

The University of Adelaide
School of Economics

Research Paper No. 2010-10
May 2010

Semiparametric Trending Panel Data Models with Cross-Sectional Dependence

Jia Chen, Jiti Gao, and Degui Li

Semiparametric Trending Panel Data Models with Cross-Sectional Dependence

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

School of Economics, The University of Adelaide, Adelaide, Australia

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 semiparametric profile likelihood approach based on the first–stage local linear fitting is developed to estimate both the parameter vector and the time trend function. As both the time series length T and the cross–sectional size N tend to infinity simultaneously, the resulting semiparametric estimator of the parameter vector is asymptotically normal with a rate of convergence of $O_P\left(\frac{1}{\sqrt{NT}}\right)$. Meanwhile, an asymptotic distribution for the estimate of the nonlinear time 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 behavior of the proposed estimation method. In addition, the proposed model and estimation method is applied to the analysis of two sets of real data.

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*: Dr Degui Li, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: degui.li@adelaide.edu.au

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 semiparametric 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 (2009) propose a semiparametric 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 (2008) 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 of 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

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

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

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, and α_i are fixed effects satisfying

$$\sum_{i=1}^{N} \alpha_i = 0. \tag{1.3}$$

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 $\{v_{it}\}$.

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.

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 estimate 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 estimator of β and estimate of $f(\cdot)$.

The rest of the paper is organized as follows. A semiparametric pooled profile likelihood method for β and $f(\cdot)$ is proposed in Section 2. The asymptotic theory of the proposed estimation method is established in Section 3. Some related discussions, such as estimation of some covariance matrices, an 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 applications are provided in Section 5. Section 6 concludes the paper. The mathematical proofs of the main results are given in Appendices A and B.

2. Estimation method

Several existing semiparametric estimation methods can be developed to estimate the parameter vector β and the time trend function $f(\cdot)$. Among such estimation 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 2009). As we discuss in Section 4.3, the averaged profile likelihood estimation method is not so efficient for our semiparametric setting. Thus, we propose using a semiparametric pooled profile likelihood method associated with a dummy variable to estimate β and $f(\cdot)$.

Before we propose the estimation method, we need to introduce the following notation:

$$
\begin{aligned}
&\widetilde{Y} = (Y_{11}, \cdots, Y_{1T}, Y_{21}, \cdots, Y_{2T}, Y_{N1}, \cdots, Y_{NT})^\top, \\
&\widetilde{X} = (X_{11}, \cdots, X_{1T}, X_{21}, \cdots, X_{2T}, X_{N1}, \cdots, X_{NT})^\top, \\
&\alpha = (\alpha_2, \cdots, \alpha_N)^\top, \quad D = (-i_{N-1}, I_{N-1})^\top \otimes i_T, \\
&\widetilde{f} = i_N \otimes (f_1, \cdots, f_T)^\top, \quad \widetilde{e} = (e_{11}, \cdots, e_{1T}, e_{21}, \cdots, e_{2T}, e_{N1}, \cdots, e_{NT})^\top,
\end{aligned}
$$

where i_k is the $k \times 1$ vector of ones and I_k is the $k \times k$ identity matrix. As $\sum_{i=1}^{N} \alpha_i = 0$, 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,

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

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

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

$$\begin{pmatrix} \widehat{f}_{\alpha,\beta}(\tau) \\ h\widehat{f}'_{\alpha,\beta}(\tau) \end{pmatrix} = \arg\min_{(a,b)^\top} \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{Z}(\tau)(a,b)^\top\right)^\top \widetilde{W}(\tau) \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{Z}(\tau)(a,b)^\top\right).$$

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

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

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

(ii) Denote

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

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

$$\begin{aligned} (\widehat{\alpha}^\top, \widehat{\beta}^\top)^\top &= \arg\min_{(\alpha^\top,\beta^\top)^\top} \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{f}_{\alpha,\beta}\right)^\top \left(\widetilde{Y} - \widetilde{X}\beta - D\alpha - \widetilde{f}_{\alpha,\beta}\right) \\ &= \arg\min_{(\alpha^\top,\beta^\top)^\top} \left(\widetilde{Y}^* - \widetilde{X}^*\beta - D^*\alpha\right)^\top \left(\widetilde{Y}^* - \widetilde{X}^*\beta - D^*\alpha\right). \end{aligned} \tag{2.3}$$

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):

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

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

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

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

Note that our study in Sections 3 and 5 below shows that the proposed pooled profile likelihood method associated with a dummy variable 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 assumptions and 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 independent and identically distributed (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 $N \to \infty$,

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

where $\delta > 2$ is a positive constant.

A3. The trend functions $f(\cdot)$ and $g(\cdot)$ have continuous derivatives of 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 $N \to \infty$,

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

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

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

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

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

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

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 technical proofs. In A2, we impose some moment conditions on $\{v_{it}\}$ and allow for cross–sectional dependence of $\{v_{it}\}$ and thus $\{X_{it}\}$. 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 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 $1 \le i_1, i_2, \cdots, i_4 \le N$ 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 $1 \le i \le 3$. In this case, equation (3.1) can be verified because

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

when $\sum_{j=1}^{3}\delta_j \ge 2$. Obviously, the conventional Euclidean metric is covered. One may also show that equation (3.1) can also be verified when some other distance functions, including an 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.. This can be relaxed by allowing both $\{e_t\}$ and $\{v_t\}$ to be stationary and α–mixing (see, for example, Gao 2007). In A4, we also do need the mutual independence between v_{it} and e_{it} in this paper. When v_{it} and e_{it} are dependent 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 Conditions A1–A5 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 $0 \leq \lambda \leq \infty$. For example, when N is proportional to T, A5 reduces to $Th^4 \to 0$ and $\frac{T^{1-\frac{2}{\delta}}h}{\log(T)} \to \infty$. For the case of $N = [T^c]$, A5 reduces to $T^{1+c}h^8 \to 0$ and $\frac{T^{1-\frac{2}{\delta}}h}{\log(T)} \to \infty$ when $c \geq 1 - \frac{4}{\delta}$ for $\delta > 4$.

3.2. Asymptotic theory

We first establish an 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*

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

Remark 3.2. The above theorem shows that the proposed pooled profile likelihood estimator of β can achieve the root–NT convergence rate. 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 for 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.

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 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 of the form

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

The above bandwidth selection method cannot be implemented 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 propose using a semiparametric "leave–one–out" cross validation method that 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 $\Sigma_{v,e}$. Define

$$\widehat{\Sigma}_v(i) = \frac{1}{T} \sum_{t=1}^{T} \widehat{v}_{it} \widehat{v}_{it}^\top \quad \text{and} \quad \widehat{v}_{it} = X_{it} - \widehat{g}_t, \tag{4.1}$$

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

$$\widehat{\Sigma}_v = \frac{1}{N} \sum_{i=1}^{N} \widehat{\Sigma}_v(i). \tag{4.2}$$

By the uniform consistency of the pooled local linear estimate (see the proofs in Appendix B) and $g(\cdot)$ is independent of i, it is easy to check that $\widehat{\Sigma}_v(i)$ is a consistent estimator of $E\left[v_{i1} v_{i1}^\top\right]$ for each i, which implies that $\widehat{\Sigma}_v$ is a consistent estimator of Σ_v.

Let $\widehat{v}_{it}$ be defined as in (4.1) and

$$\widehat{e}_{it} = Y_{it} - X_{it}^\top \widehat{\beta} - \widehat{f}_t, \tag{4.3}$$

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

$$\widehat{\rho}_{ij}(v) = \frac{1}{T} \sum_{i=1}^{T} \widehat{v}_{it} \widehat{v}_{jt}^\top \quad \text{and} \quad \widehat{\rho}_{ij}(e) = \frac{1}{T} \sum_{i=1}^{T} \widehat{e}_{it} \widehat{e}_{jt}^\top, \tag{4.4}$$

respectively. Let φ_N be some positive integer such that $\varphi_N \leq N$ 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

$$\widehat{\sigma}_e^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{\sigma}_e^2(t) \quad \text{and} \quad \widehat{\sigma}_e^2(t) = \frac{1}{\varphi_N} \sum_{i=1}^{\varphi_N} \sum_{j=1}^{\varphi_N} \widehat{e}_{it} \widehat{e}_{jt}. \tag{4.6}$$

Following Theorems 3.1 and 3.2, one may show that the resulting estimators are all consistent.

4.2. Averaged profile likelihood estimation method

As $\sum_{i=1}^{N} \alpha_i = 0$, 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, $Y_{At} = \frac{1}{N} \sum_{i=1}^{N} Y_{it}$, $X_{At} = \frac{1}{N} \sum_{i=1}^{N} X_{it}$ and $e_{At} == \frac{1}{N} \sum_{i=1}^{N} e_{it}$. 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, model (4.7) can be rewritten as

$$Y_A = X_A \beta + f + e_A. \tag{4.8}$$

Then, applying the profile likelihood estimation approach to model (4.8), one can obtain the averaged profile likelihood estimator of β and estimate of $f(\cdot)$ by

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

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

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

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 the 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

In this section, we adopt the "leave–one–out" cross validation method to select the bandwidth for both the pooled and averaged profile likelihood estimation. The selection procedure can be described as follows.

Let $\widehat{\beta}_{(-1)}$, $\widehat{\alpha}_{(-1)}$ and $\widehat{f}_{(-1)}(\cdot)$ be the leave–one–out versions of $\widehat{\beta}$, $\widehat{\alpha}$ and $\widehat{f}(\cdot)$ in (2.4)–(2.6), respectively. The leave–one–out estimator of h, $\widehat{h}_{\rm cv}$, is chosen such that

$$\widehat{h}_{\rm cv} = \arg\min \frac{1}{NT} \left(\widetilde{Y} - \widetilde{X}\widehat{\beta}_{(-1)} - \widetilde{\widehat{f}}_{(-1)} - D\widehat{\alpha}_{(-1)} \right)^{\mathsf{T}} \left(\widetilde{Y} - \widetilde{X}\widehat{\beta}_{(-1)} - \widetilde{\widehat{f}}_{(-1)} - D\widehat{\alpha}_{(-1)} \right), \quad (4.9)$$

where $\widetilde{\widehat{f}}_{(-1)}$ is defined in the same way as $\widetilde{f}$ in (2.1) with $f(\cdot)$ being replaced by $\widehat{f}_{(-1)}(\cdot)$.

5. Examples of implementation

We next carry out simulations to compare the small sample behavior of the two profile likelihood estimation methods: the pooled and the averaged methods. Meanwhile, two real–data examples are provided to show that our estimation method performs well in the empirical analysis of a consumer price index data from Australia and a temperature series data from the United Kingdom. We find significant increasing trends in both of the data sets.

5.1. Simulated Examples

Example 5.1. Consider one data generating process of the form

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

where $\beta = 2$, $f(u) = u^3 + u$, $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ 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 $\widetilde{e}_{\cdot t} = (e_{1t}, \varepsilon_{2t}, \cdots, e_{Nt})$, 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

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

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

$$\begin{aligned} \mathrm{E}(e_{it}e_{js}) &= 0 \quad \text{for} \quad 1 \le i, j \le N,\ t \ne s, \\ \mathrm{E}(e_{it}e_{jt}) &= 0.8^{|j-i|} \quad \text{for} \quad 1 \le i, j \le N,\ 1 \le t \le T. \end{aligned}$$

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

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

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 $1 \le i, j \le N$.

With $R = 500$ replications, we compare the average square–root of mean squared errors (ASMSE) of the pooled profile likelihood estimator (PPLE) of β and estimate of $f(\cdot)$ with that of the averaged profile likelihood estimator of β and estimate of $f(\cdot)$ (APLE). For a $p \times 1$ parameter $\beta = (\beta_1, \cdots, \beta_p)^\top$ and a nonparametric function $f(\cdot)$ defined on $[0, 1]$, the ASMSE's of their estimators $\widehat{\beta}$ and $f(\cdot)$ are defined as

$$\mathrm{ASMSE}(\widehat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{p}\sum_{l=1}^{p}\left(\widehat{\beta}_l^{(r)} - \beta_l\right)^2\right)^{1/2}, \tag{5.4}$$

$$\mathrm{ASMSE}(\widehat{f}) = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{f}^{(r)}(t/T) - f(t/T)\right)^2\right)^{1/2}, \tag{5.5}$$

where $\widehat{\beta}_l^{(r)}$ and $\widehat{f}^{(r)}(\cdot)$ are the estimates of β_l and $\widehat{f}(\cdot)$ in the r-th replication for $1 \le l \le p$ and $1 \le r \le R$.

Table 5.1(a). ASMSE for the profile likelihood estimators of β in Example 5.1

$N\backslash T$		5	10	20	30
5	PPLE	0.2464 (0.1916)	0.1532 (0.1129)	0.1125 (0.0817)	0.0946 (0.0656)
	APLE	0.4158 (0.3980)	0.2381 (0.1864)	0.1775 (0.1312)	0.1439 (0.1012)
10	PPLE	0.1860 (0.1407)	0.1387 (0.1055)	0.0973 (0.0673)	0.0747 (0.0543)
	APLE	0.3626 (0.2911)	0.2439 (0.1973)	0.2441 (0.1899)	0.1603 (0.1099)
20	PPLE	0.1511 (0.1137)	0.1149 (0.0867)	0.0733 (0.0524)	0.0462 (0.0338)
	APLE	0.3257 (0.2562)	0.2237 (0.1692)	0.1709 (0.1228)	0.1680 (0.1281)
30	PPLE	0.1196 (0.0945)	0.1003 (0.0703)	0.0479 (0.0358)	0.0432 (0.0305)
	APLE	0.3123 (0.2368)	0.2053 (0.1450)	0.1744 (0.1073)	0.1506 (0.0985)

Table 5.1(b). ASMSE for the profile likelihood estimates of $f(\cdot)$ in Example 5.1

$N\backslash T$		5	10	20	30
5	PPLE	0.5008 (0.2716)	0.3489 (0.1693)	0.2597 (0.1023)	0.2289 (0.0947)
	APLE	0.6829 (0.5000)	0.4285 (0.2531)	0.3230 (0.1529)	0.2735 (0.1291)
10	PPLE	0.4318 (0.2124)	0.3101 (0.1414)	0.2306 (0.0943)	0.1910 (0.0751)
	APLE	0.6165 (0.3949)	0.4150 (0.2382)	0.3136 (0.1570)	0.2697 (0.1355)
20	PPLE	0.3446 (0.1667)	0.2622 (0.1096)	0.1969 (0.0759)	0.1561 (0.0564)
	APLE	0.5199 (0.3338)	0.3651 (0.2052)	0.2895 (0.1479)	0.2628 (0.1530)
30	PPLE	0.2871 (0.1301)	0.2404 (0.0915)	0.1587 (0.0582)	0.1374 (0.0557)
	APLE	0.4740 (0.2900)	0.3394 (0.1712)	0.2805 (0.1246)	0.2465 (0.1192)

Tables 5.1(a) and 5.1(b) reveal that the PPLE outperforms the APLE uniformly. We can further find that an increase in either N or T results in an obvious improvement in the performances of the PPLE's of both β and $f(\cdot)$. In contrast, the increase in N does not necessarily result in the improvement of the performance of the APLE of β, although the increase in T seems to lead to better performance of the APLE of β (although this improvement might be slight). This indicates that the APLE can not estimate the parameter β well for small T.

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

$$Y_{it} = X_{it}^\top \beta + f\left(\frac{t}{T}\right) + \alpha_i + e_{it}, \quad 1 \le i \le N,\ 1 \le t \le T, \tag{5.6}$$

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

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

and $\alpha_N = -\sum_{i=1}^{N-1} \alpha_i$. Letting $\widetilde{e}_{\cdot t} = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{Nt})$, 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 $c_{ij}^* = \frac{2}{(i-j)^2+1}$.

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

$$\mathbf{g}(u) = \left(1 + \sin(\pi u),\ \frac{1}{2}u,\ -u\right)^\top, \tag{5.7}$$

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 $E\mathbf{v}_{it} = \mathbf{0}$,

$$E\left(\mathbf{v}_{it}\mathbf{v}_{jt}^\top\right) = \frac{1}{(j-i)^2+1}\begin{pmatrix} 4 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \ 1 \le i, j \le N, \ 1 \le t \le T,$$

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

Table 5.2(a). ASMSE for the profile likelihood estimators of β

$N\backslash T$		5	10	20	30
5	PPLE	0.3785 (0.2036)	0.2537 (0.1221)	0.1723 (0.0819)	0.1336 (0.0690)
	APLE	5.2104 (23.9788)	0.5987 (0.3523)	0.3628 (0.1884)	0.2712 (0.1421)
10	PPLE	0.2693 (0.1395)	0.1717 (0.0857)	0.1159 (0.0579)	0.0960 (0.0473)
	APLE	5.6469 (19.3887)	0.6227 (0.3878)	0.3696 (0.1997)	0.2710 (0.1421)
20	PPLE	0.1807 (0.0908)	0.1248 (0.0639)	0.0861 (0.0400)	0.0685 (0.0347)
	APLE	5.3776 (17.3639)	0.8067 (0.5337)	0.3609 (0.1922)	0.2798 (0.1326)
30	PPLE	0.1533 (0.0756)	0.0976 (0.0486)	0.0709 (0.0327)	0.0555 (0.0284)
	APLE	5.0500 (11.6489)	0.6438 (0.3892)	0.3683 (0.1966)	0.2705 (0.1493)

Table 5.2(b). ASMSE for the profile likelihood estimates of $f(\cdot)$

$N\backslash T$		5	10	20	30
5	PPLE	0.7554 (0.4229)	0.4977 (0.2360)	0.3805 (0.1629)	0.2824 (0.1308)
	APLE	6.9920 (46.0699)	0.7909 (0.5229)	0.4888 (0.2523)	0.3812 (0.1824)
10	PPLE	0.5457 (0.2773)	0.3846 (0.1709)	0.2762 (0.0994)	0.2406 (0.0783)
	APLE	5.4341 (20.3767)	0.7601 (0.4996)	0.4373 (0.2525)	0.3576 (0.1781)
20	PPLE	0.3906 (0.1805)	0.2900 (0.1190)	0.2212 (0.0842)	0.2035 (0.0538)
	APLE	6.0141 (22.3286)	0.9452 (0.7246)	0.4360 (0.2441)	0.3475 (0.1749)
30	PPLE	0.3290 (0.1477)	0.2919 (0.0977)	0.2073 (0.0553)	0.1796 (0.0422)
	APLE	5.5615 (17.4042)	0.6778 (0.4486)	0.3916 (0.2400)	0.2902 (0.1797)

The ASMSE's of the proposed pooled profile likelihood estimators and the averaged profile likelihood estimators of β and $f(\cdot)$ are calculated over 500 realizations. The results for the estimator of β and estimate of $f(\cdot)$ are given in Tables 5.2(a) and 5.2(b), respectively. While one may draw the same conclusions as those from Tables 5.1(a) and 5.1(b): The PPLE performs uniformly better than the APLE and its performance improves as either N or T increases, the performance of the APLE for the case of $T = 5$ is much worse than that for the PPLE in each individual case. This is mainly because the cross–sectional dependence imposed in Example 5.2 is stronger than that imposed in Example 5.1.

5.2. An application in modeling consumer price index

This data set consists of the 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). Here 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 pair of semiparametric models of the form

$$\begin{aligned} Y_{it} &= X_{it}\beta + f_t + \alpha_i + e_{it}, \\ X_{it} &= g_t + v_{it}, \quad 1 \leq i \leq 8,\ 1 \leq t \leq 60, \end{aligned} \tag{5.8}$$

where α_i are individual effects, and both f_t and g_t are the trend functions.

By applying the proposed pooled profile likelihood estimation procedure to the above data set, we have the estimate for β: $\widehat{\beta} = 0.6617$. The estimate of the trend function is given in Figure 5.1. It follows from this figure that there is a significant upward trend in f_t, which is consistent with the observation that the food CPI series for each city generally increases with time.

5.3 Trend modeling of a climatic data set

The second data set contains monthly mean maximum temperatures (in Celsius degrees), mean minimum temperatures (in Celsius degrees), total rainfall (in millimeters) and total sunshine duration (in hours) from 37 stations covering the United Kingdom (available from the UK Met Office at www.metoffice.gov.uk/climate/uk/stationdata). We use monthly data from January 1999–December 2008 to see if there exists a significant common trend in the mean maximum temperatures during this period among the stations.

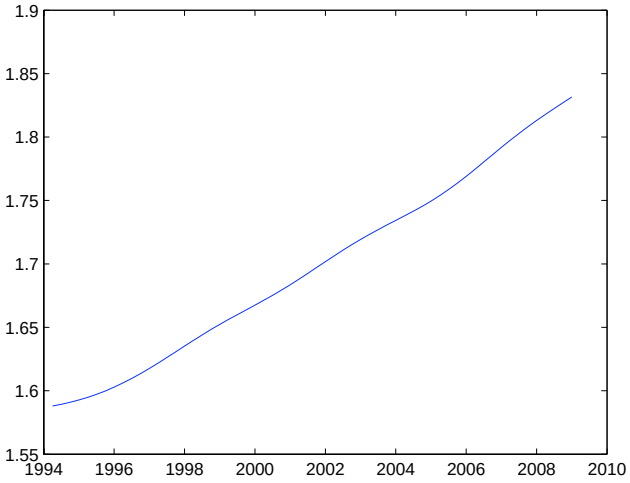


Figure 5.1. The PPLE of the trend function f_t in model (5.8).

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 January 1999–December 2008 are missing at some stations).

To illustrate, we plot three data series of monthly mean maximum temperatures, total rainfall and total sunshine duration from one of the selected stations: station Armagh. As we can see from Figures 5.2–5.4, the data series exhibit obvious seasonal effects. In this respect, we decompose the raw data series into three parts: the seasonal, the trend and the residuals. The raw data series, the seasonal, the trend and the residuals in the temperature, rainfall and sunshine series are plotted in Figures 5.2–5.4.

We first remove the seasonality from the raw data and then fit the data with the model below. To investigate the relationship between the mean maximum temperatures and total rainfall and total sunshine duration, we choose Y_{it} as the mean maximum temperatures for station i at time t, and $X_{it} = (X1_{it}, X2_{it})^\top$ as the vector consisting of the total rainfall and total sunshine duration. We then assume that $\{(Y_{it}, X_{it})\}$ satisfies a pair of semiparametric models of the form

$$\begin{aligned} Y_{it} &= f_t + X_{it}^\top \beta + e_{it}, \\ X_{it} &= g_t + v_{it}, \quad 1 \le i \le 16,\ 1 \le t \le 120. \end{aligned} \tag{5.9}$$

Applying the pooled profile likelihood estimation method given in Section 2, we have the estimate of beta as $\widehat{\beta} = (0.0014, 0.0184)^\top$. This indicates that the total sunshine

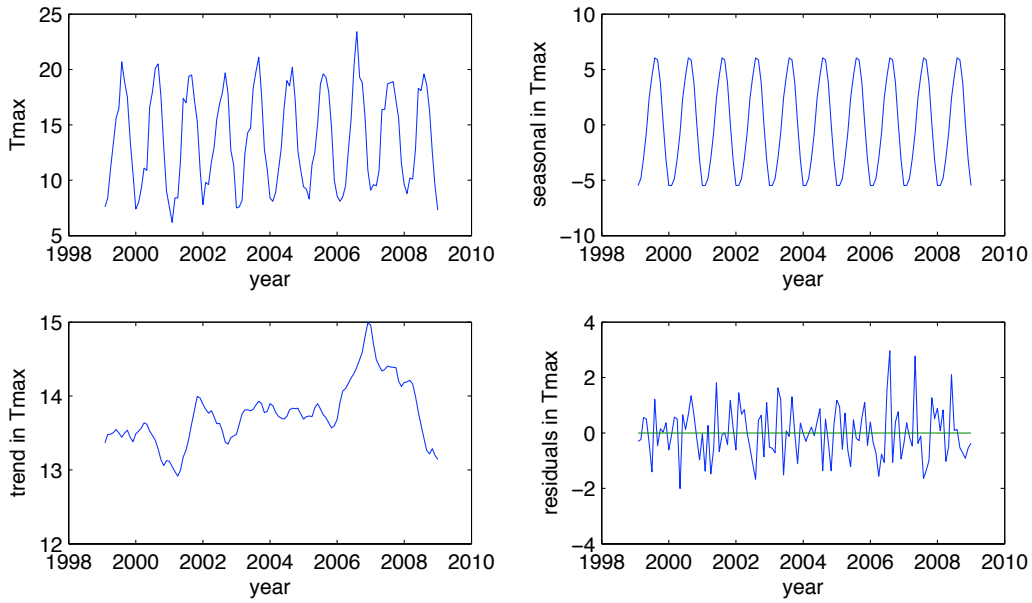


Figure 5.2. Plots of the monthly mean maximum temperatures (°C) series from station Armagh: the raw data, the seasonality, the trend and the residuals.

duration has a more significant influence than the total rainfall on the mean maximum temperatures.

The estimate of the common trend function f_t is plotted in Figure 5.5. Figure 5.5 shows that from the beginning of 1999 to the end of 2000, there is a decrease in the trend (from about 11.5 at the beginning of 1999 to about 11 at the end of 2000), which may be a result of an abnormally strong El Nino in 1998 that caused high temperatures throughout the globe. Then in the two years that followed 1998, the temperatures went from this extreme down to average. Thereafter, there is an overall increasing trend from the beginning of 2001 to the end of 2006 (from about 11 at the beginning of 2001 to about 11.8 at the end of 2006). Then from the beginning of 2007 to the end of 2008, there is a drop in the trend (from about 11.8 at the beginning of 2007 to about 10.5 at the end of 2008), which may be attributed to the La Nina conditions that have a cooling effect on temperatures.

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 semiparametric profile

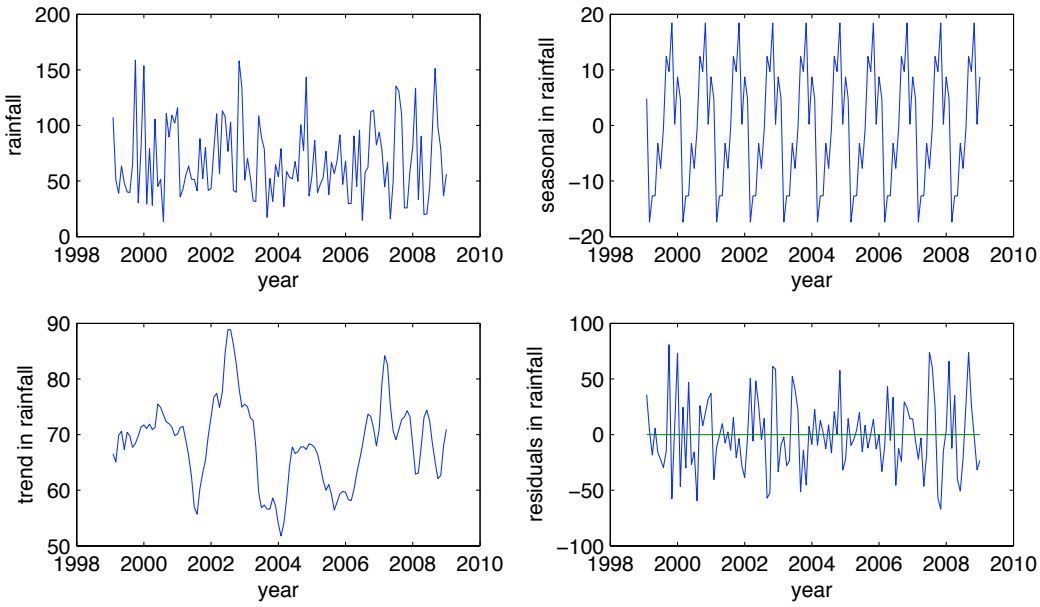


Figure 5.3. Plots of the monthly total rainfall (mm) series from station Armagh: the raw data, the seasonality, the trend and the residuals.

likelihood based estimation method has been proposed to estimate both the parameter vector and the time trend function. An asymptotically normal distribution with possible optimal rate of convergence has been established for each of the proposed estimates for the case where both the time series length T and the cross–sectional size N tend to infinity simultaneously.

This paper has used two simulated examples to evaluate the finite–sample performance of the proposed estimation method. As shown in the theory, the proposed semiparametric profile likelihood based estimation method uniformly outperforms an averaged profile likelihood based estimation method, which is commonly used in the literature. In addition, we have discussed empirical applications of the proposed theory and estimation method to 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 climatic data set from the United Kindgom.

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

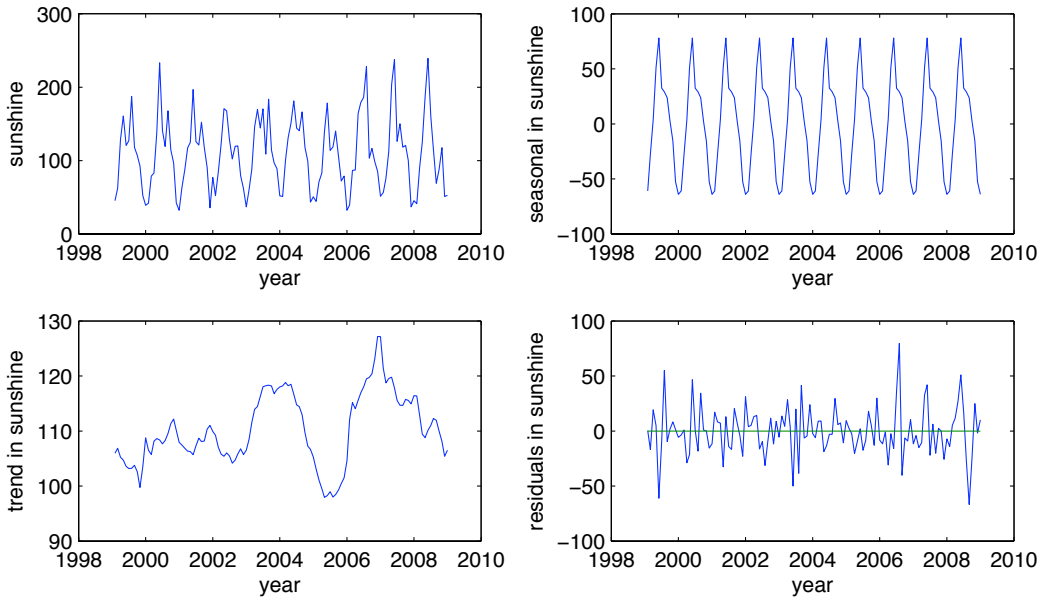


Figure 5.4. Plots of the monthly total sunshine duration (hr) series from station Armagh: the raw data, the seasonality, the trend and the residuals.

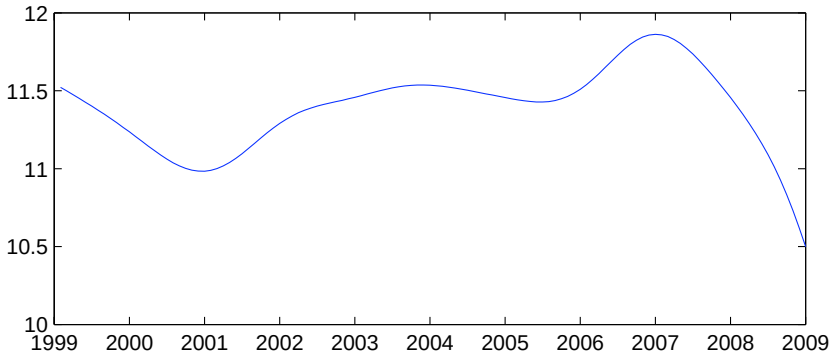


Figure 5.5. The PPLE of the trend function f_t in model (5.9).

The authors acknowledge the financial support from the Australian Research Council Discovery Grants Program under Grant Number: DP0879088. Thanks also go to Dr Alev Atak for providing us with information about the UK temperature data set.

Appendix A: Proofs of the main results

Let C is 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

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

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

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

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

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

$$\frac{1}{NT}\widetilde{X}^{*\top}\widetilde{X}^*=\Sigma_v+o_P(1), \qquad (A.3)$$

$$\frac{1}{NT}\widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}\widetilde{X}^*=o_P(1), \qquad (A.4)$$

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

Step (i). Proof of (A.3). By the definition of $\widetilde{X}^*$, we have

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

$$
\begin{aligned}
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}\left(g_t-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)^{\top}\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(g_t-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)\left(g_t-\widetilde{X}^{\top}s^{\top}\left(\frac{t}{T}\right)\right)^{\top}\\
=:\ & \Pi^*_{NT}(1)+\Pi^*_{NT}(2)+\Pi^*_{NT}(3)+\Pi^*_{NT}(4). \qquad \text{(A.6)}
\end{aligned}
$$

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

$$
\begin{aligned}
\Pi^*_{NT}(1) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}v_{it}^{\top}=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}v_{it}^{\top}\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}v_{it}^{\top}-E\left[\frac{1}{N}\sum_{i=1}^{N} v_{it}v_{it}^{\top}\right]\right)+\frac{1}{T}\sum_{t=1}^{T}E\left[\frac{1}{N}\sum_{i=1}^{N} v_{it}v_{it}^{\top}\right]\\
&=: \Pi^*_{NT}(1,1)+\Pi^*_{NT}(1,2). \qquad \text{(A.7)}
\end{aligned}
$$

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

$$
\begin{aligned}
&P\left\{|\Pi^*_{NT}(1,1)|>\epsilon\right\}\le\frac{1}{\epsilon^2}E\left[\Pi^*_{NT}(1,1)\right]^2\\
=\ &\frac{1}{\epsilon^2T^2}\sum_{t=1}^{T}\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} v_{it}v_{it}^{\top}\right)=\frac{1}{\epsilon^2TN^2}\mathrm{Var}\left(\sum_{i=1}^{N} v_{it}v_{it}^{\top}\right)=O\left(\frac{1}{T}\right).
\end{aligned}
$$

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

$$
\Pi^*_{NT}(1,1)=o_P(1). \qquad \text{(A.8)}
$$

By A2, it is easy to check that

$$
\Pi^*_{NT}(1,2)=\Sigma_v+o_P(1) \qquad \text{(A.9)}
$$

as $N,T\to\infty$ simultaneously. By (A.7)–(A.9), we have

$$
\Pi^*_{NT}(1)=\Sigma_v+o_P(1). \qquad \text{(A.10)}
$$

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

$$
\sup_{0<\tau<1}\left\|g(\tau)-\widetilde{X}^{\top}s^{\top}(\tau)\right\|=O_P\left(h^2+\sqrt{\frac{\log(NT)}{NTh}}\right). \qquad \text{(A.11)}
$$

The detailed proof of (A.11) will be given in Appendix B. From (A.11), it is easy to show

$$
\Pi^*_{NT}(4)=o_P(1). \qquad \text{(A.12)}
$$

By (A.10), (A.12) and the Cauchy–Schwarz inequality,

$$
\Pi^*_{NT}(2)=o_P(1)\quad\text{and}\quad\Pi^*_{NT}(3)=o_P(1). \qquad \text{(A.13)}
$$

With (A.6), (A.10), (A.12) and (A.13), we have shown that (A.3) holds.

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

$$D^{*\top}D^* = D^\top \left(I_{NT} - \widetilde{S}\right)^\top \left(I_{NT} - \widetilde{S}\right) D = D^\top D. \tag{A.14}$$

Furthermore,

$$D^\top D = \begin{pmatrix} 2T & T & \cdots & T \\ T & 2T & \cdots & T \\ \vdots & \vdots & \cdots & \vdots \\ T & T & \cdots & 2T \end{pmatrix} = \begin{pmatrix} T & T & \cdots & T \\ T & T & \cdots & T \\ \vdots & \vdots & \cdots & \vdots \\ T & T & \cdots & T \end{pmatrix} + \mathrm{diag}(T, \cdots, T). \tag{A.15}$$

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

$$\left(D^{*\top}D^*\right)^{-1} = \left(D^\top D\right)^{-1} = \begin{pmatrix} \frac{1}{T} - \frac{1}{NT} & -\frac{1}{NT} & \cdots & -\frac{1}{NT} \\ -\frac{1}{NT} & \frac{1}{T} - \frac{1}{NT} & \cdots & -\frac{1}{NT} \\ \vdots & \vdots & \cdots & \vdots \\ -\frac{1}{NT} & -\frac{1}{NT} & \cdots & \frac{1}{T} - \frac{1}{NT} \end{pmatrix}. \tag{A.16}$$

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

$$\widetilde{X}^{*\top}D^* = \widetilde{X}^{*\top}D = \left(A_T(2), \cdots, A_T(N)\right), \tag{A.17}$$

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

By (A.16) and (A.17), we then have

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

where $A_T^*(k) = \frac{1}{T}A_T(k) - \frac{1}{NT}\sum_{k=2}^{N} A_T(k)$.

Following the proof of (A.11) in Appendix B, we can show that for each k,

$$\frac{1}{T}A_T(k) \xrightarrow{P} 0 \quad \text{as } T \to \infty,$$

which implies

$$\left\| \frac{1}{NT}\widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}\widetilde{X}^* \right\| = o_P(1).$$

Hence, (A.4) holds.

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

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

Similarly to the proofs of (A.3) and (A.4), we can show that the leading term on the right hand side of the above equation is $\widetilde{X}^{*\top} \widetilde{f}^*$. Hence, to prove (A.5), we need only to show

$$\widetilde{X}^{*\top} \widetilde{f}^* = o_P(\sqrt{NT}). \tag{A.18}$$

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

$$\begin{aligned}
\widetilde{X}^{*\top} \widetilde{f}^* &= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \widetilde{X}^{\top} s^{\top} \left(\frac{t}{T} \right) \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(v_{it} + \left[g \left(\frac{t}{T} \right) - \widetilde{X}^{\top} s^{\top} \left(\frac{t}{T} \right) \right] \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) - \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{v}^{\top} s^{\top} \left(\frac{t}{T} \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} \left(g \left(\frac{t}{T} \right) - \widetilde{g}^{\top} s^{\top} \left(\frac{t}{T} \right) \right) \left(f \left(\frac{t}{T} \right) - s \left(\frac{t}{T} \right) \widetilde{f} \right) \\
&=: \Pi_{NT}^*(5) + \Pi_{NT}^*(6) + \Pi_{NT}^*(7),
\end{aligned} \tag{A.19}$$

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

Following the argument in the proof of (A.11) in Appendix B and by A2 and A3, we have

$$\begin{aligned}
&\sup_{0<\tau<1} \left| f(\tau) - s(\tau) \widetilde{f} \right| = O\left(h^2 \right), \\
&\sup_{0<\tau<1} \left\| g(\tau) - \widetilde{g}^{\top} s^{\top}(\tau) \right\| = O\left(h^2 \right), \\
&\sup_{0<\tau<1} \left\| \widetilde{v}^{\top} s^{\top}(\tau) \right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}} \right),
\end{aligned}$$

which, together with A5, imply

$$\Pi_{NT}^*(5) = O_P\left(\sqrt{NT} h^2 \right) = o_P(\sqrt{NT}), \tag{A.20}$$

$$\Pi_{NT}^*(6) = O_P\left(\sqrt{NT \log(NT) h^3} \right) = o_P(\sqrt{NT}), \tag{A.21}$$

$$\Pi_{NT}^*(7) = O_P\left(NTh^4 \right) = o_P(\sqrt{NT}). \tag{A.22}$$

By (A.19)–(A.22), we have shown that (A.18) holds.

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

$$\sqrt{NT}\Pi_{NT}(3) \xrightarrow{d} N\left(\mathbf{0}_d, \Sigma_v^{-1}\Sigma_{v,e}\Sigma_v^{-1}\right). \tag{A.23}$$

Proof. To prove (A.23), we need only to show

$$\frac{1}{NT}\widetilde{X}^{*\top}M^*\widetilde{X}^* \xrightarrow{P} \Sigma_v \tag{A.24}$$

and

$$\frac{1}{\sqrt{NT}}\widetilde{X}^{*\top}M^*\widetilde{e}^* \xrightarrow{d} N\left(0, \Sigma_{v,e}\right). \tag{A.25}$$

By (A.3) and (A.4) in the proof of Proposition A.1, we can easily obtain (A.24). For the proof of (A.25), observe that

$$\widetilde{X}^{*\top}M^*\widetilde{e}^* = \widetilde{X}^{*\top}\widetilde{e}^* - \widetilde{X}^{*\top}D^*\left(D^{*\top}D^*\right)^{-1}D^{*\top}\widetilde{e}^* =: \Pi_{NT}^*(8) - \Pi_{NT}^*(9). \tag{A.26}$$

For $\Pi_{NT}^*(8)$, we have

$$\begin{aligned}\frac{1}{\sqrt{NT}}\Pi_{NT}^*(8) &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T v_{it}e_{it} + \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\left(g\left(\frac{t}{T}\right) - \widetilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)e_{it} \\ &\quad - \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T v_{it}s\left(\frac{t}{T}\right)\widetilde{e} - \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\left(g\left(\frac{t}{T}\right) - \widetilde{X}^\top s^\top\left(\frac{t}{T}\right)\right)s\left(\frac{t}{T}\right)\widetilde{e} \\ &=: \Pi_{NT}^*(10) + \Pi_{NT}^*(11) - \Pi_{NT}^*(12) - \Pi_{NT}^*(13).\end{aligned} \tag{A.27}$$

Following the proof of (A.11) in Appendix B, and by A2, A4 and A5, we have

$$\Pi_{NT}^*(11) = O_P\left(h^2 + \sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1), \tag{A.28}$$

$$\Pi_{NT}^*(12) = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1), \tag{A.29}$$

$$\Pi_{NT}^*(13) = O_P\left(\sqrt{NT}\left(h^2 + \sqrt{\frac{\log(NT)}{NTh}}\right)\sqrt{\frac{\log(NT)}{NTh}}\right) = o_P(1). \tag{A.30}$$

If we can prove

$$\Pi_{NT}^*(10) \xrightarrow{d} N\left(0, \Sigma_{v,e}\right), \tag{A.31}$$

then by (A.28)–(A.30), we will show that

$$\frac{1}{\sqrt{NT}}\Pi_{NT}^*(8) \xrightarrow{d} N\left(0, \Sigma_{v,e}\right). \tag{A.32}$$

As both T and N tend to infinity, we next prove (A.31) by the joint limit approach (see Phillips and Moon 1999 for example).

Letting $Z_{t,N}(v,e) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{it} e_{it}$, then $\Pi^*_{NT}(10)$ can be rewritten as

$$\Pi^*_{NT}(10) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,N}(v,e).$$

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.31). For any $\epsilon > 0$,

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

as $N, T \to \infty$ simultaneously, which implies that the Lindeberg condition is satisfied, which in turn implies the validity of (A.31).

By (A.16), (A.17) and a standard calculation, we have

$$\Pi^*_{NT}(9) = \frac{1}{T} \sum_{k=2}^{N} A_T(k) B_T(k) - \frac{1}{NT} \left(\sum_{k=2}^{N} A_T(k) \right) \left(\sum_{k=2}^{N} B_T(k) \right),$$

where $A_T(k) = \sum_{t=1}^{T} v_{kt} - \sum_{t=1}^{T} v_{1t}$ and $B_T(k) = \sum_{t=1}^{T} e_{kt} - \sum_{t=1}^{T} e_{1t}$.

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,

$$\begin{aligned} \Pi^*_{NT}(9) &= \frac{1}{T} \sum_{k=2}^{N} A_T(k) B_T(k) - \frac{1}{NT} \left(\sum_{k=2}^{N} A_T(k) \right) \left(\sum_{k=2}^{N} B_T(k) \right) \\ &= \frac{1}{T} \sum_{k=2}^{N} \overline{A}_T(k) \overline{B}_T(k) - \frac{1}{T} \overline{B}_T(1) \sum_{k=2}^{N} \overline{A}_T(k) - \frac{1}{T} \overline{A}_T(1) \sum_{k=2}^{N} \overline{B}_T(k) \\ &+ \frac{N-1}{T} \overline{A}_T(1) \overline{B}_T(1) - \frac{1}{NT} \left(\sum_{k=2}^{N} \overline{A}_T(k) \right) \left(\sum_{k=2}^{N} \overline{B}_T(k) \right) \\ &+ \frac{N-1}{NT} \overline{B}_T(1) \sum_{k=2}^{N} \overline{A}_T(k) + \frac{N-1}{NT} \overline{A}_T(1) \sum_{k=2}^{N} \overline{B}_T(k) - \frac{(N-1)^2}{NT} \overline{A}_T(1) \overline{B}_T(1) \\ &= \frac{1}{T} \sum_{k=2}^{N} \overline{A}_T(k) \overline{B}_T(k) - \frac{1}{NT} \left(\sum_{k=2}^{N} \overline{A}_T(k) \right) \left(\sum_{k=2}^{N} \overline{B}_T(k) \right) \\ &- \frac{1}{NT} \overline{B}_T(1) \sum_{k=2}^{N} \overline{A}_T(k) - \frac{1}{NT} \overline{A}_T(1) \sum_{k=2}^{N} \overline{B}_T(k) + \frac{N-1}{NT} \overline{A}_T(1) \overline{B}_T(1) \\ &=: \sum_{j=1}^{5} \Pi^*_{NT}(9,j). \end{aligned} \tag{A.33}$$

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

$$E\left(\Pi^*_{NT}(9,1) \Pi^{*\top}_{NT}(9,1) \right)$$

$$
\begin{aligned}
&= \frac{1}{T^2}E\left[\left(\sum_{k=2}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{kt}v_{ks}\right)\left(\sum_{k=2}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{kt}v_{ks}\right)^{\top}\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k_1=2}^{N}\sum_{k_2=2}^{N}E\left(e_{k_1t}e_{k_2t}\right)E\left(v_{k_1s}v_{k_2s}^{\top}\right) \\
&= \sum_{k_1=2}^{N}\sum_{k_2=2}^{N}E\left(e_{k_1,1}e_{k_2,1}\right)E\left(v_{k_1,1}v_{k_2,1}^{\top}\right) = O(N), \\
&\quad E\left(\Pi_{NT}^{*}(9,2)\Pi_{NT}^{*\top}(9,2)\right) \\
&= \frac{1}{N^2T^2}E\left(\sum_{k_1=2}^{N}\sum_{t_1=1}^{T}e_{k_1t_1}\right)^2 E\left[\left(\sum_{k_2=2}^{N}\sum_{t_2=1}^{T}v_{k_2t_2}\right)\left(\sum_{k_2=2}^{N}\sum_{t_2=1}^{T}v_{k_2t_2}\right)^{\top}\right] \\
&= \frac{1}{N^2T^2}\left[\sum_{k_1=2}^{N}\sum_{j_1=2}^{N}\sum_{t_1=1}^{T}E\left(e_{k_1t_1}e_{j_1t_1}\right)\right]\left[\sum_{k_2=2}^{N}\sum_{j_2=2}^{N}\sum_{t_2=1}^{T}E\left(v_{k_2t_2}v_{j_2t_2}^{\top}\right)\right] \\
&= \frac{1}{N^2}\left[\sum_{k_1=2}^{N}\sum_{j_1=2}^{N}E\left(e_{k_1,1}e_{j_1,1}\right)\right]\left[\sum_{k_2=2}^{N}\sum_{j_2=2}^{N}E\left(v_{k_2,1}v_{j_2,1}^{\top}\right)\right] = O(1),
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
E\left(\Pi_{NT}^{*}(9,3)\Pi_{NT}^{*\top}(9,3)\right) &= O\left(\frac{1}{N}\right), \\
E\left(\Pi_{NT}^{*}(9,4)\Pi_{NT}^{*\top}(9,4)\right) &= O\left(\frac{1}{N}\right), \\
E\left(\Pi_{NT}^{*}(9,5)\Pi_{NT}^{*\top}(9,5)\right) &= O(1).
\end{aligned}
$$

Hence,

$$
\Pi_{NT}^{*}(9,j) = o_P(\sqrt{NT}), \quad j = 1,\cdots,5. \tag{A.34}
$$

Combining (A.33) and (A.34), we have

$$
\Pi_{NT}^{*}(9) = o_P(\sqrt{NT}). \tag{A.35}
$$

By (A.26), (A.32) and (A.35), (A.25) holds. 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

$$
\begin{aligned}
\widehat{f}(\tau) - f(\tau) &= s(\tau)\left(\widetilde{Y} - \widetilde{X}\widehat{\beta} - D\widehat{\alpha}\right) - f(\tau) \\
&= s(\tau)\left(I_{NT} - D\left(D^{*\top}D^{*}\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)(\widetilde{Y} - \widetilde{X}\widehat{\beta}) - f(\tau) \\
&= \left(s(\tau)\left(I_{NT} - D\left(D^{*\top}D^{*}\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{f} - f(\tau)\right) \\
&\quad + s(\tau)\left(I_{NT} - D\left(D^{*\top}D^{*}\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{e} \\
&\quad + s(\tau)\left(I_{NT} - D\left(D^{*\top}D^{*}\right)^{-1}D^{*\top}(I_{NT} - \widetilde{S})\right)\widetilde{X}(\beta - \widehat{\beta}) \\
&=: \Pi_{NT}^{*}(14) + \Pi_{NT}^{*}(15) + \Pi_{NT}^{*}(16).
\end{aligned} \tag{A.36}
$$

Note that

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

Hence, by (A.11) and standard argument for local linear fitting, we have

$$\Pi^*_{NT}(14) = s(\tau)\widetilde{f} - f(\tau) = \frac{1}{2}f''(\tau)\mu_2 h^2 + o_P(h^2). \tag{A.37}$$

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

$$\sqrt{NTh}\Pi^*_{NT}(15) = \sqrt{NTh}s(\tau)\widetilde{e} \xrightarrow{d} N(0, \nu_0\sigma_e^2). \tag{A.38}$$

By Theorem 3.1, we have

$$\Pi^*_{NT}(16) = s(\tau)\widetilde{X}(\beta - \widehat{\beta}) = O_P\left((NT)^{-1/2}\right) = o_P\left((NTh)^{-1/2}\right). \tag{A.39}$$

Hence, by (A.36)–(A.39), we have completed the proof of (3.8).

Appendix B

Proof of (A.11). Note that

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

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 now prove

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

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

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

We first prove

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

Note that

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

By the definition of Riemann integral, we have

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

uniformly for $0 < \tau < 1$.

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

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

Note that

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

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

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

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

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

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

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

We next 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

$$\begin{aligned} \sup_{0<\tau<1} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(\tau) \overline{Q}_{t,N}(v) \right\| &\leq \max_{1\leq l\leq U_{NT}} \sup_{\tau\in B_l} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(\tau) \overline{Q}_{t,N}(v) - \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \overline{Q}_{t,N}(v) \right\| \\ &+ \max_{1\leq l\leq U_{NT}} \left\| \sum_{t=1}^{T} \overline{K}_{t,j}(b_l) \overline{Q}_{t,N}(v) \right\| =: \Xi_{NT,3} + \Xi_{NT,4}. \end{aligned} \tag{B.8}$$

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

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

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

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

Note that

$$\begin{aligned}\Xi_{NT,4} &\leq \max_{1\leq l\leq U_{NT}}\left\|\sum_{t=1}^{T}\overline{K}_{t,j}(b_l)\widetilde{Q}_{t,N}(v)\right\| + \max_{1\leq l\leq U_{NT}}\left\|\sum_{t=1}^{T}\overline{K}_{t,j}(b_l)\widetilde{Q}^c_{t,N}(v)\right\| \\ &=: \Xi_{NT,5} + \Xi_{NT,6}. \end{aligned} \tag{B.10}$$

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

$$\begin{aligned} & P\left\{\Xi_{NT,6} > \epsilon l(NT)\sqrt{\frac{\log(NT)}{NTh}}\right\} \leq P\left\{\max_{1\leq t\leq T}\left\|\widetilde{Q}^c_{t,N}(v)\right\| > 0\right\} \\ & \leq \sum_{t=1}^{T} P\left\{\left\|\overline{Q}_{t,N}(v)\right\| > N^{-1/2}T^{1/\delta}l(NT)\right\} = O\left(E\left\|\overline{Q}_{t,N}(v)\right\|^{\delta} N^{\delta/2}(l(NT))^{-\delta}\right) \\ = \;& O\left((l(NT))^{-\delta}\right) = o(1), \end{aligned}$$

which implies

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

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 variables, we have

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

where $C > 0$ is a constant and M is a sufficiently large positive constant. The second inequality above holds because of

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

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

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

In view of (B.1), (B.2) and (B.14), it has been shown that (A.11) holds.

References

Atak, A., Linton, O., Xiao, Z., 2009. A semiparametric panel model for climate change in the United Kingdom. Working paper available from http://personal.lse.ac.uk/lintono.

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

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. and Gao, J., 2010. Nonparametric time–varying coefficient panel data models with fixed effects. Working paper available at http://www.adelaide.edu.au/directory/jiti.gao.

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., 2008. Nonparametric trending regression with cross-sectional dependence. Working paper available at http://www.ihs.ac.at/etsast/presentations/Robinson.pdf.

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