The University of Adelaide
School of Economics

Research Paper No. 2010-09
May 2010

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

Jia Chen, Jiti Gao, and Degui Li

Estimation in Single–Index Panel Data Models with Heterogeneous Link Functions

By Jia Chen, Jiti Gao and Degui Li

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

Abstract

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 so–called refined minimum average variance estimation based on a local linear smoothing method to estimate both the parameters in the single–index and the average link function. As the cross–section dimension N and the time series dimension T tend to infinity simultaneously, we establish asymptotic distributions for the proposed parametric and nonparametric estimates. In addition, we provide two real–data examples to illustrate the finite sample behavior of the proposed estimation method in this paper.

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 exists a huge literature on parametric linear and nonlinear panel data modeling as the double–index models enable researchers to extract information that may be difficult to obtain through purely cross–section or time–series data models. We refer to the books by Baltagi (1995), Arellano (2003) and Hsiao (2003) for an overview of statistical inference and econometric analysis of parametric panel data models. As in both the 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, see Li and Stengos (1996), Ullah and Roy (1998), Abrevaya (1999), Hjellvik, Chen and Tjøstheim (2004), Cai and Li (2008), Henderson, Carroll and Li (2008) and Mammen, Støve and Tjøstheim (2009) for example.

There is a growing interest in using single–index models in both the cross–sectional and time series cases (see, for example, Härdle, Hall and Ichimura 1993; Carroll et al. 1997; Xia et al. 2002; Yu and Ruppert 2002; Xia 2006; Gao 2007). So far as we know, however, there is little study in the theoretical and empirical analysis of single–index models for panel data. Single–index models search for a linear combination of the multi–dimensional covariate $\{X_{it}\}$ which can capture most information about the relationship between the response variable $\{Y_{it}\}$ and covariate $\{X_{it}\}$. For a real data, there may exist individual effects. For example, in the US cigarette demand data set given in Section 5, there are state-specific effects such as religion, race, education and tourism. To reflect the individual effects, we assume that the nonlinear link function $g(\cdot)$ varies across the individuals. The model we study in this paper is given as follows:

$$Y_{it} = g_i(\theta_0^\top X_{it}) + \varepsilon_{it}, \qquad 1 \le i \le N, \quad 1 \le t \le T, \tag{1.1}$$

where $g(\cdot)$ is an unknown link function and θ_0 is a $p \times 1$ vector of unknown parameters. For identifiability, we require $\|\theta_0\| = 1$ throughout the paper.

This paper is interested in the case that the cross–section dimension N and the time–series dimension T tend to infinity simultaneously. Model (1.1) is call a single–index panel data model with heterogeneous link functions and it is more flexible than a homogeneous single–index panel data model. In this paper, we assume that

$\{X_{it}, \varepsilon_{it},\ t \geq 1\}$ 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 can be satisfied for some stationary time series and Markov chains under certain conditions. This means that we can apply model (1.1) to the dynamic panel data case, which will be discussed in Section 4.

In Section 2, we extend the so–called refined minimum average (conditional) variance estimation (RMAVE) method for the time series case to estimate the parameter θ_0 in model (1.1). The RMAVE was introduced by Xia et al. (2002) and its asymptotic distribution was established by Xia (2006) in the time series case. As there are two indices involved in our case and the nonlinear link functions are heterogeneous, the establishment of our asymptotic theory is much more complicated than that for the time series case. We show, in Section 3, that under certain regularity conditions, the RMAVE of θ_0 is asymptotically normal with $\sqrt{NT}$ rate of convergence as $N, T \to \infty$ simultaneously. This is called the joint limiting distribution (see Phillips and Moon 1999 for detail). Meanwhile, since the link functions $g_i(\cdot)$ vary across the sections, it is reasonable to study a nonparametric estimate of the average link function of the form

$$\overline{g}(x) = \frac{1}{N}\sum_{i=1}^{N} g_i(x). \tag{1.2}$$

In Section 3, we also establish an asymptotic distribution for the local linear estimate of $\overline{g}(x)$.

When $\{X_{it}\}$ contains lagged values of Y_{it}, (1.1) becomes a dynamic panel data model. Section 4 discusses some conditions that ensure $\{Y_{it},\ t \geq 1\}$ to be a geometrically ergodic time series for each i. In this case, the stationarity and mixing conditions and thus the asymptotic properties in Section 3 still hold for such a dynamic model. We include two empirical examples in Section 5 to illustrates the applicability of the proposed models and estimation method. One is the US cigarette demand data for 46 states from 1963 to 1992, to which we fit a single index model whose covariates X_{it} contain a lagged value of Y_{it}. We compare our RMAVE results with the ordinary least squares (OLS) estimation results for a linear panel data model from Baltagi, Griffin and Xiong (2000), and find that our estimated covariate coefficients are more significant than the OLS estimates. We then discuss an empirical application to a climatic date set from the UK by examining the relationship between the monthly average maximum temperatures and the number of millimeters of rainfall and hours of sunshine duration. The heterogenous link functions used allow us to take into account the state or station specific effects.

The rest of the paper is organized as follows. In Section 2, we develop the detailed algorithm of a RMAVE method. Section 3 establishes the asymptotic theory for both the parameter estimator and nonparametric estimate. Section 4 discusses the 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. Sections 5 includes a brief discuss on the bandwidth selection problem and two real data exmples. Section 6 concludes this paper. Some technical lemmas and the detailed proofs of the main results are given in Appendices A and B.

2. Semiparametric estimation method

In this section, we develop a RMAVE method to estimate both the parameter θ_0 in the single–index and the averaged link function defined in Section one. As the link functions are heterogeneous, the RMAVE method originally studied in Xia (2006) for the time series case will need to be extended substantially to deal with our case.

Given $\theta^\top X_{it} = u$, define

$$\sigma^2_{\theta,i}(u) = E\left[\left(Y_{it} - g_i(\theta^\top X_{it})\right)^2 | \theta^\top X_{it} = u\right] \tag{2.1}$$

for $1 \leq i \leq N$. Note that

$$E\left(Y_{it} - g_i(\theta^\top X_{it})\right)^2 = E_u\left[\sigma^2_{\theta,i}(u)\right]. \tag{2.2}$$

Based on (2.1) and (2.2), the estimator of θ_0 can be obtained by minimizing

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

As the link functions $g_i(\cdot)$ are unknown for the single–index panel data case, we estimate them by the local linear method. It is well–known that the local linear fitting has advantages over the Nadaraya–Watson kernel method, such as high asymptotic efficiency, design adaption and automatic boundary correction (see Fan and Gijbels 1996 for example). For X_{it} close to the point x, by Taylor expansion, we have

$$Y_{it} - g_i(\theta^\top X_{it}) = Y_{it} - g_i(\theta^\top x) - g_i'(\theta^\top x)\theta^\top (X_{it} - x).$$

Let $\widehat{\theta}_1$ be an initial estimator of θ_0. Based on the above local linear approximation, we describe the detailed algorithm as follows.

Step 1. Let $\theta = \widehat{\theta}_1$. Calculate

$$\begin{pmatrix} a_{is} \\ b_{is} \end{pmatrix} = \left\{ \frac{1}{T} \sum_{t=1}^{T} K_h(\theta^\top X_{its}) \begin{pmatrix} 1 \\ \theta^\top X_{its} \end{pmatrix} \begin{pmatrix} 1 \\ \theta^\top X_{its} \end{pmatrix}^\top \right\}^{-1} \times \left\{ \frac{1}{T} \sum_{t=1}^{T} K_h(\theta^\top X_{its}) \begin{pmatrix} 1 \\ \theta^\top X_{its} \end{pmatrix} Y_{it} \right\}, \tag{2.3}$$

where h is a bandwidth, $K(\cdot)$ is a kernel function, $K_h(\cdot) = \frac{1}{h} K(\cdot/h)$, and $X_{its} = X_{it} - X_{is}$.

Step 2. Obtain

$$\widetilde{\theta} = \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} K_h(\theta^\top X_{its}) b_{is}^2 X_{its} X_{its}^\top / \widehat{f}_i^\theta(\theta^\top X_{is}) \right\}^+ \times \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} K_h(\theta^\top X_{its}) b_{is} X_{its} (Y_{it} - a_{is}) / \widehat{f}_i^\theta(\theta^\top X_{is}) \right\}, \tag{2.4}$$

where $\widehat{f}_{\theta,i}(\theta^\top X_{is}) = \frac{1}{T} \sum_{t=1}^{T} K_h(\theta^\top X_{its})$, $\theta = \widehat{\theta}_1$, and A^+ stands for the pseudoinverse of A.

Step 3. Update θ with $\theta = \widetilde{\theta}/\|\widetilde{\theta}\|$. Repeat Step 1 and Step 2 until convergence.

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

Let $\widehat{g}_i(x) = a_{i,x}$, where $a_{i,x}$ is defined as a_{is} in (2.3) with θ and X_{is} replaced by $\widehat{\theta}$ and x, respectively. As in Hjellvik, Chen and Tjøstheim (2004), the nonparametric estimate of $\overline{g}(x)$ is defined as

$$\widehat{g}(x) = \frac{1}{N} \sum_{i=1}^{N} \widehat{g}_i(x).$$

An asymptotic distribution of $\widehat{g}(x)$, as $T, N \to \infty$ simultaneously, is established in Section 3 below.

3. Asymptotic theory

In this section, we establish asymptotic distributions for $\widehat{\theta}$ and $\widehat{g}(\cdot)$. Before giving some regularity assumptions, we introduce following notation. Define $\mu_{\theta,i}(u) = E(X_{it}|\theta^\top X_{it} = u)$ and $\nu_{\theta,i}(x) = \mu_i^\theta(\theta^\top x) - x$. We then introduce the following assumptions.

A1. $K(\cdot)$ is a symmetric and continuous kernel function with some bounded support, and its derivative is bounded. Furthermore, $\int K(u)du = 1$.

A2 (i). Let $X_i = \{X_{it},\ t \geq 1\}$ and $\varepsilon_i = \{\varepsilon_{it},\ t \geq 1\}$. Suppose that $\{X_i, \varepsilon_i\}$, $i \geq 1$, are independent.

(ii). For each i, $\{(X_{it}, \varepsilon_{it}),\ t \geq 1\}$ is a stationary sequence of α–mixing random vectors with $E(\varepsilon_{it}|\theta^\top X_{it}) = 0$, $\max_i E\left[|\varepsilon_{it}|^{2+\delta}\right] < \infty$, $\max_i E\left[\|X_{it}\|^{2+\delta}\right] < \infty$ and mixing coefficient $\alpha_i(\cdot)$ satisfying $\max_i \alpha_i(t) = O(t^{-\kappa})$ for $\kappa > \frac{(2+\delta)}{\delta}$.

A3 (i). Let $f_{\theta,i}(\cdot)$ be the density function of $\{\theta^\top X_{it},\ \ t \geq 1\}$. Suppose that $f_{\theta,i}(\cdot)$ is continuous and its derivatives of up to the third order are bounded. Uniformly for θ in a neighborhood of θ_0,

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

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). $\mu_{\theta,i}(\cdot)$ is continuous and has bounded derivatives of up to the second order.

A4. The bandwidth h satisfies $NTh \to \infty$, $NTh^6 \to 0$, $\alpha_{T,h}/h \to 0$, $NT\alpha_{T,h}^2 h^2 \to 0$, $\alpha_{T,h}(NT)^{1/(2+\delta)} \to 0$ and $(NT)^{1+(p+\kappa+2)/(2+\delta)}\alpha_{T,h}^{\kappa-p} h^{-1-p} \to 0$, where $\alpha_{T,h} = \sqrt{\frac{\log T}{Th}}$ and κ is as defined in A2(ii) above.

Remark 3.1. A1 is a set of some mild conditions on the kernel function, which have been used by many authors in the time series case (see Fan and Yao 2003; Gao 2007 for example). In A2, we assume that (X_i, ε_i), $i \geq 1$, are cross–sectional independence (see Cai and Li 2008 for example) and each time series is α–mixing, which 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 is about some commonly–used conditions in single–index models (see Xia 2006 for example). In A4, the condition $\alpha_{T,h}/h \to 0$ implies $Th^3 \to \infty$. On the other hand, $NT\alpha_{T,h}^2 h^2 \to 0$ implies $Nh \to 0$. 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 due to the fact that we use individual time series to estimate the individual–specific link functions $g_i(\cdot)$ $(1 \le i \le N)$ and use the pooled data to estimate the index parameter θ_0.

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

Before stating an asymptotic distribution for $\widehat{\theta}$ defined in Section 2, we introduce some notation. Let $W_{it} = \left(X_{it} - \mu_{\theta_0,i}(\theta_0^\top X_{it})\right) g_i'(\theta_0^\top X_{it})\varepsilon_{it}$. By A2 (ii), we know that for each i,

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

Let

$$D_{\theta_0,i} = E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2 \nu_{\theta_0,i}(X_{is})\nu_{\theta_0,i}^\top(X_{is})\right].$$

In order to establish the asymptotic normality of $\widehat{\theta}$, we need to assume that there is an initial estimator $\widehat{\theta}_1$ such that $\|\widehat{\theta}_1 - \theta_0\| = O_P\left((NT)^{-1/2}\right)$. The proof of Theorem 3.1 below is 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*

$$\frac{1}{N}\sum_{i=1}^N \Lambda_{i,T} \to \Sigma_{\theta_0} \tag{3.1}$$

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

$$\frac{1}{N}\sum_{i=1}^N D_{\theta_0,i} \to D_{\theta_0} \quad \text{as } N \to \infty. \tag{3.2}$$

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

$$\frac{1}{N}\left(\max_{1\le i\le N}\Lambda_{i,T}\right) \to 0. \tag{3.3}$$

If the initial estimator $\widehat{\theta}_1$ is $\sqrt{NT}$–consistent, then we have

$$\sqrt{NT}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} N(\mathbf{0}, D_{\theta_0}^{+}\Sigma_{\theta_0}D_{\theta_0}^{+\top}) \tag{3.4}$$

as $N, T \to \infty$ simultaneously, where $D_{\theta_0}^{+}$ is the pseudoinverse of D_{θ_0}.

Remark 3.2. The above theorem shows that the estimator $\widehat{\theta}$ is asymptotically normal with $\sqrt{NT}$ rate of convergence even when the link functions may be heterogeneous. Equations (3.1) and (3.3) are imposed to make sure that the Lindeberg condition holds when we prove 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 one–index case (see Härdle, Hall and Ichimura 1993 and Carroll *et al* 1997 for example). As a matter of the fact, this restriction 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).

Let $b_{g,i}(u) = \frac{1}{2}\mu_2 g_i''(u)$ and $\sigma_i^2(u) = \nu_0\sigma_{\theta_0,i}^2(u)/f_{\theta_0,i}(u)$, where $\mu_k = \int u^k K(u)du$, $\nu_k = \int u^k K^2(u)du$ and $\sigma_{\theta_0,i}^2(u) = E\left(\varepsilon_{it}^2|\theta_0^\top X_{it} = u\right)$. We next establish an asymptotic distribution for $\widehat{g}(x)$ in the following theorem; its proof is given in Appendix B.

Theorem 3.2. *Assume that the conditions of Theorem 3.1 are satisfied. If, in addition, $NTh^4 \to \infty$,*

$$\frac{1}{N}\sum_{i=1}^{N} b_{g,i}(u) \to b_g(u), \quad \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2(u) \to \sigma_g^2(u) \quad \text{as } N \to \infty, \tag{3.5}$$

and

$$\max_{1\le i\le N}\sigma_i^2(u) = o(N),$$

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

$$\sqrt{NTh}\left(\widehat{g}(u) - \overline{g}(u) - b_g(u)h^2\right) \xrightarrow{d} N\left(0, \sigma_g^2(u)\right). \tag{3.6}$$

Remark 3.3. Note that Theorem 3.2 covers the case that $g(\cdot)$ can be consistently estimated by $\widehat{g}(\cdot)$ when model (1.1) reduces to the case where the link functions are all homogeneous (i.e., $g_i(\cdot) \equiv g(\cdot)$).

4. Dynamic single-index panel data models

We next consider the case that X_{it} contains lagged values of Y_{it}. If $X_{it} = \widetilde{Y}_{i,t-1} = (Y_{i,t-1}, \cdots, Y_{i,t-p})^\top$, then model (1.1) becomes

$$Y_{it} = g_i\left(\theta_0^\top \widetilde{Y}_{i,t-1}\right) + \varepsilon_{it}, \tag{4.1}$$

where, for each i, $\{\epsilon_{it},\ t \geq 1\}$ is a sequence of i.i.d. random variables and ε_{it} is 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 $i \geq 1$. This implies that $\{Y_{it},\ t \geq 1\}$ satisfies the stationarity and 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 $\phi_i(x_1, \cdots, x_p) = g_i(\theta^\top x)$ with $x = (x_1, \cdots, x_p)^\top$.*
(i). Suppose that

$$\sup_{\|x\| \leq C} |\phi_i(x)| < \infty \quad \text{for any } C > 0 \text{ and } i \geq 1 \tag{4.2}$$

$$\lim_{\|x\| \to \infty} \frac{\left|\phi_i(x) - \alpha_i^\top x\right|}{\|x\|} = 0 \quad \text{for each } i \geq 1, \tag{4.3}$$

where $\alpha_i = (\alpha_{i,1}, \cdots, \alpha_{i,p})^\top$ satisfies

$$x^p - \alpha_{i,1} x^{p-1} - \cdots - \alpha_{i,p-1} x - \alpha_{i,p} \neq 0 \quad \text{for all } |x| > 1. \tag{4.4}$$

Then, $\{Y_{it},\ t \geq 1\}$ defined by (4.1) is geometrically ergodic for each $i \geq 1$.
(ii). Suppose that there exists a positive number $\lambda_i < 1$ and a constant C_i for each i, such that

$$|\phi_i(x)| \leq \lambda_i \max\{|x_1|, \cdots, |x_p|\} + C_i. \tag{4.5}$$

Then $\{Y_{it},\ t \geq 1\}$ defined by (4.1) is geometrically ergodic for each $i \geq 1$.

The detailed proof of Proposition 4.1 follows from the same arguments as used in An and Huang (1996). Similar results about geometrical ergodicity are available from Masry and Tjøstheim (1995), and Lu (1998).

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

Example 4.1. Let $\theta_0 = (\theta_{01}, \theta_{02})^\top = (0.6, 0.8)^\top$ and $g_i(u) = \frac{1}{\sqrt{2}}u + \sin(\frac{2\pi}{i}u)$. Then the dynamic panel data model (4.1) reduces to

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

As $\sin(\cdot)$ is a bounded function, by letting $\alpha_i = \left(\frac{0.6}{\sqrt{2}}, \frac{0.8}{\sqrt{2}}\right)^\top$, it is easy to show that (4.2) and (4.3) 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 the link functions $g_i(\cdot)$ satisfy

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

where κ_i and ρ_i are positive constants, $\rho_i < 1$, and p is the dimension of θ_0 in (4.1). Following the same arguments as used 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 $g_i(\cdot)$ satisfies

$$\lim_{|u|\to\infty} \frac{|g_i(u) - c_i^* u|}{|u|} = 0, \quad \text{for each } i \geq 1,$$

where c_i^* satisfies (4.4) in Proposition 4.1 (ii), then we also can show that $\{Y_{it},\ t \geq 1\}$ is geometrically ergodic for each $i \geq 1$.

5. Empirical examples

We give a brief discussion on the bandwidth selection and then give two real data examples to illustrate the proposed estimation method.

5.1. Bandwidth selection

Bandwidth selection is important for nonparametric estimation. Consider the estimate of $\overline{g}$ at the final step of the iterations. It follows from (3.6) that the asymptotic integrated mean squared error of $\widehat{g}(\cdot)$ is given by

$$h^4 \int b_g^2(u)du + \frac{\int \sigma_g^2(u)du}{NTh}$$

and an optimal global bandwidth is of the form

$$h_{opt} = \frac{\int \sigma_g^2(u)du}{4\int b_g^2(u)du}(NT)^{-1/5}. \tag{5.1}$$

Based on (5.1), we can use the plug–in method (see Ruppert, Sheather and Wand 1995 for detail) to choose an optimal bandwidth for the implementation of (5.1) in practice. In the real data application below, we instead propose using a semiparametric leave-one-out cross validation method to select the bandwidth.

Suppose that $\widehat{\theta}(h)$ is an estimate of θ_0 via the iterative procedure described in Section 2 with bandwidth h. For each $1 \leq i \leq N$ and $1 \leq t \leq T$, we calculate

$$a_{it}^{(-t)}(h) = \left\{ \sum_{s=1, s\neq t}^{T} K_h\left(\widehat{\theta}(h)^\top X_{its}\right) \right\}^{-1} \left\{ \sum_{s=1, s\neq t}^{T} K_h\left(\widehat{\theta}(h)^\top X_{its}\right) Y_{is} \right\}, \tag{5.2}$$

and let

$$CV(h) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - a_{it}^{(-t)}(h) \right)^2. \tag{5.3}$$

Then, we choose $\widehat{h} = \arg\min_h CV(h)$ as an optimal bandwidth in our implementation in the rest of this section.

5.2. Real data examples

Example 5.1. The first example is about the cigarettes demand in 46 American states over the period 1963–1992. The data set is from Baltagi, Griffin and Xiong (2000). The data set contains 7 variables: average retail price per pack of cigarettes, population, population above the age of 16, consumer price indices, real per capita disposable income, real per capita sales of cigarettes and minimum real price of cigarettes in any neighboring state.

As in Baltagi, Griffin and Xiong (2000) and Mammen, Støve and Tjøstheim (2009), we use only four variables to model cigarettes demand: real per capita sales of cigarettes (denoted as $Y_{i,t}$), average retail price per pack of cigarettes (denoted as $X_{i,t,2}$), real per capita disposable income (denoted as $X_{i,t,3}$) and minimum real price of cigarettes in any neighboring state (denoted as $X_{i,t,4}$). Denote $X_{i,t,1} = Y_{i,t-1}$, $i = 1, \cdots, 46$, $t = 1, \cdots, 29$. Baltagi, Griffin and Xiong (2000) modeled the data with the following log-linear dynamic demand model

$$\ln Y_{i,t} = \alpha + \beta_1 \ln X_{i,t,1} + \beta_2 \ln X_{i,t,2} + \beta_3 \ln X_{i,t,3} + \beta_4 \ln X_{i,t,4} + u_{i,t}, \tag{5.4}$$

where $u_{i,t} = \mu_i + \lambda_t + v_{i,t}$, μ_i denotes a state-specific effect, and λ_t denotes a year-specific effect. In this paper, we use a single–index panel data model with heterogeneous link functions. By allowing the link functions to vary across states, we can also incorporate state-specific effects such as religion, race, tourism, tax, and education into our model.

As all the four variables exhibit a time trend, we first remove the trend from the data. Similarly to Mammen, Støve and Tjøstheim (2009), we make the following transformation

$$\widetilde{Y}_{i,t} = \ln Y_{i,t} - s_Y(t) \quad \text{and} \quad \widetilde{X}_{i,t,l} = \ln X_{i,t,l} - s_{X_l}(t),\ l = 1, \cdots, 4,$$

where $s_Y(t)$ is the nonparametric estimator of the time trend in observations $\ln Y_{i,t}$, and $s_{X_l}(t)$ is the nonparametric estimator of the trend in observations $\ln X_{i,t,l}$. $s_Y(t)$ can also be seen as time-specific effects λ_t in model (5.4), which may include policy interventions, health warnings and so on. We then assume that

$$\widetilde{Y}_{i,t} = g_i(\theta^\top \widetilde{X}_{i,t}) + \varepsilon_{i,t}, \quad i = 1, \cdots, 46,\ t = 1, \cdots, 29, \tag{5.5}$$

where $\widetilde{X}_{i,t} = \left(\widetilde{X}_{i,t,1}, \widetilde{X}_{i,t,2}, \widetilde{X}_{i,t,3}, \widetilde{X}_{i,t,4}\right)^\top$, and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ is the vector of parameters to be estimated. We apply the RMAVE estimation method proposed in Section 2 to the transformed observations. For the initial estimator $\widehat{\theta}_1$ of θ, we use a normalized version of Baltagi, Griffin and Xiong (2000)'s OLS estimate of $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$ from model (5.4): $\widehat{\theta}_1 = \frac{\widehat{\beta}}{\|\widehat{\beta}\|} = (0.9765, -0.2029, 0.0456, 0.0575)^\top$.

Our semiparametric estimate of θ is then $\widehat{\theta} = (0.9171, -0.3478, 0.1764, 0.0823)^\top$. Comparison of our estimate with Baltagi, Griffin and Xiong (2000)'s OLS estimate sees a drop in the coefficient for the lagged consumption from 0.9765 to 0.9171, and increases in all the other covariate coefficients, especially in the coefficient for disposable income ($X_{i,t,3}$) which sees almost a threefold increase from 0.0456 to 0.1764. Mammen, Støve and Tjøstheim (2009) used a nonparametric additive model to fit the data and found a similar result: the nonparametric estimates of the elasticities for retail price, disposable income, and minimum price in any neighboring state ($X_{i,t,2}$, $X_{i,t,3}$, and $X_{i,t,4}$) are more significant than Baltagi, Griffin and Xiong (2000)'s OLS estimates.

To see whether the estimates of the link functions vary across states, we plotted the estimated link functions for the first two states in Figure 5.1. The figure shows that there does exist some level of difference between the nonparametric estimates of the link functions for the two states.

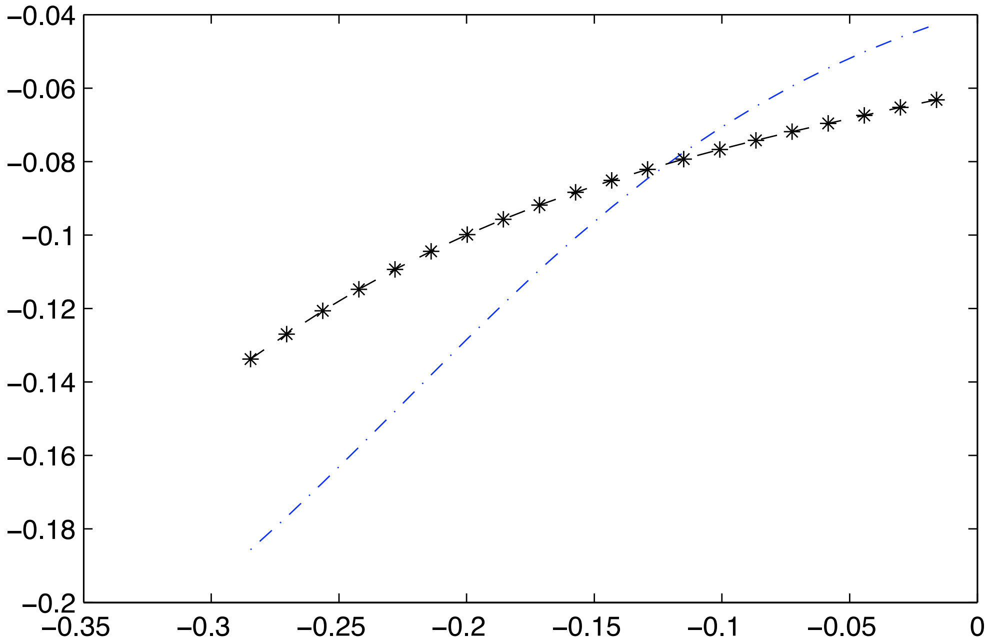

Figure 5.1. Estimated link functions for the first (dash-dotted line) and second (dash-starred line) states.

Example 5.2. The second data set, which is available from the UK Met Office website http://www.metoffice.gov.uk/climate/uk/stationdata/, contains monthly data of the average maximum temperature (TMAX), the average minimum temperature (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 and we focus on investigating the relationship between the TMAX and RAIN and SUN. For the i–th station, denote the seasonality and trend removed TMAX at time t as $Y_{i,t}$, and the seasonality and trend removed RAIN and SUN 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

$$Y_{i,t} = g_i(\theta^\top X_{i,t}) + \varepsilon_{i,t},\ \ i = 1, \cdots, 16,\ \ t = 1, \cdots, 120, \tag{5.6}$$

where $X_{it} = (X_{i,t,1}, X_{i,t,2})^\top$. We first use a least squares (LS) estimation method to

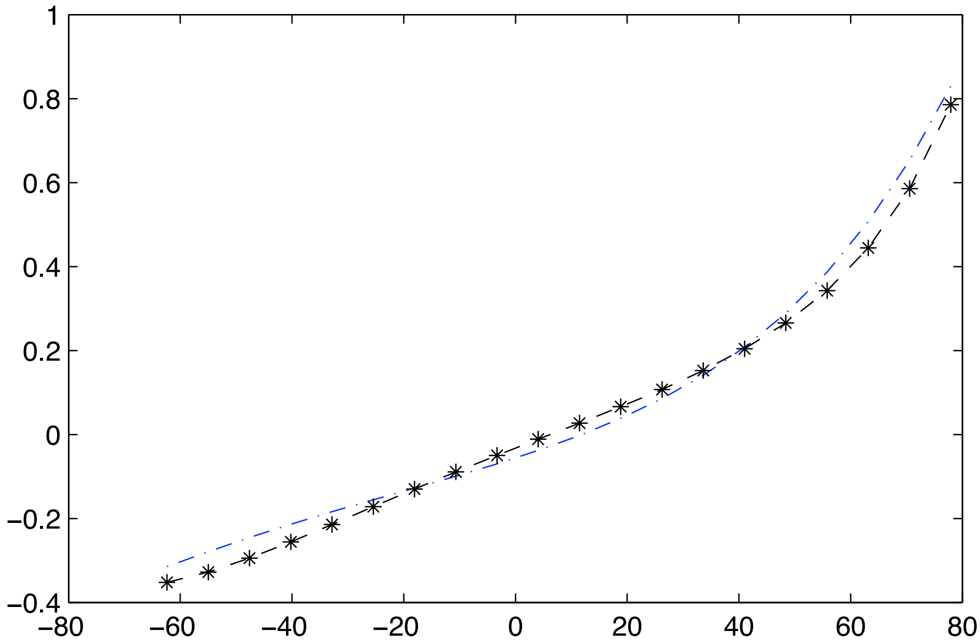

Figure 5.2. Estimated link functions for stations Armagh (dash-dotted line) and Bradford (dash-starred line).

estimate β in a linear model of the form

$$Y_{i,t} = X_{i,t}^{\top}\beta + \alpha_i + e_{i,t}, \tag{5.7}$$

where α_i are station-specific effects. Then, we use the normalized LS estimate $\frac{\widehat{\beta}}{\|\widehat{\beta}\|} = (0.1931, 0.9812)^{\top}$ as the initial estimate for θ in the RMAVE estimation of (5.6). The resulting RMAVE estimator of θ is $\widehat{\theta} = (0.1046, 0.9945)^{\top}$, which sees a drop from 0.1931 to 0.1046 in the coefficient for the rainfall covariate and a slight increase in the coefficient of sunshine.

As in Example 5.1, plots of the link functions for the first two stations are given in Figure 5.2. The figure shows the two estimated functions almost coincide which indicates that the difference between the two link functions is small.

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 and an average of the link functions. An asymptotically normal distribution has been established for each of the proposed estimates. In addition, we have included two real data examples to show how the proposed theory and estimation method is illustrated and implemented in practice.

The paper has some limitations and several extensions may be done. One of the topics is to establish some corresponding theory for the case where the residuals are cross–sectionally dependent. Another of the topics is whether the established theory may be extended to the case where $\{X_{it}\}$ is nonstationary in t and cross–sectionally dependent in i. Such topics may be discussed in future research.

7. Acknowledgments

The authors would like to acknowledge the financial support from the Australian Research Council Discovery Grants Program under Grant Number: DP0879088. Thanks also go to Dr Bård Støve for providing us with the data set used in Example 5.1.

Appendix A: Technical lemmas

Let $X_{it,x} = X_{it} - x$. We 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

$$\Theta_{NT} = \left\{\theta : \ |\theta - \theta_0| \le C_\theta (NT)^{-1/2}\right\},$$

$$\mathcal{X}_{NT} = \left\{x : \|x\| \le M(NT)^{1/(2+\delta)}\right\},$$

and

$$\mathcal{F}_{NT} = \{(x, \theta) : \ x \in \mathcal{X}_{NT}, \ \theta \in \Theta_{NT}\},$$

where C_θ and M are two positive constants. Define

$$Z_h(x, X_{it}) = K_h(\theta^\top X_{it,x}) \left(\frac{\theta^\top X_{it,x}}{h}\right)^k X_{it,x}$$

and

$$Z_h^*(x, X_{it}) = K_h(\theta^\top X_{it,x}) \left(\frac{\theta^\top X_{it,x}}{h}\right)^k .$$

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

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

we have

$$\max_{1\le i\le N}\sup_{x\in\mathcal{X}_{NT}}\left|\frac{1}{T}\sum_{t=1}^{T}Z_h^*(x,X_{it})-f_{\theta,i}(\theta^\top x)\mu_k-f'_{\theta,i}(\theta^\top x)\mu_{k+1}h\right|=O_P(h^2+\alpha_{T,h}) \quad (A.1)$$

and

$$\max_{1\le i\le N}\sup_{x\in\mathcal{X}_{NT}}\left|\frac{1}{T}\sum_{t=1}^{T}Z_h(x,X_{it})-f_{\theta,i}(\theta^\top x)\nu_{\theta,i}(x)\mu_k-\left[f_{\theta,i}(\theta^\top x)\mu_{\theta,i}(\theta^\top x)\right]'\mu_{k+1}h\right|$$
$$= O_P(h^2+\alpha_{T,h}), \quad (A.2)$$

where $f'_{\theta,i}(\theta^\top x)$ *is the derivative of* $f_{\theta,i}(\theta^\top x)$.

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

$$\max_{1\le i\le N}\sup_{x\in\mathcal{X}_{NT}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(Z_h(x,X_{it})-E\left[Z_h(x,X_{it})\right]\right)\right|=O_P(\alpha_{T,h}). \quad (A.3)$$

We first partition the set $\mathcal{X}_{NT}$ into $\mathcal{B}$ balls B_k, $1\le k\le\mathcal{B}$, each centered at x_k with radius $r=O(h\alpha_{T,h})$. By a simple calculation, we have

$$\mathcal{B}=O\left((NT)^{p/(2+\delta)}h^{-p}\alpha_{T,h}^{-p}\right).$$

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

$$\begin{aligned}
&\max_{1\le i\le N}\sup_{x\in\mathcal{X}_{NT}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(Z_h(x,X_{it})-E\left[Z_h(x,X_{it})\right]\right)\right|\\
\le\ &\max_{1\le k\le\mathcal{B}}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left(Z_h(x_k,X_{it})-E\left[Z_h(x_k,X_{it})\right]\right)\right|\\
+\ &\max_{1\le k\le\mathcal{B}}\sup_{x\in B_k}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\left[Z_h(x,X_{it})-Z_h(x_k,X_{it})\right]-E\left[Z_h(x,X_{it})-Z_h(x_k,X_{it})\right]\right)\right|\\
=:\ &\max_{1\le k\le\mathcal{B}}\max_{1\le i\le N}\left|H_{T,1}(k,i)\right|+\max_{1\le k\le\mathcal{B}}\sup_{x\in B_k}\max_{1\le i\le N}\left|H_{T,2}(k,i,x)\right|. \quad (A.4)
\end{aligned}$$

We first consider $\max_{1\le k\le\mathcal{B}}\max_{1\le i\le N}H_{T,1}(k,i)$. Let

$$\begin{aligned}
\overline{Z}_h(x_k,X_{it})&=Z_h(x_k,X_{it})I\left\{\|X_{it}\|\le\Delta_{NT}\right\},\\
Z_h^c(x_k,X_{it})&=Z_h(x_k,X_{it})-\overline{Z}_h(x_k,X_{it}),
\end{aligned}$$

where $\Delta_{NT}=(NT)^{1/(2+\delta)}L(NT)$, and $L(\cdot)$ is a positive slowly–varying function satisfying

$$\begin{aligned}
&L(NT)\to\infty,\ \ (NT)^{1/(2+\delta)}L(NT)\alpha_{T,h}\to 0,\\
&(NT)^{1+(p+\kappa+2)/(2+\delta)}\alpha_{T,h}^{\kappa-p}h^{-1-p}L^{\kappa+2}(NT)\to 0, \quad (A.5)
\end{aligned}$$

as $N,T\to\infty$.

It is easy to check that

$$H_{T,1}(k,i) = \frac{1}{T}\sum_{t=1}^{T}\left\{\overline{Z}_h(x_k,X_{it}) - E\left[\overline{Z}_h(x_k,X_{it})\right]\right\} + \frac{1}{T}\sum_{t=1}^{T}\left\{Z_h^c(x_k,X_{it}) - E\left[Z_h^c(x_k,X_{it})\right]\right\}.$$

By the first term in (A.5) and $E\left[\|X_{it}\|^{2+\delta}\right] < \infty$ in A2(ii), we can show that for any $\eta > 0$,

$$\begin{aligned} & P\left(\max_{1\le k\le \mathcal{B}}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{Z_h^c(x_k,X_{it}) - E\left[Z_h^c(x_k,X_{it})\right]\right\}\right| > \eta\alpha_{T,h}\right) \\ \le\ & \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{E\|X_{it}\|^{2+\delta}}{NTL^{2+\delta}(NT)} \le \frac{1}{L^{2+\delta}(NT)} = o(1), \end{aligned}$$

which implies that

$$\max_{1\le k\le \mathcal{B}}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{Z_h^c(x_k,X_{it}) - E\left[Z_h^c(x_k,X_{it})\right]\right\}\right| = O_P(\alpha_{T,h}). \tag{A.6}$$

Furthermore, by A1, A2(ii), A3(i) and the standard argument for the variance of α–mixing nonparametric kernel statistic, we have

$$\max_{1\le k\le \mathcal{B}}\max_{1\le i\le N}\operatorname{Var}\left(\sum_{t=1}^{T}\overline{Z}_h(x_k,X_{it})\right) = O\left(Th^{-1}\right). \tag{A.7}$$

Then, by Bernstein inequality for α–mixing processes (see Theorem 2.18 in Fan and Yao 2003 for example),

$$\begin{aligned} & P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\left\{\overline{Z}_h(x_k,X_{it}) - E\overline{Z}_h(x_k,X_{it})\right\}\right| > \eta\alpha_{T,h}\right) \\ =\ & P\left(\left|\sum_{t=1}^{T}\left\{\overline{Z}_h(x_k,X_{it}) - E\overline{Z}_h(x_k,X_{it})\right\}\right| > \eta T\alpha_{T,h}\right) \\ \le\ & 4\exp\left(-C\eta^2\log T\right) + CT\alpha_{Th}^{\kappa}\Delta_{NT}^{\kappa+2}h^{-1}, \end{aligned}$$

where C is some positive constant. Hence, as $NT\alpha_{Th}^{\kappa-p}h^{-p} = o(1)$,

$$\begin{aligned} & P\left(\max_{1\le k\le \mathcal{B}}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{Z_h(x_k,X_{it}) - EZ_h(x_k,X_{it})\right\}\right| > \eta\alpha_{T,h}\right) \\ \le\ & O\left(\mathcal{B}NT^{-C\eta^2} + \mathcal{B}NT\alpha_{Th}^{\kappa}\Delta_{NT}^{\kappa+2}h^{-1}\right) \\ =\ & O\left(\alpha_{T,h}^{-p}h^{-p}N^{1+p/(2+\delta)}T^{p/(2+\delta)-C\eta^2} + (NT)^{1+(p+\kappa+2)/(2+\delta)}\alpha_{Th}^{\kappa-p}h^{-1-p}L^{\kappa+2}(NT)\right) \\ =\ & o(1) \end{aligned}$$

when η is large enough, which, together with (A.6), implies

$$\max_{1\le k\le \mathcal{B}}\max_{1\le i\le N}H_{T,1}(k,i) = O_P(\alpha_{T,h}). \tag{A.8}$$

Meanwhile, by A1 we have

$$\begin{aligned} & \max_{1\le k\le \mathcal{B}} \sup_{x\in B_k} \max_{1\le i\le N} |Z_h(x, X_{it}) - Z_h(x_k, X_{it})| \\ \le \quad & Ch^{-1}\left(\max_{1\le k\le \mathcal{B}} \sup_{x\in B_k} |x - x_k|\right) \le Ch^{-1}h\alpha_{T,h} = O(\alpha_{T,h}), \end{aligned}$$

which implies

$$\max_{1\le k\le \mathcal{B}} \sup_{x\in B_k} \max_{1\le i\le N} H_{T,2}(k, i, x) = O_P(\alpha_{T,h}). \tag{A.9}$$

Combining (A.4), (A.8), (A.9) and the proof of Lemma 6.7 in Xia (2006), we obtain (A.3).

Moreover, by A1 and A3 (i)(iii), we obtain

$$E\left[K_h(\theta^\top X_{it,x})\left(\frac{\theta^\top X_{it,x}}{h}\right)^k\right] = f_{\theta,i}(\theta^\top x)\mu_k + f_{\theta,i}(\theta^\top x)\mu_{k+1}h + O(h^2),$$

and

$$E\left[K_h(\theta^\top X_{it,x})\left(\frac{\theta^\top X_{it,x}}{h}\right)^k X_{it,x}\right] = f_{\theta,i}(\theta^\top x)\nu_{\theta,i}(x)\mu_k + \left[f_{\theta,i}(\theta^\top x)\mu_{\theta,i}(\theta^\top x)\right]'\mu_{k+1}h + O(h^2).$$

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

Lemma A.2. *Letting $a_{i,x}$ and $b_{i,x}$ be defined as a_{is} and b_{is} with X_{its} replaced by $X_{it,x}$ in (2.3), then under A1–A3,*

$$\begin{aligned} a_{i,x} \quad = \quad & g_i(\theta_0^\top x) + g_i'(\theta_0^\top x)(\theta_0 - \theta)^\top \nu_{\theta,i}(x) + \frac{1}{2}g_i''(\theta_0^\top x)h^2 + \frac{1}{Tf_{\theta,i}(\theta^\top x)}\sum_{t=1}^T K_h(\theta^\top X_{it,x})\varepsilon_{it} \\ & + O\left(\left(|\delta_\theta|^2 + h(h^2 + \alpha_{T,h}) + \alpha_{T,h}^2 + (h + \alpha_{T,h})|\delta_\theta|\right)\right) \end{aligned} \tag{A.10}$$

and

$$\begin{aligned} b_{i,x} \quad = \quad & g_i'(\theta_0^\top x) + \frac{1}{Thf_i^\theta(\theta^\top x)}\sum_{t=1}^T K_h(\theta^\top X_{it,x})\left(\frac{\theta^\top X_{it,x}}{h}\right)\varepsilon_{it} \\ + \quad & O\left(\left(h^2 + \alpha_{T,h} + \alpha_{T,h}^2/h + (h + \alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right), \end{aligned} \tag{A.11}$$

uniformly hold for $x \in \mathcal{X}_{NT}$, where $\delta_\theta = \theta - \theta_0$.

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

$$a_{i,x} = \left(S_{i,0}^\theta S_{i,2}^\theta - \left(S_{i,1}^\theta\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^T K_h(\theta^\top X_{it,x})\left(S_{i,2}^\theta - S_{i,1}^\theta(\theta^\top X_{it,x})\right)Y_{it}\right\} \tag{A.12}$$

and

$$b_{i,x} = \left(S^\theta_{i,0}S^\theta_{i,2} - \left(S^\theta_{i,1}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^T K_h(\theta^\top X_{it,x})\left(S^\theta_{i,0}(\theta^\top X_{it,x}) - S^\theta_{i,1}\right)Y_{it}\right\}. \tag{A.13}$$

By Lemma A.1, we have uniformly for $x \in \mathcal{X}_{NT}$,

$$S^\theta_{i,0} = f_{\theta,i}(\theta^\top x) + O_P(h^2 + \alpha_{T,h}), \tag{A.14}$$

$$S^\theta_{i,1} = O_P(h(h + \alpha_{T,h})) = O_P(h^2 + h\alpha_{T,h}), \tag{A.15}$$

$$S^\theta_{i,2} = f_{\theta,i}(\theta^\top x)h^2 + O_P\left(h^2(h^2 + \alpha_{T,h})\right), \tag{A.16}$$

$$S^\theta_{i,3} = O_P(h^3(h + \alpha_{T,h})) = O_P(h^4 + h^3\alpha_{T,h}). \tag{A.17}$$

Hence, by (A.14)–(A.16),

$$S^\theta_{i,0}S^\theta_{i,2} - \left(S^\theta_{i,1}\right)^2 = \left(f_{\theta,i}(\theta^\top x)\right)^2 h^2 + O_P\left(h^2(h^2 + \alpha_{T,h})\right). \tag{A.18}$$

By Taylor expansion, we have

$$\begin{aligned}
Y_{it} &= g_i(\theta_0^\top X_{it}) + \varepsilon_{it} \\
&= \varepsilon_{it} + g_i(\theta_0^\top x) + g_i'(\theta_0^\top x)\theta_0^\top X_{it,x} + \frac{1}{2}g_i''(\theta_0^\top x)(\theta_0^\top X_{it,x})^2 + O(|\theta_0^\top X_{it,x}|^3) \\
&= \varepsilon_{it} + g_i(\theta_0^\top x) + g_i'(\theta_0^\top x)\theta^\top X_{it,x} + \frac{1}{2}g_i''(\theta_0^\top x)(\theta^\top X_{it,x})^2 \\
&\quad + g_i'(\theta_0^\top x)(\theta_0 - \theta)^\top X_{it,x} + \frac{1}{2}g_i''(\theta_0^\top x)\left[(\theta_0^\top X_{it,x})^2 - (\theta^\top X_{it,x})^2\right] \qquad \text{(A.19)} \\
&\quad + O\left(|\theta_0^\top X_{it,x}|^3\right) \\
&= \varepsilon_{it} + g_i(\theta_0^\top x) + g_i'(\theta_0^\top x)\theta^\top X_{it,x} + \frac{1}{2}g_i''(\theta_0^\top x)(\theta^\top X_{it,x})^2 \\
&\quad + g_i'(\theta_0^\top x)(\theta_0 - \theta)^\top X_{it,x} + \Delta_{it,x},
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{it,x} &= O\left(\left[(\theta_0^\top X_{it,x})^2 - (\theta^\top X_{it,x})^2\right] + |\theta_0^\top X_{it,x}|^3\right) \\
&= O\Big(\left[(\theta_0 - \theta)^\top X_{it,x}\right]^2 + 2\theta^\top X_{it,x}(\theta_0 - \theta)^\top X_{it,x} + (\theta^\top X_{it,x})^3 \\
&\quad + \left[(\theta_0 - \theta)^\top X_{it,x}\right]^3 + 3(\theta_0 - \theta)^\top X_{it,x}(\theta^\top X_{it,x})^2 + 3\left[(\theta_0 - \theta)^\top X_{it,x}\right]^2 \theta^\top X_{it,x}\Big) \\
&= O\Big(|\delta_\theta|^2|X_{it,x}|^2 + |\delta_\theta||X_{it,x}||\theta^\top X_{it,x}| + |\theta^\top X_{it,x}|^3 + |\delta_\theta|^3|X_{it,x}|^3 \\
&\quad + |\delta_\theta||X_{it,x}||\theta^\top X_{it,x}|^2 + |\delta_\theta|^2|X_{it,x}|^2|\theta^\top X_{it,x}|\Big). \qquad \text{(A.20)}
\end{aligned}$$

Meanwhile, by (A.14)–(A.18) we have

$$\begin{aligned}
&\left(S^\theta_{i,0}S^\theta_{i,2} - \left(S^\theta_{i,1}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^T K_h(\theta^\top X_{it,x})\left(S^\theta_{i,2} - S^\theta_{i,1}(\theta^\top X_{it,x})\right)g_i(\theta_0^\top x)\right\} \\
&= g_i(\theta_0^\top x)\left(S^\theta_{i,0}S^\theta_{i,2} - (S^\theta_{i,1})^2\right)^{-1}\left(S^\theta_{i,0}S^\theta_{i,2} - (S^\theta_{i,1})^2\right) = g_i(\theta_0^\top x), \qquad \text{(A.21)}
\end{aligned}$$

$$\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S_{i,2}^{\theta}-S_{i,1}^{\theta}(\theta^\top X_{it,x})\right)g_i'(\theta_0^\top x)\theta^\top X_{it,x}\right\}$$

$$= g_i'(\theta_0^\top x)\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left(S_{i,1}^{\theta}S_{i,2}^{\theta}-S_{i,1}^{\theta}S_{i,2}^{\theta}\right)=0 \tag{A.22}$$

and

$$\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S_{i,2}^{\theta}-S_{i,1}^{\theta}(\theta^\top X_{it,x})\right)\left(\frac{1}{2}g_i''(\theta_0^\top x)(\theta^\top X_{it,x})^2\right)\right\}$$

$$= \frac{1}{2}g_i''(\theta_0^\top x)\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left((S_{i,2}^{\theta})^2-S_{i,1}^{\theta}S_{i,3}^{\theta}\right) \tag{A.23}$$

$$= \frac{1}{2}g_i''(\theta_0^\top x)h^2+O_P\left(h^2(h^2+\alpha_{T,h})\right).$$

Let $Q_{i,k}^{\theta}=\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})(\theta^\top X_{it,x})^k X_{it,x}$ for $k=0,\,1,\,2$. By Lemma A.1, we have

$$Q_{i,0}^{\theta}=f_{\theta,i}(\theta^\top x)\nu_{\theta,i}(x)+O_P(h^2+\alpha_{T,h}), \tag{A.24}$$

$$Q_{i,1}^{\theta}=O_P(h(h+\alpha_{T,h})). \tag{A.25}$$

As a result, we have

$$\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S_{i,2}^{\theta}-S_{i,1}^{\theta}(\theta^\top X_{it,x})\right)g_i'(\theta_0^\top x)(\theta_0-\theta)^\top X_{it,x}\right\}$$

$$= g_i'(\theta_0^\top x)(\theta_0-\theta)^\top\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left\{S_{i,2}^{\theta}Q_{i,0}^{\theta}-S_{i,1}^{\theta}Q_{i,1}^{\theta}\right\} \tag{A.26}$$

$$= g_i'(\theta_0^\top x)(\theta_0-\theta)^\top\nu_{\theta,i}(x)+O_P\left((h^2+\alpha_{T,h})|\delta_\theta|\right).$$

Furthermore, by (A.20) we have

$$\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S_{i,2}^{\theta}-S_{i,1}^{\theta}(\theta^\top X_{it,x})\right)\Delta_{it,x}\right)$$

$$= O_P\left(\left[|\delta_\theta|^2+h|\delta_\theta|+h^3+|\delta_\theta|^3\ +\ h^2|\delta_\theta|+h|\delta_\theta|^2\right]\right) \tag{A.27}$$

$$= O_P\left((|\delta_\theta|^2+h^3+h|\delta_\theta|)\right).$$

Since

$$\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\varepsilon_{it}=O_P(\alpha_{T,h}),$$

$$\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})(\theta^\top X_{it,x})\varepsilon_{it}=O_P(h\alpha_{T,h}), \tag{A.28}$$

we have

$$\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S_{i,2}^{\theta}-S_{i,1}^{\theta}(\theta^\top X_{it,x})\right)\varepsilon_{it}\right\}$$

$$
\begin{aligned}
&= \left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1} S_{i,2}^{\theta}\left(\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\varepsilon_{it}\right) \\
&\quad -\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1} S_{i,1}^{\theta}\left(\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})(\theta^{\top}X_{it,x})\varepsilon_{it}\right) \qquad \text{(A.29)}\\
&= \frac{1}{Tf_{\theta,i}(\theta^{\top}x)}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\varepsilon_{it}+O_P\left(\alpha_{T,h}(h+\alpha_{T,h})\right).
\end{aligned}
$$

From (A.12), (A.18), (A.19), (A.21)–(A.23), (A.26), (A.27) and (A.29), we have proved (A.10).

Meanwhile, it is straightforward to have

$$
\begin{aligned}
&\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\left(S_{i,0}^{\theta}(\theta^{\top}X_{it,x})-S_{i,1}^{\theta}\right)g_i(\theta_0^{\top}x)\right\}\\
&= g_i(\theta_0^{\top}x)\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left(S_{i,0}^{\theta}S_{i,1}^{\theta}-S_{i,0}^{\theta}S_{i,1}^{\theta}\right)=0, \qquad \text{(A.30)}
\end{aligned}
$$

$$
\begin{aligned}
&\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\left(S_{i,0}^{\theta}(\theta^{\top}X_{it,x})-S_{i,1}^{\theta}\right)g_i'(\theta_0^{\top}x)\theta^{\top}X_{it,x}\right\}\\
&= g_i'(\theta_0^{\top}x)\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-(S_{i,1}^{\theta})^{2}\right)=g_i'(\theta_0^{\top}x), \qquad \text{(A.31)}
\end{aligned}
$$

$$
\begin{aligned}
&\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\left(S_{i,0}^{\theta}(\theta^{\top}X_{it,x})-S_{i,1}^{\theta}\right)\left(\frac{1}{2}g_i''(\theta_0^{\top}x)(\theta^{\top}X_{it,x})^2\right)\right\}\\
&= \frac{1}{2}g_i''(\theta_0^{\top}x)\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left(S_{i,0}^{\theta}S_{i,3}^{\theta}-S_{i,1}^{\theta}S_{i,2}^{\theta}\right) \qquad \text{(A.32)}\\
&= O_P(h(h+\alpha_{T,h})),
\end{aligned}
$$

$$
\begin{aligned}
&\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\left(S_{i,0}^{\theta}(\theta^{\top}X_{it,x})-S_{i,1}^{\theta}\right)g_i'(\theta_0^{\top}x)(\theta_0-\theta)^{\top}X_{it,x}\right\}\\
&= g_i'(\theta_0^{\top}x)(\theta_0-\theta)^{\top}\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left(S_{i,0}^{\theta}Q_{i,1}^{\theta}-S_{i,1}^{\theta}Q_{i,0}^{\theta}\right) \qquad \text{(A.33)}\\
&= O_P\left((h+\alpha_{T,h})|\delta_\theta|/h\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\left(S_{i,0}^{\theta}S_{i,2}^{\theta}-\left(S_{i,1}^{\theta}\right)^{2}\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^{\top}X_{it,x})\left(S_{i,0}^{\theta}(\theta^{\top}X_{it,x})-S_{i,1}^{\theta}\right)\Delta_{it,x}\right\}\\
&= O_P\left(\left(|\delta_\theta|^2/h+|\delta_\theta|+h^2+|\delta_\theta|^3/h+h|\delta_\theta|+|\delta_\theta|^2\right)\right) \qquad \text{(A.34)}\\
&= O_P\left((|\delta_\theta|^2/h+|\delta_\theta|+h^2)\right).
\end{aligned}
$$

Again from (A.14), (A.15) and (A.28), we can obtain

$$
\begin{aligned}
& \left(S^{\theta}_{i,0}S^{\theta}_{i,2}-\left(S^{\theta}_{i,1}\right)^2\right)^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\left(S^{\theta}_{i,0}(\theta^\top X_{it,x})-S^{\theta}_{i,1}\right)\varepsilon_{it}\right\} \\
= & \left(S^{\theta}_{i,0}S^{\theta}_{i,2}-\left(S^{\theta}_{i,1}\right)^2\right)^{-1}S^{\theta}_{i,0}\left(\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})(\theta^\top X_{it,x})\varepsilon_{it}\right) \\
- & \left(S^{\theta}_{i,0}S^{\theta}_{i,2}-\left(S^{\theta}_{i,1}\right)^2\right)^{-1}S^{\theta}_{i,1}\left(\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})\varepsilon_{it}\right) \qquad \text{(A.35)} \\
= & \frac{1}{Th^2 f_{\theta,i}(\theta^\top x)}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})(\theta^\top X_{it,x})\varepsilon_{it}+O_P\left(\alpha_{T,h}(h+\alpha_{T,h})/h\right).
\end{aligned}
$$

In view of (A.13), (A.18), (A.19) and (A.30)–(A.35), the proof of (A.11) is completed.

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

$$
\begin{aligned}
& \frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}K_h(\theta^\top X_{its})b_{is}^2X_{its}X_{its}^\top/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \qquad \text{(A.36)} \\
= & \frac{2}{N}\sum_{i=1}^{N}D_{\theta_0,i}+O_P\left(h^2+\alpha_{T,h}/h+|\delta_\theta|+|\delta_\theta|^2/h+(NT)^{-1/2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}K_h(\theta^\top X_{its})b_{is}X_{its}\left(Y_{it}-a_{is}-b_{is}\theta_0^\top X_{its}\right)/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_{\theta_0,i}(\theta_0^\top X_{it}))g_i'(\theta_0^\top X_{it})\varepsilon_{it}+\frac{1}{N}\sum_{i=1}^{N}D_{\theta_0,i}(\theta-\theta_0) \\
& +O_P\Big(\left(h+\alpha_{T,h}+|\delta_\theta|\right)\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right) \\
& +(NT)^{-1/2}(h^2+|\delta_\theta|)\Big), \qquad \text{(A.37)}
\end{aligned}
$$

where $D_{\theta_0,i}=E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2\nu_{\theta_0,i}(X_{is})\nu_{\theta_0,i}^\top(X_{is})\right]$ *and* $\widehat{f}_{\theta,i}(\theta^\top x)=\frac{1}{T}\sum_{t=1}^{T}K_h(\theta^\top X_{it,x})$.

Proof. Note that $E\left[\|X_{it}\|^{2+\delta}\right]\leq\infty$ by A2 (ii). It is easy to check that for any small $\epsilon>0$,

$$
\begin{aligned}
& P\left(\max_{1\leq i\leq N}\max_{1\leq t\leq T}\|X_{it}\|>M(NT)^{1/(2+\delta)}\right) \\
\leq & \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{E\|X_{it}\|^{2+\delta}}{M^{2+\delta}NT}=\frac{1}{M^{2+\delta}}<\epsilon
\end{aligned}
$$

if taking $M>\sqrt{1/\epsilon}$.

Hence, in the rest of the proof, we need only to consider the case of $\max_{1\leq i\leq N}\max_{1\leq t\leq T}\|X_{it}\|\leq M(NT)^{1/(2+\delta)}$.

Define $\widetilde{\omega}_{\theta,i}(x) = E\left[(X_{it}-x)(X_{it}-x)^\top \middle| \theta^\top X_{it} = \theta^\top x\right]$. By Lemma A.2, we have uniformly for $x \in \mathcal{X}_{NT}$,

$$\frac{1}{T}\sum_{t=1}^{T} K_h(\theta^\top X_{it,x}) X_{it,x} X_{it,x}^\top = f_{\theta,i}(\theta^\top x)\widetilde{\omega}_{\theta,i}(x) + O_P\left((h^2+\alpha_{T,h})\right), \tag{A.38}$$

and

$$\widehat{f}_{\theta,i}(\theta^\top x) = f_{\theta,i}(\theta^\top x) + O_P\left(h^2+\alpha_{T,h}\right). \tag{A.39}$$

By (A.38) and (A.39),

$$\frac{1}{T}\sum_{t=1}^{T} K_h(\theta^\top X_{it,x}) X_{it,x} X_{it,x}^\top / \widehat{f}_{\theta,i}(\theta^\top x) = \widetilde{\omega}_{\theta,i}(x) + O_P\left(h^2+\alpha_{T,h}\right). \tag{A.40}$$

Meanwhile, by Lemma A.2, we have

$$\begin{aligned} b_{i,x} &= g_i'(\theta_0^\top x) + \tfrac{1}{Thf_{\theta,i}(\theta^\top x)}\sum_{t=1}^{T} K_h(\theta^\top X_{it,x})\left(\tfrac{\theta^\top X_{it,x}}{h}\right)\varepsilon_{it} \\ &\quad + O\left(\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h + (h+\alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right) \\ &= g_i'(\theta_0^\top x) + O\left((h^2+\alpha_{T,h}/h + |\delta_\theta| + |\delta_\theta|^2/h)\right) \end{aligned} \tag{A.41}$$

uniformly for $x \in \mathcal{X}_{NT}$.

By (A.40) and (A.41), we have

$$\begin{aligned} &\tfrac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T} K_h(\theta^\top X_{its}) b_{is}^2 X_{its}X_{its}^\top/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\ &= \tfrac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T} K_h(\theta^\top X_{its})\left(g_i'(\theta_0^\top X_{is})\right)^2 X_{its}X_{its}^\top/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\ &+ \left(\tfrac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T} K_h(\theta^\top X_{its}) g_i'(\theta_0^\top X_{is}) X_{its}X_{its}^\top\left(\widehat{f}_{\theta,i}(\theta^\top X_{is})\right)^{-1}\right) \\ &\times O\left(h^2+\alpha_{T,h}/h + |\delta_\theta| + |\delta_\theta|^2/h\right) \\ &= \tfrac{1}{TN}\sum_{i=1}^{N}\sum_{s=1}^{T}\left(g_i'(\theta_0^\top X_{is})\right)^2\left(\tfrac{1}{T}\sum_{t=1}^{T} K_h(\theta^\top X_{its}) X_{its}X_{its}^\top/\widehat{f}_{\theta,i}(\theta^\top X_{is})\right) \\ &+ O_P\left(h^2+\alpha_{T,h}/h + |\delta_\theta| + |\delta_\theta|^2/h\right) \\ &= \tfrac{1}{TN}\sum_{i=1}^{N}\sum_{s=1}^{T}\left(g_i'(\theta_0^\top X_{is})\right)^2\widetilde{\omega}_{\theta,i}(X_{is}) + O_P\left(h^2+\alpha_{T,h}/h + |\delta_\theta| + |\delta_\theta|^2/h\right) \\ &= \tfrac{1}{N}\sum_{i=1}^{N} E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2\widetilde{\omega}_{\theta_0,i}(X_{is})\right] + O_P\left(h^2+\alpha_{T,h}/h + |\delta_\theta| + |\delta_\theta|^2/h + (NT)^{-1/2}\right) \end{aligned} \tag{A.42}$$

In the meantime, we have

$$\begin{aligned} &E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2\widetilde{\omega}_{\theta_0,i}(X_{is})\right] \\ &= E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2 E\left(\widetilde{\omega}_{\theta_0,i}(X_{is})|\,\theta_0^\top X_{is}\right)\right] \end{aligned}$$

$$= 2E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2\left(E\left[X_{is}X_{is}^\top\Big|\theta_0^\top X_{is}\right]-\mu_{\theta_0,i}(\theta_0^\top X_{is})\mu_{\theta_0,i}^\top(\theta_0^\top X_{is})\right)\right]$$
$$= 2E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2 E\left(\nu_{\theta_0,i}(X_{is})\nu_{\theta_0,i}^\top(X_{is})\Big|\theta_0^\top X_{is}\right)\right]$$
$$= 2E\left[\left(g_i'(\theta_0^\top X_{is})\right)^2 \nu_{\theta_0,i}(X_{is})\nu_{\theta_0,i}^\top(X_{is})\right] = 2D_{\theta_0,i},$$

which, combined with (A.42), implies (A.36).

We next turn to the proof of (A.37). Observe that by Lemma A.2,

$$\begin{aligned}
& Y_{it}-a_{i,x}-b_{i,x}\theta_0^\top X_{it,x} \\
=\ & \varepsilon_{it}+g_i(\theta_0^\top X_{it})-g_i(\theta_0^\top x)-g_i'(\theta_0^\top x)(\theta_0-\theta)^\top\nu_{\theta,i}(x)-\frac{1}{2}g_i''(\theta_0^\top x)h^2-g_i'(\theta_0^\top x)(\theta_0^\top X_{it,x}) \\
-\ & \frac{1}{Tf_{\theta,i}(\theta^\top x)}\sum_{l=1}^T K_h(\theta^\top X_{il,x})\varepsilon_{il}-\left[\frac{1}{Tf_{\theta,i}(\theta^\top x)}\sum_{l=1}^T K_h(\theta^\top X_{il,x})\varepsilon_{il}\left(\frac{\theta^\top X_{il,x}}{h}\right)\right]\left(\frac{\theta_0^\top X_{it,x}}{h}\right) \\
+\ & O\left(\left(h(h^2+\alpha_{T,h})+\alpha_{T,h}^2+|\delta_\theta|^2+(h+\alpha_{T,h})|\delta_\theta|\right)(1+|\theta_0^\top X_{it,x}|/h)\right) \\
=\ & \varepsilon_{it}+\frac{1}{2}g_i''(\theta_0^\top x)\left((\theta_0^\top X_{it,x})^2-h^2\right)-g_i'(\theta_0^\top x)(\theta_0-\theta)^\top\nu_{\theta,i}(x) \qquad \text{(A.43)} \\
-\ & \frac{1}{Tf_{\theta,i}(\theta^\top x)}\sum_{l=1}^T K_h(\theta^\top X_{il,x})\varepsilon_{il}-\left[\frac{1}{Tf_{\theta,i}(\theta^\top x)}\sum_{l=1}^T K_h(\theta^\top X_{il,x})\varepsilon_{il}\left(\frac{\theta^\top X_{il,x}}{h}\right)\right]\left(\frac{\theta^\top X_{it,x}}{h}\right) \\
+\ & O\left(\left(h(h^2+\alpha_{T,h})+\alpha_{T,h}^2+|\delta_\theta|^2+(h+\alpha_{T,h})|\delta_\theta|\right)(1+|\theta_0^\top X_{it,x}|/h)\right)
\end{aligned}$$

uniformly for $x\in\mathcal{X}_{NT}$.

Therefore, we have

$$\begin{aligned}
& \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}\left(Y_{it}-a_{is}-b_{is}\theta_0^\top X_{its}\right)/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
=\ & \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}\varepsilon_{it}/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
+\ & \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}g_i'(\theta_0^\top X_{is})\nu_{\theta,i}^\top(X_{is})(\theta-\theta_0)/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
+\ & \frac{1}{2T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}g_i''(\theta_0^\top X_{is})\left[(\theta_0^\top X_{its})^2-h^2\right]/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
-\ & \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}\left\{\frac{1}{Tf_{\theta,i}(\theta^\top X_{is})}\sum_{l=1}^T K_h(\theta^\top X_{ils})\varepsilon_{il}\right\}/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
-\ & \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})b_{is}X_{its}\left(\frac{\theta^\top X_{its}}{h}\right) \\
\times\ & \left\{\frac{1}{Tf_{\theta,i}(\theta^\top X_{is})}\sum_{l=1}^T K_h(\theta^\top X_{ils})\left(\frac{\theta^\top X_{ils}}{h}\right)\varepsilon_{il}\right\}/\widehat{f}_{\theta,i}(\theta^\top X_{is}) \\
+\ & O_P\left(\left(h(h^2+\alpha_{T,h})+\alpha_{T,h}^2+|\delta_\theta|^2+(h+\alpha_{T,h})|\delta_\theta|\right)(1+|\delta_\theta|/h)\right) \qquad \text{(A.44)}
\end{aligned}$$

$$
\begin{aligned}
&=: \Pi_{N,T}^{1} + \Pi_{N,T}^{2} + \Pi_{N,T}^{3} - \Pi_{N,T}^{4} - \Pi_{N,T}^{5} \\
&+ O_P\left(\left(h(h^2+\alpha_{T,h}) + \alpha_{T,h}^2 + |\delta_\theta|^2 + (h+\alpha_{T,h})|\delta_\theta|\right)(1+|\delta_\theta|/h)\right).
\end{aligned}
$$

Define $E_{it}[G(X_{it},Y_{it},X_{is},Y_{is})] = E\left[G(X_{it},Y_{it},X_{is},Y_{is})|X_{it},Y_{it}\right]$. It then follows that

$$
\begin{aligned}
\Pi_{N,T}^{1} &= \frac{1}{T^2N}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T} K_h(\theta^\top X_{its}) g_i'(\theta_0^\top X_{is}) X_{its}\varepsilon_{it}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1} \\
&\quad + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h + (h+\alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\varepsilon_{it}\right\} \\
&\quad + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h + (h+\alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\right.\right. \\
&\quad \left.\left. - E_{it}\left(K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\right)\right]\varepsilon_{it}\right\} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E_{it}\left(K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\right)\right\}\varepsilon_{it} \\
&\quad + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h + (h+\alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right) \qquad \text{(A.45)} \\
&= \Pi_{N,T}^{1,1} + \Pi_{N,T}^{1,2} + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h + (h+\alpha_{T,h})|\delta_\theta|/h + |\delta_\theta|^2/h\right)\right).
\end{aligned}
$$

Similarly to the proof of Lemma 6.7 in Xia (2006), we have

$$
\Pi_{N,T}^{1,1} = O_P\left(\alpha_{T,h}^2\right). \qquad \text{(A.46)}
$$

Additionally, since

$$
\begin{aligned}
& E_{it}\left(K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\right) \\
&= E\left\{E\left[K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is}) X_{its}\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\middle| X_{it},\theta^\top X_{is}\right]\middle| X_{it}\right\} \\
&= E\left\{K_h(\theta^\top X_{its}) g_i'(\theta^\top X_{is})\left(X_{it}-\mu_{\theta,i}(\theta^\top X_{is})\right)\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\middle| X_{it}\right\} \\
&= (X_{it}-\mu_{\theta,i}(\theta^\top X_{it}))g_i'(\theta^\top X_{it}) + O_P(h^2),
\end{aligned}
$$

we have

$$
\Pi_{N,T}^{1,2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_{\theta,i}(\theta^\top X_{it}))g_i'(\theta^\top X_{it})\varepsilon_{it} + O_P\left((NT)^{-1/2}h^2\right). \qquad \text{(A.47)}
$$

By (A.45)–(A.47), we obtain

$$
\Pi_{N,T}^{1} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_{\theta,i}(\theta^\top X_{it}))g_i'(\theta^\top X_{it})\varepsilon_{it} + O_P\left((NT)^{-1/2}h^2\right)
$$

$$
\begin{aligned}
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right) \qquad \text{(A.48)}\\
= \; & \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (X_{it}-\mu_{\theta_0,i}(\theta_0^\top X_{it}))g_i'(\theta_0^\top X_{it})\varepsilon_{it} + O_P\left((NT)^{-1/2}(h^2+|\delta_\theta|)\right)\\
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\Pi_{N,T}^2 \; = \; & \frac{1}{T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T K_h(\theta^\top X_{its})\left(g_i'(\theta_0^\top X_{is})\right)^2 X_{its}\nu_{\theta,i}^\top(X_{is})(\theta-\theta_0)\left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1}\\
& + O_P\left(|\delta_\theta|(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h)\right)\\
= \; & \frac{1}{TN}\sum_{i=1}^N\sum_{s=1}^T \left(g_i'(\theta_0^\top X_{is})\right)^2 \nu_{\theta,i}(X_{is})\nu_{\theta,i}^\top(X_{is})(\theta-\theta_0) \qquad \text{(A.49)}\\
& + O_P\left(|\delta_\theta|(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h)\right)\\
= \; & \frac{1}{N}\sum_{i=1}^N D_{\theta_0,i}(\theta-\theta_0)\\
& + O_P\left(|\delta_\theta|\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h+(NT)^{-1/2}\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\Pi_{N,T}^3 \; = \; & \frac{1}{2T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T g_i'(\theta_0^\top X_{is})g_i''(\theta_0^\top X_{is})(f_{\theta,i}(\theta^\top X_{is}))^{-1}K_h(\theta^\top X_{its})X_{its}\left[(\theta_0^\top X_{its})^2-h^2\right]\\
& + O_P\left(h^2\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right)\\
= \; & \frac{1}{2T^2N}\sum_{i=1}^N\sum_{s=1}^T\sum_{t=1}^T g_i'(\theta_0^\top X_{is})g_i''(\theta_0^\top X_{is})(f_{\theta,i}(\theta^\top X_{is}))^{-1}K_h(\theta^\top X_{its})X_{its}\left[(\theta^\top X_{its})^2-h^2\right]\\
& + O_P\left(h^2\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h\right)+|\delta_\theta|^2+|\delta_\theta|h\right)\\
= \; & O_P\left(h^2\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h\right)+|\delta_\theta|^2+|\delta_\theta|h\right), \qquad \text{(A.50)}
\end{aligned}
$$

where the last equality is due to the fact that

$$
\frac{1}{T}\sum_{t=1}^T K_h(\theta^\top X_{its})X_{its}\left[(\theta^\top X_{its})^2-h^2\right] = O_P\left(h^2(h^2+\alpha_{T,h})\right).
$$

As $E_{it}\left[K_h(\theta^\top X_{its})g_i'(\theta_0^\top X_{is})\nu_{\theta,i}(X_{is})(f_{\theta,i}(\theta^\top X_{is}))^{-1}\right]=0$, by similar arguments to the proof of Lemma 6.7 of Xia (2006), we have

$$
\begin{aligned}
\Pi_{N,T}^4 \; = \; & \frac{1}{TN}\sum_{i=1}^N\sum_{s=1}^T \left(f_{\theta,i}(\theta^\top X_{is})\right)^{-1} g_i'(\theta_0^\top X_{is})\left\{\frac{1}{Tf_{\theta,i}(\theta^\top X_{is})}\sum_{l=1}^T K_h(\theta^\top X_{ils})\varepsilon_{il}\right\}\\
& \times \left\{\frac{1}{T}\sum_{t=1}^T K_h(\theta^\top X_{its})X_{its}\right\}
\end{aligned}
$$

$$
\begin{aligned}
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right) \\
= & \frac{1}{TN}\sum_{i=1}^{N}\sum_{s=1}^{T} g_i'(\theta_0^\top X_{is})\nu_{\theta,i}(X_{is})\left\{\frac{1}{Tf_{\theta,i}(\theta^\top X_{is})}\sum_{l=1}^{T}K_h(\theta^\top X_{ils})\varepsilon_{il}\right\} \\
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right) \\
= & \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{s=1}^{T}\sum_{l=1}^{T}\left(K_h(\theta^\top X_{ils})g_i'(\theta_0^\top X_{is})\nu_{\theta,i}(X_{is})(f_{\theta,i}(\theta^\top X_{is}))^{-1}\right)\varepsilon_{il}\right\} \\
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right) \\
= & \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(K_h(\theta^\top X_{its})g_i'(\theta_0^\top X_{is})\nu_{\theta,i}(X_{is})(f_{\theta,i}(\theta^\top X_{is}))^{-1}\right)\varepsilon_{it}\right\} \\
& + O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right) \\
= & O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right). \qquad \text{(A.51)}
\end{aligned}
$$

Analogously, we have

$$
\Pi_{N,T}^5 = O_P\left(\alpha_{T,h}\left(h^2+\alpha_{T,h}+\alpha_{T,h}^2/h+(h+\alpha_{T,h})|\delta_\theta|/h+|\delta_\theta|^2/h\right)\right). \qquad \text{(A.52)}
$$

It therefore follows from (A.44), (A.48)–(A.52) that the proof of (A.37) is completed.

Appendix B: Proofs of the main results

We now provide the detailed proofs of the main results in Section 3.

Proof of Theorem 3.1. Denote

$$
S_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_{\theta_0,i}(\theta_0^\top X_{it})\right)g_i'(\theta_0^\top X_{it})\varepsilon_{it}.
$$

Let $\theta = \widehat{\theta}_1$ be an initial estimator of θ_0, then after one iteration, we have

$$
\begin{aligned}
\widetilde{\theta}-\theta_0 = & \left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}K_h(\theta^\top X_{its})b_{is}^2X_{its}X_{its}^\top/\widehat{f}_{\theta,i}(\theta^\top X_{is})\right\}^{-1} \\
& \times\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}K_h(\theta^\top X_{its})b_{is}X_{its}(Y_{it}-a_{is}-b_{is}\theta_0^\top X_{its})/\widehat{f}_{\theta,i}(\theta^\top X_{is})\right\}.
\end{aligned}
$$

This, combined with Lemma A.3, implies

$$
\begin{aligned}
\widetilde{\theta}-\theta_0 = & \frac{1}{2}\left\{\frac{1}{N}\sum_{i=1}^{N}D_{\theta_0,i}\right\}^{+}S_{NT}+\frac{1}{2}\left\{\frac{1}{N}\sum_{i=1}^{N}D_{\theta_0,i}\right\}^{+}\left\{\frac{1}{N}\sum_{i=1}^{N}D_{\theta_0,i}\right\}(\theta-\theta_0) \\
& + O_P\left(h(h^2+\alpha_{T,h})+\alpha_{T,h}^2+\frac{\alpha_{T,h}^3}{h}+\frac{(h^2+\alpha_{T,h})}{h}|\delta_\theta|+|\delta_\theta|^2+\frac{|\delta_\theta|^3}{h}\right) \\
= & \frac{1}{2}D_{\theta_0}^{+}S_{NT}+\frac{1}{2}D_{\theta_0}^{+}D_{\theta_0}(\theta-\theta_0)+o_P\left((NT)^{-1/2}+|\delta_\theta|\right) \qquad \text{(B.1)} \\
& + O_P\left(h(h^2+\alpha_{T,h})+\alpha_{T,h}^2+\frac{\alpha_{T,h}^3}{h}+\frac{(h^2+\alpha_{T,h})}{h}|\delta_\theta|+|\delta_\theta|^2+\frac{|\delta_\theta|^3}{h}\right).
\end{aligned}
$$

Let $\theta^{(k)}$ be the value of the estimator of θ_0 after k iterations, $k \geq 1$. Recursing the above equation and by A4, we have

$$\begin{aligned}
\theta^{(k)} - \theta_0 &= \left\{\sum_{l=1}^{k} \frac{1}{2^l}\right\} D_{\theta_0}^{+} S_{NT} + \frac{1}{2^k} D_{\theta_0}^{+} D_{\theta_0}(\theta - \theta_0) + o_P\left((NT)^{-1/2} + |\delta_\theta|\right) \\
&\quad + O_P\left(h(h^2 + \alpha_{T,h}) + \alpha_{T,h}^2 + \frac{\alpha_{T,h}^3}{h} + \frac{(h^2 + \alpha_{T,h})}{h}|\delta_\theta| + |\delta_\theta|^2 + \frac{|\delta_\theta|^3}{h}\right) \\
&= \left\{\sum_{l=1}^{k} \frac{1}{2^l}\right\} D_{\theta_0}^{+} S_{NT} + \frac{1}{2^k} D_{\theta_0}^{+} D_{\theta_0}(\theta - \theta_0) + o_P\left((NT)^{-1/2}\right) \qquad \text{(B.2)}
\end{aligned}$$

Letting $k \to \infty$, we have

$$\widehat{\theta} - \theta_0 = D_{\theta_0}^{+} S_{NT} + o_P\left((NT)^{-1/2}\right). \qquad \text{(B.3)}$$

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

$$B_{iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(X_{it} - \mu_{\theta_0,i}(\theta_0^\top X_{it})\right) g_i'(\theta_0^\top X_{it})\varepsilon_{it} =: \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{it}.$$

Note that

$$\sqrt{NT} S_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{it}\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} B_{iT}. \qquad \text{(B.4)}$$

We adopt the same argument as in the proof of Theorem 2 in Phillips and Moon (1999) to prove the joint asymptotic normality of $\sqrt{NT} S_{NT}$. As $\{B_{iT},\ 1 \leq i \leq N\}$ is independent by A2 (i) and

$$E\left[B_{iT} B_{iT}^\top\right] = E\left[W_{i1} W_{i1}^\top\right] + 2\sum_{t=2}^{T} [1 - (t-1)/T] E\left[W_{i1} W_{it}^\top\right] = \Lambda_{i,T},$$

it is enough for us to justify the Lindeberg condition.

By (3.1), we need to show that

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

for any $\epsilon > 0$. Equation (B.5) follows directly from (3.3). Then, by (3.1) and (B.5), we have

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

By (B.3) and (B.6), we have therefore shown that (3.4) holds.

Proof of Theorem 3.2. Note that

$$\widehat{g}(u) - \overline{g}(u) = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{g}_i(u) - g_i(u)\right) \qquad \text{(B.7)}$$

and

$$\widehat{g}_i(u) = \sum_{t=1}^{T} w_{it}(\widehat{\theta}) Y_{it} = \sum_{t=1}^{T} w_{it}(\widehat{\theta}) \left(g_i(\theta_0^\top X_{it}) + \varepsilon_{it} \right), \tag{B.8}$$

where

$$\begin{aligned} w_{it}(\theta) &= (1,0) \left\{ \frac{1}{T} \sum_{t=1}^{T} K_h(\theta^\top X_{it} - u) \begin{pmatrix} 1 \\ \theta^\top X_{it} - u \end{pmatrix} \begin{pmatrix} 1 \\ \theta^\top X_{it} - u \end{pmatrix}^\top \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{T} \sum_{t=1}^{T} K_h(\theta^\top X_{it} - u) \begin{pmatrix} 1 \\ \theta^\top X_{it} - u \end{pmatrix} \right\}. \end{aligned}$$

By (B.8), we have for each $i \geq 1$,

$$\begin{aligned} \widehat{g}_i(u) - g_i(u) &= \left(\sum_{t=1}^{T} w_{it}(\widehat{\theta}) g_i(\theta_0^\top X_{it}) - g_i(u) \right) + \sum_{t=1}^{T} w_{it}(\widehat{\theta}) \varepsilon_{it} \\ &=: V_{iT}(1) + V_{iT}(2). \end{aligned}$$

Observe that

$$\begin{aligned} V_{iT}(1) &= \left(\sum_{t=1}^{T} w_{it}(\widehat{\theta}) \left(g_i(\widehat{\theta}^\top X_{it}) - g_i(\theta_0^\top X_{it}) \right) \right) - \left(\sum_{t=1}^{T} w_{it}(\widehat{\theta}) g_i(\widehat{\theta}^\top X_{it}) - g_i(u) \right) \\ &=: V_{iT}(1,1) + V_{iT}(1,2). \end{aligned} \tag{B.9}$$

By Theorem 3.1, we can show that

$$V_{iT}(1,1) = O_P((NT)^{-1/2}). \tag{B.10}$$

Following the proofs of Lemmas A.1 and A.2, we have

$$\max_i \left| V_{iT}(1,2) - h^2 \mu_2 g_i''(u) \right| = o_P(h^2). \tag{B.11}$$

In view of (B.9)–(B.11) and noting $NTh^4 \to \infty$, we have

$$\frac{1}{N} \sum_{i=1}^{N} V_{iT}(1) = b_g(u) h^2 + o_P(h^2) \tag{B.12}$$

Meanwhile, following the proof of Theorem 3.1, we can show

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sqrt{Th} V_{iT}(2) \xrightarrow{d} N\left(0, \sigma_g^2(u)\right). \tag{B.13}$$

Therefore, equation (3.6) follows from (B.7), (B.12) and (B.13).

We have therefore completed the proofs of Theorems 3.1 and 3.2.

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 Econometics*. 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*. John Wiley & Sons: New York.

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*. Chapman & Hall: London.

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

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

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*. Cambridge University Press: New York.

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. Marcel Dekker, New York, pp. 579-604.

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.

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.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single–index models. *Journal of the American Statistical Association* **97**, 1042-1054.