



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Estimation in Single-Index Panel Data Models
with Heterogeneous Link Functions**

Jia Chen, Jiti Gao and Degui Li

September 2011

Working Paper 12/11

ESTIMATION IN SINGLE–INDEX PANEL DATA MODELS WITH HETEROGENEOUS LINK FUNCTIONS

Jia Chen^{1,2}, Jiti Gao^{1*}, and Degui Li¹

¹ *Department of Econometrics and Business Statistics, Monash University, Australia*

² *School of Mathematics, University of Queensland, Australia*

□ *In this paper, we study semiparametric estimation for a single–index panel data model where the nonlinear link function varies among the individuals. We propose using the refined minimum average variance estimation method to estimate the parameter in the single–index. As the cross–section dimension N and the time series dimension T tend to infinity simultaneously, we establish asymptotic distributions for the proposed estimator. In addition, we provide a real–data example to illustrate the finite sample behavior of the proposed estimation method.*

JEL subject classifications C13; C14; C23.

Keywords Asymptotic distribution; local linear smoother; minimum average variance estimation; panel data; semiparametric estimation; single–index models.

1. INTRODUCTION

During the last two decades or so, there has been an increasing interest in parametric linear and nonlinear panel data modeling as the double–index panel data models enable researchers to extract information that may be difficult to obtain

*Address correspondence to Jiti Gao, Department of Econometrics and Business Statistics, Monash University Caulfield Campus, VIC 3145, Australia; E-mail: jiti.gao@monash.edu

through purely using cross-section or time-series data models. We refer to the books by Baltagi (1995), Arellano (2003) and Hsiao (2003) for an overview of the statistical inference and econometric analysis of parametric panel data models. As in cross-section and time-series analysis, however, parametric models may be misspecified and estimators obtained from such misspecified parametric models are often inconsistent. To address such issues, some nonparametric and semiparametric models have been proposed. Existing literature on nonparametric and semiparametric panel data modelling includes Li and Stengos (1996), Ullah and Roy (1998), Abrevaya (1999), Hjellvik et al. (2004), Cai and Li (2008), Henderson et al. (2008) and Mammen et al. (2009).

Single-index modelling is one of the most commonly used semiparametric modelling techniques. It searches for a linear combination of the multi-dimensional covariates \mathbf{X} which can capture most information about the relationship between the response variable Y and covariates \mathbf{X} , while still allowing for curvature between Y and \mathbf{X} through link functions. Single-index models have been studied in depth in time series setting. Existing literature includes Härdle et al. (1993), Carroll et al. (1997), Xia et al. (2002), Yu and Ruppert (2002), Xia (2006), Gao (2007), and Li and Racine (2007). However, so far as we know, little work has been done in the theoretical and empirical analysis of single-index models for panel data. In this paper, we will study a single-index panel data model with heterogeneous link functions. The proposed model is more flexible than homogeneous single-index panel data models. By letting the link functions to vary across individuals, we allow the individual specific effects, which include locations of the weather stations in the climate data example in Section 5, being reflected in our model. More specifically, we consider the following panel data model:

$$Y_{it} = g_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) + \varepsilon_{it}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq T, \quad (1.1)$$

where $g_i(\cdot)$ are unknown link functions that vary across individuals, \mathbf{X}_{it} are p -dimensional covariates, $\boldsymbol{\theta}_0 = (\theta_{0,1}, \dots, \theta_{0,p})^\top$ is a $p \times 1$ vector of unknown parameters, and \top symbolizes the transpose of a vector or matrix. For identifiability, we assume $\|\boldsymbol{\theta}_0\| = 1$ and that the first nonzero entity of $\boldsymbol{\theta}_0$ is positive.

In Section 2, we extend a so-called refined minimum average (conditional) variance estimation (RMAVE) method to estimate the parameter $\boldsymbol{\theta}_0$ in model (1.1).

The RMAVE method was introduced for time–series single–index models by Xia et al. (2002), and its asymptotic properties were established by Xia (2006). As there are two indices (i.e. i and t) involved and the nonlinear link functions are heterogeneous, the establishment of the asymptotic theory for the RMAVE method for model (1.1) is much more complicated than that for the time series case. In Section 3, we derive an asymptotic distribution for the RMAVE based estimator when the cross–section dimension N and the time–series dimension T tend to infinity simultaneously. Such an asymptotic distribution is called the joint limiting distribution (see Phillips and Moon 1999 for details). We show that under certain regularity conditions, the RMAVE of θ_0 is asymptotically normal with \sqrt{NT} rate of convergence.

We assume that $\{(\mathbf{X}_{it}^\top, \varepsilon_{it})^\top, t \geq 1\}$ is stationary α –mixing for each i . It is well–known that α –mixing dependence is one of the weakest mixing conditions for weakly dependent processes and it is satisfied by many stationary time series and Markov chains under certain regularity conditions. Studying model (1.1) under the α –mixing dependence allows us to extend it to a dynamic panel data model where the covariates \mathbf{X}_{it} contain lagged values of Y_{it} . Section 4 discusses some conditions that ensure $\{Y_{it}, t \geq 1\}$ to be a geometrically ergodic process for each i . Thus the stationarity and mixing conditions and thus the asymptotic properties in Section 3 still hold for such a dynamic model.

We provide an empirical example in Section 5 to illustrate the applicability of the proposed model and estimation method. We consider an application to a climate data set from UK by examining the relationship between the monthly average maximum temperatures and the number of millimeters of rainfall and hours of sunshine duration. The heterogeneous link functions used allow us to take into account individual specific effects such as the effects of the location of each weather station. We compare the results from the proposed single–index model and RMAVE method with those from a linear model and the ordinary least squares (OLS) estimation method.

The rest of the paper is organized as follows. In Section 2, we develop a detailed algorithm of the RMAVE method. Section 3 establishes an asymptotic theory for the proposed estimator of the single–index parameter. Section 4 discusses certain

conditions for a dynamic single-index model to be geometrically ergodic, which ensures that the asymptotic properties in Section 3 are still valid for the dynamic model. Section 5 discusses bandwidth selection and a real data example. Section 6 concludes this paper. Some technical lemmas and the detailed proof of the main result are given in Appendices A and B, respectively.

2. SEMIPARAMETRIC ESTIMATION METHODS

In this section, we develop a RMAVE method to estimate both the parameter θ_0 in the single-index part and the averaged link function defined in (1.1). As the link functions are heterogeneous, the RMAVE method originally studied in Xia (2006) for the time series case will have to be extended substantially when applied to model (1.1).

Given $\theta^\top \mathbf{X}_{it} = u$, define

$$\sigma_{\theta,i}^2(u) = E \left[\left(Y_{it} - g_i(\theta^\top \mathbf{X}_{it}) \right)^2 \mid \theta^\top \mathbf{X}_{it} = u \right], \quad 1 \leq i \leq N.$$

It can be easily seen that

$$E \left(Y_{it} - g_i(\theta^\top \mathbf{X}_{it}) \right)^2 = E_u \left[\sigma_{\theta,i}^2(u) \right]. \quad (2.1)$$

Based on (2.1), the estimator of θ_0 can be obtained by minimizing

$$\sum_{i=1}^N E \left(Y_{it} - g_i(\theta^\top \mathbf{X}_{it}) \right)^2 = \sum_{i=1}^N E_u \left[\sigma_{\theta,i}^2(u) \right].$$

As the link functions $g_i(\cdot)$ are unknown, we estimate them by the local linear method. It is well-known that the local linear fitting has certain advantages over the Nadaraya-Watson kernel method (see Fan and Gijbels 1996 for example). We assume that for each $1 \leq i \leq N$, $g_i(\cdot)$ is differentiable up to the third order. Then, by Taylor expansion, we have the following local linear approximation when \mathbf{X}_{it} are close to \mathbf{x} :

$$Y_{it} - g_i(\theta^\top \mathbf{X}_{it}) \approx Y_{it} - g_i(\theta^\top \mathbf{x}) - g_i'(\theta^\top \mathbf{x}) \theta^\top (\mathbf{X}_{it} - \mathbf{x}),$$

where the symbol ' denote the derivative of a function. We then propose estimating θ_0 by the values of θ that minimizes

$$\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K \left(\frac{\theta^\top \mathbf{X}_{its}}{h} \right) \left(Y_{it} - g_i(\theta^\top \mathbf{X}_{is}) - g_i'(\theta^\top \mathbf{X}_{is}) \theta^\top \mathbf{X}_{its} \right)^2, \quad (2.2)$$

where $\mathbf{X}_{its} = \mathbf{X}_{it} - \mathbf{X}_{is}$, $K(\cdot)$ is a scalar kernel function, and h is the bandwidth. As $g_i(\cdot)$ and $g'_i(\cdot)$ are all unknown, an iterative procedure has to be used to obtain the solution to the above minimization problem.

We start the iterations with an initial estimate of θ_0 , which is denoted $\tilde{\theta}_0$. Letting $\theta = \tilde{\theta}_0$, the algorithm for estimating θ_0 is described as follows:

Step 1. For given θ , calculate

$$\begin{aligned} \begin{pmatrix} a_{is} \\ b_{is} \end{pmatrix} &= \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\theta^\top \mathbf{X}_{its}) \begin{pmatrix} 1 \\ \theta^\top \mathbf{X}_{its} \end{pmatrix} \begin{pmatrix} 1 \\ \theta^\top \mathbf{X}_{its} \end{pmatrix}^\top \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\theta^\top \mathbf{X}_{its}) \begin{pmatrix} 1 \\ \theta^\top \mathbf{X}_{its} \end{pmatrix} Y_{it} \right\}, \end{aligned} \quad (2.3)$$

for each $1 \leq i \leq N$ and $1 \leq t \leq T$, where $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$.

Step 2. With the above a_{is} and b_{is} replacing $g_i(\theta^\top \mathbf{X}_{is})$ and $g'_i(\theta^\top \mathbf{X}_{is})$ and by minimizing the resulting weighted least squares in (2.2) with respect to θ , obtain

$$\begin{aligned} \tilde{\theta} &= \left\{ \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K_h(\theta^\top \mathbf{X}_{its}) b_{is}^2 \mathbf{X}_{its} \mathbf{X}_{its}^\top / \widehat{f}_i^\theta(\theta^\top \mathbf{X}_{is}) \right\}^+ \\ &\quad \times \left\{ \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K_h(\theta^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} (Y_{it} - a_{is}) / \widehat{f}_i^\theta(\theta^\top \mathbf{X}_{is}) \right\}, \end{aligned} \quad (2.4)$$

where $\widehat{f}_{\theta,i}(\theta^\top \mathbf{X}_{is}) = \frac{1}{T} \sum_{t=1}^T K_h(\theta^\top \mathbf{X}_{its})$, and A^+ stands for the pseudoinverse of a matrix A .

Step 3. Update θ with $\theta = \tilde{\theta} / \|\tilde{\theta}\|$, and then repeat Step 1 and Step 2 until the values of θ converge.

We denote the final estimate of θ_0 by $\widehat{\theta}$. To implement the above algorithm, we need to choose a suitable initial estimator $\tilde{\theta}_0$ and an optimal bandwidth h . Such issues will be discussed in Sections 3 and 4 below.

Let $\widehat{g}_i(u) = a_{i,u}$, where $a_{i,u}$ is defined as a_{is} in (2.3) with θ and $\theta^\top \mathbf{X}_{is}$ replaced by $\widehat{\theta}$ and u , respectively. Then, for each $1 \leq i \leq N$, $\widehat{g}_i(u)$ is an estimator of $g_i(u)$.

3. ASYMPTOTIC THEORY

In this section, we establish asymptotic distributions for $\widehat{\boldsymbol{\theta}}$ and $\widehat{g}(\cdot)$. To this end, we impose some regularity conditions. We assume, throughout the paper, that $\{\mathbf{X}_{it}\}$ is stationary over t , and define $\boldsymbol{\mu}_{\boldsymbol{\theta},i}(u) = E(\mathbf{X}_{it}|\boldsymbol{\theta}^\top \mathbf{X}_{it} = u)$ and $\boldsymbol{\nu}_{\boldsymbol{\theta},i}(u) = \boldsymbol{\mu}_{\boldsymbol{\theta},i}(u) - \mathbf{x}$.

A1. $K(\cdot)$ is a symmetric and continuous probability kernel function with a compact support, and its derivative is bounded.

A2(i). $\{\mathbf{X}_{it}\}$ and $\{\varepsilon_{it}\}$ are cross sectionally independent.

(ii). For each i , $\{(\mathbf{X}_{it}^\top, \varepsilon_{it})^\top, t \geq 1\}$ is a stationary sequence of α -mixing random vectors with

$$E[\varepsilon_{it}|\boldsymbol{\theta}^\top \mathbf{X}_{it}] = 0, \quad \max_i E|\varepsilon_{it}|^{2+\delta} < \infty \quad \text{and} \quad \max_i E\|\mathbf{X}_{it}\|^{2+\delta} < \infty$$

for some $\delta > 0$, and the mixing coefficient $\alpha_i(\cdot)$ satisfying $\max_i \alpha_i(t) = O(t^{-\kappa})$ with $\kappa > \frac{(2+\delta)}{\delta}$.

A3(i). For each i , let $f_{\boldsymbol{\theta},i}(\cdot)$ be the density function of $\{\boldsymbol{\theta}^\top \mathbf{X}_{it}, t \geq 1\}$. Then, $f_{\boldsymbol{\theta},i}(\cdot)$ has bounded derivatives of up to the third order. Moreover,

$$\min_i \inf_{\|\mathbf{x}\| \leq C_{NT}} f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) > 0,$$

uniformly for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, where $C_{NT} = C(NT)^{\frac{1}{2+\delta}}$ for some $C > 0$.

(ii). For $1 \leq i \leq N$, each of the link functions $g_i(\cdot)$ has bounded derivatives of up to the third order.

(iii). $\boldsymbol{\mu}_{\boldsymbol{\theta},i}(\cdot)$ has bounded derivatives of up to the second order.

A4. The bandwidth h satisfies

$$\begin{aligned} NTh &\rightarrow \infty, & NTh^6 &\rightarrow 0, & \gamma_{NT}h^{-1} &\rightarrow 0, & NTh^2\gamma_{NT}^2 &\rightarrow 0, \\ (NT)^{1/(2+\delta)}\gamma_{NT} &\rightarrow 0 & \text{and} & (NT)^{1+(p+\kappa+2)/(2+\delta)}h^{-1-p}\gamma_{NT}^{\kappa-p} &\rightarrow 0, \end{aligned}$$

where $\gamma_{NT} = \sqrt{\frac{\log T}{Th}}$ and κ was defined in **A2(ii)**.

Remark 3.1. **A1** is a set of mild conditions on the kernel function, and has been used by many authors in the time series case (see Fan and Yao 2003; Gao 2007; Li and Racine 2007 for example). In **A2**, we assume that $\{\mathbf{X}_{it}\}$ and $\{\varepsilon_{it}\}$ are cross-sectionally independent (see Cai and Li 2008 for example) and time series α -mixing. The α -mixing conditions can be satisfied by many linear and nonlinear time series models (see, for example, Auestad and Tjøstheim 1990, Chen and Tsay 1993 for example). **A3** includes commonly-used conditions for single-index models (see Xia 2006 for example). In **A4**, the condition $\gamma_{NT}h^{-1} \rightarrow 0$ implies $Th^3 \rightarrow \infty$. On the other hand, $NTh^2\gamma_{NT}^2 \rightarrow 0$ implies $Nh \rightarrow 0$. Therefore, $T \gg h^{-3} \gg N^3$, which indicates that the limiting theory in this paper holds under the condition that the rate of T tending to infinity is faster than that of N^3 . This is a rigorous condition and is required due to the fact that each individual-specific link function $g_i(\cdot)$ ($1 \leq i \leq N$) is estimated using T time series observations from each individual and the index parameter $\boldsymbol{\theta}_0$ is estimated using all NT observations.

Remark 3.2. Note that the condition $(NT)^{1+(p+\kappa+2)/(2+\delta)}h^{-1-p}\gamma_{NT}^{\kappa-p} \rightarrow 0$ in **A4** is close to $\gamma_{NT}(NT)^{1/(2+\delta)} \rightarrow 0$ as $\kappa \rightarrow \infty$. In addition, if $\delta \rightarrow \infty$, $(NT)^{1/(2+\delta)}\gamma_{NT} \rightarrow 0$ is close to $\gamma_{NT} \rightarrow 0$, which is a conventional condition for uniform consistency of nonparametric kernel-based statistics in the time series case. When $T \sim N^4$ and $h \sim (NT)^{-\tau}$, it can be shown that $NTh \rightarrow \infty$, $NTh^6 \rightarrow 0$, $\gamma_{NT}h^{-1} \rightarrow 0$ and $NTh^2\gamma_{NT}^2 \rightarrow 0$ are all satisfied when $\frac{1}{6} < \tau < \frac{1}{5}$.

Let $\mathbf{W}_{it} = (\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}_0, i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}))g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it})\varepsilon_{it}$. Then, by **A2(ii)**, it is easy to show that for each i ,

$$\Lambda_{i,T} := \frac{1}{T} \text{Var} \left[\sum_{t=1}^T \mathbf{W}_{it} \right] = \mathbb{E} \left[\mathbf{W}_{i1} \mathbf{W}_{i1}^\top \right] + 2 \sum_{t=2}^T \left[1 - \frac{(t-1)}{T} \right] \mathbb{E} \left[\mathbf{W}_{i1} \mathbf{W}_{it}^\top \right] < \infty. \quad (3.1)$$

Further denote $D_{\boldsymbol{\theta}_0, i} = \mathbb{E} \left[\left(g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right)^2 \boldsymbol{\nu}_{\boldsymbol{\theta}_0, i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta}_0, i}^\top(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]$. Theorem 3.1 below establishes the asymptotic distribution of the final estimator $\widehat{\boldsymbol{\theta}}$ under the condition that there is an initial estimator $\widetilde{\boldsymbol{\theta}}_0$ that is root- NT consistent, i.e. $\|\widetilde{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\| = O_P((NT)^{-1/2})$. The proof of the theorem will be given in Appendix B.

Theorem 3.1. *Assume that conditions **A1**–**A4** hold and that there exist two positive*

definite matrices Σ_{θ_0} and D_{θ_0} such that

$$\frac{1}{N} \sum_{i=1}^N \Lambda_{i,T} \rightarrow \Sigma_{\theta_0} \quad \text{and} \quad \frac{1}{N} \sum_{i=1}^N D_{\theta_0,i} \rightarrow D_{\theta_0} \quad (3.2)$$

as $N, T \rightarrow \infty$ simultaneously. If, in addition,

$$\frac{1}{N} \left(\max_{1 \leq i \leq N} \Lambda_{i,T} \right) \rightarrow 0, \quad (3.3)$$

and the initial estimator $\tilde{\theta}_0$ is \sqrt{NT} -consistent, then we have, as $N, T \rightarrow \infty$ simultaneously,

$$\sqrt{NT} (\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, D_{\theta_0}^+ \Sigma_{\theta_0} D_{\theta_0}^{+\top}) \quad (3.4)$$

where $D_{\theta_0}^+$ is the pseudoinverse of D_{θ_0} .

Remark 3.2. The above theorem shows that the estimator $\hat{\theta}$ is asymptotically normal with \sqrt{NT} rate of convergence even when the link functions are heterogeneous across individuals. (3.2) and (3.3) are imposed to make sure that the Lindeberg condition holds for the joint central limit theorem. In the meantime, the condition that the initial estimate is \sqrt{NT} -consistent is similar to the \sqrt{T} -consistency condition in the time series case (see Härdle, Hall and Ichimura 1993 and Carroll *et al* 1997 for example). As a matter of fact, this condition is feasible as such an initial estimator can be obtained by using some existing methods (see, for example, Härdle and Stoker 1989; Horowitz and Härdle 1996).

4. DYNAMIC SINGLE-INDEX PANEL DATA MODELS

We next consider the case when \mathbf{X}_{it} contains lagged values of Y_{it} . If $\mathbf{X}_{it} = (Y_{i,t-1}, \dots, Y_{i,t-p})^\top$, then model (1.1) becomes

$$Y_{it} = g_i \left(\sum_{j=1}^p \theta_{0,j} Y_{i,t-j} \right) + \varepsilon_{it}, \quad (4.1)$$

where ε_{it} are independent of $Y_{i,s}$ for all $s < t$. To ensure that the asymptotic distributions in Section 3 still hold for this dynamic model, we provide some sufficient conditions for $\{Y_{it}, t \geq 1\}$ to be geometrically ergodic for each $i \geq 1$. As a consequence, $\{Y_{it}, t \geq 1\}$ satisfies the stationarity and α -mixing conditions. Motivated by Theorems 3.1 and 3.2 in An and Huang (1996), we give two kinds of conditions on the link functions g_i that ensure the geometrical ergodicity of $\{Y_{it}, t \geq 1\}$.

Proposition 4.1. Let $\mathbf{x} = (x_1, \dots, x_p)^\top$ be a vector in \mathcal{R}^p , and assume that $\{Y_{it}\}$ is generated by model (4.1).

(i). If for each i , the link function $g_i(\cdot)$ in (4.1) satisfies

$$\sup_{\|\mathbf{x}\| \leq C} |g_i(\boldsymbol{\theta}_0^\top \mathbf{x})| < \infty \quad \text{for any } C > 0, \quad (4.2)$$

and that there exists a p -dimensional vector $\boldsymbol{\beta}_i = (\beta_{i,1}, \dots, \beta_{i,p})^\top$ such that

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{|g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) - \boldsymbol{\beta}_i^\top \mathbf{x}|}{\|\mathbf{x}\|} = 0 \quad (4.3)$$

and

$$u^p - \beta_{i,1}u^{p-1} - \dots - \beta_{i,p-1}u - \beta_{i,p} \neq 0 \quad \text{for all } |u| \geq 1, \quad (4.4)$$

then $\{Y_{it}, t \geq 1\}$ is geometrically ergodic for each $i \geq 1$.

(ii). If for each i , there exists a positive number $\lambda_i < 1$ and a constant C_i such that

$$|g_i(\boldsymbol{\theta}_0^\top \mathbf{x})| \leq \lambda_i \max\{|x_1|, \dots, |x_p|\} + C_i, \quad (4.5)$$

then $\{Y_{it}, t \geq 1\}$ is geometrically ergodic for each $i \geq 1$.

The proof of Proposition 4.1 is similar to that of Theorems 3.1 and 3.2 in An and Huang (1996). There are also similar results about geometrical ergodicity in Masry and Tjøstheim (1995), and Lu (1998).

We next provide two examples that satisfy the conditions in the above proposition.

Example 4.1. Let $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02})^\top = (0.6, 0.8)^\top$ and $g_i(u) = u/\sqrt{2} + \sin(2\pi u/i)$. Then the dynamic panel data model (4.1) reduces to

$$Y_{it} = (0.6Y_{i,t-1} + 0.8Y_{i,t-2})/\sqrt{2} + \sin\{2\pi(0.6Y_{i,t-1} + 0.8Y_{i,t-2})/i\} + \varepsilon_{it}.$$

As $\sin(\cdot)$ is a bounded function, by letting $\boldsymbol{\beta}_i = (0.6, 0.8)^\top/\sqrt{2}$, it is easy to show that (4.2)–(4.4) are satisfied. Hence, by Proposition 4.1(i), $\{Y_{it}, t \geq 1\}$ is geometrically ergodic for each $i \geq 1$.

Example 4.2. Assume that each link functions $g_i(\cdot)$ in model (4.1) satisfies

$$|g_i(u)| \leq \rho_i |u|/\sqrt{p} + \kappa_i, \quad \text{for any } u \in \mathcal{R},$$

where κ_i and ρ_i are two positive constants with $\rho_i < 1$, and p is the dimension of θ_0 in (4.1). Then following the arguments in Example 3.5 of An and Huang (1996), we can show that (4.5) holds with $\lambda_i = \rho_i$ and $C_i = \kappa_i$. And hence, $\{Y_{it}, t \geq 1\}$ is geometrically ergodic for each $i \geq 1$. On the other hand, if for each i , there exists a constant $|c_i^*| < 1$, such that

$$\lim_{|u| \rightarrow \infty} \frac{|g_i(u) - c_i^* u|}{|u|} = 0,$$

then we also can show that $\{Y_{it}, t \geq 1\}$ is geometrically ergodic for each $i \geq 1$.

5. EMPIRICAL EXAMPLES

In this section, we discuss the choice of the bandwidth and give one real data example to illustrate the proposed estimation method.

5.1. Bandwidth Selection

The choice of bandwidth is an important consideration in practice. As the link functions vary over i and each $g_i(\cdot)$ is estimated using observations from section i , we use a leave-one-unit-out cross validation method in this section. The basic idea is that for each $1 \leq t \leq T$, we use $\{(Y_{it}, \mathbf{X}_{it}) : 1 \leq i \leq N\}$ (i.e. observations from the unit t) as the testing data and use the rest of the observations as the training data. Then, we choose the optimal bandwidth as the one that minimizes the mean squared estimation error which is defined by

$$\sum_{t=1}^T \sum_{i=1}^N \left[Y_{it} - \widehat{g}_i^{(-t)}(\mathbf{x}_{it}^\top \widehat{\boldsymbol{\theta}}^{(-t)}) \right]^2, \quad (5.1)$$

where $\widehat{g}_i^{(-t)}(\cdot)$ and $\widehat{\boldsymbol{\theta}}^{(-t)}$ are estimates of $g_i(\cdot)$ and $\boldsymbol{\theta}_0$ from data with observations $\{(Y_{it}, \mathbf{X}_{it}) : 1 \leq i \leq N\}$ from unit t being removed. This bandwidth selection method is similar to the one used in Sun et al. (2009).

To see how the choice of bandwidth affects the performance of the proposed estimation in the empirical example below, we provide a plot of the mean squared estimation errors against values of the bandwidth. See details in Figure 5.1.

5.2. An empirical example

The data set we use for this empirical application is available from the UK Met Office website <http://www.metoffice.gov.uk/climate/uk/stationdata/>. This data set contains monthly data on average maximum temperatures (TMAX), average minimum temperatures (TMIN), the number of days of air frost (AF), the number of millimeters of rainfall (RAIN), and the number of hours of sunshine (SUN). The data were collected from 37 stations across the UK. We select data over the decade of January 1999–December 2008 from 16 stations according to data availability.

Both seasonality and trend are first removed from the data before fitting into a particular model, and we focus on investigating the relationship between the TMAX and RAIN and SUN. For station i , denote the seasonally adjusted and detrended TMAX at time t as $Y_{i,t}$, and the seasonally adjusted and detrended SUN and RAIN as $X_{i,t,1}$ and $X_{i,t,2}$, respectively. We then use the proposed semiparametric RMAVE method to estimate the parameter θ in the model

$$Y_{i,t} = g_i(\theta^\top \mathbf{X}_{i,t}) + \varepsilon_{i,t}, \quad i = 1, \dots, 16, \quad t = 1, \dots, 120, \quad (5.2)$$

where $\mathbf{X}_{i,t} = (X_{i,t,1}, X_{i,t,2})^\top$. We first use a least squares (OLS) estimation method to estimate β in a linear model of the form

$$Y_{i,t} = \mathbf{X}_{i,t}^\top \boldsymbol{\vartheta} + \zeta_i + \varepsilon_{i,t}, \quad (5.3)$$

where ζ_i are station-specific effects.

Then, we use the normalized OLS estimate of $\boldsymbol{\vartheta}$ as the initial estimate for θ in the RMAVE estimation of (5.2). The resulting RMAVE estimate of θ is $\widehat{\theta} = (0.9793, -0.2023)^\top$ with standard deviation $(0.0131, 0.0296)$. The calculated coefficient of determination for the single index model (5.2) is $R^2 = 0.6199$ and that for the linear model (5.3) is $R^2 = 0.1694$, which indicates that the single index model (5.2) fits the data much better than the simple linear model (5.3).

The bandwidth used in the computation of the RMAVE in this example is selected by the cross validation method introduced in Section 5.1. A plot of the mean squared estimation errors against values of the bandwidth is given in Figure 5.1, which shows that the estimation error does not vary dramatically with the change of the bandwidth. We also present, as an illustration, the estimated curves of the link functions for the first two stations as well as their 95% confidence intervals in Figure 5.2.

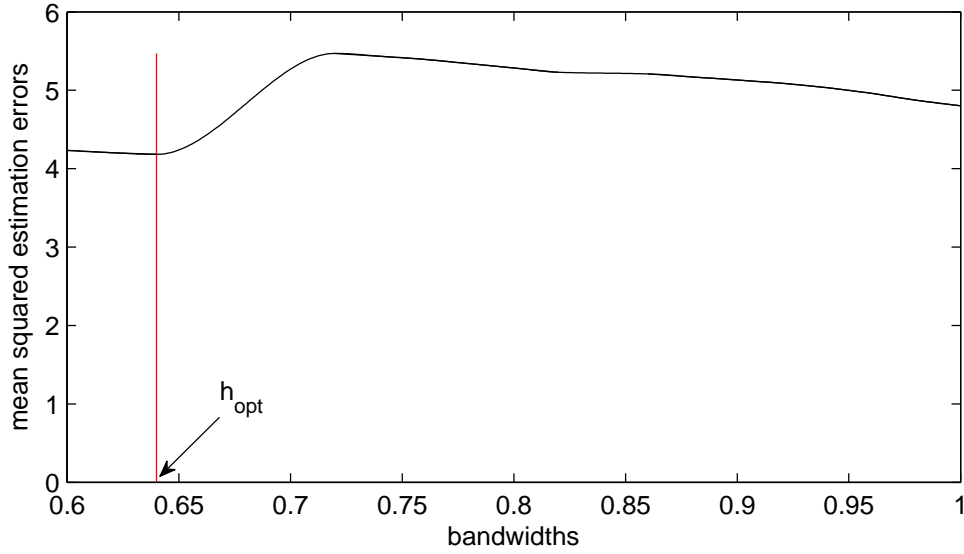


Figure 5.1. The solid curve is the mean squared estimation errors plotted against bandwidths; the vertical line is the bandwidth selected by the leave-one-unit-out cross validation.

6. CONCLUSION

We have considered an estimation problem in a single-index panel data model with heterogeneous link functions. A nonparametric local linear based minimum average variance estimation method has been proposed to estimate the parameter vector. An asymptotically normal distribution has been established for the proposed estimator. In addition, we have provided one real data example to show how the proposed model and estimation method can be used in practice.

The paper has some limitations and further extensions are left to be explored. One direction of future work is to establish the theory for the case where the residuals are cross-sectionally dependent. Another direction is to see whether the established theory can be extended to the case where $\{X_{it}\}$ is nonstationary over t and cross-sectionally dependent over i .

7. Acknowledgments

The authors acknowledge the financial support from the Australian Research Council Discovery Grants Program under Grant Number: DP0879088.

APPENDIX A: TECHNICAL LEMMAS

Define $\mathbf{X}_{it,x} = \mathbf{X}_{it} - \mathbf{x}$, $\mu_l^* = \int u^l K(u) du$ and $\nu_l^* = \int u^l K^2(u) du$ for $l = 0, 1, 2, \dots$. We

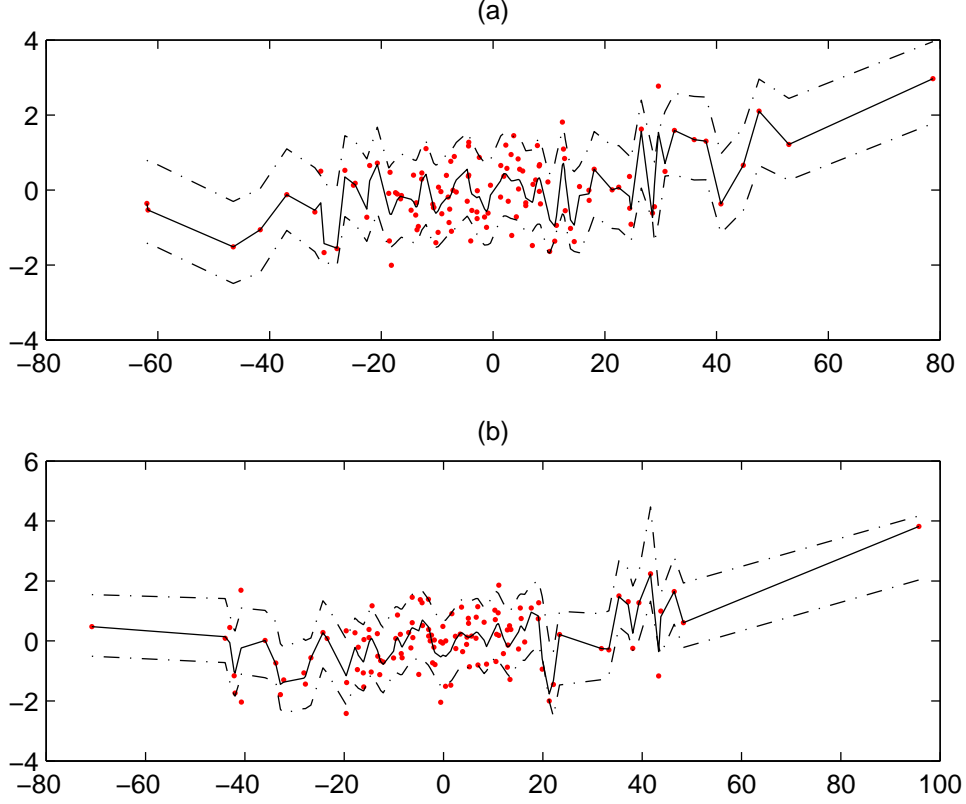


Figure 5.2. (a) & (b): curve estimates for weather stations 1 and 2, respectively. The dots denote Y_{it} plotted against $\widehat{\boldsymbol{\theta}}^\top \mathbf{X}_{it}$; the solid lines denote the estimated link functions $\widehat{g}_i(\widehat{\boldsymbol{\theta}}^\top \mathbf{X}_{it})$; the dash-dotted lines represent the 95% confidence intervals.

assume, without loss of generality, that $\mu_2^* = 1$ (otherwise, we can let $K(u) = \mu_2^{*1/2} K(\mu_2^{*1/2} u)$). Denote

$$\begin{aligned}\boldsymbol{\theta}_{NT} &= \left\{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq C_\theta (NT)^{-1/2} \right\}, \\ \mathcal{X}_{NT} &= \left\{ \mathbf{x} : \|\mathbf{x}\| \leq M(NT)^{1/(2+\delta)} \right\}, \\ \mathcal{F}_{NT} &= \{(\mathbf{x}, \boldsymbol{\theta}) : \mathbf{x} \in \mathcal{X}_{NT}, \boldsymbol{\theta} \in \boldsymbol{\theta}_{NT}\},\end{aligned}$$

where C_θ and M are two positive constants. Furthermore, define

$$\begin{aligned}\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) &= K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}}{h} \right)^l \mathbf{X}_{it,\mathbf{x}}, \\ \mathbf{Z}_{it}^*(\mathbf{x}, \boldsymbol{\theta}, l) &= K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}}{h} \right)^l.\end{aligned}$$

Lemma A.1. Assume that **A1**, **A2**, **A3 (i)(iii)** hold. If, in addition,

$$h \rightarrow 0, \quad (NT)^{1/(2+\delta)} \gamma_{NT} \rightarrow 0, \quad (NT)^{1+(p+\kappa+2)/(2+\delta)} h^{-1-p} \gamma_{NT}^{\kappa-p} \rightarrow 0,$$

then for $l = 0, 1, 2, \dots$, we have

$$\max_{1 \leq i \leq N} \sup_{\mathbf{x} \in \mathcal{X}_{NT}} \left| \frac{1}{T} \sum_{t=1}^T Z_{it}^*(\mathbf{x}, \boldsymbol{\theta}, l) - f_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_l^* - f'_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_{l+1}^* h \right| = O_P(h^2 + \gamma_{NT}) \quad (\text{A.1})$$

and

$$\begin{aligned} & \max_{1 \leq i \leq N} \sup_{\mathbf{x} \in \mathcal{X}_{NT}} \left| \frac{1}{T} \sum_{t=1}^T \mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - f_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_l^* \right. \\ & \quad \left. - [f'_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) + f_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}'_{\boldsymbol{\theta}, i}(\boldsymbol{\theta}^\top \mathbf{x})] \mu_{l+1}^* h \right| \\ = & O_P(h^2 + \gamma_{NT}), \end{aligned} \quad (\text{A.2})$$

where $f'_{\boldsymbol{\theta}, i}(u)$ and $\boldsymbol{\nu}'_{\boldsymbol{\theta}, i}(u)$ denote the derivatives of $f_{\boldsymbol{\theta}, i}(u)$ and $\boldsymbol{\nu}_{\boldsymbol{\theta}, i}(u)$, respectively.

Proof. We only prove (A.2) as the proof of (A.1) is similar. To prove (A.2), we first show that

$$\max_{1 \leq i \leq N} \sup_{\mathbf{x} \in \mathcal{X}_{NT}} \left| \frac{1}{T} \sum_{t=1}^T \{\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - \mathbb{E}[\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l)]\} \right| = O_P(\gamma_{NT}). \quad (\text{A.3})$$

We partition the set \mathcal{X}_{NT} into B_{NT} balls \mathcal{B}_j , $1 \leq j \leq B_{NT}$, each centered at \mathbf{x}_j with radius $r_{NT} = O(h\gamma_{NT})$. By a simple calculation, we have

$$B_{NT} = O\left((NT)^{p/(2+\delta)} h^{-p} \gamma_{NT}^{-p}\right).$$

Then, for each $\boldsymbol{\theta} \in \boldsymbol{\theta}_{NT}$ we have

$$\begin{aligned} & \max_{1 \leq i \leq N} \sup_{\mathbf{x} \in \mathcal{X}_{NT}} \left| \frac{1}{T} \sum_{t=1}^T \{\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - \mathbb{E}[\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l)]\} \right| \\ \leq & \max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T \{\mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E}[\mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l)]\} \right| \\ & + \max_{1 \leq j \leq B_{NT}} \sup_{\mathbf{x} \in \mathcal{B}_j} \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T \{[\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - \mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l)] - \mathbb{E}[\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - \mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l)]\} \right| \\ =: & \max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} |H_{T,i}(\mathbf{x}_j, \boldsymbol{\theta}, l)| + \max_{1 \leq j \leq B_{NT}} \sup_{\mathbf{x} \in \mathcal{B}_j} \max_{1 \leq i \leq N} |H_{T,i}^*(\mathbf{x}, \boldsymbol{\theta}, l)|. \end{aligned} \quad (\text{A.4})$$

We first consider $\max_{1 \leq k \leq B_{NT}} \max_{1 \leq i \leq N} |H_{T,i}(\mathbf{x}_j, \boldsymbol{\theta}, l)|$. Let

$$\bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) = \mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) I\{\|\mathbf{X}_{it}\| \leq \Delta_{NT}\} \quad \text{and} \quad \mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) = \mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l),$$

where $\Delta_{NT} = (NT)^{1/(2+\delta)} l(NT)$, in which $l(\cdot)$ is a positive slowly-varying function satisfying

$$l(NT) \rightarrow \infty, \quad (NT)^{1/(2+\delta)} l(NT) \gamma_{NT} \rightarrow 0, \quad (\text{A.5})$$

$$(NT)^{1+(p+\kappa+2)/(2+\delta)} \gamma_{NT}^{\kappa-p} h^{-1-p} l^{\kappa+2}(NT) \rightarrow 0, \quad (\text{A.6})$$

as $N, T \rightarrow \infty$. It is easy to check that

$$H_{T,i}(\mathbf{x}_j, \boldsymbol{\theta}, l) = \frac{1}{T} \sum_{t=1}^T \left\{ \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} + \frac{1}{T} \sum_{t=1}^T \left\{ \mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\}.$$

As $l(NT) \rightarrow \infty$ and $\mathbb{E} \left[\|\mathbf{X}_{it}\|^{2+\delta} \right] < \infty$, we have, for any $M > 0$,

$$\begin{aligned} & P \left\{ \max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T \left\{ \mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} \right| > M \gamma_{NT} \right\} \\ & \leq \sum_{i=1}^N \sum_{t=1}^T \frac{\mathbb{E} \|\mathbf{X}_{it}\|^{2+\delta}}{NT l^{2+\delta} (NT)} = O \left(\frac{1}{l^{2+\delta} (NT)} \right) = o(1), \end{aligned}$$

which implies

$$\max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T \left\{ \mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\mathbf{Z}_{it}^c(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} \right| = o_P(\gamma_{NT}). \quad (\text{A.7})$$

Furthermore, by **A1**, **A2(ii)**, **A3(i)** and the standard calculation for the variance of α -mixing nonparametric kernel statistics, we have

$$\max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} \text{Var} \left[\sum_{t=1}^T \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] = O(Th^{-1}). \quad (\text{A.8})$$

By Bernstein inequality for α -mixing processes (see Theorem 2.18 in Fan and Yao 2003 for example), it follows that

$$\begin{aligned} & P \left\{ \left| \frac{1}{T} \sum_{t=1}^T \left\{ \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} \right| > M \gamma_{NT} \right\} \\ & = P \left\{ \left| \sum_{t=1}^T \left\{ \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} \right| > MT \gamma_{NT} \right\} \\ & \leq 4 \exp(-CM^2 \log T) + CT \gamma_{NT}^\kappa \Delta_{NT}^{\kappa+2} h^{-1}, \end{aligned}$$

where C is a generic positive constant. This implies

$$\begin{aligned} & P \left\{ \max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T \left\{ \bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) - \mathbb{E} \left[\bar{\mathbf{Z}}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l) \right] \right\} \right| > M \gamma_{NT} \right\} \\ & = O \left(B_{NT} N T^{-CM^2} + B_{NT} N T \gamma_{NT}^\kappa \Delta_{NT}^{\kappa+2} h^{-1} \right) \\ & = O \left(\gamma_{NT}^{-p} h^{-p} N^{1+p/(2+\delta)} T^{p/(2+\delta)-CM^2} + (NT)^{1+(p+\kappa+2)/(2+\delta)} \gamma_{NT}^{\kappa-p} h^{-1-p} l^{\kappa+2} (NT) \right) \\ & = o(1) \end{aligned} \quad (\text{A.9})$$

when M is large enough. Because of $B_{NT} = O \left((NT)^{p/(2+\delta)} h^{-p} \gamma_{NT}^{-p} \right)$ and $\Delta_{NT} = (NT)^{1/(2+\delta)} l(NT)$, the second equality in (A.9) holds. The third equality holds as $(NT)^{1+(p+\kappa+2)/(2+\delta)} h^{-1-p} \gamma_{NT}^{\kappa-p} \rightarrow 0$ (see **A4**). It follows from (A.7) and (A.9) that

$$\max_{1 \leq j \leq B_{NT}} \max_{1 \leq i \leq N} H_{T,i}(\mathbf{x}_j, \boldsymbol{\theta}, l) = O_P(\gamma_{NT}). \quad (\text{A.10})$$

Meanwhile, by **A1** we have

$$\begin{aligned} & \max_{1 \leq j \leq B_{NT}} \sup_{\mathbf{x} \in \mathcal{B}_j} \max_{1 \leq i \leq N} |\mathbf{Z}_{it}(\mathbf{x}, \boldsymbol{\theta}, l) - \mathbf{Z}_{it}(\mathbf{x}_j, \boldsymbol{\theta}, l)| \\ & \leq Ch^{-1} \left(\max_{1 \leq j \leq B_{NT}} \sup_{\mathbf{x} \in \mathcal{B}_j} |\mathbf{x} - \mathbf{x}_j| \right) \leq Ch^{-1} h \gamma_{NT} = O(\gamma_{NT}), \end{aligned}$$

which implies

$$\max_{1 \leq j \leq B_{NT}} \sup_{\mathbf{x} \in \mathcal{B}_j} \max_{1 \leq i \leq N} H_{T,i}^*(\mathbf{x}, \boldsymbol{\theta}, l) = O_P(\gamma_{NT}). \quad (\text{A.11})$$

By (A.4), (A.10) and (A.11), equation (A.3) follows.

Meanwhile, by **A1** and **A3(i)(iii)**, we have

$$\mathbb{E} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}}{h} \right)^l \right] = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_l^* + f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_{l+1}^* h + O(h^2),$$

and

$$\begin{aligned} & \mathbb{E} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}}{h} \right)^l \mathbf{X}_{it,\mathbf{x}} \right] \\ & = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \mu_l^* + \left[f'_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}'_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \right] \mu_{l+1}^* h + O(h^2). \end{aligned}$$

Lemma A.1 follows from (A.3) and the above two equations.

Lemma A.2. Let $a_{i,\mathbf{x}}$ and $b_{i,\mathbf{x}}$ be defined as $a_{i,s}$ and $b_{i,s}$ with \mathbf{X}_{its} replaced by $\mathbf{X}_{it,\mathbf{x}}$ respectively in (2.3), and define

$$\Xi_{i,l}^\theta(\mathbf{x}) = \frac{1}{T} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{x}) \sum_{l=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{il,\mathbf{x}}) \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{il,\mathbf{x}}}{h} \right)^l \varepsilon_{il}, \quad l = 0, 1, 2, \dots$$

Then, under **A1–A3**, we have

$$\begin{aligned} a_{i,\mathbf{x}} & = g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) (\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + \frac{1}{2} g_i''(\boldsymbol{\theta}_0^\top \mathbf{x}) h^2 + \Xi_{i,0}^\theta(\mathbf{x}) \\ & \quad + O_P(\|\boldsymbol{\delta}_\theta\|^2 + h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\|) \end{aligned} \quad (\text{A.12})$$

and

$$b_{i,\mathbf{x}} = g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) + h^{-1} \Xi_{i,1}^\theta(\mathbf{x}) + O_P(h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}) \quad (\text{A.13})$$

uniformly in $\mathbf{x} \in \mathcal{X}_{NT}$, where $\boldsymbol{\delta}_\theta = \boldsymbol{\theta} - \boldsymbol{\theta}_0$.

Proof. Define $S_{i,l}^\theta = \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})^l$, $l = 0, 1, 2, 3$. By simple calculation, we have

$$a_{i,\mathbf{x}} = \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \right] Y_{it} \right\} \quad (\text{A.14})$$

and

$$b_{i,x} = \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,x}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,x}) - S_{i,1}^\theta \right] Y_{it} \right\}. \quad (\text{A.15})$$

Furthermore, it follows from Lemma A.1 that

$$S_{i,0}^\theta = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P(h^2 + \gamma_{NT}), \quad (\text{A.16})$$

$$S_{i,1}^\theta = O_P(h[h + \gamma_{NT}]) = O_P(h^2 + h\gamma_{NT}), \quad (\text{A.17})$$

$$S_{i,2}^\theta = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x})h^2 + O_P(h^2[h^2 + \gamma_{NT}]), \quad (\text{A.18})$$

$$S_{i,3}^\theta = O_P(h^3[h + \gamma_{NT}]) = O_P(h^4 + h^3\gamma_{NT}), \quad (\text{A.19})$$

uniformly in $\mathbf{x} \in \mathcal{X}_{NT}$. Equations (A.16)–(A.18) imply

$$S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 = f_{\boldsymbol{\theta},i}^2(\boldsymbol{\theta}^\top \mathbf{x})h^2 + O_P(h^2[h^2 + \gamma_{NT}]). \quad (\text{A.20})$$

Meanwhile, by Taylor expansion, we have, for $|\boldsymbol{\theta}^\top \mathbf{X}_{it}| \leq Ch$,

$$\begin{aligned} Y_{it} &= g_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) + \varepsilon_{it} \\ &= \varepsilon_{it} + g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x} + \frac{1}{2}g''_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x})^2 + O(|\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x}|^3) \\ &= \varepsilon_{it} + g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})\boldsymbol{\theta}^\top \mathbf{X}_{it,x} + \frac{1}{2}g''_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^2 + g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} \\ &\quad + \frac{1}{2}g''_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x})^2 - (\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^2 \right] + O(|\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x}|^3) \\ &= \varepsilon_{it} + g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})\boldsymbol{\theta}^\top \mathbf{X}_{it,x} + \frac{1}{2}g''_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^2 \\ &\quad + g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} + \Delta_{it,x}, \end{aligned} \quad (\text{A.21})$$

where

$$\begin{aligned} \Delta_{it,x} &= O\left(\left[(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x})^2 - (\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^2 \right] + |\boldsymbol{\theta}_0^\top \mathbf{X}_{it,x}|^3 \right) \\ &= O\left(\left[(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} \right]^2 + \left[(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} \right]^3 + (\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^3 + 2\boldsymbol{\theta}^\top \mathbf{X}_{it,x}(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} \right. \\ &\quad \left. + 3(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x}(\boldsymbol{\theta}^\top \mathbf{X}_{it,x})^2 + 3\left[(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,x} \right]^2 \boldsymbol{\theta}^\top \mathbf{X}_{it,x} \right) \\ &= O\left(\|\delta_\theta\|^2 \|\mathbf{X}_{it,x}\|^2 + \|\delta_\theta\|^3 \|\mathbf{X}_{it,x}\|^3 + h^3 + \|\delta_\theta\| \|\mathbf{X}_{it,x}\| h \right. \\ &\quad \left. + \|\delta_\theta\| \|\mathbf{X}_{it,x}\| h^2 + \|\delta_\theta\|^2 \|\mathbf{X}_{it,x}\|^2 h \right). \end{aligned} \quad (\text{A.22})$$

It follows from (A.16)–(A.20) that

$$\begin{aligned} &\left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,x}) \left[S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,x}) \right] g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \right\} \\ &= g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right] \\ &= g_i(\boldsymbol{\theta}_0^\top \mathbf{x}), \end{aligned} \quad (\text{A.23})$$

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) [S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})] g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \right\} \\
&= g'_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,1}^\theta S_{i,2}^\theta - S_{i,1}^\theta S_{i,2}^\theta) \\
&= 0,
\end{aligned} \tag{A.24}$$

and

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) [S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})] \left[\frac{1}{2} g''_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})^2 \right] \right\} \\
&= \frac{1}{2} g''_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} [(S_{i,2}^\theta)^2 - S_{i,1}^\theta S_{i,3}^\theta] \\
&= \frac{1}{2} g''_i(\boldsymbol{\theta}_0^\top \mathbf{x}) h^2 + O_P(h^2(h^2 + \gamma_{NT})).
\end{aligned} \tag{A.25}$$

Define $Q_{i,l}^\theta(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})^l \mathbf{X}_{it,\mathbf{x}}$ for $l = 0, 1, 2$. Then, it follows from Lemma A.1 that

$$Q_{i,0}^\theta(\mathbf{x}) = f_{\theta,i}(\boldsymbol{\theta}^\top \mathbf{x}) \boldsymbol{\nu}_{\theta,i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P(h^2 + \gamma_{NT}), \tag{A.26}$$

$$Q_{i,1}^\theta(\mathbf{x}) = O_P(h(h + \gamma_{NT})). \tag{A.27}$$

Consequently, we have

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) [S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})] g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,\mathbf{x}} \right\} \\
&= g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,2}^\theta Q_{i,0}^\theta(\mathbf{x}) - S_{i,1}^\theta Q_{i,1}^\theta(\mathbf{x})) \\
&= g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \boldsymbol{\nu}_{\theta,i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P((h^2 + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\|).
\end{aligned} \tag{A.28}$$

Furthermore, by (A.22) we have

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) [S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})] \Delta_{it,\mathbf{x}} \right\} \\
&= O_P(\|\boldsymbol{\delta}_\theta\|^2 + h\|\boldsymbol{\delta}_\theta\| + h^3 + \|\boldsymbol{\delta}_\theta\|^3 + h^2\|\boldsymbol{\delta}_\theta\| + h\|\boldsymbol{\delta}_\theta\|^2) \\
&= O_P(\|\boldsymbol{\delta}_\theta\|^2 + h^3 + h\|\boldsymbol{\delta}_\theta\|).
\end{aligned} \tag{A.29}$$

Moreover, by noting that

$$\Xi_{i,0}^\theta(\mathbf{x}) = O_P(\gamma_{NT}) \quad \text{and} \quad \Xi_{i,1}^\theta(\mathbf{x}) = O_P(\gamma_{NT}), \tag{A.30}$$

we have

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) [S_{i,2}^\theta - S_{i,1}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})] \varepsilon_{it} \right\} \\
&= \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} f_{\theta,i}(\boldsymbol{\theta}^\top \mathbf{x}) [S_{i,2}^\theta \Xi_{i,0}^\theta(\mathbf{x}) - h S_{i,1}^\theta \Xi_{i,1}^\theta(\mathbf{x})] \\
&= \Xi_{i,0}^\theta(\mathbf{x}) + O_P(\gamma_{NT}(h + \gamma_{NT})).
\end{aligned} \tag{A.31}$$

By (A.14), (A.20), (A.21), (A.23)–(A.25), (A.28), (A.29) and (A.31), equation (A.12) follows.

On the other hand, we have

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta \right] g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \right\} \\
&= g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,0}^\theta S_{i,1}^\theta - S_{i,0}^\theta S_{i,1}^\theta) \\
&= 0,
\end{aligned} \tag{A.32}$$

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta \right] g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) \boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}} \right\} \\
&= g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2) \\
&= g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}),
\end{aligned} \tag{A.33}$$

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) (S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta) \left(\frac{1}{2} g_i''(\boldsymbol{\theta}_0^\top \mathbf{x}) (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})^2 \right) \right\} \\
&= \frac{1}{2} g_i''(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,0}^\theta S_{i,3}^\theta - S_{i,1}^\theta S_{i,2}^\theta) \\
&= O_P(h(h + \gamma_{NT})),
\end{aligned} \tag{A.34}$$

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta \right] g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) (\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \mathbf{X}_{it,\mathbf{x}} \right\} \\
&= g_i'(\boldsymbol{\theta}_0^\top \mathbf{x}) (\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} (S_{i,0}^\theta Q_{i,1}^\theta(\mathbf{x}) - S_{i,1}^\theta Q_{i,0}^\theta(\mathbf{x})) \\
&= O_P((h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| h^{-1})
\end{aligned} \tag{A.35}$$

and

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta \right] \Delta_{it,\mathbf{x}} \right\} \\
&= O_P(\|\boldsymbol{\delta}_\theta\|^2 h^{-1} + \|\boldsymbol{\delta}_\theta\| + h^2 + \|\boldsymbol{\delta}_\theta\|^3 h^{-1} + h \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2) \\
&= O_P(\|\boldsymbol{\delta}_\theta\|^2 h^{-1} + \|\boldsymbol{\delta}_\theta\| + h^2).
\end{aligned} \tag{A.36}$$

And from (A.16), (A.17) and (A.30), it follows that

$$\begin{aligned}
& \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} \left\{ \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \left[S_{i,0}^\theta (\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) - S_{i,1}^\theta \right] \varepsilon_{it} \right\} \\
&= \left[S_{i,0}^\theta S_{i,2}^\theta - (S_{i,1}^\theta)^2 \right]^{-1} f_{\theta,i}(\boldsymbol{\theta}^\top \mathbf{x}) \left[h S_{i,0}^\theta \Xi_{i,1}^\theta(\mathbf{x}) - S_{i,1}^\theta \Xi_{i,0}^\theta(\mathbf{x}) \right] \\
&= h^{-1} \Xi_{i,1}^\theta(\mathbf{x}) + O_P(\gamma_{NT}(h + \gamma_{NT})h^{-1}).
\end{aligned} \tag{A.37}$$

Equation (A.13) then follows from (A.15), (A.20), (A.21) and (A.32)–(A.37).

Lemma A.3. *Under the conditions of Lemma A.2, we have*

$$\begin{aligned} & \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is}^2 \mathbf{X}_{its} \mathbf{X}_{its}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\ = & \frac{2}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0,i} + O_P\left(h^2 + \gamma_{NT} h^{-1} + \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 h^{-1} + (NT)^{-1/2}\right) \end{aligned} \quad (\text{A.38})$$

and

$$\begin{aligned} & \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} (Y_{it} - a_{is} - b_{is} \boldsymbol{\theta}_0^\top \mathbf{X}_{its}) \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\ = & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{it})] g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \varepsilon_{it} + \frac{1}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ & + O_P\left((h + \gamma_{NT} + \|\boldsymbol{\delta}_\theta\|) \left[h^2 + \gamma_{NT} + \gamma_{NT}^2 h^{-1} + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| h^{-1} + \|\boldsymbol{\delta}_\theta\|^2 h^{-1} \right] \right. \\ & \left. + (NT)^{-1/2} (h^2 + \|\boldsymbol{\delta}_\theta\|) \right), \end{aligned} \quad (\text{A.39})$$

where $D_{\boldsymbol{\theta}_0,i} = E\left[\left(g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is})\right)^2 \boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}^\top(\boldsymbol{\theta}_0^\top \mathbf{X}_{is})\right]$ and $\widehat{f}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}})$.

Proof. By $C \equiv E[\|\mathbf{X}_{it}\|^{2+\delta}] < \infty$ in A2 (ii), it follows that for any small $\epsilon > 0$ and $M > \sqrt{1/\epsilon}$, we have

$$P\left\{\max_{1 \leq i \leq N} \max_{1 \leq t \leq T} \|\mathbf{X}_{it}\| > M(NT)^{1/(2+\delta)}\right\} \leq \sum_{i=1}^N \sum_{t=1}^T \frac{E\|\mathbf{X}_{it}\|^{2+\delta}}{M^{2+\delta} NT} = \frac{C}{M^{2+\delta}} < \epsilon.$$

Hence, we need only to prove that (A.38) and (A.39) hold for $\max_{1 \leq i \leq N} \max_{1 \leq t \leq T} \|\mathbf{X}_{it}\| \leq M(NT)^{1/(2+\delta)}$.

Define $\widetilde{\omega}_{\boldsymbol{\theta},i}(u) = E[(\mathbf{X}_{it} - \mathbf{x})(\mathbf{X}_{it} - \mathbf{x})^\top | \boldsymbol{\theta}^\top \mathbf{X}_{it} = u]$. Then, it follows from Lemma A.2 that uniformly in $\mathbf{x} \in \mathcal{X}_{NT}$,

$$\frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \mathbf{X}_{it,\mathbf{x}} \mathbf{X}_{it,\mathbf{x}}^\top = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) \widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P(h^2 + \gamma_{NT}), \quad (\text{A.40})$$

$$\widehat{f}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) = f_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P(h^2 + \gamma_{NT}), \quad (\text{A.41})$$

both of which imply

$$\frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) \mathbf{X}_{it,\mathbf{x}} \mathbf{X}_{it,\mathbf{x}}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{x}) = \widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) + O_P(h^2 + \gamma_{NT}). \quad (\text{A.42})$$

Meanwhile, by Lemma A.2 and (A.30), we have

$$\begin{aligned} b_{i,\mathbf{x}} &= g'_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + h^{-1} \Xi_{i,1}^\theta(\mathbf{x}) + O\left(h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2\right] h^{-1}\right) \\ &= g'_i(\boldsymbol{\theta}_0^\top \mathbf{x}) + O\left(h^2 + \gamma_{NT} h^{-1} + \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 h^{-1}\right) \end{aligned} \quad (\text{A.43})$$

uniformly in $\mathbf{x} \in \mathcal{X}_{NT}$. It follows from (A.42) and (A.43) that

$$\begin{aligned}
& \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is}^2 \mathbf{X}_{its} \mathbf{X}_{its}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
= & \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \mathbf{X}_{its} \mathbf{X}_{its}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
= & \frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \left[\frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) \mathbf{X}_{its} \mathbf{X}_{its}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] \\
& + O_P(h^2 + \gamma_{NT} h^{-1} + \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 h^{-1}) \\
= & \frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) + O_P(h^2 + \gamma_{NT} h^{-1} + \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 h^{-1}) \\
= & \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right\} \\
& + O_P(h^2 + \gamma_{NT} h^{-1} + \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 h^{-1} + (NT)^{-1/2}). \tag{A.44}
\end{aligned}$$

Meanwhile, note that

$$\begin{aligned}
& \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right\} \\
= & \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \mathbb{E} \left[\widetilde{\omega}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \mid \boldsymbol{\theta}_0^\top \mathbf{X}_{is} \right] \right\} \\
= & 2 \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \left[\mathbb{E}(\mathbf{X}_{is} \mathbf{X}_{is}^\top \mid \boldsymbol{\theta}_0^\top \mathbf{X}_{is}) - \boldsymbol{\mu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\mu}_{\boldsymbol{\theta}_0,i}^\top(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right] \right\} \\
= & 2 \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \mathbb{E} \left[\boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}^\top(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \mid \boldsymbol{\theta}_0^\top \mathbf{X}_{is} \right] \right\} \\
= & 2 \mathbb{E} \left\{ \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta}_0,i}^\top(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right\} \\
= & 2D_{\boldsymbol{\theta}_0,i},
\end{aligned}$$

which, in conjunction with (A.44), implies (A.38).

We next turn to the proof of (A.39). Observe that by (A.21) and Lemma A.2, we have

$$\begin{aligned}
& Y_{it} - a_{i,\mathbf{x}} - b_{i,\mathbf{x}} \boldsymbol{\theta}_0^\top \mathbf{X}_{it,\mathbf{x}} \tag{A.45} \\
= & \varepsilon_{it} + g_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) - g_i(\boldsymbol{\theta}_0^\top \mathbf{x}) - g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) - \frac{1}{2} g''_i(\boldsymbol{\theta}_0^\top \mathbf{x}) h^2 \\
& - g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,\mathbf{x}}) - \Xi_{i,0}^\theta(\mathbf{x}) - \Xi_{i,1}^\theta(\mathbf{x})(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,\mathbf{x}}) h^{-1} \\
& + O(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \|\boldsymbol{\delta}_\theta\|^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\|) \\
= & \varepsilon_{it} + \frac{1}{2} g''_i(\boldsymbol{\theta}_0^\top \mathbf{x}) \left[(\boldsymbol{\theta}_0^\top \mathbf{X}_{it,\mathbf{x}})^2 - h^2 \right] - g'_i(\boldsymbol{\theta}_0^\top \mathbf{x})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})^\top \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{x}) - \Xi_{i,0}^\theta(\mathbf{x}) \\
& - \Xi_{i,1}^\theta(\mathbf{x})(\boldsymbol{\theta}^\top \mathbf{X}_{it,\mathbf{x}}) h^{-1} + O(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \|\boldsymbol{\delta}_\theta\|^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\|)
\end{aligned}$$

uniformly in $\mathbf{x} \in \mathcal{X}_{NT}$. It then follows that

$$\begin{aligned}
& \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} (Y_{it} - a_{is} - b_{is} \boldsymbol{\theta}_0^\top \mathbf{X}_{its}) \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
= & \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} \varepsilon_{it} \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
& + \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}^\top(\boldsymbol{\theta}^\top \mathbf{X}_{is}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
& + \frac{1}{2T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} g''_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) [(\boldsymbol{\theta}_0^\top \mathbf{X}_{its})^2 - h^2] \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\
& - \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \Xi_{i,0}^\theta(\mathbf{X}_{is}) \\
& - \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} \left(\frac{\boldsymbol{\theta}^\top \mathbf{X}_{its}}{h} \right) \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \Xi_{i,1}^\theta(\mathbf{X}_{is}) \\
& + O_P(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \|\boldsymbol{\delta}_\theta\|^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\|) \\
=: & \Pi_{N,T}^1 + \Pi_{N,T}^2 + \Pi_{N,T}^3 - \Pi_{N,T}^4 - \Pi_{N,T}^5 \tag{A.46} \\
& + O_P(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \|\boldsymbol{\delta}_\theta\|^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\|).
\end{aligned}$$

Define $E_{it}[G(\mathbf{X}_{it}, Y_{it}, \mathbf{X}_{is}, Y_{is})] = E[G(\mathbf{X}_{it}, Y_{it}, \mathbf{X}_{is}, Y_{is}) | \mathbf{X}_{it}, Y_{it}]$. Then, we have

$$\begin{aligned}
\Pi_{N,T}^1 &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \mathbf{X}_{its} \varepsilon_{it} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \tag{A.47} \\
&+ O_P(\gamma_{NT} \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}) \\
&= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \varepsilon_{it} \\
&+ O_P(\gamma_{NT} \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}) \\
=: & \Pi_{N,T}^{1,1} + \Pi_{N,T}^{1,2} + O_P(\gamma_{NT} \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT})\|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}),
\end{aligned}$$

where

$$\begin{aligned}
\Pi_{N,T}^{1,1} &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \left\{ K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right. \\
&\quad \left. - E_{it} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] \right\} \varepsilon_{it}
\end{aligned}$$

and

$$\Pi_{N,T}^{1,2} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left\{ E_{it} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] \right\} \varepsilon_{it}$$

In a similar way to the proof of Lemma 6.7 in Xia (2006), we have

$$\Pi_{N,T}^{1,1} = O_P(\gamma_{NT}^2). \quad (\text{A.48})$$

Furthermore, by noting that

$$\begin{aligned} & \mathbb{E}_{it} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] \\ &= \mathbb{E} \left\{ \mathbb{E} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mathbf{X}_{its} f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mid \mathbf{X}_{it}, \boldsymbol{\theta}^\top \mathbf{X}_{is} \right] \mid \mathbf{X}_{it} \right\} \\ &= \mathbb{E} \left\{ K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \mid \mathbf{X}_{it} \right\} \\ &= \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{it}) \right] g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{it}) + O_P(h^2), \end{aligned}$$

we have

$$\Pi_{N,T}^{1,2} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{it}) \right] g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{it}) \varepsilon_{it} + O_P((NT)^{-1/2} h^2). \quad (\text{A.49})$$

It follows from (A.47)–(A.49) that

$$\begin{aligned} \Pi_{N,T}^1 &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{it}) \right] g'_i(\boldsymbol{\theta}^\top \mathbf{X}_{it}) \varepsilon_{it} + O_P((NT)^{-1/2} h^2) \\ &\quad + O_P(\gamma_{NT} [h^2 + \gamma_{NT} + \gamma_{NT}^2/h + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\|/h + \|\boldsymbol{\delta}_\theta\|^2/h]) \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \right] g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \varepsilon_{it} + O_P((NT)^{-1/2} [h^2 + \|\boldsymbol{\delta}_\theta\|]) \\ &\quad + O_P(\gamma_{NT} \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}). \end{aligned} \quad (\text{A.50})$$

Similarly, we have

$$\begin{aligned} \Pi_{N,T}^2 &= \frac{1}{T^2 N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \mathbf{X}_{its} \boldsymbol{\nu}_{\boldsymbol{\theta},i}^\top(\boldsymbol{\theta}^\top \mathbf{X}_{is}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \\ &\quad + O_P(\|\boldsymbol{\delta}_\theta\| \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}) \\ &= \frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T \left[g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \right]^2 \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}^\top(\boldsymbol{\theta}^\top \mathbf{X}_{is}) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &\quad + O_P(\|\boldsymbol{\delta}_\theta\| \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}) \\ &= \frac{1}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + O_P(\|\boldsymbol{\delta}_\theta\| \{h^2 + \gamma_{NT} + [\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2] h^{-1}\}) \end{aligned} \quad (\text{A.51})$$

and

$$\begin{aligned}
& \Pi_{N,T}^3 \\
= & \frac{1}{2T^2N} \sum_{i=1}^N \sum_{s=1}^T \sum_{t=1}^T g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) g''_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) \mathbf{X}_{its} \left[(\boldsymbol{\theta}_0^\top \mathbf{X}_{its})^2 - h^2 \right] \\
& + O_P \left(h^2 \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right) \\
= & O_P \left(h^2 \left(h^2 + \gamma_{NT} + \gamma_{NT}^2 h^{-1} \right) + \|\boldsymbol{\delta}_\theta\|^2 + \|\boldsymbol{\delta}_\theta\| h \right). \tag{A.52}
\end{aligned}$$

The last equality in (A.52) holds because

$$\frac{1}{T} \sum_{t=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) \mathbf{X}_{its} \left[(\boldsymbol{\theta}^\top \mathbf{X}_{its})^2 - h^2 \right] = O_P \left(h^2 (h^2 + \gamma_{NT}) \right).$$

Furthermore, as $E_{it} \left[K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right] = \mathbf{0}$, by (A.26), Lemma A.2 and similar arguments to those in the proof of Lemma 6.7 in Xia (2006), we have

$$\begin{aligned}
\Pi_{N,T}^4 &= \frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \Xi_{i,0}^\theta(\mathbf{X}_{is}) f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) Q_{i,0}^\theta(\mathbf{X}_{is}) \\
&+ O_P \left(\gamma_{NT} \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right) \\
&= \frac{1}{TN} \sum_{i=1}^N \sum_{s=1}^T g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \Xi_{i,0}^\theta(\mathbf{X}_{is}) \\
&+ O_P \left(\gamma_{NT} \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right) \\
&= \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T^2} \sum_{s=1}^T \sum_{l=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{ils}) g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{is}) \boldsymbol{\nu}_{\boldsymbol{\theta},i}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) f_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \varepsilon_{il} \right] \\
&+ O_P \left(\gamma_{NT} \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right) \\
&= O_P \left(\gamma_{NT} \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right). \tag{A.53}
\end{aligned}$$

Analogously, we have

$$\Pi_{N,T}^5 = O_P \left(\gamma_{NT} \left\{ h^2 + \gamma_{NT} + \left[\gamma_{NT}^2 + (h + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| + \|\boldsymbol{\delta}_\theta\|^2 \right] h^{-1} \right\} \right). \tag{A.54}$$

Equation (A.39) thus follows from (A.46) and (A.50)–(A.54).

APPENDIX B: PROOFS OF MAIN RESULTS

We now provide the detailed proof of Theorem 3.1.

Proof of Theorem 3.1

Define

$$S_{NT} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}_0,i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \right] g'_i(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \varepsilon_{it}.$$

Let $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}_0$ be a \sqrt{NT} -consistent initial estimator of $\boldsymbol{\theta}_0$ and $\widetilde{\boldsymbol{\theta}}^{(k)}$, $k \geq 1$, be the estimator of $\boldsymbol{\theta}_0$ after k iterations. Then, from (2.4), we have

$$\begin{aligned} & \widetilde{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0 \\ = & \left[\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is}^2 \mathbf{X}_{its} \mathbf{X}_{its}^\top \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right]^{-1} \\ & \times \left[\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T K_h(\boldsymbol{\theta}^\top \mathbf{X}_{its}) b_{is} \mathbf{X}_{its} (Y_{it} - a_{is} - b_{is} \boldsymbol{\theta}_0^\top \mathbf{X}_{its}) \widehat{f}_{\boldsymbol{\theta},i}^{-1}(\boldsymbol{\theta}^\top \mathbf{X}_{is}) \right]. \end{aligned}$$

This, together with Lemma A.3, implies

$$\begin{aligned} & \widetilde{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0 \\ = & \frac{1}{2} \left\{ \frac{1}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0, i} \right\}^+ S_{NT} + \frac{1}{2} \left\{ \frac{1}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0, i} \right\}^+ \left\{ \frac{1}{N} \sum_{i=1}^N D_{\boldsymbol{\theta}_0, i} \right\} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ & + O_P(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \gamma_{NT}^3 h^{-1} + (h^2 + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| h^{-1} + \|\boldsymbol{\delta}_\theta\|^2 + \|\boldsymbol{\delta}_\theta\|^3 h^{-1}) \\ = & \frac{1}{2} D_{\boldsymbol{\theta}_0}^+ S_{NT} + \frac{1}{2} D_{\boldsymbol{\theta}_0}^+ D_{\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_P((NT)^{-1/2} + \|\boldsymbol{\delta}_\theta\|) \\ & + O_P(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \gamma_{NT}^3 h^{-1} + (h^2 + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| h^{-1} + \|\boldsymbol{\delta}_\theta\|^2 + \|\boldsymbol{\delta}_\theta\|^3 h^{-1}). \end{aligned} \quad (\text{B.1})$$

By A4 and recursing the above equation $k - 1$ times, we get

$$\begin{aligned} & \widetilde{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0 \\ = & \left\{ \sum_{l=1}^k \frac{1}{2^l} \right\} D_{\boldsymbol{\theta}_0}^+ S_{NT} + \frac{1}{2^k} D_{\boldsymbol{\theta}_0}^+ D_{\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_P((NT)^{-1/2} + \|\boldsymbol{\delta}_\theta\|) \\ & + O_P(h(h^2 + \gamma_{NT}) + \gamma_{NT}^2 + \gamma_{NT}^3 h^{-1} + (h^2 + \gamma_{NT}) \|\boldsymbol{\delta}_\theta\| h^{-1} + \|\boldsymbol{\delta}_\theta\|^2 + \|\boldsymbol{\delta}_\theta\|^3 h^{-1}) \\ = & \left\{ \sum_{l=1}^k \frac{1}{2^l} \right\} D_{\boldsymbol{\theta}_0}^+ S_{NT} + \frac{1}{2^k} D_{\boldsymbol{\theta}_0}^+ D_{\boldsymbol{\theta}_0} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_P((NT)^{-1/2}). \end{aligned} \quad (\text{B.2})$$

Letting $k \rightarrow \infty$ in (B.2), we have

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = D_{\boldsymbol{\theta}_0}^+ S_{NT} + o_P((NT)^{-1/2}). \quad (\text{B.3})$$

We next prove the joint central limit theorem for $\sqrt{NT} S_{NT}$. Let

$$B_{iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^T (\mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}_0, i}(\boldsymbol{\theta}_0^\top \mathbf{X}_{it})) g_i'(\boldsymbol{\theta}_0^\top \mathbf{X}_{it}) \varepsilon_{it} =: \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_{it}.$$

Then, it is obvious that

$$\sqrt{NT} S_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{W}_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{W}_{it} \right] = \frac{1}{\sqrt{N}} \sum_{i=1}^N B_{iT}. \quad (\text{B.4})$$

To prove the joint asymptotic normality of $\sqrt{NT}S_{NT}$, we use the same argument as that used in the proof of Theorem 2 in Phillips and Moon (1999). It follows from **A2(i)** that $\{B_{iT}, 1 \leq i \leq N\}$ is an independent sequence and

$$E(B_{iT}B_{iT}^\top) = E(\mathbf{W}_{i1}\mathbf{W}_{i1}^\top) + 2 \sum_{t=2}^T [1 - (t-1)/T] E(\mathbf{W}_{i1}\mathbf{W}_{it}^\top) = \Lambda_{i,T} < \infty.$$

Thus, to prove the joint asymptotic normality of $\sqrt{NT}S_{NT}$, it suffices to verify the Lindeberg condition, i.e. we need only to show that

$$\frac{1}{N} \sum_{i=1}^N E[\|B_{iT}\|^2 I\{\|B_{iT}\| > \sqrt{N}\epsilon\}] \rightarrow 0 \quad (\text{B.5})$$

for any $\epsilon > 0$. Equation (B.5) follows immediately from (3.3). It then follows from (3.1) and (B.5) that

$$\sqrt{NT}S_{NT} \xrightarrow{d} N(\mathbf{0}, \Sigma_{\theta_0}). \quad (\text{B.6})$$

From (B.3) and (B.6), it is easily seen that (3.4) holds.

REFERENCES

- Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics* 93:203-228.
- An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6:943-956.
- Arellano, M. (2003). *Panel data Econometrics*. Oxford University Press: New York.
- Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* 77, 669-687.
- Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. New York: John Wiley & Sons.
- Baltagi, B. H., Griffin, J. M. and Xiong, W. (2000). To pool or not to pool: Homogenous versus heterogenous estimators applied to cigarette demand. *Review of Economics and Statistics* 82:117-126.
- Cai, Z. and Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* 24:1321-1342.
- Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92:477-489.
- Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88:298-308.

- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman & Hall.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.
- Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. London: Chapman & Hall/CRC.
- Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics* 21:157–178.
- Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by method of average derivatives. *Journal of the American Statistical Association* 84:986–995.
- Henderson, D., Carroll, R. and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144:257–275.
- Hjellvik, V., Chen, R. and Tjøstheim, D. (2004). Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis* 25:831–872.
- Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association* 91:1632–1640.
- Hsiao, C. (2003). *Analysis of Panel Data*. New York: Cambridge University Press.
- Li, Q. and Racine, J. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.
- Li, Q. and Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics* 71:389–397.
- Lu, Z. (1998). On the geometric ergodicity of a nonlinear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* 8:1205–1217.
- Mammen, E., Støve, B. and Tjøstheim, D. (2009). Nonparametric additive models for panels of time series. *Econometric Theory* 25:442–481.
- Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH models. *Econometric Theory* 11:258–289.
- Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67:1057–1111.
- Ruppert, D., Sheather, J. and Wand, P. M. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90:1257–1270.
- Ullah, A. and Roy, N. (1998). Nonparametric and semiparametric econometrics of panel data. In: Ullah, A., Giles, D.E.A. (Eds.), *Handbook of Applied Economics Statistics*, vol. 1. New York: Marcel Dekker, pp. 579–604.

- Xia, Y. C. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory* 22:1112–1137.
- Xia, Y. C. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* 97:1162–1184.
- Xia, Y. C., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society Series B* 64:363–410.
- Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* 97:1042–1054.