profile likelihood estimator and the cross–validation bandwidth selection method, are given in Section 4. Two simulated examples as well as two real–data examples are provided in Section 5. Section 6 concludes the paper. The mathematical proofs of the main results are relegated to Appendices A and B.

2. Estimation method

There are several semiparametric methods that can be adopted to estimate the parameter vector β and the time trend function $f(\cdot)$. Among these methods, the averaged profile likelihood estimation method is a commonly–used method and has been investigated by some authors in both the time series and panel data cases (see, for example, Fan and Huang 2005; Su and Ullah 2006; Atak, Linton and Xiao 2010). In this paper, we propose using a pooled semiparametric profile likelihood dummy variable method to estimate β and $f(\cdot)$, which is more efficient than the averaged profile likelihood method (which will be introduced in Section 4.2).

Before giving the estimation method, we first introduce the following notations:

$$\widetilde{Y} = (Y_{11}, \cdots, Y_{1T}, Y_{21}, \cdots, Y_{2T}, \cdots, Y_{N1}, \cdots, Y_{NT})^{\top},$$

$$\widetilde{X} = (X_{11}, \cdots, X_{1T}, X_{21}, \cdots, X_{2T}, \cdots, X_{N1}, \cdots, X_{NT})^{\top},$$

$$\alpha = (\alpha_2, \cdots, \alpha_N)^{\top}, \quad D = (-i_{N-1}, I_{N-1})^{\top} \otimes i_T,$$

$$\widetilde{f} = i_N \otimes (f_1, \cdots, f_T)^{\top}, \quad \widetilde{e} = (e_{11}, \cdots, e_{1T}, e_{21}, \cdots, e_{2T}, \cdots, e_{N1}, \cdots, e_{NT})^{\top},$$

where $\otimes$ denotes the Kronecker product, i_k is the $k \times 1$ vector of ones and I_k is the $k \times k$ identity matrix. As $\sum\limits_{i=1}^{N} \alpha_i = 0$, model (1.1) can be rewritten as

$$\widetilde{Y} = \widetilde{X}\beta + \widetilde{f} + D\alpha + \widetilde{e}. \tag{2.1}$$

Let $K(\cdot)$ denote a kernel function and h is a bandwidth. Denote $Z(\tau) = \begin{pmatrix} 1 & \frac{1-\tau T}{Th} \\ \vdots & \vdots \\ 1 & \frac{T-\tau T}{Th} \end{pmatrix}$

and $\widetilde{Z}(\tau) = i_N \otimes Z(\tau)$. Then by Taylor expansion, we have

$$\tilde{f} \approx \widetilde{Z}(\tau) \begin{pmatrix} f(\tau) \\ hf'(\tau) \end{pmatrix}.$$

Let $W(\tau) = \text{diag}\left(K(\frac{1-\tau T}{Th}), \cdots, K\left(\frac{T-\tau T}{Th}\right)\right)$ and $\widetilde{W}(\tau) = I_N \otimes W(\tau)$. The pooled semi-parametric profile likelihood dummy variable estimation method is given as follows.

(i) Define the following loss function:

$$L(a, b) = \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{Z}(\tau)(a, b)^\top\right)^\top \widetilde{W}(\tau) \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{Z}(\tau)(a, b)^\top\right).$$

For given α and β, we estimate $f(\tau)$ and $f'(\tau)$ by

$$\begin{pmatrix} \widehat{f}_{\alpha,\beta}(\tau) \\ h\widehat{f'}_{\alpha,\beta}(\tau) \end{pmatrix} = \arg\min_{(a,b)^\top} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - X_{it}^\top\beta - \alpha_i - a - b\left(\frac{t}{T} - \tau\right)\right)^2 K\left(\frac{t - \tau T}{Th}\right)$$

$$= \arg\min_{(a,b)^\top} L(a, b).$$

If we denote $S(\tau) = \left(\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{Z}(\tau)\right)^{-1} \widetilde{Z}^\top(\tau)\widetilde{W}(\tau)$, then by simple calculation, we have

$$\widehat{f}_{\alpha,\beta}(\tau) = (1, 0)S(\tau)(\widetilde{Y} - \widetilde{X}\beta - D\alpha) = s(\tau)(\widetilde{Y} - \widetilde{X}\beta - D\alpha), \qquad (2.2)$$

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

(ii) Denote

$$\tilde{f}_{\alpha,\beta} = i_N \otimes \left(\widehat{f}_{\alpha,\beta}(1/T), \cdots, \widehat{f}_{\alpha,\beta}(T/T)\right)^\top = \widetilde{S}(\widetilde{Y} - \widetilde{X}\beta - D\alpha),$$

where $\widetilde{S} = i_N \otimes \left(s^\top(1/T), \cdots, s^\top(T/T)\right)^\top$. Then we estimate α and β by

$$\begin{aligned} (\widehat{\alpha}^\top, \widehat{\beta}^\top)^\top &= \arg\min_{(\alpha^\top,\beta^\top)^\top} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - X_{it}\beta - \alpha_i - \widehat{f}_{\alpha,\beta}\left(\frac{t}{T}\right)\right)^2 \\ &= \arg\min_{(\alpha^\top,\beta^\top)^\top} \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \tilde{f}_{\alpha,\beta}\right)^\top \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \tilde{f}_{\alpha,\beta}\right) \\ &= \arg\min_{(\alpha^\top,\beta^\top)^\top} \left(\widetilde{Y}^* - \widetilde{X}^*\beta - D^*\alpha\right)^\top \left(\widetilde{Y}^* - \widetilde{X}^*\beta - D^*\alpha\right), \qquad (2.3) \end{aligned}$$

where $\widetilde{Y}^* = \left(I_{NT} - \widetilde{S}\right)\widetilde{Y}$, $\widetilde{X}^* = \left(I_{NT} - \widetilde{S}\right)\widetilde{X}$ and $D^* = \left(I_{NT} - \widetilde{S}\right)D$.

Define $M^* = I_{NT} - D^*\left(D^{*\top}D^*\right)^{-1} D^{*\top}$. Simple calculation leads to the solution of the minimization problem (2.3):

$$\widehat{\beta} = \left(\widetilde{X}^{*\top} M^* \widetilde{X}^*\right)^{-1} \widetilde{X}^{*\top} M^* \widetilde{Y}^*, \qquad (2.4)$$

$$\widehat{\alpha} = \left(D^{*\top}D^*\right)^{-1} D^{*\top}\left(\widetilde{Y}^* - \widetilde{X}^*\widehat{\beta}\right). \qquad (2.5)$$

(iii) Plug (2.4) and (2.5) into (2.2) to obtain the estimate of $f(\tau)$ by

$$\widehat{f}(\tau) = s(\tau)\left(\widetilde{Y} - \widetilde{X}\widehat{\beta} - D\widehat{\alpha}\right).\qquad(2.6)$$

Note that our study in Sections 3 and 5 below shows that the proposed pooled profile likelihood dummy variable method has both theoretical and numerical advantages over the averaged profile likelihood estimation method,.

3. The main results

In this section, we first introduce some regularity conditions and then establish asymptotic distributions for $\widehat{\beta}$ and $\widehat{f}(\cdot)$.

3.1. Assumptions

A1. The kernel function $K(\cdot)$ is continuous and symmetric with compact support.

A2. Let $v_t = (v_{1t}, \cdots, v_{Nt})^\top$, $1 \le t \le T$. Suppose that $\{v_t,\ t \ge 1\}$ is a sequence of i.i.d. $N \times d$ random matrices with zero mean and $E\left[\|v_{it}\|^4\right] < \infty$. There exist $d \times d$ positive definite matrices Σ_v and Σ_v^*, such that as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} E\left[v_{it}v_{it}^\top\right] \longrightarrow \Sigma_v, \quad \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[v_{it}v_{jt}^\top\right] \longrightarrow \Sigma_v^*, \quad E\left\|\sum_{i=1}^{N} v_{it}\right\|^\delta = O\left(N^{\delta/2}\right),\quad(3.1)$$

where $\delta > 2$ is a positive constant. Furthermore, $\{x_i\}$ satisfies $\max_i \|x_i\| = O_P(1)$, where $\|\cdot\|$ is the Euclidean norm.

A3. The trend functions $f(\cdot)$ and $g(\cdot)$ have continuous derivatives up to the second order.

A4. Let $e_t = \{e_{it},\ 1 \le i \le N\}$. Suppose that $\{e_t,\ t \ge 1\}$ is a sequence of i.i.d. random errors with zero mean and independent of $\{v_{it}\}$. There exists a $d \times d$ positive definite matrix $\Sigma_{v,e}$ such that as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[v_{i1}v_{j1}^\top\right] E\left[e_{i1}e_{j1}\right] \to \Sigma_{v,e}.\qquad(3.2)$$

Furthermore, there is some $0 < \sigma_e^2 < \infty$ such that as $N \to \infty$

$$\frac{1}{N} E\left(\sum_{i=1}^{N} e_{it}\right)^2 \to \sigma_e^2 \quad \text{and} \quad E\left|\sum_{i=1}^{N} e_{it}\right|^\delta = O\left(N^{\delta/2}\right),\qquad(3.3)$$

where $\delta > 2$ is as defined in A2.

6

A5. The bandwidth h satisfies as $T \to \infty$ and $N \to \infty$ simultaneously,

$$NTh^8 \to 0, \quad \frac{\sqrt{NT}h}{\log(NT)} \to \infty \text{ and } \frac{T^{1-\frac{2}{\delta}}h}{\log(NT)} \to \infty.$$

Furthermore, $N^2/T^\delta = o(1)$ for $\delta > 2$ defined as in A2.

Remark 3.1. A1 is a mild condition on the kernel function and many commonly–used kernels, including the Epanechnikov kernel, satisfy A1. Furthermore, the compact support condition for the kernel function can be relaxed at the cost of more tedious proofs.

In A2, we impose some moment conditions on $\{v_{it}\}$ and allow for cross–sectional dependence of $\{v_{it}\}$ and thus $\{X_{it}\}$. Such condition of cross–sectional dependence is similar to the condition A4 in Chen, Gao and Li (2010). When $\{v_{it}\}$ is also i.i.d. across individuals, it is easy to check that (3.1) holds. Since there is no natural ordering for cross–sectional indices, it may not be appropriate to impose any kind of mixing or martingale difference conditions on $\{v_{it}\}$ when v_{it} and v_{jt} are dependent. Equation (3.1) instead imposes certain conditions on the measurement of the "distance" between cross–sections i_j and i_k. To explain this in some detail, let us consider the case of $d = 1$ and define a kind of "distance function" among the cross–sections of the form

$$\rho(i_1, i_2, \cdots, i_k) = E\left[v_{i_1,1}^{j_1} \cdots v_{i_k,1}^{j_k}\right], \tag{3.4}$$

and then consider one of the cases where $k = 4$ and $j_1 = j_2 = \cdots = j_4 = 1$. In addition, we focus on the case where all $1 \leq i_1, i_2, \cdots, i_4 \leq N$ are different. Consider a distance function of the form

$$\rho(i_1, i_2, \cdots, i_4) = \frac{1}{|i_4 - i_3|^{\delta_3}|i_3 - i_2|^{\delta_2} \cdots |i_2 - i_1|^{\delta_1}} \tag{3.5}$$

for $\delta_i > 0$ for all $1 \leq i \leq 3$. In this case, equation (3.1) can be verified because

$$\sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_4=1}^{N} E[v_{i_1,1} \cdots v_{i_4,1}] = O\left(N^{4-\sum\limits_{j=1}^{3}\delta_j}\right) = O\left(N^2\right) \tag{3.6}$$

when $\sum\limits_{j=1}^{3} \delta_j \geq 2$. Obviously, the conventional Euclidean metric is covered. One may also show that equation (3.1) can be satisfied when some other distance functions, including the exponential distance function, are considered.

A3 is a commonly used condition in local linear fitting. In A2 and A4, we assume that both $\{e_t\}$ and $\{v_t\}$ are i.i.d., which a stronger than the corresponding condition in Robinson

(2010). Our experience suggests however that the asymptotic theory in this paper still holds when both $\{e_t\}$ and $\{v_t\}$ are uncorrelated. Furthermore, the dependence condition on $\{e_t\}$ and $\{v_t\}$ can be relaxed by allowing both $\{e_t\}$ and $\{v_t\}$ to be stationary and α–mixing (see, for example, Gao 2007). However, the resulting asymptotic variance involved in (3.7) below will be more complicated under the stationary α–mixing dependence condition than that for the i.i.d. case. To avoid further technicality, this paper focuses on the i.i.d. case. Such extensions are left for future research.

In A4, we impose the mutual independence between v_{it} and e_{it}. When v_{it} and e_{it} are dependent with each other, we do not necessarily have $E[v_{it}e_{it}] = 0$. In this case, a modified estimation method, such as an instrumental variable based method may be needed to construct a consistent estimator for β. To emphasize the main ideas, the proposed estimation method and the resulting theory as well as to avoid involving further technicality, we establish the main results under the exogenous condition throughout this paper. However, such extensions are left for future discussion. The cross–sectional dependence conditions in (3.2) and (3.3) are similar to those in (3.1).

A5 is required for establishing the asymptotic theory without involving too much technicality. A5 covers the case of $\frac{N}{T} \to \lambda$ for $0 \le \lambda \le \infty$. For example, when N is proportional to T, A5 reduces to $Th^4 \to 0$ and $\frac{T^{1-\frac{2}{\delta}}h}{\log(T)} \to \infty$, which can be satisfied if we take the bandwidth $h \propto (NT)^{-1/5} \propto T^{-2/5}$ and assume that $\delta > \frac{10}{3}$ ($a_n \propto b_n$ means that $|a_n/b_n| \le C$ for some positive constant C as n is large enough). For the general case of $N = [T^c]$, A5 reduces to $T^{1+c}h^8 \to 0$ and $\frac{T^{1-\frac{2}{\delta}}h}{\log(T)} \to \infty$ ($c \ge 1 - \frac{4}{\delta}, \delta > 4$), and it can be satisfied if we take a bandwidth of the form $h \propto (NT)^{-1/5} \propto T^{-(1+c)/5}$ for some $c < 4 - \frac{10}{\delta}$.

3.2. Asymptotic theory

We first give the asymptotic distribution for $\widehat{\beta}$ in the following theorem.

Theorem 3.1. *Let Conditions A1–A5 hold. Then as $T \to \infty$ and $N \to \infty$ simultaneously*

$$\sqrt{NT}\left(\widehat{\beta} - \beta\right) \overset{d}{\longrightarrow} N\left(\mathbf{0}_d, \ \Sigma_v^{-1}\Sigma_{v,e}\Sigma_v^{-1}\right). \tag{3.7}$$

Remark 3.2. The above theorem shows that the proposed pooled profile likelihood estimator of β can achieve the convergence rate of $O_P\left(\frac{1}{\sqrt{NT}}\right)$. As both T and N tend to infinity jointly, the asymptotic variance in (3.7) is simplified, compared with some existing literature on the profile likelihood estimation of semiparametric panel data models with fixed effects (see, for example, Su and Ullah 2006). A consistent estimation method for Σ_v and $\Sigma_{v,e}$ will be proposed in Section 4.1 below.

(ii) Note that the proposed pooled profile likelihood estimator of β is asymptotically efficient under cross–sectional independence. Under this independence, the asymptotic variance reduces to $\Sigma_v^{-1}\sigma_e^2$, which is the smallest possible variance when both e_{it} and v_{it} are normally distributed.

Define $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$. An asymptotic distribution of $\widehat{f}(\tau)$ is established in the following theorem.

Theorem 3.2. *Let Conditions A1–A5 hold. Then, for $\tau \in (0,1)$, as $T \to \infty$ and $N \to \infty$ simultaneously*

$$\sqrt{NTh}\left(\widehat{f}(\tau) - f(\tau) - b_f(\tau)h^2 + o_P(h^2)\right) \xrightarrow{d} N\left(0, \ \nu_0\sigma_e^2\right), \tag{3.8}$$

where $b_f(\tau) = \frac{1}{2}\mu_2 f''(\tau)$.

Remark 3.3. The asymptotic distribution in (3.8) is a standard result for local linear fitting of nonlinear time trend function. From (3.8), we can obtain the mean integrated square error (MISE) of $\widehat{f}(\cdot)$

$$\text{MISE}(\widehat{f}(\tau)) = E\int_0^1 \left(\widehat{f}(\tau) - f(\tau)\right)^2 d\tau \approx \frac{\nu_0\sigma_e^2}{NTh} + \int_0^1 b_f^2(\tau)d\tau \ h^4, \tag{3.9}$$

where the symbol "$a_n \approx b_n$" denotes that $\frac{a_n}{b_n} \to 1$ as $n \to \infty$. From (3.9), we can obtain an optimal bandwidth

$$h_{opt} = \left(\frac{\nu_0\sigma_e^2}{4\int_0^1 b_f^2(\tau)d\tau}\right)^{1/5} (NT)^{-1/5}. \tag{3.10}$$

The above bandwidth cannot be used directly as both σ_e^2 and $b_f^2(\tau)$ in (3.10) are unknown. Hence, in the simulation study in Section 5, we use a semiparametric "leave–one–out" cross validation method which will be introduced in Section 4.3 below.

4. Some related discussions

In Section 4.1, consistent estimators are constructed for Σ_v, $\Sigma_{v,e}$ and σ_e^2 which are involved in Theorems 3.1 and 3.2. Then, an averaged profile likelihood estimation is introduced in Section 4.2. The so–called "leave–one–out" cross validation bandwidth selection criterion is provided in Section 4.3.

4.1. Estimation of Σ_v, $\Sigma_{v,e}$ and σ_e^2

To make the proposed estimation method practically implementable, we also need to construct consistent estimators for Σ_v and $\Sigma_{v,e}$. By (1.2), and letting $\Delta v_{it} = v_{it} - v_{i,t-1}$,

$\Delta X_{it} = X_{it} - X_{i,t-1}$ and $\Delta g_t = g_t - g_{t-1}$, we have for $t \geq 2$,

$$\Delta v_{it} = \Delta X_{it} - \Delta g_t.$$

Noting that $\{v_{it}\}$ is i.i.d. across time by A2, it is easy to check that

$$\text{Var}\left(\Delta v_{it}\right) = 2\text{Var}\left(v_{it}\right) =: 2\Sigma_v(i).$$

Define

$$\widehat{v}_{it} = X_{it} - \widehat{g}_t \quad \text{and} \quad \Delta\widehat{v}_{it} = \widehat{v}_{it} - \widehat{v}_{i,t-1},$$

where $\widehat{g}_t := \widehat{g}\left(\frac{t}{T}\right)$ is the local linear estimate of $g\left(\frac{t}{T}\right)$. Then, Σ_v can be estimated by

$$\widehat{\Sigma}_v = \frac{1}{N}\sum_{i=1}^{N}\widehat{\Sigma}_v(i), \tag{4.1}$$

where

$$\widehat{\Sigma}_v(i) = \frac{1}{2[T/2]}\sum_{t=1}^{[T/2]}\Delta\widehat{v}_{i,2t}\Delta\widehat{v}_{i,2t}^\top. \tag{4.2}$$

By the uniform consistency of the local linear estimate (see the proofs in Appendix B) and the fact that $g(\cdot)$ is independent of i, it is easy to check that $\widehat{\Sigma}_v(i)$ is a consistent estimator of $\Sigma_v(i)$ uniformly for i, which implies that $\widehat{\Sigma}_v$ is a consistent estimator of Σ_v.

Let $\Delta\widehat{v}_{it}$ be defined as in (4.2) and

$$\Delta\widehat{e}_{it} = \widehat{e}_{it} - \widehat{e}_{i,t-1} \quad \text{and} \quad \widehat{e}_{it} = Y_{it} - X_{it}^\top\widehat{\beta} - \widehat{f}_t, \tag{4.3}$$

where $\widehat{f}_t := \widehat{f}\left(\frac{t}{T}\right)$. Then, $\rho_{ij}(v) := E\left[v_{i1}v_{j1}^\top\right]$ and $\rho_{ij}(e) := E\left[e_{i1}e_{j1}\right]$ can be estimated by

$$\widehat{\rho}_{ij}(v) = \frac{1}{2[T/2]}\sum_{t=1}^{[T/2]}\Delta\widehat{v}_{i,2t}\Delta\widehat{v}_{j,2t}^\top \quad \text{and} \quad \widehat{\rho}_{ij}(e) = \frac{1}{2[T/2]}\sum_{t=1}^{T}\Delta\widehat{e}_{i,2t}\Delta\widehat{e}_{j,2t}^\top, \tag{4.4}$$

respectively. Let φ_N be a positive integer satisfying $\varphi_N \leq N$ and $\varphi_N \to \infty$. By (3.2), $\Sigma_{v,e}$ can be consistently estimated by

$$\widehat{\Sigma}_{v,e} = \frac{1}{\varphi_N}\sum_{i=1}^{\varphi_N}\sum_{j=1}^{\varphi_N}\widehat{\rho}_{ij}(v)\widehat{\rho}_{ij}(e). \tag{4.5}$$

Similarly, by (3.3), σ_e^2 can be consistently estimated by

$$\widehat{\sigma}_e^2 = \frac{1}{\varphi_N}\sum_{i=1}^{\varphi_N}\sum_{j=1}^{\varphi_N}\widehat{\rho}_{ij}(e). \tag{4.6}$$

4.2. Averaged profile likelihood estimation method

As $\sum\limits_{i=1}^{N} \alpha_i = 0$, another way to eliminate the individual effects α_i from model (1.1) is to take averages over i

$$Y_{At} = X_{At}^\top \beta + f_t + e_{At}, \tag{4.7}$$

where the subscript A indicates averaging with respect to i, $Y_{At} = \frac{1}{N}\sum\limits_{i=1}^{N} Y_{it}$, $X_{At} = \frac{1}{N}\sum\limits_{i=1}^{N} X_{it}$ and $e_{At} = \frac{1}{N}\sum\limits_{i=1}^{N} e_{it}$. Denote $Y_A = (Y_{A1}, \cdots, Y_{AT})^\top$, $X_A = (X_{A1}, \cdots, X_{AT})^\top$, $f = (f_1, \cdots, f_T)^\top$ and $e_A = (e_{A1}, \cdots, e_{AT})^\top$. Then, by applying the profile likelihood estimation approach to model (4.7), one can obtain averaged profile likelihood estimators of β and $f(\cdot)$ as

$$\widehat{\beta}_A = \left(X_A^{*\top} X_A^*\right)^{-1} X_A^{*\top} Y_A^*,$$
$$\widehat{f}_A(\tau) = (1,0) \left(Z^\top(\tau)W(\tau)Z(\tau)\right)^{-1} Z^\top(\tau)W(\tau)(Y_A - X_A\widehat{\beta}_A),$$

where $X_A^* = X_A - MX_A = (I_T - M)X_A$, $Y_A^* = (I_T - M)Y_A$,

$$M = \begin{pmatrix} (1,0) \left(Z^\top(1/T)W(1/T)Z(1/T)\right)^{-1} Z^\top(1/T)W(1/T) \\ \vdots \\ (1,0) \left(Z^\top(T/T)W(T/T)Z(T/T)\right)^{-1} Z^\top(T/T)W(T/T) \end{pmatrix},$$

in which $W(\tau)$ and $Z(\tau)$ are defined in Section 2, I_T is the $T \times T$ identity matrix.

It can be shown that the rate of convergence of $\widehat{\beta}_A$ to β is of order $\sqrt{T}$, while the rate of convergence of $\widehat{f}_A(\tau)$ to $f(\tau)$ is of the same order of $\sqrt{NTh}$ as that for $\widehat{f}(\tau)$. This is clearly illustrated in Tables 5.1 and 5.2 below.

4.3. Bandwidth Selection

Due to the existence of the fixed effects, the traditional method of cross validation may not provide satisfactory results in selecting the optimal bandwidth. Hence, we use an extension of the conventional leave–one–out cross validation method, which is proposed by Sun et al (2009) and is called leave–one–unit–out cross validation method. The idea is to remove $\{(X_{it}, Y_{it}), \ 1 \le t \le T\}$ from the data for each $1 \le i \le N$ and use the rest of the $(N-1)T$ observations as the training data to obtain estimates of β and $f(\cdot)$, which are denoted as $\widehat{\beta}_{(-i)}$ and $\widehat{f}_{(-i)}(\cdot)$. Then, the optimal bandwidth is chosen such that it minimizes a weighted squared prediction error of the form

$$\left(\widetilde{Y} - B\left(\widetilde{X}, \widehat{\beta}_{(-)}\right) - \widehat{f}_{(-)}\right)^\top M^\top M \left(\widetilde{Y} - B\left(\widetilde{X}, \widehat{\beta}_{(-)}\right) - \widehat{f}_{(-)}\right), \tag{4.8}$$

where $M = I_{N \times T} - \frac{1}{T}I_N \otimes (i_T i_T^\top)$,

$$B\left(\widetilde{X}, \widehat{\beta}_{(-)}\right) = \left(X_{11}^\top\widehat{\beta}_{(-1)}, \cdots, X_{1T}^\top\widehat{\beta}_{(-1)}, X_{21}^\top\widehat{\beta}_{(-2)}, \cdots, X_{2T}^\top\widehat{\beta}_{(-2)}, \cdots, X_{N1}^\top\widehat{\beta}_{(-N)}, \cdots, X_{NT}^\top\widehat{\beta}_{(-N)}\right)^\top$$

11

and $\widehat{f}_{(-)} = \left(\widehat{f}_{(-1)}(1/T), \cdots, \widehat{f}_{(-1)}(T/T), \widehat{f}_{(-2)}(1/T), \cdots, \widehat{f}_{(-2)}(T/T), \cdots, \widehat{f}_{(-N)}(1/T), \cdots, \widehat{f}_{(-N)}(T/T) \right)^{\top}$.

The weight matrix M is constructed to satisfy $MD = \mathbf{0}$ so that the fixed effects α_i are eliminated from (4.8). In effect, M removes a cross–time average from each variable. For example, $M\widetilde{Y} = (Y_{11} - Y_{1A}, \cdots, Y_{1T} - Y_{1A}, Y_{21} - Y_{2A}, \cdots, Y_{NT} - Y_{NA})^{\top}$, where $Y_{iA} = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$ for $i = 1, \cdots, N$.

5. Examples of implementation

We next carry out simulation studies to compare the small sample behavior of the two profile likelihood estimation methods: the pooled and the averaged methods. The Epanechnikove kernel $K(u) = \frac{3}{4}(1 - u^2)I\{|u| \le 1\}$ is used throughout the section. We also illustrate the proposed pooled profile likelihood dummy variable method through applications to a CPI data set and an input–output data set.

5.1. Simulated Examples

Example 5.1. Consider a data generating process of the form

$$Y_{it} = X_{it}\beta + f\left(\frac{t}{T}\right) + \alpha_i + e_{it}, \quad 1 \le i \le N, \ 1 \le t \le T, \tag{5.1}$$

where $\beta = 2$, $f(u) = 2u^3 + u$, $\alpha_i = \frac{1}{T} \sum_{t=1}^{T} X_{it}$ for $i = 1, \cdots, N - 1$, and $\alpha_N = -\sum_{i=1}^{N-1} \alpha_i$. The error terms e_{it} are generated as follows. For each $1 \le t \le T$, let $\widetilde{e}_{\cdot t} = (e_{1t}, e_{2t}, \cdots, e_{Nt})$, which is a N–dimensional vector. Then, $\{\widetilde{e}_{\cdot t}, \ 1 \le t \le T\}$ is generated as a N–dimensional vector of independent Gaussian variables with zero mean and covariance matrix $(c_{ij})_{N \times N}$, where

$$c_{ij} = 0.8^{|j-i|}, \quad 1 \le i, j \le N. \tag{5.2}$$

From the way e_{it} are generated, it is easy to see that

$$\mathrm{E}(e_{it}e_{js}) = 0 \quad \text{for} \quad 1 \le i, j \le N, \ t \ne s,$$

$$\mathrm{E}(e_{it}e_{jt}) = 0.8^{|j-i|} \quad \text{for} \quad 1 \le i, j \le N, \ 1 \le t \le T.$$

The above equations imply that $\{e_{it}\}$ is cross–sectionally dependent and time independent. The explanatory variables X_{it} are generated by

$$X_{it} = g\left(\frac{t}{T}\right) + x_i + v_{it}, \quad 1 \le i \le N, \ 1 \le t \le T, \tag{5.3}$$

where $g(u) = 2\sin(\pi u)$, $\{v_{it}\}$ is independent of $\{e_{it}\}$ and is generated in the same way as $\{e_{it}\}$ but with a different covariance matrix $(d_{ij})_{N \times N}$, where $d_{ij} = 0.5^{|j-i|}$ for $1 \le i, j \le N$, $x_i \sim U(-0.2, 0.2)$ for $1 \le i \le N - 1$ and $x_N = -\sum_{i=1}^{N-1} x_i$.

We compare the pooled profile likelihood estimator (PPLE) with the averaged profile likelihood estimator (APLE). The means and standard deviations (SDs) (in parentheses) of the two estimators of the parameter $\beta = 2$ based on $R = 500$ replications are reported in Table 5.1(a). Table 5.1(b) compares the mean squared errors (MSEs) of the two estimators of the trend function and their SDs (in parentheses), where for an estimator $\widehat{f}$, its MSE is defined as

$$\text{MSE}(\widehat{f}) = \frac{1}{T} \sum_{t=1}^{T} \left[\widehat{f}\left(\frac{t}{T}\right) - f\left(\frac{t}{T}\right) \right]^2.   \tag{5.4}$$

Tables 5.1(a)–(b) indicate that the PPLE estimates both the parameter and the trend function much more accurately than the APLE, and it is also more efficient than the APLE. Another important finding from the simulation results is that an increase in either N or T results in a decrease in the SD of the PPLE of the parameter β or the MSE of the PPLE of the trend function $f(\cdot)$. However, this does not apply to the APLE. While the SD of the APLE of the parameter β or the MSE of the APLE of $f(\cdot)$ decreases as T increases, an increase in N does not necessarily imply a decrease in them when T is fixed.

Table 5.1(a). *Means and SDs of estimators for $\beta = 2$ in Example 5.1*

$N\backslash T$	10		20		30	
	PPLE	APLE	PPLE	APLE	PPLE	APLE
10	1.9963	1.7543	1.9977	1.7848	1.9982	1.7752
	(0.1313)	(0.3339)	(0.0978)	(0.2279)	(0.0813)	(0.1821)
20	1.9969	1.6722	2.0018	1.7177	1.9986	1.7270
	(0.0982)	(0.3114)	(0.0715)	(0.3071)	(0.0576)	(0.1613)
30	1.9972	1.7774	1.9996	1.7001	1.9992	1.7087
	(0.0873)	(0.3124)	(0.0567)	(0.1791)	(0.0487)	(0.1371)

Table 5.1(b). *Means and SDs of the MSEs of estimators for $f(u) = 2u^3 + u$ in Example 5.1*

$N\backslash T$	10		20		30	
	PPLE	APLE	PPLE	APLE	PPLE	APLE
10	0.2272	0.4482	0.1361	0.2876	0.0900	0.2144
	(0.1548)	(0.5848)	(0.0983)	(0.2585)	(0.0593)	(0.1798)
20	0.1576	0.4567	0.0902	0.2949	0.0639	0.2431
	(0.1050)	(0.5285)	(0.0570)	(0.2543)	(0.0418)	(0.1948)
30	0.1182	0.3284	0.0649	0.2780	0.0457	0.1645
	(0.0736)	(0.3620)	(0.0377)	(0.2202)	(0.0288)	(0.2026)

Example 5.2. Consider another the data generating process of the form

$$Y_{it} = X_{it}^\top \beta + f\left(\frac{t}{T}\right) + \alpha_i + e_{it}, \quad 1 \le i \le N, \ 1 \le t \le T, \tag{5.5}$$

where $\beta = (2, 1, 2)^\top$, $X_{it} = (X_{it,1}, X_{it,2}, X_{it,3})^\top$, $f(u) = 1 + 2\sin(\pi u)$,

$$\alpha_i = \max\left\{\frac{1}{T}\sum_{t=1}^T X_{it,1}, \ \frac{1}{T}\sum_{t=1}^T X_{it,2}, \ \frac{1}{T}\sum_{t=1}^T X_{it,3}, \right\}, \quad i = 1, \cdots, N-1,$$

and $\alpha_N = -\sum_{i=1}^{N-1} \alpha_i$.

Letting $\widetilde{e}_{\cdot t} = (e_{1t}, e_{2t}, \cdots, e_{Nt})$, then we generate $\{\widetilde{e}_{\cdot t}, \ 1 \le t \le T\}$ as a N–dimensional vector of Gaussian variables with zero mean and covariance matrix $(c_{ij}^*)_{N \times N}$, where $c_{ij}^* = \frac{1}{(i-j)^2+1}$.

The explanatory variables X_{it} are generated by $X_{it} = \mathbf{g}\left(\frac{t}{T}\right) + x_i + \mathbf{v}_{it}$, where

$$\mathbf{g}(u) = \left(2(u - 1/2)^2, \ u, \ -2u\right)^\top, \tag{5.6}$$

and $\{\mathbf{v}_{it} = (v_{it,1}, v_{it,2}, v_{it,3})^\top : \ 1 \le i \le N, 1 \le t \le T\}$ satisfies $E\mathbf{v}_{it} = \mathbf{0}$,

$$E\left(\mathbf{v}_{it}\mathbf{v}_{jt}^\top\right) = \frac{1}{(j-i)^2+1}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad 1 \le i, j \le N, \ 1 \le t \le T,$$

and $E\left(\mathbf{v}_{it}\mathbf{v}_{js}^\top\right) = \mathbf{0}$ for $1 \le i, j \le N$ and $s \ne t$, $x_i = 0.5N(\mathbf{0}, I_3)$ for $1 \le i \le N-1$, and $x_N = -\sum_{i=1}^{N-1} x_i$.

We also compare the estimates from the two estimation methods, and the simulation results based on 500 replications are given in Tables 5.2(a)–(b). Similar conclusions can be drawn from Tables 5.2(a)–(b) as those from Tables 5.1(a)–(b): the PPLE outperforms APLE in terms of accuracy and efficiency, and its performance improves as either N or T increases, however, the performance of the APLE does not necessarily improve as N increases while we hold T fixed.

Table 5.2(a). *Means and SDs of estimators for $\beta = (2,1,2)^\top$ in Example 5.2*

$N\backslash T$		10		20		30	
		PPLE	APLE	PPLE	APLE	PPLE	APLE
10	$\beta_1 = 2$	1.9945	1.9114	1.9967	1.9956	1.9985	1.9737
		(0.1283)	(0.5906)	(0.0929)	(0.2917)	(0.0687)	(0.2108)
	$\beta_2 = 1$	1.0036	1.0177	0.9953	0.9856	0.9996	0.9930
		(0.0973)	(0.4170)	(0.0622)	(0.2019)	(0.0513)	(0.1519)
	$\beta_3 = 2$	2.0020	1.9552	1.9987	1.9992	2.0035	2.0175
		(0.1279)	(0.6229)	(0.0914)	(0.2783)	(0.0761)	(0.2113)
20	$\beta_1 = 2$	2.0013	1.3617	1.9996	2.0052	2.0019	1.9819
		(0.0867)	(0.7619)	(0.0605)	(0.3178)	(0.0525)	(0.2294)
	$\beta_2 = 1$	0.9979	1.0125	0.9989	0.9888	1.0010	0.9979
		(0.0685)	(0.6986)	(0.0449)	(0.2147)	(0.0362)	(0.1576)
	$\beta_3 = 2$	1.9952	2.0234	2.0016	2.0017	1.9994	2.0193
		(0.0927)	(0.9640)	(0.0645)	(0.2845)	(0.0509)	(0.2391)
30	$\beta_1 = 2$	2.0013	1.7425	2.0011	1.9274	1.9990	1.9402
		(0.0735)	(0.5787)	(0.0575)	(0.2901)	(0.0413)	(0.2283)
	$\beta_2 = 1$	1.0041	0.9905	1.0016	1.0049	1.0000	0.9965
		(0.0522)	(0.4563)	(0.0354)	(0.1918)	(0.0302)	(0.1533)
	$\beta_3 = 2$	1.9985	1.9685	2.0011	2.0084	1.9998	2.0275
		(0.0777)	(0.6875)	(0.0524)	(0.2804)	(0.0434)	(0.2142)

Table 5.2(b). *Means and SDs of MSEs of estimators for $f(u) = 1 + 2\sin(\pi u)$ in Example 5.2*

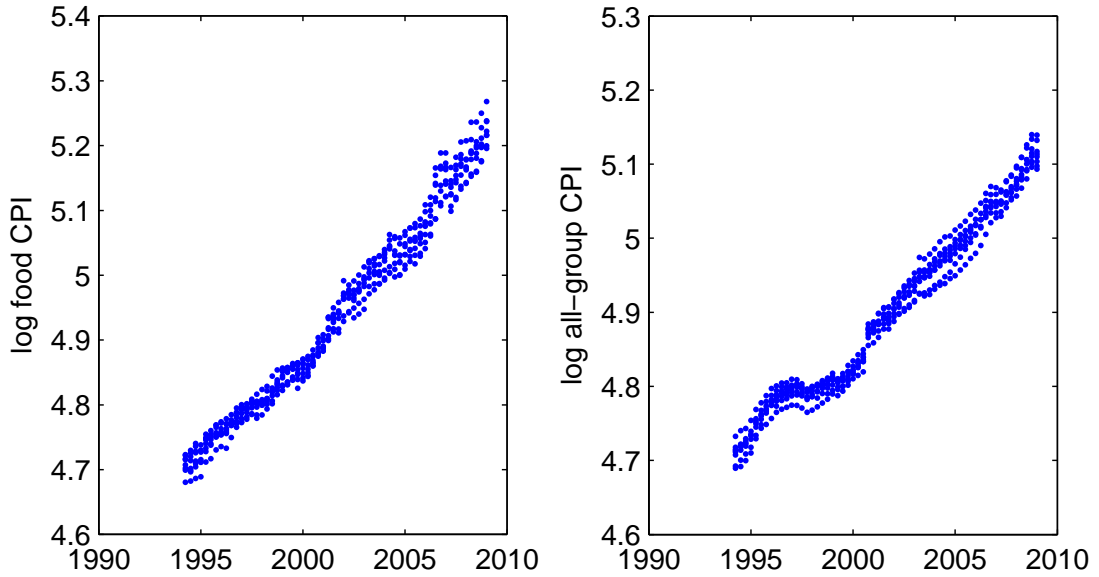| $N\backslash T$	10		20		30	
	PPLE	APLE	PPLE	APLE	PPLE	APLE
10	0.1290	0.9614	0.0668	0.1900	0.0500	0.1223
	(0.1018)	(1.5668)	(0.0465)	(0.2340)	(0.0315)	(0.1424)
20	0.0767	2.0194	0.0429	0.1792	0.0271	0.1138
	(0.0541)	(3.2566)	(0.0286)	(0.2362)	(0.0166)	(0.1342)
30	0.0539	0.9126	0.0286	0.1713	0.0204	0.0994
	(0.0409)	(2.3934)	(0.0179)	(0.2104)	(0.0118)	(0.1349)

Figure 5.1. From left to right: scatter plots of the observations of log food CPI and log all-group CPI.

5.2. Empirical Applications

Example 5.3. The first data set consists of quarterly consumer price index (CPI) numbers of 11 classes of commodities for 8 Australian capital cities spanning from 1994 to 2008 (available from the Australian Bureau of Statistics at www.abs.gov.au). The raw data were seasonally adjusted before being used. We study the empirical relationship between the log food CPI and the log all–group CPI. Let Y_{it} be the log food CPI and X_{it} be the log all–group CPI for city i at time t, where $1 \leq i \leq 8$ and $1 \leq t \leq 60$. We then assume that $\{(Y_{it}, X_{it})\}$ satisfies a semiparametric model of the form

$$Y_{it} = X_{it}\beta + f_t + \alpha_i + e_{it}, \tag{5.7}$$

where α_i are individual effects, and f_t is the trend in the log food CPI. The scatter plots of the observations of both Y_{it} and X_{it} are given in Figure 5.1.

By applying the proposed pooled profile likelihood estimation procedure to the above data set, we have the estimate of β as $\widehat{\beta} = 0.8160$ with standard deviation 9.1×10^{-5}, which indicates that this estimated coefficient is significant. The semiparametric estimate of the trend function as well as its 95% confidence band is given in Figure 5.2. The estimated trend curve exhibits an obvious upward pattern, which is consistent with the observation that the food CPI series for each city generally increases with time. We also fit the data with
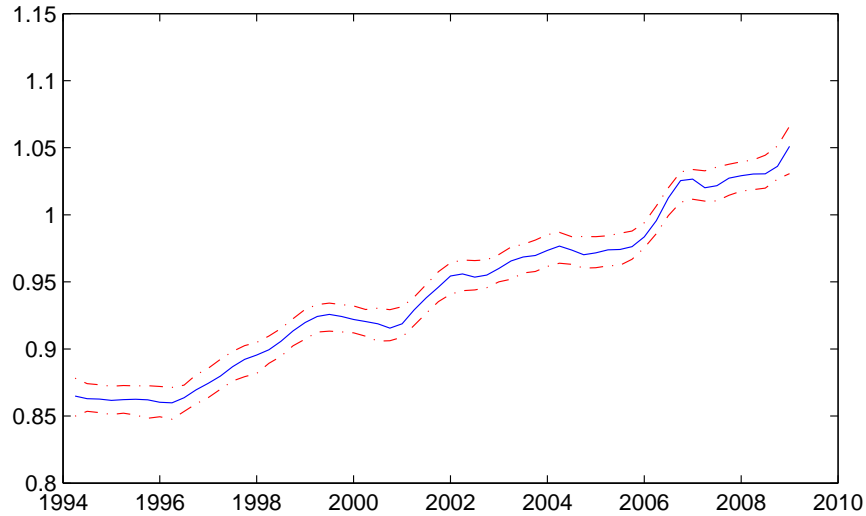
16

Figure 5.2. The estimated common trend in the log food CPI of the 8 Australian capital cities from Q1 1994 to Q4 2008 as well as its 95% confidence band.

a linear trend and use the ordinary least squares (OLS) method to estimate the resulting linear model. The OLS estimate of β is $\widehat{\beta}_{OLS} = 0.7583$ with standard deviation 0.0370, and the linear trend is estimated as $1.1143 + 0.0036t$, which also implies that there is an upward common trend in the 8 capital cities' food CPI. The coefficient of determination for the semiparametric model (5.7) is $R^2 = 0.9959$ and that for the linear model is $R^2 = 0.9877$, which implies that the semiparametric model fits the data a little better than the linear model.

Example 5.4. As a second empirical example, we use economic data extracted from the World Bank's STAR database on 82 countries over the period 1960 to 1987. This database contains data on measures of gross domestic product (GDP) and the aggregate physical capital stock, both of which are denominated in constant, end of period 1987, local currency units. The database also provides data on the number of individuals in the workforce between the ages of 15 to 64 and the mean years of schooling of members in the workforce. For cross–country comparability purposes and common econometric issues, the raw data on GDP and the physical capital stock are transformed into constant 1987 U.S. dollars. The papers of Duffy and Papageorgiou (DP) (2000) and You, Zhou and Zhou (2011) contain more details about the data. DP (2000) used the data to estimate a general constant–elasticity–of–substitution production function specification, and found that the Cobb–Douglas specifi-
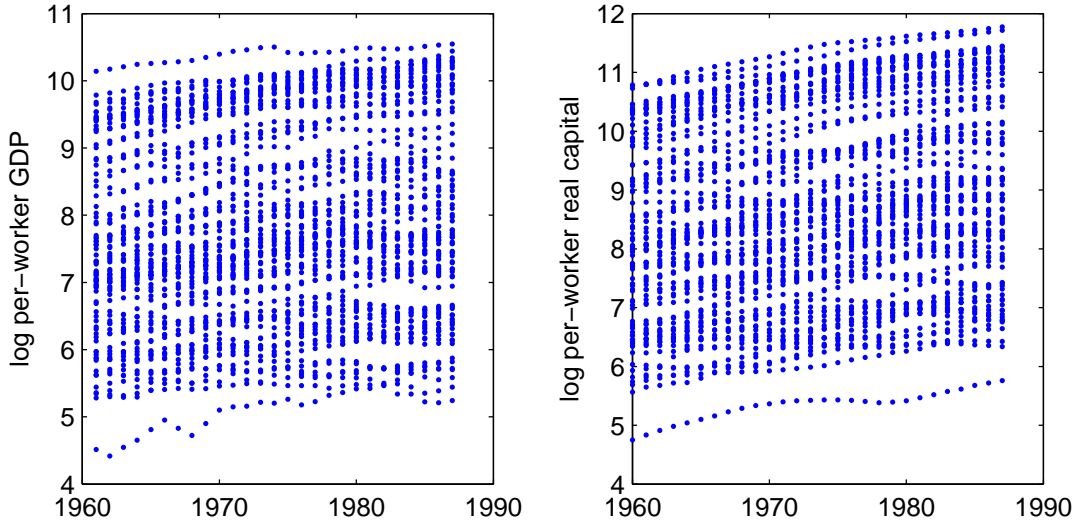
17

Figure 5.3. From left to right: scatter plots of the observations of log per–worker GDP and log per–worker real capital.

cation could be rejected for the entire sample of countries.

Let G_{it} denote real GDP, K_{it} denote real capital stock (in constant 1987 U.S. dollars), and L_{it} denote the number of people in the workforce, where $i = 1, 2, \cdots, 82$ indexes 82 countries and $t = 1, 2, \cdots, 28$ indexes the 28 years from 1960 to 1987. Denote $GL_{it} = G_{it}/L_{it}$ and $KL_{it} = K_{it}/L_{it}$, which represent real GDP per worker and real capital per worker, respectively. DP (2000) considered the following specification to characterize the input–output production relationship:

$$\log(GL_{it}) = \gamma + \lambda t + \beta_1 \log(KL_{it}) + \beta_2 [\log(KL_{it})]^2 + \alpha_i + \epsilon_{it}, \qquad (5.8)$$

where a common linear trend λt was used. We generalize this specification by replacing the linear trend $\gamma + \lambda t$ with a nonparametric trend function f_t:

$$\log(GL_{it}) = f_t + \beta_1 \log(KL_{it}) + \beta_2 [\log(KL_{it})]^2 + \alpha_i + \epsilon_{it}. \qquad (5.9)$$

The nonparametric trend f_t brings more flexibility into model (5.9) and the inclusion of the fixed effects α_i allows us to capture country–specific characteristics that might affect real GDP. Figure 5.3 gives the scatter plots of the observations of both log per–worker GDP and log per–worker real capital.

Applying the pooled profile likelihood method given in Section, we obtain the estimates of $(\beta_1, \beta_2)^\top$ as $(0.1203, 0.0237)^\top$ with standard deviations $(0.0017, 0.0002)^\top$, which implies
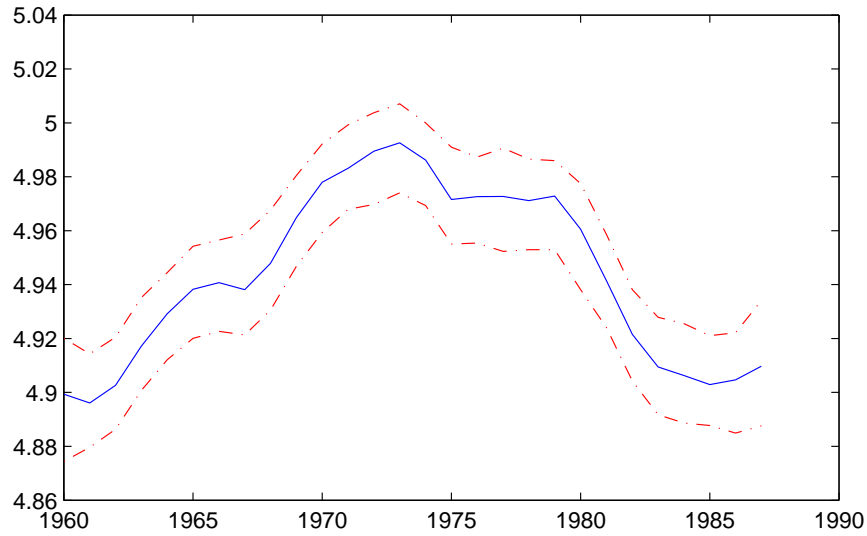
18

Figure 5.4. The estimated common trend in the per–worker GDP of the 82 countries from 1960 to 1987.

that both of the estimated coefficients are significantly different from zeros. In contrast, the OLS estimates of $(\beta_1, \beta_2)^\top$ of model (5.8) in DP (2000) is $(0.3201, 0.0318)^\top$ with standard deviations $(0.0520, 0.0029)$, which also indicates that these two coefficients are significant.

Figure 5.4 shows the pooled profile likelihood estimate of the common trend in per–worker GDP of the 82 countries as well as its 95% confidence band. The estimated curve is bell shaped and far from linear. It shows an upward trend over the period from 1960 to 1973. However, in the two years following 1973, there is a drop in the per–worker GDP, which may be a result of the oil crisis in 1973. During 1975 to 1979, the per–worker GDP remains relatively steady, which indicates that the countries were recovering from the crisis. However, a sharp decrease is seen for the period from 1979 to 1983, which shows the adverse impact of the 1979 energy crisis on these 82 countries' economy. After 1983, the per–work GDP stabilizes and starts to rise from 1985. In comparison, if we model the trend by a linear function as in (5.8), then the estimated trend resulting from the OLS would be $2.7689 - 0.0119t$, which indicates a decrease in the per-worker output from 1960 to 1987. Hence, the semiparametric model (5.9) allows us to extract information that cannot be obtained by using a simple linear trend. For example, the two turning points of 1973 and 1979 in Figure 5.4 are indicative of the two crisis in the 1970s. The linear–trend–model (5.8), however, gives no such information except a downward trend in the per-worker output from 1960 to 1987. A comparison of the coefficients of determination, which are $R^2 = 0.9932$ for

19

(5.9) and $R^2 = 0.9400$ for (5.8), also shows model (5.9) fits the data better than (5.8).

6. Conclusions and discussion

We have considered a semiparametric fixed effects panel data model with cross–sectional dependence in both the regressors and the residuals. A pooled semiparametric profile likelihood dummy variable method has been proposed to estimate both the parameter vector and the time trend function. Asymptotic distributions for the estimators of the two components have been derived with possible optimal rate of convergence when both the time series length T and the cross–sectional size N tend to infinity simultaneously.

We have also used two simulated examples to evaluate the finite–sample performance of the proposed estimation method. The simulation results have shown that the proposed pooled semiparametric profile likelihood method uniformly outperforms the averaged profile likelihood method, which is commonly used in literature. In addition, we have illustrated the proposed method through two sets of real–data with the first one being an Australian consumer price index data set and the second one being a set of input–output data from 82 countries.

There are some limitations in this paper. This paper assumes that there is no endogeneity between $\{e_{it}\}$ and $\{X_{it}\}$ while allowing for cross–sectional dependence between them. A future topic is to accommodate such endogeneity in a semiparametric model.

7. Acknowledgments

Appendix A: Proofs of the main results

Let C be a generic positive constant whose value may vary from place to place throughout the rest of this paper.

Proof of Theorem 3.1. Note that

$$\widehat{\beta} - \beta \;=\; \left(\widetilde{X}^{*\top} M^* \widetilde{X}^*\right)^{-1} \widetilde{X}^{*\top} M^* \widetilde{Y}^* - \beta$$

$$
\begin{aligned}
&= \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}M^*\left(I_{NT}-\widetilde{S}\right)\left(\widetilde{X}\beta+\widetilde{f}+D\alpha+\widetilde{e}\right)-\beta \\
&= \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}M^*\widetilde{f}^* + \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}M^*D^*\alpha \\
&\quad + \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}M^*\widetilde{e}^* \\
&=: \ \Pi_{NT}(1) + \Pi_{NT}(2) + \Pi_{NT}(3),
\end{aligned}
\tag{A.1}
$$

where $\widetilde{f}^* = \left(I_{NT}-\widetilde{S}\right)\widetilde{f}$ and $\widetilde{e}^* = \left(I_{NT}-\widetilde{S}\right)\widetilde{e}$.

As $\sum\limits_{i=1}^{N}\alpha_i = 0$, we have

$$
\begin{aligned}
\Pi_{NT}(2) &= \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}M^*D^*\alpha \\
&= \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}D^*\alpha - \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}D^*\alpha \\
&= \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}D^*\alpha - \left(\widetilde{X}^{*\top}M^*\widetilde{X}^*\right)^{-1}\widetilde{X}^{*\top}D^*\alpha = \mathbf{0}_d,
\end{aligned}
\tag{A.2}
$$

where $\mathbf{0}_d$ is a $d \times 1$ vector of zeros.

The asymptotic distribution in Theorem 3.1 can be proved via the following two propositions.

Proposition A.1. *Under A1–A3 and A5, we have*

$$
\Pi_{NT}(1) = o_P\left((NT)^{-1/2}\right).
$$

Proof. Note that $\widetilde{X}^{*\top}M^*\widetilde{X}^* = \widetilde{X}^{*\top}\widetilde{X}^* - \widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}\widetilde{X}^*$. Hence, to prove Proposition A.1, it suffices for us to prove

$$
\frac{1}{NT}\widetilde{X}^{*\top}\widetilde{X}^* = \Sigma_v + \frac{1}{N}\sum_{i=1}^{N}x_i x_i^{\top} + o_P(1),
\tag{A.3}
$$

$$
\frac{1}{NT}\widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}\widetilde{X}^* = \frac{1}{N}\sum_{i=1}^{N}x_i x_i^{\top} + o_P(1),
\tag{A.4}
$$

$$
\widetilde{X}^{*\top}M^*\widetilde{f}^* = o_P(\sqrt{NT}).
\tag{A.5}
$$

Step (i). Proof of (A.3). By the definition of $\widetilde{X}^*$ and the notation $s(\tau) = (1,0)S(\tau)$ in (2.2), we have

$$
\begin{aligned}
\frac{1}{NT}\widetilde{X}^{*\top}\widetilde{X}^* &= \frac{1}{NT}\widetilde{X}^{\top}\left(I_{NT}-\widetilde{S}\right)^{\top}\left(I_{NT}-\widetilde{S}\right)\widetilde{X} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)\left(X_{it}-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)^{\top} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}+x_i)(v_{it}+x_i)^{\top} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_t-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)(v_{it}+x_i)^{\top}
\end{aligned}
$$

21

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}+x_i)\left(g_t-\widetilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)^\top$$

$$+\frac{1}{T}\sum_{t=1}^{T}\left(g_t-\widetilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)\left(g_t-\widetilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)^\top$$

$$=: \ \Pi_{NT}^*(1)+\Pi_{NT}^*(2)+\Pi_{NT}^*(3)+\Pi_{NT}^*(4). \tag{A.6}$$

We first consider $\Pi_{NT}^*(1)$. Note that

$$\begin{aligned}
\Pi_{NT}^*(1) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}+x_i)(v_{it}+x_i)^\top\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}v_{it}v_{it}^\top\right)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_ix_i^\top+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it}x_i^\top+x_iv_{it}^\top\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}v_{it}v_{it}^\top-E\left[\frac{1}{N}\sum_{i=1}^{N}v_{it}v_{it}^\top\right]\right)+\frac{1}{T}\sum_{t=1}^{T}E\left[\frac{1}{N}\sum_{i=1}^{N}v_{it}v_{it}^\top\right]\\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_ix_i^\top+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it}x_i^\top+x_iv_{it}^\top\right)\\
&=: \ \Pi_{NT}^*(1,1)+\Pi_{NT}^*(1,2)+\Pi_{NT}^*(1,3)+\Pi_{NT}^*(1,4). \tag{A.7}
\end{aligned}$$

By the moment condition in A2 and the Markov inequality, we have, for any $\epsilon>0$,

$$P\left\{|\Pi_{NT}^*(1,1)|>\epsilon\right\}\leq\frac{1}{\epsilon^2}E\left[\Pi_{NT}^*(1,1)\right]^2$$

$$= \frac{1}{\epsilon^2T^2}\sum_{t=1}^{T}\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}v_{it}v_{it}^\top\right)=\frac{1}{\epsilon^2TN^2}\mathrm{Var}\left(\sum_{i=1}^{N}v_{it}v_{it}^\top\right)=O\left(\frac{1}{T}\right).$$

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

$$\Pi_{NT}^*(1,1)=o_P(1). \tag{A.8}$$

By A2, it is easy to check that

$$\Pi_{NT}^*(1,2)=\Sigma_v+o_P(1) \tag{A.9}$$

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

On the other hand, as $\{v_{it},\ t\geq 1\}$ is a sequence of i.i.d. random vectors for each $i\geq 1$, we have

$$\max_{1\leq i\leq N}\left\|\sum_{t=1}^{T}v_{it}\right\|=o_P(T). \tag{A.10}$$

The detailed proof of (A.10) is relegated to Appendix B. In view of (A.10) and $\max_i\|x_i\|=O_P(1)$, we get

$$\|\Pi_{NT}^*(1,4)\|\leq\frac{2}{NT}\sum_{i=1}^{N}\left(\|x_i\|\left\|\sum_{t=1}^{T}v_{it}\right\|\right)=o_P(1).$$

This combined with (A.7)–(A.9) and $\Pi_{NT}^*(1,3)=\frac{1}{N}\sum_{i=1}^{N}x_ix_i^\top$ yields

$$\Pi_{NT}^*(1)=\Sigma_v+\frac{1}{N}\sum_{i=1}^{N}x_ix_i^\top+o_P(1). \tag{A.11}$$

For $\Pi^*_{NT}(4)$, we use the uniform consistency result:

$$\sup_{0 \le \tau \le 1} \left\| g(\tau) - \widetilde{X}^\top s^\top(\tau) \right\| = O_P \left(h^2 + \sqrt{\frac{\log(NT)}{NTh}} \right). \tag{A.12}$$

The detailed proof of (A.12) will be given in Appendix B. From (A.12), it is easy to show

$$\Pi^*_{NT}(4) = o_P(1). \tag{A.13}$$

Noting that $\sum_{i=1}^{N} x_i x_i^\top = O_P(N)$, by (A.11), (A.13) and the Cauchy–Schwarz inequality,

$$\Pi^*_{NT}(2) = o_P(1) \quad \text{and} \quad \Pi^*_{NT}(3) = o_P(1). \tag{A.14}$$

With (A.6), (A.11), (A.13) and (A.14), we have shown that (A.3) holds.

Step (ii). Proof of (A.4). As $\widetilde{S}D = 0$, we have

$$D^{*\top} D^* = D^\top \left(I_{NT} - \widetilde{S} \right)^\top \left(I_{NT} - \widetilde{S} \right) D = D^\top D. \tag{A.15}$$

Furthermore,

$$D^\top D = \begin{pmatrix} 2T & T & \cdots & T \\ T & 2T & \cdots & T \\ \vdots & \vdots & \cdots & \vdots \\ T & T & \cdots & 2T \end{pmatrix} = \begin{pmatrix} T & T & \cdots & T \\ T & T & \cdots & T \\ \vdots & \vdots & \cdots & \vdots \\ T & T & \cdots & T \end{pmatrix} + \mathrm{diag}(T, \cdots, T). \tag{A.16}$$

Letting $A = \mathrm{diag}(T, \cdots, T)$, $B = (1, \cdots, 1)^\top$, $C = T$ and $P = (1, \cdots, 1)$, and applying the result about the inverse matrix (Poirier 1995):

$$(A + BCP)^{-1} = A^{-1} - A^{-1} B \left(PA^{-1}B + C^{-1} \right)^{-1} PA^{-1},$$

we have

$$\left(D^{*\top} D^* \right)^{-1} = \left(D^\top D \right)^{-1} = \begin{pmatrix} \frac{1}{T} - \frac{1}{NT} & -\frac{1}{NT} & \cdots & -\frac{1}{NT} \\ -\frac{1}{NT} & \frac{1}{T} - \frac{1}{NT} & \cdots & -\frac{1}{NT} \\ \vdots & \vdots & \cdots & \vdots \\ -\frac{1}{NT} & -\frac{1}{NT} & \cdots & \frac{1}{T} - \frac{1}{NT} \end{pmatrix}. \tag{A.17}$$

Meanwhile, by the definitions of $\widetilde{X}^*$ and D^*, we have

$$\widetilde{X}^{*\top} D^* = \widetilde{X}^{*\top} D = \left(A_T(2), \cdots, A_T(N) \right), \tag{A.18}$$

where $A_T(k) = -\sum_{t=1}^{T} X_{1t} + \sum_{t=1}^{T} X_{kt}$, $k \ge 2$.

23

Letting $A_T^*(k) = \frac{1}{T} A_T(k) - \frac{1}{NT} \sum_{k=2}^{N} A_T(k)$, by (A.17) and (A.18), we then have

$$\frac{1}{NT} \widetilde{X}^{*\top} D^* \left(D^{*\top} D^* \right)^{-1} D^{*\top} \widetilde{X}^* = \frac{1}{NT} \left(\sum_{k=2}^{N} A_T^*(k) A_T^\top(k) \right)$$

$$= \frac{1}{N} \sum_{k=2}^{N} \left(\frac{1}{T} A_T(k) \right) \left(\frac{1}{T} A_T^\top(k) \right) - \left(\frac{1}{N} \sum_{k=2}^{N} \frac{1}{T} A_T(k) \right) \left(\frac{1}{N} \sum_{k=2}^{N} \frac{1}{T} A_T^\top(k) \right).$$

By (A.10) and the definition of $A_T(k)$, we can show that

$$\frac{1}{T} A_T(k) = x_k - x_1 + o_P(1) \tag{A.19}$$

uniformly for $2 \le k \le N$, which implies

$$\frac{1}{N} \sum_{k=2}^{N} \left(\frac{1}{T} A_T(k) \right) \left(\frac{1}{T} A_T^\top(k) \right) = \frac{1}{N} \sum_{k=2}^{N} x_k x_k^\top + \frac{N+1}{N} x_1 x_1^\top + o_P(1) \tag{A.20}$$

and

$$\left(\frac{1}{N} \sum_{k=2}^{N} \frac{1}{T} A_T(k) \right) \left(\frac{1}{N} \sum_{k=2}^{N} \frac{1}{T} A_T^\top(k) \right) = x_1 x_1^\top + o_P(1). \tag{A.21}$$

With (A.20) and (A.21), we have show that (A.4) holds.

Step (iii). Proof of (A.5). Note that

$$\widetilde{X}^{*\top} M^* \widetilde{f}^* = \widetilde{X}^{*\top} \widetilde{f}^* - \widetilde{X}^{*\top} D^* \left(D^{*\top} D^* \right)^{-1} D^{*\top} \widetilde{f}^*.$$

We first prove that

$$\widetilde{X}^{*\top} \widetilde{f}^* = o_P(\sqrt{NT}). \tag{A.22}$$

By the definition of $\widetilde{X}^*$ and $\widetilde{f}^*$, we have

$$
\begin{aligned}
\widetilde{X}^{*\top} \widetilde{f}^* &= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \widetilde{X}^\top s^\top \left(\frac{t}{T} \right) \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(v_{it} + \left[g \left(\frac{t}{T} \right) - \widetilde{X}^\top s^\top \left(\frac{t}{T} \right) \right] \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) - \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{v}^\top s^\top \left(\frac{t}{T} \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g \left(\frac{t}{T} \right) - \widetilde{g}^\top s^\top \left(\frac{t}{T} \right) \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} x_i \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) + \sum_{i=1}^{N} \sum_{t=1}^{T} x^\top s^\top \left(\frac{t}{T} \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&=: \Pi_{NT}^*(5) + \Pi_{NT}^*(6) + \Pi_{NT}^*(7) + \Pi_{NT}^*(8) + \Pi_{NT}^*(9), \tag{A.23}
\end{aligned}
$$

where $\widetilde{g} = i_N \otimes (g_1, g_2, \cdots, g_T)^\top$, $\widetilde{v} = (v_{11}, \cdots, v_{1T}, v_{21}, \cdots, v_{2T}, \cdots, v_{NT})^\top$ and $x = (x_1, \cdots, x_N)^\top \otimes i_T$.

Following the argument in the proof of (A.12) in Appendix B and by A2 and A3, we have

$$\sup_{0 \leq \tau \leq 1} \left| f(\tau) - s(\tau)\widetilde{f} \right| = O\left(h^2\right),$$

$$\sup_{0 \leq \tau \leq 1} \left\| g(\tau) - \widetilde{g}^\top s^\top(\tau) \right\| = O\left(h^2\right),$$

$$\sup_{0 \leq \tau \leq 1} \left\| \widetilde{v}^\top s^\top(\tau) \right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right),$$

which, together with A5, imply

$$\Pi_{NT}^*(5) = O_P\left(\sqrt{NT}h^2\right) = o_P(\sqrt{NT}), \tag{A.24}$$

$$\Pi_{NT}^*(6) = O_P\left(\sqrt{NT\log(NT)h^3}\right) = o_P(\sqrt{NT}), \tag{A.25}$$

$$\Pi_{NT}^*(7) = O_P\left(NTh^4\right) = o_P(\sqrt{NT}). \tag{A.26}$$

On the other hand, noticing $\sum_{i=1}^{N} x_i = \mathbf{0}_d$ and the fact that $s\left(\frac{t}{T}\right)$ is independent of i, we have

$$\Pi_{NT}^*(8) = \mathbf{0}_d = o_P(\sqrt{NT}) \quad \text{and} \quad \Pi_{NT}^*(9) = \mathbf{0}_d = o_P(\sqrt{NT}). \tag{A.27}$$

By (A.23)–(A.27), we have shown that (A.22) holds. Following the argument in Step (ii), we can also show

$$\widetilde{X}^{*\top} D^* \left(D^{*\top} D^*\right)^{-1} D^{*\top} \widetilde{f}^* = o_P(\sqrt{NT}). \tag{A.28}$$

Then, with (A.22) and (A.28), we prove (A.5).

In view of (A.3)–(A.5), the proof of Proposition A.1 is completed.

Proposition A.2. *Let A1–A5 hold. Then we have*

$$\sqrt{NT}\Pi_{NT}(3) \xrightarrow{d} N\left(\mathbf{0}_d, \Sigma_v^{-1}\Sigma_{v,e}\Sigma_v^{-1}\right). \tag{A.29}$$

Proof. To prove (A.29), it suffices for us to show

$$\frac{1}{NT}\widetilde{X}^{*\top} M^* \widetilde{X}^* \xrightarrow{P} \Sigma_v \tag{A.30}$$

and

$$\frac{1}{\sqrt{NT}}\widetilde{X}^{*\top} M^* \widetilde{e}^* \xrightarrow{d} N\left(0, \Sigma_{v,e}\right). \tag{A.31}$$

By (A.3) and (A.4) in the proof of Proposition A.1, we can easily obtain (A.30). For the proof of (A.31), observe that

$$\widetilde{X}^{*\top} M^* \widetilde{e}^* = \widetilde{X}^{*\top} \widetilde{e}^* - \widetilde{X}^{*\top} D^* \left(D^{*\top} D^*\right)^{-1} D^{*\top} \widetilde{e}^* =: \Pi_{NT}^*(10) + \Pi_{NT}^*(11). \tag{A.32}$$

For $\Pi^*_{NT}(10)$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\Pi^*_{NT}(10) \;=\;& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}e_{it} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_i e_{it} \\
& + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g\left(\frac{t}{T}\right) - \tilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)e_{it} \\
& - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}s\left(\frac{t}{T}\right)\tilde{e} - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_i s\left(\frac{t}{T}\right)\tilde{e} \\
& - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g\left(\frac{t}{T}\right) - \tilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)s\left(\frac{t}{T}\right)\tilde{e} \\
=:\;& \sum_{j=1}^{6}\Pi^*_{NT}(10,j). \tag{A.33}
\end{aligned}
$$

Following the proof of (A.12) in Appendix B, and by A2, A4 and A5, we have

$$
\Pi^*_{NT}(10,3) = O_P\left(h^2 + \sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1), \tag{A.34}
$$

$$
\Pi^*_{NT}(10,4) = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1), \tag{A.35}
$$

$$
\Pi^*_{NT}(10,6) = O_P\left(\sqrt{NT}\left(h^2 + \sqrt{\frac{\log(NT)}{NTh}}\right)\sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1). \tag{A.36}
$$

Similarly to the proof of (A.27), we have

$$
\Pi^*_{NT}(10,5) = \mathbf{0}_d = o_P(1). \tag{A.37}
$$

By (A.33)–(A.37), we have

$$
\frac{1}{\sqrt{NT}}\Pi^*_{NT}(10) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}e_{it} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_i e_{it} + o_P(1). \tag{A.38}
$$

On the other hand, by (A.17), (A.18) and a standard calculation, we have

$$
\Pi^*_{NT}(11) = \frac{1}{T}\sum_{k=2}^{N}A_T(k)B_T(k) - \frac{1}{NT}\left(\sum_{k=2}^{N}A_T(k)\right)\left(\sum_{k=2}^{N}B_T(k)\right),
$$

where $A_T(k) = \sum\limits_{t=1}^{T}X_{kt} - \sum\limits_{t=1}^{T}X_{1t}$ and $B_T(k) = \sum\limits_{t=1}^{T}e_{kt} - \sum\limits_{t=1}^{T}e_{1t}$. Observe that

$$
\begin{aligned}
A_T(k) \;=\;& \sum_{t=1}^{T}X_{kt} - \sum_{t=1}^{T}X_{1t} \\
=\;& \left(\sum_{t=1}^{T}v_{kt} - \sum_{t=1}^{T}v_{1t}\right) + T(x_k - x_1) \\
=:\;& A_T(k,1) + A_T(k,2).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\Pi_{NT}^*(11) &= \frac{1}{T}\sum_{k=2}^{N}\left(A_T(k,1)+A_T(k,2)\right)B_T(k)\\
&\quad -\frac{1}{NT}\left(\sum_{k=2}^{N}A_T(k,1)+A_T(k,2)\right)\left(\sum_{k=2}^{N}B_T(k)\right)\\
&= \frac{1}{T}\sum_{k=2}^{N}A_T(k,1)B_T(k)-\frac{1}{NT}\left(\sum_{k=2}^{N}A_T(k,1)\right)\left(\sum_{k=2}^{N}B_T(k)\right)\\
&\quad +\frac{1}{T}\sum_{k=2}^{N}A_T(k,2)B_T(k)-\frac{1}{NT}\left(\sum_{k=2}^{N}A_T(k,2)\right)\left(\sum_{k=2}^{N}B_T(k)\right)\\
&=: \Pi_{NT}^*(12)+\Pi_{NT}^*(13).
\end{aligned}
\tag{A.39}
$$

By the identification condition on $\{x_i\}$, we have

$$
\Pi_{NT}^*(13)=\sum_{i=1}^{N}\sum_{t=1}^{T}x_i e_{it}.
$$

We then consider $\Pi_{NT}^*(12)$. Define $\overline{A}_T(k)=\sum_{t=1}^{T}v_{kt}$ and $\overline{B}_T(k)=\sum_{t=1}^{T}e_{kt}$ for $k=1,\cdots,N$. Then,

$$
\begin{aligned}
\Pi_{NT}^*(12) &= \frac{1}{T}\sum_{k=2}^{N}A_T(k,1)B_T(k)-\frac{1}{NT}\left(\sum_{k=2}^{N}A_T(k,1)\right)\left(\sum_{k=2}^{N}B_T(k)\right)\\
&= \frac{1}{T}\sum_{k=2}^{N}\overline{A}_T(k)\overline{B}_T(k)-\frac{1}{T}\overline{B}_T(1)\sum_{k=2}^{N}\overline{A}_T(k)-\frac{1}{T}\overline{A}_T(1)\sum_{k=2}^{N}\overline{B}_T(k)\\
&\quad +\frac{N-1}{T}\overline{A}_T(1)\overline{B}_T(1)-\frac{1}{NT}\left(\sum_{k=2}^{N}\overline{A}_T(k)\right)\left(\sum_{k=2}^{N}\overline{B}_T(k)\right)\\
&\quad +\frac{N-1}{NT}\overline{B}_T(1)\sum_{k=2}^{N}\overline{A}_T(k)+\frac{N-1}{NT}\overline{A}_T(1)\sum_{k=2}^{N}\overline{B}_T(k)-\frac{(N-1)^2}{NT}\overline{A}_T(1)\overline{B}_T(1)\\
&= \frac{1}{T}\sum_{k=2}^{N}\overline{A}_T(k)\overline{B}_T(k)-\frac{1}{NT}\left(\sum_{k=2}^{N}\overline{A}_T(k)\right)\left(\sum_{k=2}^{N}\overline{B}_T(k)\right)\\
&\quad -\frac{1}{NT}\overline{B}_T(1)\sum_{k=2}^{N}\overline{A}_T(k)-\frac{1}{NT}\overline{A}_T(1)\sum_{k=2}^{N}\overline{B}_T(k)+\frac{N-1}{NT}\overline{A}_T(1)\overline{B}_T(1)\\
&=: \sum_{j=1}^{5}\Pi_{NT}^*(12,j).
\end{aligned}
\tag{A.40}
$$

By A2 and A4, we have, as $N,T\to\infty$ simultaneously,

$$
\begin{aligned}
&E\left(\Pi_{NT}^*(12,1)\Pi_{NT}^{*\top}(12,1)\right)\\
&= \frac{1}{T^2}E\left[\left(\sum_{k=2}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{kt}v_{ks}\right)\left(\sum_{k=2}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{kt}v_{ks}\right)^{\top}\right]
\end{aligned}
$$

$$= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k_1=2}^{N} \sum_{k_2=2}^{N} E\left(e_{k_1 t} e_{k_2 t}\right) E\left(v_{k_1 s} v_{k_2 s}^\top\right)$$

$$= \sum_{k_1=2}^{N} \sum_{k_2=2}^{N} E\left(e_{k_1,1} e_{k_2,1}\right) E\left(v_{k_1,1} v_{k_2,1}^\top\right) = O(N),$$

$$E\left(\Pi_{NT}^*(12,2) \Pi_{NT}^{*\top}(12,2)\right)$$

$$= \frac{1}{N^2 T^2} E\left(\sum_{k_1=2}^{N} \sum_{t_1=1}^{T} e_{k_1 t_1}\right)^2 E\left[\left(\sum_{k_2=2}^{N} \sum_{t_2=1}^{T} v_{k_2 t_2}\right)\left(\sum_{k_2=2}^{N} \sum_{t_2=1}^{T} v_{k_2 t_2}\right)^\top\right]$$

$$= \frac{1}{N^2 T^2} \left[\sum_{k_1=2}^{N} \sum_{j_1=2}^{N} \sum_{t_1=1}^{T} E\left(e_{k_1 t_1} e_{j_1 t_1}\right)\right] \left[\sum_{k_2=2}^{N} \sum_{j_2=2}^{N} \sum_{t_2=1}^{T} E\left(v_{k_2 t_2} v_{j_2 t_2}^\top\right)\right]$$

$$= \frac{1}{N^2} \left[\sum_{k_1=2}^{N} \sum_{j_1=2}^{N} E\left(e_{k_1,1} e_{j_1,1}\right)\right] \left[\sum_{k_2=2}^{N} \sum_{j_2=2}^{N} E\left(v_{k_2,1} v_{j_2,1}^\top\right)\right] = O(1),$$

and similarly,

$$E\left(\Pi_{NT}^*(12,3) \Pi_{NT}^{*\top}(12,3)\right) = O\left(\frac{1}{N}\right),$$

$$E\left(\Pi_{NT}^*(12,4) \Pi_{NT}^{*\top}(12,4)\right) = O\left(\frac{1}{N}\right),$$

$$E\left(\Pi_{NT}^*(12,5) \Pi_{NT}^{*\top}(12,5)\right) = O(1).$$

Hence,

$$\Pi_{NT}^*(12,j) = o_P(\sqrt{NT}), \quad j = 1, \cdots, 5. \tag{A.41}$$

Combining (A.39)–(A.41), we have

$$\Pi_{NT}^*(11) = \sum_{i=1}^{N} \sum_{t=1}^{T} x_i e_{it} + o_P(\sqrt{NT}). \tag{A.42}$$

By (A.32), (A.38) and (A.42), we have

$$\widetilde{X}^{*\top} M^* \widetilde{e}^* = \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} e_{it} + o_P(\sqrt{NT}). \tag{A.43}$$

We next prove that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} e_{it} \xrightarrow{d} N\left(0, \Sigma_{v,e}\right). \tag{A.44}$$

As both T and N tend to infinity, we next prove (A.44) by the joint limit approach (see Phillips and Moon 1999 for example). Letting $Z_{t,N}(v,e) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{it} e_{it}$, we have

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} e_{it} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,N}(v,e).$$

By A2 and A4, $\{Z_{t,N}(v,e), \ t \geq 1\}$ is a sequence of i.i.d. random vectors. Hence, we apply the Lindeberg–Feller central limit theorem to prove (A.44). For any $\epsilon > 0$,

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\|Z_{t,N}(v,e)\|^2 I\left\{\|Z_{t,N}(v,e)\| \geq \epsilon\sqrt{T}\right\}\right)$$
$$= E\left(\|Z_{1,N}(v,e)\|^2 I\left\{\|Z_{1,N}(v,e)\| \geq \epsilon\sqrt{T}\right\}\right) \to 0$$

as $N, T \to \infty$ simultaneously, which implies that the Lindeberg condition is satisfied, which in turn implies the validity of (A.44). Hence, the proof of Proposition A.2 is completed.

Proof of Theorem 3.2. By the definition of $\widehat{f}(\tau)$ in (2.6), we have

$$
\begin{aligned}
\widehat{f}(\tau) - f(\tau) &= s(\tau)\left(\widetilde{Y} - \widetilde{X}\widehat{\beta} - D\widehat{\alpha}\right) - f(\tau) \\
&= s(\tau)\left(I_{NT} - D\left(D^{*\top}D^*\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\left(\widetilde{Y} - \widetilde{X}\widehat{\beta}\right) - f(\tau) \\
&= \left(s(\tau)\left(I_{NT} - D\left(D^{*\top}D^*\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{f} - f(\tau)\right) \\
&\quad + s(\tau)\left(I_{NT} - D\left(D^{*\top}D^*\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{e} \\
&\quad + s(\tau)\left(I_{NT} - D\left(D^{*\top}D^*\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{X}(\beta - \widehat{\beta}) \\
&=: \Pi_{NT}^*(14) + \Pi_{NT}^*(15) + \Pi_{NT}^*(16).
\end{aligned}
\tag{A.45}
$$

Note that

$$s(\tau)D = (1,0)\left(\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{Z}(\tau)\right)^{-1}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)D = \mathbf{0}_{N-1}^\top.$$

Hence, by (A.12) and standard argument for local linear fitting, we have

$$\Pi_{NT}^*(14) = s(\tau)\widetilde{f} - f(\tau) = \frac{1}{2}f''(\tau)\mu_2 h^2 + o_P(h^2).\tag{A.46}$$

By the Lindeberg–Feller central limit theorem and following the proof of Proposition A.2, we have

$$\sqrt{NTh}\Pi_{NT}^*(15) = \sqrt{NTh}s(\tau)\widetilde{e} \xrightarrow{d} N(0, \nu_0\sigma_e^2).\tag{A.47}$$

By Theorem 3.1, we have

$$\Pi_{NT}^*(16) = s(\tau)\widetilde{X}(\beta - \widehat{\beta}) = O_P\left((NT)^{-1/2}\right) = o_P\left((NTh)^{-1/2}\right).\tag{A.48}$$

Hence, by (A.45)–(A.48), we have completed the proof of (3.8).

Appendix B: Proof of some uniform consistency results

Proof of (A.12). By the identification condition $\sum_{i=1}^{N} x_i = 0$, we have

$$g(\tau) - \widetilde{X}^\top s^\top(\tau) = \left(g(\tau) - \widetilde{g}^\top s^\top(\tau)\right) - \widetilde{v}^\top s^\top(\tau) =: \Xi_{NT,1}(\tau) - \Xi_{NT,2}(\tau),\tag{B.1}$$

where $\widetilde{g} = i_N \otimes (g_1, g_2, \cdots, g_T)^\top$, in which i_N is the $N \times 1$ vector of ones.

We first prove that

$$\sup_{0 \leq \tau \leq 1} \|\Xi_{NT,2}(\tau)\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right). \tag{B.2}$$

By the definition of $s(\tau)$ in Section 2, we have

$$\Xi_{NT,2}^\top(\tau) = (1,0)S(\tau)\widetilde{v} = (1,0)\left(\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{Z}(\tau)\right)^{-1}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{v}.$$

Note that

$$\frac{1}{NTh}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{Z}(\tau) = \begin{pmatrix} \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{t-\tau T}{Th}\right) & \frac{1}{Th}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)K\left(\frac{t-\tau T}{Th}\right) \\ \frac{1}{Th}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)K\left(\frac{t-\tau T}{Th}\right) & \frac{1}{Th}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^2 K\left(\frac{t-\tau T}{Th}\right) \end{pmatrix}.$$

By the definition of Riemann integral, we have

$$\frac{1}{Th}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^j K\left(\frac{t-\tau T}{Th}\right) = \mu_j(\tau) + O\left(\frac{1}{Th}\right),$$

uniformly for $0 \leq \tau \leq 1$, where $\mu_j(\tau) = \mu_j = \int_{-1}^{1} u^j K(u)du$ if $h \leq \tau \leq 1-h$; $\mu_j(\tau) = \int_{-c}^{1} u^j K(u)du$ if $\tau = ch$ $(0 \leq c \leq 1)$; and $\mu_j(\tau) = \int_{-1}^{c} u^j K(u)du$ if $\tau = 1-ch$ $(0 \leq c \leq 1)$. Hence, we have

$$\sup_{0 \leq \tau \leq 1}\left\|\frac{1}{NTh}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{Z}(\tau) - \Lambda_\mu\right\| = O\left(\frac{1}{Th}\right) = o(1), \tag{B.3}$$

where

$$\Lambda_\mu = \begin{pmatrix} \mu_0(\tau) & \mu_1(\tau) \\ \mu_1(\tau) & \mu_2(\tau) \end{pmatrix}.$$

In view of (B.3), to prove (B.2), we need only to prove

$$\sup_{0 \leq \tau \leq 1}\left\|\frac{1}{NTh}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{v}\right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right). \tag{B.4}$$

Note that

$$\frac{1}{NTh}\widetilde{Z}^\top(\tau)\widetilde{W}(\tau)\widetilde{v} = \begin{pmatrix} \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T} K\left(\frac{t-\tau T}{Th}\right)v_{it} \\ \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)K\left(\frac{t-\tau T}{Th}\right)v_{it} \end{pmatrix}.$$

Hence, to prove (B.4), it suffices to show that for $j = 0, 1$,

$$\sup_{0 \leq \tau \leq 1}\left\|\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^j K\left(\frac{t-\tau T}{Th}\right)v_{it}\right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right). \tag{B.5}$$

Define $\overline{Q}_{t,N}(v) = \frac{1}{N}\sum_{i=1}^{N} v_{it}$. It is easy to see that (B.5) is equivalent to

$$\sup_{0 \leq \tau \leq 1}\left\|\frac{1}{Th}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^j K\left(\frac{t-\tau T}{Th}\right)\overline{Q}_{t,N}(v)\right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right). \tag{B.6}$$

Let $l(\cdot)$ be any positive function that satisfies $l(n) \to \infty$ as $n \to \infty$. Then to prove (B.6), it suffices to prove

$$\sup_{0 \leq \tau \leq 1} \left\| \frac{1}{Th} \sum_{t=1}^{T} \left(\frac{t - \tau T}{Th} \right)^j K \left(\frac{t - \tau T}{Th} \right) \overline{Q}_{t,N}(v) \right\| = o_P \left(l(NT) \sqrt{\frac{\log(NT)}{NTh}} \right). \qquad (B.7)$$

We then cover the interval $[0, 1]$ by a finite number of subintervals $\{B_l\}$ that are centered at b_l and of length $\delta_{NT} = o(h^2)$. Denoting U_{NT} the number of such subintervals, then $U_{NT} = O\left(\delta_{NT}^{-1} \right)$. Define $\overline{K}_{t,j}(\tau) = \frac{1}{Th} \left(\frac{t - \tau T}{Th} \right)^j K \left(\frac{t - \tau T}{Th} \right)$. Observe that

$$\sup_{0 \leq \tau \leq 1} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(\tau) \overline{Q}_{t,N}(v) \right\| \leq \max_{1 \leq l \leq U_{NT}} \sup_{\tau \in B_l} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(\tau) \overline{Q}_{t,N}(v) - \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \overline{Q}_{t,N}(v) \right\|$$

$$+ \max_{1 \leq l \leq U_{NT}} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \overline{Q}_{t,N}(v) \right\| =: \Xi_{NT,3} + \Xi_{NT,4}. \qquad (B.8)$$

By A1 and taking $\delta_{NT} = O\left((l(NT))^{1+\delta} \sqrt{\frac{\log(NT)}{NTh}} h^2 \right)$, we have

$$\Xi_{NT,3} = O_P \left(\frac{\delta_{NT}}{h^2} E \left\| \overline{Q}_{1,N}(v) \right\| \right) = o_P \left(l(NT) \sqrt{\frac{\log(NT)}{NTh}} \right). \qquad (B.9)$$

For $\Xi_{NT,4}$, we apply the truncation technique. Define

$$\widetilde{Q}_{t,N}(v) = \overline{Q}_{t,N}(v) I \left\{ \left\| \overline{Q}_{t,N}(v) \right\| \leq N^{-1/2} T^{1/\delta} l(NT) \right\},$$
$$\widetilde{Q}_{t,N}^c(v) = \overline{Q}_{t,N}(v) - \widetilde{Q}_{t,N}(v).$$

Note that

$$\Xi_{NT,4} \leq \max_{1 \leq l \leq U_{NT}} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \widetilde{Q}_{t,N}(v) \right\| + \max_{1 \leq l \leq U_{NT}} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \widetilde{Q}_{t,N}^c(v) \right\|$$

$$=: \Xi_{NT,5} + \Xi_{NT,6}. \qquad (B.10)$$

For $\Xi_{NT,6}$, applying the Markov inequality and A2, we have for any $\epsilon > 0$

$$P \left\{ \Xi_{NT,6} > \epsilon l(NT) \sqrt{\frac{\log(NT)}{NTh}} \right\} \leq P \left\{ \max_{1 \leq t \leq T} \left\| \widetilde{Q}_{t,N}^c(v) \right\| > 0 \right\}$$

$$\leq \sum_{t=1}^{T} P \left\{ \left\| \overline{Q}_{t,N}(v) \right\| > N^{-1/2} T^{1/\delta} l(NT) \right\} = O \left(E \left\| \overline{Q}_{t,N}(v) \right\|^\delta N^{\delta/2} (l(NT))^{-\delta} \right)$$

$$= O \left((l(NT))^{-\delta} \right) = o(1),$$

which implies

$$\Xi_{NT,6} = o_P \left(l(NT) \sqrt{\frac{\log(NT)}{NTh}} \right). \qquad (B.11)$$

31

For $\Xi_{NT,5}$, observe that

$$\left\| \overline{K}_{t,j}(b_l)\widetilde{Q}_{t,N}(v) \right\| \leq N^{-1/2}T^{-1+1/\delta}h^{-1}l(NT).$$

Applying A2, A5 and Bernstein's inequality for i.i.d. random vectors, we have

$$
\begin{aligned}
P &\left\{ \Xi_{NT,5} > \epsilon l(NT)\sqrt{\frac{\log(NT)}{NTh}} \right\} \\
&\leq C\delta_{NT}^{-1}\exp\left\{ -\frac{\epsilon^2 l^2(NT)\log(NT)}{C + C\epsilon T^{1/\delta - 1/2}h^{-1/2}l^2(NT)(\log(NT))^{1/2}} \right\} \\
&\leq C\delta_{NT}^{-1}\exp\{-M\log(NT)\} = o(1),
\end{aligned}
\tag{B.12}
$$

where $C > 0$ is a constant and M is a sufficiently large positive constant. The second inequality above holds by letting $l(\cdot)$ satisfy

$$l(NT) \to \infty \quad \text{and} \quad \frac{T^{1-\frac{2}{\delta}}h}{l^4(NT)\log(NT)} \to \infty.$$

Hence,

$$\Xi_{NT,5} = o_P\left(l(NT)\sqrt{\frac{\log(NT)}{NTh}} \right). \tag{B.13}$$

From (B.8)–(B.13), we can see that (B.7) holds, which in turn implies the validity of (B.2).

Meanwhile, following standard argument in local linear fitting (see, for example, Fan and Gijbels 1996), we can show

$$\sup_{0\leq\tau\leq 1} \|\Xi_{NT,1}(\tau)\| = O_P\left(h^2 \right). \tag{B.14}$$

In view of (B.1), (B.2) and (B.14), it has been shown that (A.12) holds.

Proof of (A.19). By Burkholder's inequality for i.i.d. random vectors, we have

$$E\left\| \sum_{t=1}^{T} v_{it} \right\|^{\delta} \leq C_\delta T^{\delta/2}, \tag{B.15}$$

where C_δ is a positive constant which only depends on δ and is independent of i. By A5, (B.15) and the Markov inequality, we have, for any $\epsilon > 0$,

$$
\begin{aligned}
P &\left(\max_{1\leq i\leq N} \|\sum_{t=1}^{T} v_{it}\| > \epsilon T \right) \leq \sum_{i=1}^{N} P\left(\|\sum_{t=1}^{T} v_{it}\| > \epsilon T \right) \\
&\leq \sum_{i=1}^{N} \frac{E\left\| \sum_{t=1}^{T} v_{it} \right\|^{\delta}}{\epsilon^\delta T^\delta} = O(T^{-\delta/2}N) = o(1),
\end{aligned}
\tag{B.16}
$$

which implies that (A.19) holds.

References

Atak, A., Linton, O., Xiao, Z., 2010. A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. Forthcoming in *Journal of Econometrics* (available at http://mpra.ub.uni–muenchen.de/22079/).

Cai, Z., 2007. Trending time–varying coefficient time series models with serially correlated errors. *Journal of Econometrics* **136**, 163-188.

Chen, J., Gao, J., Li, D., 2010. A new diagnostic test for cross–section uncorrelatedness in nonparametric panel data models. Working paper available at http://www.economics.adelaide.edu.au/research/papers/.

Duffy, J., Papageorgiou, C. 2000. A cross–country empirical investigation of the aggregate production function apecification. *Journal of Economic Growth* **5**, 87–120.

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

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

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

Gao, J., Hawthorne, K., 2006. Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal* **9**, 332-355.

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

Li, D., Chen, J., Gao, J., 2010. Nonparametric time–varying coefficient panel data models with fixed effects. Working paper available at http://www.economics.adelaide.edu.au/research/papers/.

Pesaran, M. H., 2006. Estimation and inference in large heterogenous panels with multifactor error. *Econometrica* 74, 967–1012.

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

Phillips, P. C. B. and Moon, H., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* **67**, 1057-1111.

Poirier, D. J., 1995. *Intermediate Statistics and Econometrics: a Comparative Approach*. MIT Press, Boston.

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

Robinson, P. M., 2010. Nonparametric trending regression with cross-sectional dependence. Forthcoming in *Journal of Econometrics*.

Su, L., Jin, S., 2011. Sieve estimation of panel data models with cross section dependence. Forthcoming in *Journal of Econometrics* (available at http://www.mysmu.edu/faculty/ljsu/research.htm).

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

You, J., Zhou, X., Zhou, Y. 2011. Series estimation in partially linear in–slide regression models. *Scandinavian Journal of Statistics* **38**, 89-107.