ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Estimation in Partially Linear Single-Index Panel Data Models with Fixed Effects

Jia Chen, Jiti Gao and Degui Li

September 2011

Working Paper 14/11

Estimation in Partially Linear Single–Index Panel Data Models with Fixed Effects

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

**University of Queensland and †Monash University*

Abstract

In this paper, we consider semiparametric estimation in a partially linear single–index panel data model with fixed effects. Without taking the difference explicitly, we propose using a semiparametric minimum average variance estimation (SMAVE) based on a dummy–variable method to remove the fixed effects and obtain consistent estimators for both the parameters and the unknown link function. As both the cross section size and the time series length tend to infinity, we not only establish an asymptotically normal distribution for the estimators of the parameters in the single index and the linear component of the model, but also obtain an asymptotically normal distribution for the nonparametric local linear estimator of the unknown link function. The asymptotically normal distributions of the proposed estimators are similar to those obtained in the random effects case. In addition, we study several partially linear single–index dynamic panel data models. The methods and results are augmented by simulation studies and illustrated by an application to a cigarette–demand data set in the US from 1963–1992.

JEL subject classifications: C13, C14, C23.

Keywords: Fixed effects, local linear smoothing, panel data, semiparametric estimation, single–index models.

[1]Corresponding author: Department of Econometrics and Business Statistics, Monash University Caulfield Campus, Caulfield East, Victoria 3145, Australia. E–mail address: degui.li08@gmail.com

1. Introduction

Panel data analysis has become increasingly popular in many fields, such as climatology, economics and finance. The double–index models enable researchers to estimate complex models and extract information that may be difficult to obtain by applying purely cross–section or time–series models. There exists rich literature on parametric linear and nonlinear panel data models. For an overview of statistical inference and econometric analysis of parametric panel data models, we refer to the books by Baltagi (1995), Arellano (2003) and Hsiao (2003). As in both the cross section and time series cases, parametric panel data models may be misspecified, and estimators obtained from such misspecified models are often inconsistent. To address such issues, some nonparametric methods have been used in both panel data model estimation and specification testing. Recent studies include Ullah & Roy (1998), Hjellvik *et al* (2004), Cai & Li (2008), Henderson *et al* (2008), and Mammen *et al* (2009).

In the multivariate setting with more than three covariates, the underlying regression function cannot be estimated with reasonable accuracy due to the so–called "curse of dimensionality". How to circumvent the curse of dimensionality is an important issue in both nonlinear time series and panel data analysis. Many approaches have been developed to address this issue (see, recent books by Fan & Yao 2003, Gao 2007, Li & Racine 2007 for example). One commonly–used approach is the semiparametric partially linear modeling. An advantage of the semiparametric partially linear modeling is that any existing information concerning possible linearity of some of the components can be taken into account in such models. This has been studied extensively in both the time series and panel data cases (see, for example, Gao 2007, Li & Racine 2007).

As is well known, however, the nonparametric components in the partially linear models may only accommodate covariates $\mathbf{X}$ with low dimension and they are also subject to the curse of dimensionality when the dimension of $\mathbf{X}$ is larger than three. To address this issue, we use the dimension reduction technique of single–index modelling. Specifically, we consider a partially linear single–index panel data model of the form

$$Y_{it} = \mathbf{Z}_{it}^{\top}\boldsymbol{\beta}_0 + \eta(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}_0) + \alpha_i + v_{it}, \quad 1 \le i \le n, \quad 1 \le t \le T, \tag{1.1}$$

where $\mathbf{Z}_{it} = (Z_{it,1}, \cdots, Z_{it,d})^\top$ and $\mathbf{X}_{it} = (X_{it,1}, \cdots, X_{it,p})^\top$ are the respective d–dimensional and p–dimensional covariate vectors, $\boldsymbol{\beta}_0 = (\beta_{0,1}, \cdots, \beta_{0,d})^\top$ and $\boldsymbol{\theta}_0 = (\theta_{0,1}, \cdots, \theta_{0,p})^\top$ are unknown parameters with dimensions d and p, respectively, $\eta(\cdot)$ is an unknown link function, α_i are unobserved time-invariant individual effects, and v_{it} are the random errors. Note that $\mathbf{Z}_{it}$ can be either continuous or discrete random variables, while $\mathbf{X}_{it}$ are assumed to be continuous random variables.

Model (1.1) is called a fixed effects model if $\{\alpha_i\}$ is correlated with $\{\mathbf{Z}_{it}\}$ and (or) $\{\mathbf{X}_{it}\}$ with an unknown correlation structure. Model (1.1) is called a random effects model if $\{\alpha_i\}$ is uncorrelated with both $\{\mathbf{Z}_{it}\}$ and $\{\mathbf{X}_{it}\}$. In this paper, we are concerned with the fixed effects case. For the purpose of identification, we assume that

$$(i)\ \sum_{i=1}^{n} \alpha_i = 0, \qquad (ii)\ \|\boldsymbol{\theta}_0\| = 1 \text{ and the first component of } \boldsymbol{\theta}_0 \text{ is positive,} \tag{1.2}$$

where $\|\cdot\| := \|\cdot\|_2$ is the L_2–distance. (i) is a commonly used identification condition on the fixed effects (see, for example, Su & Ullah 2006, Sun *et al* 2009). (ii) is an identification condition for the single–index structure in our model (see, for example, Carroll *et al* 1997, Xia *et al* 2002).

Model (1.1) covers many interesting panel data models. When $\boldsymbol{\beta}_0 \equiv \mathbf{0}$, model (1.1) reduces to a single–index panel data model (Bai *et al* 2009). When $\mathbf{X}_{it}$ are scalar, model (1.1) becomes to a partially linear panel data model with fixed effects (Su & Ullah 2006). When $\boldsymbol{\beta}_0 \equiv \mathbf{0}$ and $\eta(\cdot)$ is known, model (1.1) is a generalized linear panel data model with fixed effects (Hsiao 2003).

Existing literature mainly focuses on both nonparametric and semiparametric estimation of random effects panel data models (see, for example, Li & Stengos 1996, Ullah & Roy 1998, Henderson & Ullah 2005). Note that the random effects estimators are inconsistent if the true model is one with fixed effects. In this paper, we will develop a semiparametric estimation method associated with a local linear dummy variable approach for model (1.1). The estimation method is consistent under *either the random effects setting or the fixed effects setting.*

In this paper, we also allow that either $\mathbf{Z}_{it}$ or $\mathbf{X}_{it}$ contain time lagged values of Y_{it}. In this case, model (1.1) covers several partially linear single–index dynamic panel data models. In Section 4, we show that, for each $i \geq 1$, $\{Y_{it} : t \geq 1\}$ is geometrically ergodic under some mild conditions, when it is generated by a type of partially linear autoregressive models. This implies that stationarity and mixing conditions on the underlying process are satisfied for each $i \geq 1$. Furthermore, we apply the partially linear single–index panel data model to analyze the dynamic demand of cigarettes based on a panel data set from 46 states in the US. The data set contains the consumption of cigarettes, the lagged consumption of cigarettes, the average retail price, disposable income and the minimum price of cigarettes in any neighboring state. Baltagi *et al* (2000) and Mammen *et al* (2009) respectively used a parametric linear model and a nonparametric additive model to analyze the relationship among the variables. From the study by Mammen *et al* (2009), we can see that there is some linear relationship between the consumption of cigarettes and its lagged consumption. This suggests that model (1.1) might be a better option for such a data set (see Section 5 for detail).

The main contribution of this paper can be summarized as follows. We first propose using a semiparametric minimum average variance estimation (SMAVE) approach associated with a dummy variable method to estimate the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ as well as the unknown link function $\eta(\cdot)$. Under certain regularity conditions, we are able to establish asymptotically normal distributions for the proposed parametric estimators and nonparametric estimator when both n and T tend to infinity. Furthermore, we find that the dummy variable approach proposed for the fixed effects case enables us derive the same asymptotically normal distributions as in the case where random effects are involved.

The rest of the paper is organized as follows. In Section 2, we introduce the so–called SMAVE method to estimate $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ and $\eta(\cdot)$. Section 3 establishes the asymptotic theory for the proposed estimators. Section 4 discusses some autoregression extensions of the proposed model. Section 5 illustrates the performance of the proposed models and estimation methods using both simulated and real data examples. Technical assumptions and proofs of the main results are provided in Appendices A–C. An additional appendix as Appendix D is given in a supplemental document.

2. Dummy variable based SMAVE approach

In the time series case ($n = 1$ and $\alpha_i \equiv 0$) of model (1.1), several estimation methods have been introduced (see, for example, Carroll *et al* 1997, Liang *et al* 2010, Wang *et al* 2010 for the profile likelihood method; Yu & Ruppert 2002 for the penalized spline method; Xia and Härdle 2006 for the SMAVE method). However, these methods cannot be readily used for the panel data model (1.1) due to the presence of the fixed effects. The fixed effects, which are absent in time series models, have to be eliminated in the estimation procedure so that consistent estimators can be constructed. In linear panel data models, the conventional method of removing the fixed effects is differencing, i.e., deducting either a cross–time average or the observations for the previous time period from the observations for the current time period (Henderson *et al* 2008). However, due to the single–index structure in model (1.1), the differencing will complicate the estimation of the link function. Hence, we will develop an estimation procedure based on a local linear dummy variable approach, which is motivated by the least squares dummy variable approach used for parametric panel data analysis (Hsiao 2003). In the dummy variable approach, the unobserved fixed effects are brought explicitly into the model (1.1) and are treated as the coefficients of the model. Having re-specified model (1.1) in this way, we can estimate it by using the SMAVE method.

Apart from the fixed effects, another factor in the estimation of model (1.1) that is different from the estimation of corresponding time series models is the involvement of two indices: the time index t and the individual index i, which, as one might expect, will add further complexity to the estimation of model (1.1). We will establish asymptotic theory for the proposed estimators, as both the time–series dimension T and the cross–sectional dimension n tend to infinity, by using the joint limit approach introduced by Phillips and Moon (1999). The detailed proofs for such joint limiting distribution results are more complicated than those for the asymptotic distribution theory of time series models.

We next introduce the SMAVE method, which estimates both the parameters and the unknown link function by minimizing a single common loss function. The SMAVE method was first introduced by Xia *et al* (2002) for single–index time series models, Recently, Xia (2006) established an asymptotic theory for this approach in time series models and Xia & Härdle (2006) extended the approach and its asymptotic theory to partially linear single–index time series models. However, extending this approach to the partially linear single–index panel data model (1.1) is challenging for the reasons stated above. To address these issues, we will combine the dummy variable approach with the SMAVE method and construct root–nT consistent parametric estimators.

We first introduce some notations for brevity of the presentation of our estimation method. Let

$$\begin{aligned}
&\mathbb{Y} = (Y_{11}, \cdots, Y_{1T}, Y_{21}, \cdots, Y_{nT})^\top, \\
&\mathbb{Z} = (\mathbf{Z}_{11}, \cdots, \mathbf{Z}_{1T}, \mathbf{Z}_{21}, \cdots, \mathbf{Z}_{nT})^\top, \\
&\mathbb{V} = (v_{11}, \cdots, v_{1T}, v_{21}, \cdots, v_{nT})^\top, \\
&\eta(\mathbb{X}, \boldsymbol{\theta}) = \left(\eta(\mathbf{X}_{11}^\top \boldsymbol{\theta}), \cdots, \eta(\mathbf{X}_{1T}^\top \boldsymbol{\theta}), \eta(\mathbf{X}_{21}^\top \boldsymbol{\theta}), \cdots, \eta(\mathbf{X}_{nT}^\top \boldsymbol{\theta})\right)^\top, \\
&D_0 = I_n \otimes e_T, \quad \boldsymbol{\alpha}_0 = (\alpha_1, \cdots, \alpha_n)^\top,
\end{aligned}$$

where I_n is the $n \times n$ identity matrix, e_T is a T–dimensional vector with all elements being 1, and $\otimes$ denotes the Kronecker product. With these notations, we can rewrite model (1.1) as

$$\mathbb{Y} = \mathbb{Z}\boldsymbol{\beta}_0 + \eta(\mathbb{X}, \boldsymbol{\theta}_0) + D_0 \boldsymbol{\alpha}_0 + \mathbb{V}. \tag{2.1}$$

Furthermore, by the identification assumption $\sum_{i=1}^{n} \alpha_i = 0$, we have $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$. Letting $D = [-e_{n-1}, I_{n-1}]^\top \otimes e_T$ and $\boldsymbol{\alpha} = (\alpha_2, \cdots, \alpha_n)^\top$, (2.1) can then be rewritten as

$$\mathbb{Y} = \mathbb{Z}\boldsymbol{\beta}_0 + \eta(\mathbb{X}, \boldsymbol{\theta}_0) + D\boldsymbol{\alpha} + \mathbb{V}. \tag{2.2}$$

For $\mathbf{X}_{it}$ close to $\mathbf{x} \in \mathbb{R}^p$, we have the following local linear approximation:

$$\eta(\mathbf{X}_{it}^\top \boldsymbol{\theta}_0) \approx \eta(\mathbf{x}^\top \boldsymbol{\theta}_0) + \eta'(\mathbf{x}^\top \boldsymbol{\theta}_0)(\mathbf{X}_{it} - \mathbf{x})^\top \boldsymbol{\theta}_0,$$

where $\eta'(u)$ is the derivative of $\eta(u)$ at u. The basic idea of the SMAVE method is to minimize

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - D\boldsymbol{\alpha} - (e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta}))\,(a_{it}, b_{it})^\top\right]^\top \mathbb{W}_{it} \left[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - D\boldsymbol{\alpha} - (e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta}))\,(a_{it}, b_{it})^\top\right] \tag{2.3}$$

with respect to $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $(a_{it}, b_{it})^\top$, where

$$\mathbb{X}_{it}(\boldsymbol{\theta}) = \left((\mathbf{X}_{11} - \mathbf{X}_{it})^\top \boldsymbol{\theta}, \cdots, (\mathbf{X}_{1T} - \mathbf{X}_{it})^\top \boldsymbol{\theta}, (\mathbf{X}_{21} - \mathbf{X}_{it})^\top \boldsymbol{\theta}, \cdots, (\mathbf{X}_{nT} - \mathbf{X}_{it})^\top \boldsymbol{\theta}\right)^\top$$

and $\mathbb{W}_{it} = \mathrm{diag}(w_{11,it}, \cdots, w_{1T,it}, w_{21,it}, \cdots, w_{nT,it})$ is a diagonal matrix with its elements satisfying $\sum_{j=1}^{n} \sum_{s=1}^{T} w_{js,it} = 1$ for each pair (i, t).

To solve the minimization problem (2.3), we will use an iterative procedure, which is detailed as follows.

Step (i): For given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, minimizing

$$\left[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - D\boldsymbol{\alpha} - (e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta}))\,(a_{it}, b_{it})^\top\right]^\top \mathbb{W}_{it} \left[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - D\boldsymbol{\alpha} - (e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta}))\,(a_{it}, b_{it})^\top\right] \tag{2.4}$$

with respect to $\boldsymbol{\alpha}$, we get

$$\boldsymbol{\alpha}_{it} = \left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it} \left[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - (e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta}))\,(a_{it}, b_{it})^\top\right]. \tag{2.5}$$

Then, letting $\boldsymbol{\alpha}$ in (2.4) replaced by the right hand side of (2.5) and minimizing the resulting weighted least squares with respect to $(a_{it}, b_{it})^\top$, we obtain the local linear estimator of $\left(\eta(\mathbf{X}_{it}^\top\boldsymbol{\theta}), \eta\prime(\mathbf{X}_{it}^\top\boldsymbol{\theta})\right)^\top$:

$$(a_{it}, b_{it})^\top = \left(\overline{\mathbb{X}}_{it,*}^\top(\boldsymbol{\theta})\mathbb{W}_{it}\overline{\mathbb{X}}_{it,*}(\boldsymbol{\theta})\right)^{-1} \overline{\mathbb{X}}_{it,*}^\top(\boldsymbol{\theta})\mathbb{W}_{it}\left(\mathbb{Y}_{it,*} - \mathbb{Z}_{it,*}\boldsymbol{\beta}\right), \tag{2.6}$$

where

$$\begin{aligned}
\overline{\mathbb{X}}_{it,*}(\boldsymbol{\theta}) &= \left[I_{nT} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it}\right](e_{nT}, \mathbb{X}_{it}(\boldsymbol{\theta})), \\
\mathbb{Y}_{it,*} &= \mathbb{Y} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it}\mathbb{Y}, \\
\mathbb{Z}_{it,*} &= \mathbb{Z} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it}\mathbb{Z}.
\end{aligned}$$

Step (ii): For each pair (i, t), substitute $\boldsymbol{\alpha}$ and $(a_{it}, b_{it})^\top$ in (2.3) with the right hand sides of (2.5) and (2.6) and solve the resulting minimization problem with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to obtain

$$(\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top = \begin{pmatrix} \mathbb{Z}_*^\top\mathbb{W}\mathbb{Z}_* & \mathbb{Z}_*^\top\mathbb{W}\mathbb{X}_* \\ \mathbb{X}_*^\top\mathbb{W}\mathbb{Z}_* & \mathbb{X}_*^\top\mathbb{W}\mathbb{X}_* \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{Z}_*^\top \\ \mathbb{X}_*^\top \end{pmatrix} \mathbb{W}\left(\mathbb{Y}_* - \mathbb{A}_*\right), \tag{2.7}$$

where

$$\begin{aligned}
\mathbb{W} &= \mathrm{diag}\left(\mathbb{W}_{11}, \cdots, \mathbb{W}_{1T}, \mathbb{W}_{21}, \cdots, \mathbb{W}_{nT}\right), \\
\mathbb{Y}_* &= \left(Y_{11,*}^\top, \cdots, Y_{1T,*}^\top, Y_{21,*}^\top, \cdots, Y_{nT,*}^\top\right)^\top, \\
\mathbb{Z}_* &= \left(\mathbb{Z}_{11,*}^\top, \cdots, \mathbb{Z}_{1T,*}^\top, \mathbb{Z}_{21,*}^\top, \cdots, \mathbb{Z}_{nT,*}^\top\right)^\top, \\
\mathbb{X}_* &= \left(b_{11}\mathbb{X}_{11,*}^\top, \cdots, b_{1T}\mathbb{X}_{1T,*}^\top, b_{21}\mathbb{X}_{21,*}^\top, \cdots, b_{nT}\mathbb{X}_{nT,*}^\top\right)^\top, \\
\mathbb{X}_{it,*} &= \left[I_{nT} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it}\right]\mathbb{X}_{it}, \\
\mathbb{X}_{it} &= \left((\mathbf{X}_{11} - \mathbf{X}_{it}), \cdots, (\mathbf{X}_{1T} - \mathbf{X}_{it}), (\mathbf{X}_{21} - \mathbf{X}_{it}), \cdots, (\mathbf{X}_{nT} - \mathbf{X}_{it})\right)^\top, \\
\mathbb{A}_* &= \left(a_{11}e_{11,*}^\top, \cdots, a_{1T}e_{1T,*}^\top, a_{21}e_{21,*}^\top, \cdots, a_{nT}e_{nT,*}^\top\right)^\top, \\
e_{it,*} &= \left[I_{nT} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it}\right]e_{nT}.
\end{aligned}$$

Step (iii): With the updated values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, repeat the above two steps until convergence.

As in Xia *et al* (2002), we use two sets of weights in the above iterative procedure. The first is a set of multidimensional kernel weights defined as

$$w_{js,it} = \frac{H\left((\mathbf{X}_{js} - \mathbf{X}_{it})/h_1\right)}{\sum_{j=1}^{n}\sum_{s=1}^{T} H\left((\mathbf{X}_{js} - \mathbf{X}_{it})/h_1\right)}, \tag{2.8}$$

where $H(\cdot)$ is a p–variate symmetric kernel function and h_1 is a bandwidth. Choosing any d–dimensional vector $\boldsymbol{\beta}$ and p–dimensional vector $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\| = 1$ and following the above iterations, we can obtain initial estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$, which will later be shown to be consistent. The initial estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ are denoted $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, respectively. However, the estimators based on the p–variate kernel $H(\cdot)$ are not efficient due to the "curse of dimensionality". To improve the efficiency, we then use a set of single–index weights which are defined as

$$w^{\boldsymbol{\theta}}_{js,it} = \frac{K\left((\mathbf{X}_{js} - \mathbf{X}_{it})^{\top}\boldsymbol{\theta}/h_2\right)}{\sum_{j=1}^{n}\sum_{s=1}^{T} K\left((\mathbf{X}_{js} - \mathbf{X}_{it})^{\top}\boldsymbol{\theta}/h_2\right)}, \tag{2.9}$$

where $K(\cdot)$ is a univariate symmetric kernel function and h_2 is a bandwidth. Using the initial estimates $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ and following steps (i)–(iii) with the single–index weights, we then obtain the final estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. By substituting $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\mathbf{X}_{it}^{\top}\boldsymbol{\theta}$ in (2.6) with $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$ and u, we obtain the estimator of $\eta(u)$, which is denoted $\widehat{\eta}(u)$.

3. Asymptotic theory

In this section, we establish the weak consistency of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ and then give the asymptotically normal distributions of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$ and the nonparametric local linear estimate of the link function.

Theorem 3.1. *Let Assumptions* **A1**–**A7** *listed in Appendix A hold. Then, we have*

$$\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_P(1) \quad \text{and} \quad \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_P(1). \tag{3.1}$$

The proof of Theorem 3.1 is given in Appendix B below. Theorem 3.1 establishes the weak consistency of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$. Note that the detailed proof of Theorem 3.1 and related technical lemmas in Appendix D of the supplemental document indicate that one can possibly strengthen the weak consistency result to strong consistency. The consistency of the initial estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ will help us to establish the root-nT convergence of the final estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$.

Before giving the asymptotic distribution for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, we introduce some notations. Let $\widetilde{\mathbf{Z}}_{it,\boldsymbol{\theta}} = \mathbf{Z}_{it} - \mathbf{v}_{\boldsymbol{\theta}}(\mathbf{X}_{it})$ and $\widetilde{\mathbf{X}}_{it,\boldsymbol{\theta}} = \mathbf{X}_{it} - \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_{it})$, where $\mathbf{v}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathsf{E}\left(\mathbf{Z}_{11}|\mathbf{X}_{11}^{\top}\boldsymbol{\theta} = \mathbf{x}^{\top}\boldsymbol{\theta}\right)$ and $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathsf{E}\left(\mathbf{X}_{11}|\mathbf{X}_{11}^{\top}\boldsymbol{\theta} = \mathbf{x}^{\top}\boldsymbol{\theta}\right)$. Define

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \Sigma_0(1) & \Sigma_0(2) \\ \Sigma_0^{\top}(2) & \Sigma_0(3) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} \Sigma_1(1) & \Sigma_1(2) \\ \Sigma_1^{\top}(2) & \Sigma_1(3) \end{pmatrix}, \tag{3.2}$$

where

$$\begin{aligned}
&\Sigma_0(1) = \mathsf{E}\left(\widetilde{\mathbf{Z}}_{11}\widetilde{\mathbf{Z}}_{11}^{\top}\right), \qquad \Sigma_0(2) = \mathsf{E}\left[\widetilde{\mathbf{Z}}_{11}\eta'(\mathbf{X}_{11}^{\top}\boldsymbol{\theta}_0)\widetilde{\mathbf{X}}_{11}^{\top}\right], \\
&\Sigma_0(3) = \mathsf{E}\left[\left(\eta'(\mathbf{X}_{11}^{\top}\boldsymbol{\theta}_0)\right)^2 \widetilde{\mathbf{X}}_{11}\widetilde{\mathbf{X}}_{11}^{\top}\right], \quad \Sigma_1(1) = \sum_{t=-\infty}^{\infty} \mathsf{E}\left(\widetilde{\mathbf{Z}}_{i1}\widetilde{\mathbf{Z}}_{it}^{\top}v_{i1}v_{it}\right), \\
&\Sigma_1(2) = \sum_{t=-\infty}^{\infty} \mathsf{E}\left[\widetilde{\mathbf{Z}}_{i1}\eta'(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}_0)\widetilde{\mathbf{X}}_{it}^{\top}v_{i1}v_{it}\right] \quad \text{and} \\
&\Sigma_1(3) = \sum_{t=-\infty}^{\infty} \mathsf{E}\left[\eta'(\mathbf{X}_{i1}^{\top}\boldsymbol{\theta}_0)\eta'(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}_0)\widetilde{\mathbf{X}}_{i1}\widetilde{\mathbf{X}}_{it}^{\top}v_{i1}v_{it}\right].
\end{aligned}$$

The asymptotically normal distribution of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ is given in the following theorem.

Theorem 3.2. *Let Assumptions* **A1**–**A7** *and* **B1**–**B4** *listed in Appendix A hold. Then, as* $n,\ T \to \infty$ *simultaneously, we have*

$$\sqrt{nT}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathsf{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\right), \tag{3.3}$$

where $\mathbf{0}$ *is a null–vector of dimension* $d + p$.

Theorem 3.2 shows that the final estimators resulting from the iterative procedure associated with the second set of weights achieve the root–nT rate of convergence. The asymptotic distribution in (3.3) can be regarded as a natural and substantial extension of existing results for time series case, such as Theorems 2 and 3 in Carroll *et al* (1997), Theorem 1 in Xia & Härdle (2006) and Theorem 1 in Liang *et al* (2010). Furthermore, if we assume that the error process $\{v_{it}\}$ is independent of $\{\mathbf{Z}_{it}\}$ and $\{\mathbf{X}_{it}\}$, and v_{it} are independent and identically distributed (i.i.d.) over i and t, the asymptotic variance in (3.3) can be reduced to $\sigma^2\boldsymbol{\Sigma}_0^{-1}$, where $\sigma^2 = \mathsf{E}[v_{it}^2]$. This implies that the SMAVE method achieves an semiparametrically efficient bound (see Carroll *et al* 1997 for details).

Under some mild conditions, we can show that the joint limit as both n and T tend to infinity is identical to the sequential limit as $T \to \infty$ first and then $n \to \infty$ or the sequential limit as $n \to \infty$ first and then $T \to \infty$ (see, for example, Phillips & Moon 1999). Additionally, we also find that, as $T \to \infty$, the dummy variable approach proposed for the fixed effects case provides the same asymptotically normal distribution as in the case where random effects are involved. To the best of our knowledge, this is a set of new findings for this type of nonlinear panel data models.

Let us turn to the asymptotic distribution of the the nonparametric estimator of the link function. Let $\mu_k = \int u^k K(u)du$, $\nu_k = \int u^k K^2(u)du$, $b_\eta(u) = \frac{1}{2}\mu_2 \eta''(u) h_2^2$ and $\sigma_\eta^2(u) = \nu_0 \sigma_{\boldsymbol{\theta}_0}^2(u) f_{\boldsymbol{\theta}_0}^{-1}(u)$, where $\sigma_{\boldsymbol{\theta}_0}^2(u) = \mathsf{E}(v_{it}^2 | \mathbf{X}_{it}^\top \boldsymbol{\theta}_0 = u)$ and $f_{\boldsymbol{\theta}_0}(\cdot)$ is the density function of $\mathbf{X}_{it}^\top \boldsymbol{\theta}_0$.

Theorem 3.3. *Let the conditions of Theorem 3.2 hold. As $n, T \to \infty$ simultaneously,*

$$\sqrt{nTh_2}\left(\widehat{\eta}(\mathbf{x}^\top \widehat{\boldsymbol{\theta}}) - \eta(\mathbf{x}^\top \boldsymbol{\theta}_0) - b_\eta(\mathbf{x}^\top \boldsymbol{\theta}_0)\right) \xrightarrow{d} \mathsf{N}\left(0, \sigma_\eta^2(\mathbf{x}^\top \boldsymbol{\theta}_0)\right). \tag{3.4}$$

From the above theorem, the forms of the bias term $b_\eta(\cdot)$ and the asymptotic variance term $\sigma_\eta^2(\cdot)$ are similar to those of the local linear estimator for panel data models with random effects (see, for example, Theorem 3 in Cai & Li 2008). This implies that the dummy variable approach proposed for the fixed effects case has similar asymptotically normal distribution to that in the random effects case.

The proofs of Theorems 3.2 and 3.3 are given in Appendix C.

4. Dynamic partially linear single–index panel data models

This section introduces several dynamic models where the regressors $\mathbf{Z}_{it}$ and (or) $\mathbf{X}_{it}$ in (1.1) contain time–lagged values of Y_{it}. Three types of partially linear single–index dynamic panel data models are considered.

Case (i) Letting $\mathbf{Z}_{it} = (Y_{i,t-1}, \cdots, Y_{i,t-d})^\top$, model (1.1) then becomes

$$Y_{it} = \sum_{j=1}^{d} Y_{i,t-j}\beta_{0,j} + \eta(\mathbf{X}_{it}^\top \boldsymbol{\theta}_0) + \alpha_i + v_{it}, \ 1 \le i \le n, 1 \le t \le T. \tag{4.1}$$

For each i, suppose that $\{\mathbf{X}_{it} : \ t \ge 1\}$ and $\{v_{it} : \ t \ge 1\}$ are two i.i.d. sequences, and $(\mathbf{X}_{it}, v_{it})$ are independent of $Y_{i,t-j}$, $j \ge 1$. Then, a sufficient condition for the geometrical ergodicity of $\{Y_{it} : \ t \ge 1\}$ for each i is that

$$y^d - \beta_{0,1} y^{d-1} - \cdots - \beta_{0,d-1} y - \beta_{0,d} \ne 0 \ \text{ for any } |y| \ge 1, \tag{4.2}$$

which also leads to the stationarity of $\{Y_{it} : \ t \ge 1\}$.

Case (ii) Consider the case where $\mathbf{X}_{it}$ contain time–lagged values of Y_{it} with $\mathbf{X}_{it} = (Y_{i,t-1}, \cdots, Y_{i,t-p})^\top$. Model (1.1) then becomes

$$Y_{it} = \mathbf{Z}_{it}^\top \boldsymbol{\beta}_0 + \eta \left(\sum_{j=1}^{p} Y_{i,t-j}\theta_{0,j} \right) + \alpha_i + v_{it}, \ 1 \le i \le n, 1 \le t \le T. \tag{4.3}$$

For each i, suppose that $\{\mathbf{Z}_{it} : \ t \ge 1\}$ and $\{v_{it} : \ t \ge 1\}$ are two i.i.d. sequences, and $(\mathbf{Z}_{it}, v_{it})$ are independent of $Y_{i,t-j}$, $j \ge 1$. Furthermore, assume that for any $u \in \mathbb{R}$,

$$|\eta(u)| \le \lambda^* |u|/\sqrt{p} + c^*, \tag{4.4}$$

where $0 < \lambda^* < 1$ and $0 < c^* < \infty$. Then, following the same argument as in Example 3.5 of An and Huang (1996), we can show that $\{Y_{it} : \ t \ge 1\}$ is geometrically ergodic for each i.

Case (iii) Consider the case where both $\mathbf{Z}_{it}$ and $\mathbf{X}_{it}$ contain time–lagged values of Y_{it}. In this case, (1.1) becomes

$$Y_{it} = \sum_{j=1}^{d} Y_{i,t-j}\beta_{0,j} + \eta \left(\sum_{j=1}^{p} Y_{i,t-j}\theta_{0,j} \right) + \alpha_i + v_{it}, \ 1 \le i \le n, 1 \le t \le T. \tag{4.5}$$

Xia *et al* (1999) considered the time series case of (4.5) with $\alpha_i \equiv 0$ and gave some conditions for the model to be identifiable. We now consider the geometrical ergodicity of $\{Y_{it} : \ t \ge 1\}$ in the panel data model (4.5) with $\alpha_i \ne 0$ generally. Let $\eta_i(u) = \eta(u) + \alpha_i$. Then (4.5) can be rewritten as

$$Y_{it} = \sum_{j=1}^{d} Y_{i,t-j}\beta_{0,j} + \eta_i \left(\sum_{j=1}^{p} Y_{i,t-j}\theta_{0,j} \right) + v_{it}. \tag{4.6}$$

Suppose that $\beta_{0,1}, \ \cdots, \ \beta_{0,d}$ satisfy (4.2), $\max_i |\alpha_i| < \infty$, $\lim_{|u|\to\infty} \left| \frac{\eta_i(u)}{u} \right| = \lim_{|u|\to\infty} \left| \frac{\eta(u)}{u} \right| = 0$, and the probability density function of $\{v_{it}\}$ is positive everywhere. Then it can be shown, following the proof of Theroem 3 in Xia *et al* (1999), that $\{Y_{it} : \ t \ge 1\}$ is geometrically ergodic for each i.

5. Numerical Examples

In this section, we first carry out a Monte Carlo simulation study to examine the finite sample performance of the proposed estimation method, and then use the proposed model and method to analyze a set of US cigarette demand data.

As introduced in Section 2, we use two sets of weights: one set of multivariate weights for producing consistent initial estimates of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ and a set of single–index weights for producing final estimates. Throughout this section, we use a product kernel $H(\mathbf{x}) =$

$\prod_{j=1}^{p} K(x_j)$ for the multivariate weights, where $K(u) = \frac{3}{4}(1-u^2)I(|u| \leq 1)$. Equal bandwidth of $h_1 = \widehat{\sigma}_X (nT)^{-1/(4+p)}$ is used for each variate of the multivariate weights, where $\widehat{\sigma}_X$ is the sample standard deviation of $\mathbf{X}_{it}$, $1 \leq i \leq n$, $1 \leq t \leq T$. The bandwidth h_1 is simply chosen under the following considerations: firstly it can reduce the computational burden that we suffer from the iterations and secondly the bandwidth choice for the production of initial estimates has little effect on the performance of the final estimates.

For the single–index weights, we use the quadratic kernel $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ and apply a leave–one–unit–out cross validation method for choosing the bandwidth. The leave–one–out cross validation method was proposed in Sun *et al* (2009) and is an extension of the conventional leave–one–out cross validation method. The idea is to remove $\{(\mathbf{Z}_{it}, \mathbf{X}_{it}, Y_{it}) : 1 \leq t \leq T\}$ from the data and use the rest of the $(n-1)T$ observations as the training data to obtain estimates of $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ and $\eta(\cdot)$, which are denoted as $\widehat{\boldsymbol{\beta}}_{(-i)}$, $\widehat{\boldsymbol{\theta}}_{(-i)}$ and $\widehat{\eta}_{(-i)}(\cdot)$. We thus choose an optimal bandwidth that minimizes a weighted squared prediction error of the form

$$\left(\mathbb{Y} - B(\mathbb{Z}, \widehat{\boldsymbol{\beta}}_{(-)}) - \eta(\mathbb{X}, \widehat{\boldsymbol{\theta}}_{(-)})\right)^\top M^\top M \left(\mathbb{Y} - B(\mathbb{Z}, \widehat{\boldsymbol{\beta}}_{(-)}) - \eta(\mathbb{X}, \widehat{\boldsymbol{\theta}}_{(-)})\right), \qquad (5.1)$$

where $M = I_{n\times T} - \frac{1}{T} I_n \otimes (e_T e_T^\top)$,

$$B(\mathbb{Z}, \widehat{\boldsymbol{\beta}}_{(-)}) = \left(\mathbf{Z}_{11}^\top \widehat{\boldsymbol{\beta}}_{(-1)}, \cdots, \mathbf{Z}_{1T}^\top \widehat{\boldsymbol{\beta}}_{(-1)}, \mathbf{Z}_{21}^\top \widehat{\boldsymbol{\beta}}_{(-2)}, \cdots, \mathbf{Z}_{2T}^\top \widehat{\boldsymbol{\beta}}_{(-2)}, \cdots, \mathbf{Z}_{n1}^\top \widehat{\boldsymbol{\beta}}_{(-n)}, \cdots \mathbf{Z}_{nT}^\top \widehat{\boldsymbol{\beta}}_{(-n)}\right)^\top$$

and

$$\eta(\mathbb{X}, \widehat{\boldsymbol{\theta}}_{(-)}) = \left(\eta_{(-1)}\left(\mathbf{X}_{11}^\top \widehat{\boldsymbol{\theta}}_{(-1)}\right), \cdots, \eta_{(-1)}\left(\mathbf{X}_{1T}^\top \widehat{\boldsymbol{\theta}}_{(-1)}\right), \eta_{(-2)}\left(\mathbf{X}_{21}^\top \widehat{\boldsymbol{\theta}}_{(-2)}\right), \cdots, \eta_{(-n)}\left(\mathbf{X}_{nT}^\top \widehat{\boldsymbol{\theta}}_{(-n)}\right)\right)^\top.$$

The weight matrix M is constructed to satisfy $MD = \mathbf{0}$ so that the fixed effect term $D\boldsymbol{\alpha}$ is eliminated from (5.1). In fact, M removes a cross-time average from each variable. For example,

$$M\mathbb{Y} = (Y_{11} - Y_{1A}, \cdots, Y_{1T} - Y_{1A}, \cdots, \cdots, Y_{n1} - Y_{nA}, \cdots, Y_{nT} - Y_{nA})^\top,$$

where $Y_{iA} = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$ for $i = 1, \cdots, n$.

5.1. Simulated Examples

Example 5.1. We first use the following data generating process

$$Y_{it} = 0.3 Z_{it} + \sin\left\{\pi \left[(X_{it,1} + X_{it,2} + X_{it,3})/\sqrt{3} - A\right]/(B - A)\right\} + \alpha_i + v_{it}, \qquad (5.2)$$

where $Z_{it} = 0$ for odd t and $Z_{it} = 1$ for even t, $\mathbf{X}_{it} = (X_{it,1}, X_{it,2}, X_{it,3})^\top$ are three–dimensional random vectors with independent uniform $U(0, 1)$ components and are i.i.d. over both i and t, $A = 0.3912$ and $B = 1.3409$, $\alpha_i = 0.5Z^*_{iA} + u_i$ for $i = 1, \cdots, n-1$, and $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$, in which $Z^*_{iA} = \frac{1}{2T} \sum_{t=1}^{T} (Z_{it,1} + Z_{it,2})$ and u_i are i.i.d. $N(0, 0.1^2)$ random errors, v_{it} are i.i.d. (over both i and t) $N(0, 0.1^2)$ random variables. In addition, $\{Z_{it}\}$, $\{\mathbf{X}_{it}\}$, $\{u_i\}$ and $\{v_{it}\}$ are mutually independent.

The true parameters of model (5.2) are $\beta_0 = 0.3$ and $\boldsymbol{\theta}_0 = (1, 1, 1)^\top/\sqrt{3}$, and the link function is $\eta(u) = \sin\{\pi(u - A)/(B - A)\}$. The time series counterpart of this example was used by Carroll *et al* (1997), Xia and Härdle (2006) and Liang *et al* (2010).

We start the iterative estimation procedure described in Section 2 with $\boldsymbol{\theta} = (0, 1, 2)^\top/\sqrt{5}$ as the initial values of $\boldsymbol{\theta}_0$. The resulting estimates of the parameters over 200 realizations, as well as their corresponding mean squared errors (MSEs) for samples of sizes $n, T = 10$, 20, 30 are summarized in Table 5.1 with the MSEs parenthesized. The estimates of the link function $\eta(\cdot)$ from typical realizations of sample sizes $n, T = 10$, 20, 30 are given in Figure 5.1.

Table 5.1 indicates that the SMAVE method produces accurate estimates of both β_0 and $\boldsymbol{\theta}_0$, and as either n or T increases, the MSEs of the estimates become smaller and smaller. Comparison of the results in Table 5.1 with those in the second panel of Table 1 in Xia and Härdle (2006) also suggests that the estimates and MSEs here are comparable with those in Xia and Härdle (2006).

Example 5.2. Consider the following model

$$Y_{it} = (2Z_{it,1} + Z_{it,2})/\sqrt{5} + 2\exp\left\{-(2X_{it} + X_{i,t-1} + 2X_{i,t-2})^2/3\right\} + \alpha_i + v_{it}, \qquad (5.3)$$

where $\mathbf{Z}_{it} = (Z_{it,1}, Z_{it,2})^\top$ are two–dimensional i.i.d. (over both i and t) random vectors with independent components that have binary distribution with $P(Z_{it,j} = 0) = P(Z_{it,j} = 1) = 0.5$, $j = 1, 2$, $\mathbf{X}_{it} = (X_{it}, X_{i,t-1}, X_{i,t-2})^\top$ in which $X_{it} = 0.4X_{i,t-1} + x_{it}$ and x_{it} are i.i.d. (over i and t) and uniformly distributed with $x_{it} \sim U(-1, 1)$, v_{it} are i.i.d. (over i and t) with normal distribution $N(0, 0.5^2)$, $\alpha_i = 0.5Z^*_{iA} + u_i$ for $i = 1, \cdots, n-1$, and $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$, in which $Z^*_{iA} = \frac{1}{2T} \sum_{t=1}^{T} (Z_{it,1} + Z_{it,2})$ and $u_i \overset{i.i.d.}{\sim} N(0, 0.2^2)$. $\{Z_{it}\}$, $\{x_{it}\}$, $\{u_i\}$ and $\{v_{it}\}$ are mutually independent.

The true parameters of model (5.4) are $\boldsymbol{\beta}_0 = (2, 1)^\top/\sqrt{5}$ and $\boldsymbol{\theta}_0 = (2, 1, 2)^\top/3$, and the true link function is $\eta(u) = 2\exp\{-3u^2\}$.

The means as well as the MSEs of the estimates of the parameters over 200 replications are given in Table 5.2. These results indicate that the SMAVE method estimates the parameters accurately, and its performance (in terms of MSE) improves as n or T increases.

Table 5.1. Means and MSEs of the estimates of the parameters in Example 5.1

$n\backslash T$		10		20		30	
	True Value	Mean	MSE($\times 10^{-4}$)	Mean	MSE($\times 10^{-4}$)	Mean	MSE($\times 10^{-4}$)
10	$\beta_0 = 0.3000$	0.2989	(4.3502)	0.3011	(2.1873)	0.3002	(1.4515)
	$\theta_{0,1} = 0.5774$	0.5789	(2.6539)	0.5783	(1.2542)	0.5769	(0.8428)
	$\theta_{0,2} = 0.5774$	0.5768	(2.6542)	0.5769	(1.2403)	0.5773	(0.8118)
	$\theta_{0,3} = 0.5774$	0.5763	(3.1450)	0.5768	(1.4429)	0.5778	(0.9853)
20	$\beta_0 = 0.3000$	0.3012	(2.1887)	0.3006	(0.9868)	0.2998	(0.7108)
	$\theta_{0,1} = 0.5774$	0.5767	(1.3108)	0.5779	(0.6288)	0.5767	(0.4139)
	$\theta_{0,2} = 0.5574$	0.5786	(1.2686)	0.5766	(0.5951)	0.5771	(0.3705)
	$\theta_{0,3} = 0.5774$	0.5768	(1.4648)	0.5776	(0.6755)	0.5782	(0.4379)
30	$\beta_0 = 0.3000$	0.2993	(1.5891)	0.2994	(0.6470)	0.3001	(0.4859)
	$\theta_{0,1} = 0.5774$	0.5770	(0.8558)	0.5767	(0.4155)	0.5769	(0.2822)
	$\theta_{0,2} = 0.5774$	0.5768	(0.8354)	0.5779	(0.3981)	0.5773	(0.2338)
	$\theta_{0,3} = 0.5774$	0.5783	(0.9518)	0.5775	(0.40812)	0.5778	(0.2374)

Table 5.2. Means and MSEs of the estimates of the parameters in Example 5.2

$n\backslash T$		10		20		30	
	True Value	Mean	MSE($\times 10^{-4}$)	Mean	MSE($\times 10^{-4}$)	Mean	MSE($\times 10^{-4}$)
10	$\beta_{0,1} = 0.8944$	0.8901	(100.0000)	0.8787	(45.0000)	0.8875	(44.0000)
	$\beta_{0,2} = 0.4472$	0.4422	(105.0000)	0.4538	(48.0000)	0.4484	(36.0000)
	$\theta_{0,1} = 0.6667$	0.6683	(13.0000)	0.6612	(5.7443)	0.6642	(4.7835)
	$\theta_{0,2} = 0.3333$	0.3281	(27.0000)	0.3400	(14.0000)	0.3320	(9.1121)
	$\theta_{0,3} = 0.6667$	0.6635	(15.0000)	0.6668	(7.6202)	0.6684	(4.3630)
20	$\beta_{0,1} = 0.8944$	0.9036	(57.0000)	0.8950	(26.0000)	0.8897	(18.0000)
	$\beta_{0,2} = 0.4472$	0.4460	(52.0000)	0.4499	(33.0000)	0.4473	(19.0000)
	$\theta_{0,1} = 0.6667$	0.6651	(6.5923)	0.6639	(4.8670)	0.6662	(2.2190)
	$\theta_{0,2} = 0.3333$	0.3299	(16.0000)	0.3308	(9.4963)	0.3291	(4.4093)
	$\theta_{0,3} = 0.6667$	0.6679	(4.8119)	0.6693	(3.8523)	0.6686	(2.1373)
30	$\beta_{0,1} = 0.8944$	0.9012	(47.0000)	0.8940	(14.0000)	0.8932	(11.0000)
	$\beta_{0,2} = 0.4472$	0.4505	(37.0000)	0.4484	(17.0000)	0.4495	(14.0000)
	$\theta_{0,1} = 0.6667$	0.6662	(4.9653)	0.6647	(2.3813)	0.6671	(1.0029)
	$\theta_{0,2} = 0.3333$	0.3299	(14.0000)	0.3323	(4.7590)	0.3316	(3.3297)
	$\theta_{0,3} = 0.6667$	0.6669	(5.2189)	0.6685	(0.40812)	0.6667	(1.0682)

5.2. A Real Data Example

The real data example is about the cigarette demand in 46 states of the USA over the period 1963–1992. The data set is from Baltagi *et al* (2000), who used a linear dynamic panel data model of the form

$$\ln C_{it} = \beta_0 + \beta_1 \ln C_{i,t-1} + \theta_1 \ln DI_{it} + \theta_2 \ln P_{it} + \theta_3 \ln PN_{it} + u_{it} \tag{5.4}$$

to analyze the demand for cigarettes, where $i = 1, \cdots, 46$, denotes the i–th state, $t = 1, \cdots, 29$ denotes the t–th year, C_{it} is the real per capita sales of cigarettes (measured in packs), DT_{it} is the real per capita disposable income, P_{it} is the average retail price of a

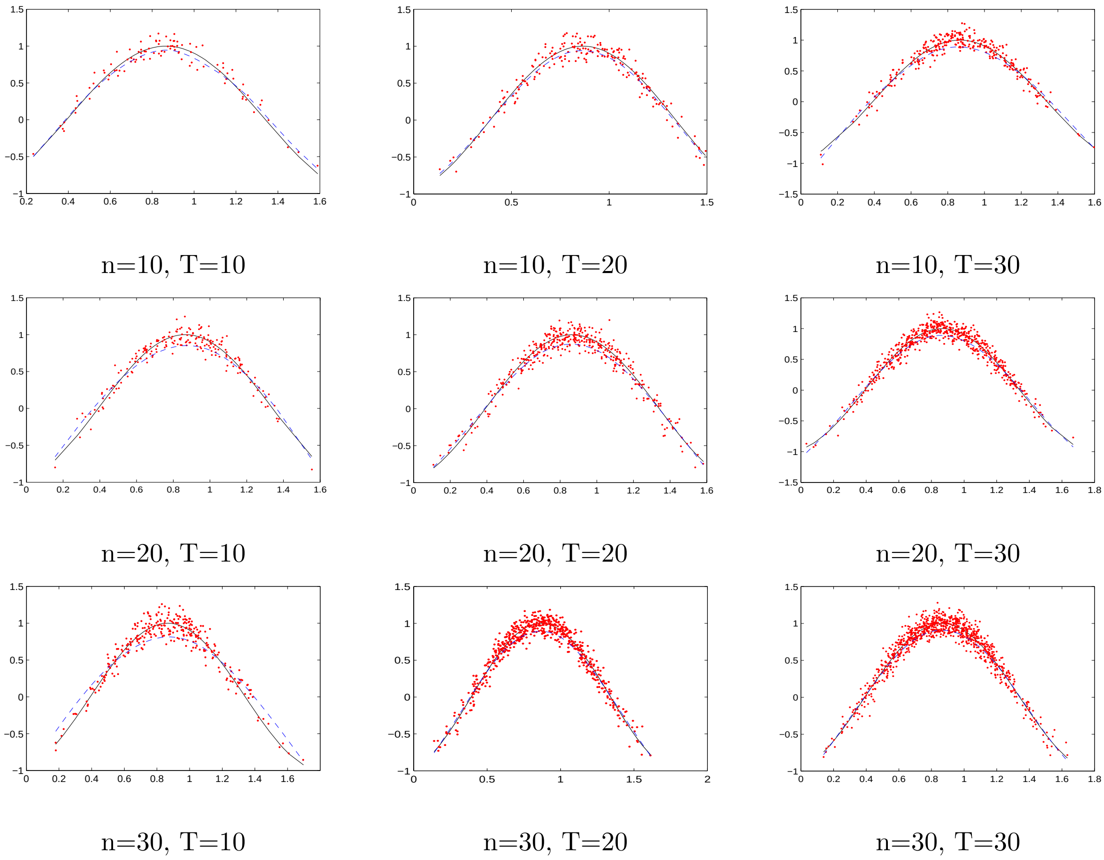

Figure 5.1. Curve estimates from single replications of the simulation study of Example 5.1. The solid curves are the true functions $\eta(\mathbf{X}_{it}^\top \boldsymbol{\theta}_0)$, the dashed curves are the corresponding estimated functions $\widehat{\eta}(\mathbf{X}_{it}^\top \widehat{\boldsymbol{\theta}})$, the dots denote $Y_{it} - \mathbf{Z}_{it}^\top \widehat{\boldsymbol{\beta}} - \widehat{\alpha}_i$ plotted against $\mathbf{X}_{it}^\top \widehat{\boldsymbol{\theta}}$.

pack of cigarettes measured in real terms, PN_{it} is the minimum real price of cigarettes in any neighboring state, and the disturbance term u_{it} in (5.4) is specified as

$$u_{it} = \mu_i + \lambda_t + v_{it}, \tag{5.5}$$

where μ_i denotes a state-specific effect, and λ_t denotes a year-specific effect, which can also be interpreted as a trend in t.

Due to the presence of the time–specific effect or trend λ_t in all the variables, we first remove the trend from the log-transformed observations as in Mammen *et al* (2009),

$$Y_{it} = \ln C_{it} - s_C(t), \quad V1_{it} = Y_{i,t-1}, \quad V2_{it} = \ln DI_{it} - s_{DI}(t),$$
$$V3_{it} = \ln P_{it} - s_P(t), \quad V4_{it} = \ln PN_{it} - s_{PN}(t),$$

where $s_C(t)$, $s_{DI}(t)$, $s_P(t)$, and $s_{PN}(t)$ are the nonparametric estimates of the trends in $\ln C_{it}$, $\ln DI_{it}$, $\ln P_{it}$ and $\ln PN_{it}$, $i = 1 \cdots, 46$, $t = 1, \cdots, 29$. In Figure 5.3, we give the

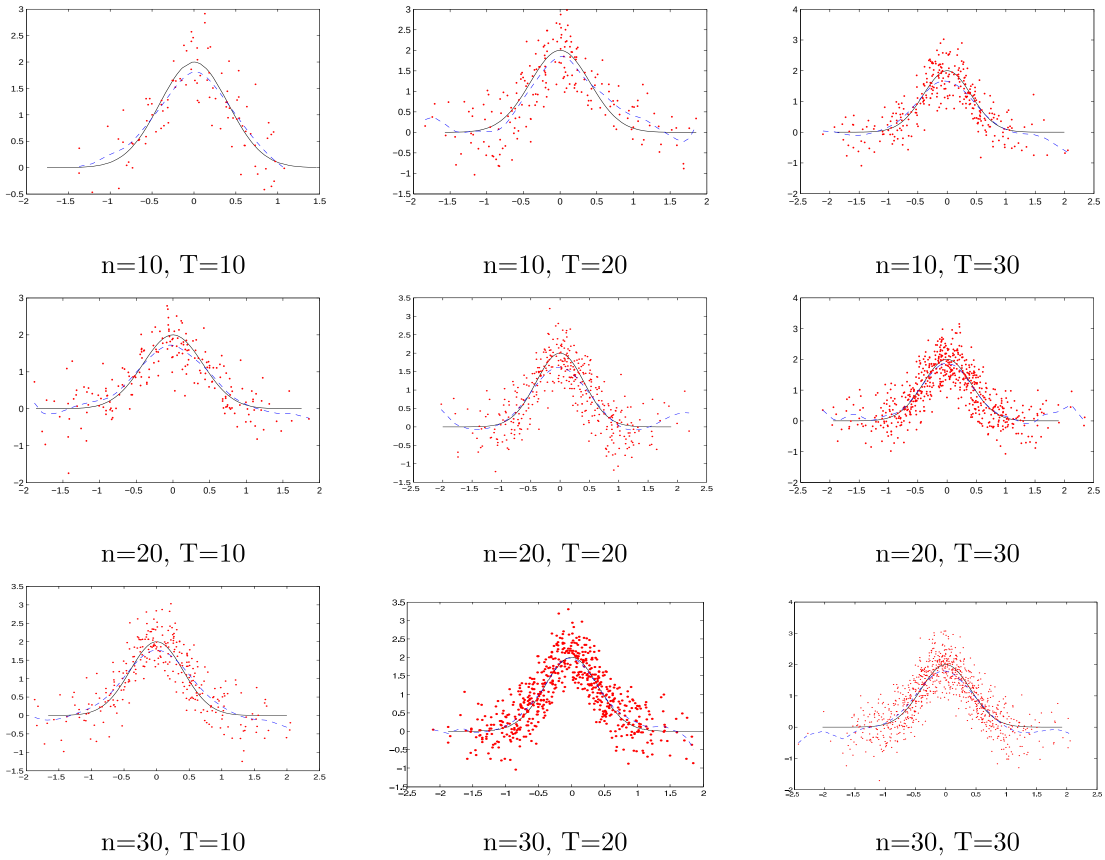

Figure 5.2. Curve estimates from single replications of the simulation study of Example 5.2. The solid curves are the true functions $\eta(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}_0)$, the dashed curves are the corresponding estimated functions $\widehat{\eta}(\mathbf{X}_{it}^{\top}\widehat{\boldsymbol{\theta}})$, the dots denote $Y_{it} - \mathbf{Z}_{it}^{\top}\widehat{\boldsymbol{\beta}} - \widehat{\alpha}_i$ plotted against $\mathbf{X}_{it}^{\top}\widehat{\boldsymbol{\theta}}$.

scatter plots of Y against $V1$, $V2$, $V3$, and $V4$. It is clear from Figure 5.3 that Y exhibits strong linearity with $V1$ (i.e. the lagged variable of Y). For the other three covariates, their linearities with Y are not as strong as that for the lagged-variable. Hence, we define $Z_{it} = V1_{it}$ and $\mathbf{X}_{it} = (V2_{it}, V3_{it}, V4_{it})^{\top}$, and put Z_{it} in the linear term and $\mathbf{X}_{it}$ in the single-index term of the following model

$$Y_{it} = Z_{it}\beta + g(\mathbf{X}_{it}^{\top}\boldsymbol{\theta}) + \alpha_i + v_{it}, \tag{5.6}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\top}$, α_i is a state–specific effect which may include religion, race, tourism, tax, and education. α_i corresponds to μ_i in model (5.4)–(5.5). Furthermore, as we detrended $\ln C_{it}$, $\ln DI_{it}$, $\ln P_{it}$ and $\ln PN_{it}$, the year-specific term λ_t that appeared in model (5.4)–(5.5) is eliminated from model (5.6).

After applying the estimation method proposed in Section 2 to the data on Y_{it}, Z_{it}, $\mathbf{X}_{it}$, we can obtain the estimates of the parameters in (5.6), which are summarized in Table

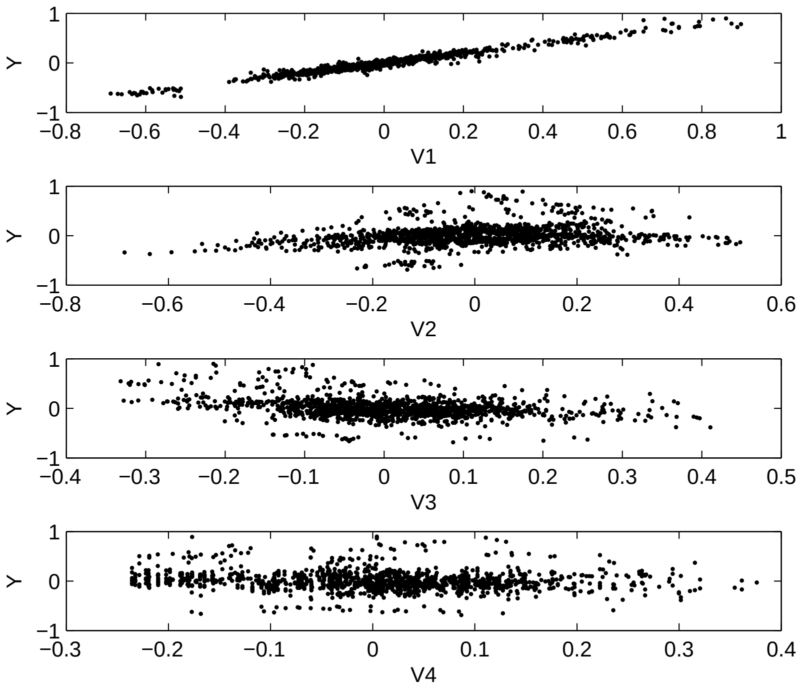

Figure 5.3. From top to bottom: the scatter plots of Y against $V1$, $V2$, V_3, and V_4.

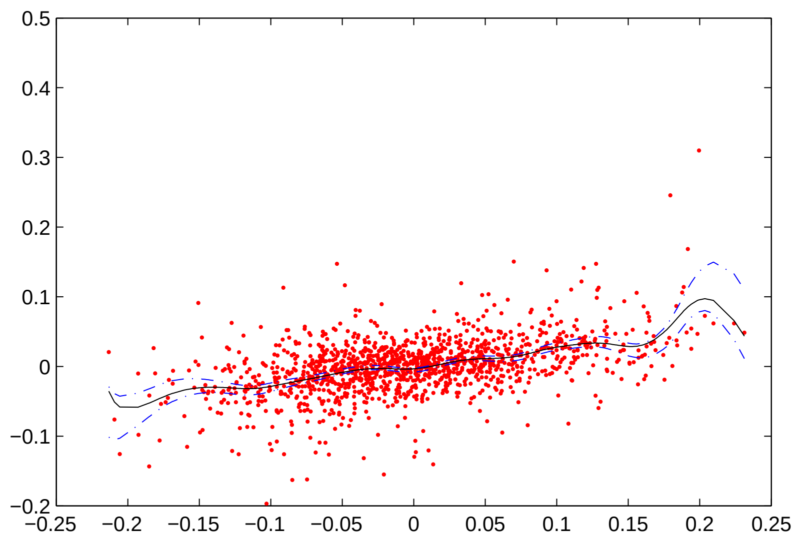

Figure 5.4. Estimated link function and its 95% confidence band. Dots denote $Y_{it} - Z_{it}\widehat{\beta} - \widehat{\alpha}_i$ plotted against $\mathbf{X}_{it}^\top\widehat{\boldsymbol{\theta}}$. The solid line denotes the estimated link function $\widehat{\eta}(\mathbf{X}_{it}^\top\widehat{\boldsymbol{\theta}})$. The dash-dotted lines represent the 95% confidence band.

5.3. The estimated curve of the link function as well as its 95% confidence band is given in Figure 5.4.

Table 5.3. Estimates of the parameters in the cigarette data example

parameter	β	θ_1	θ_2	θ_3
estimate	0.8480	0.2594	-0.8735	0.4119
(SD)	(0.0073)	(0.0217)	(0.0099)	(0.0260)

Comparison of the results in Table 5.3 with that in Baltagi *et al* (2000) indicates that our estimate of β is smaller than the estimates of the corresponding coefficient in Baltagi *et al* (2000), where a value of 0.90 from the OLS method and a value of 0.91 from the GLS method were obtained. In addition, compared with $\widehat{\boldsymbol{\theta}} = (0.2112, -0.9404, 0.2665)^\top$ from the OLS and $\widehat{\boldsymbol{\theta}} = (0.1602, -0.9503, 0.2669)^\top$ from the GLS in Baltagi *et al* (2000), the absolute value of our estimate of θ_2 is smaller, while those of θ_1 and θ_3 are larger (note that due to the identification condition $\|\boldsymbol{\theta}\| = 1$, one has to normalize the estimates of $\boldsymbol{\theta}$ in (5.4) before making comparisons). The computed coefficient of determination for model (5.6) is $R^2 = 0.9698$, which indicates a good fit to the data.

6. Conclusions and Discussion

This paper has considered a partially linear single–index panel data model with fixed effects. A semiparametric minimum average variance estimation method associated with a dummy–variable approach has been proposed to deal with the estimation of both the parametric and nonparametric components of the model. We have shown that the proposed estimators all have asymptotically normal distribution regardless of whether the effects involved are random or fixed. We have then assessed the finite–sample performance of the proposed estimation method through using both simulated and real data examples.

The paper certainly has some limitations. One question is whether the established theory may be extended to the case where both $\{\mathbf{X}_{it}\}$ amd $\{\mathbf{Z}_{it}\}$ are nonstationary over t and cross–sectional dependent over i. How to answer such a question should be left in future research.

7. Acknowledgments

The authors would like to thank Xiaohong Chen, Qi Li, Oliver Linton, Liangjun Su and the seminar participants at Monash University, University of Adelaide, University of Queensland, SETA 2011 Conference in Melbourne, and ESAM 2011 Conference in Adelaide.

This project was financially supported by the Australian Research Council Discovery Grants Program under Grant Number: DP0879088.

Appendix A: Assumptions

Let $\mathbf{Z}_i = (\mathbf{Z}_{it} : 1 \le t \le T)$, $\mathbf{X}_i = (\mathbf{X}_{it} : 1 \le t \le T)$ and $V_i = (v_{it} : 1 \le t \le T)$. To derive the consistency of the initial estimates $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, we need the following set of regularity conditions.

A1 $(\mathbf{Z}_i, \mathbf{X}_i, V_i)$, $i = 1, \cdots, n$, are i.i.d. and $\{(\mathbf{Z}_{it}, \mathbf{X}_{it}, v_{it}) : t \ge 1\}$ is a stationary α–mixing sequence with mixing coefficient $\alpha_i(t)$ for each i. Furthermore, there exists a positive coefficient function $\alpha(t)$ such that

$$\sup_i \alpha_i(t) \le \alpha(t) \quad \text{with} \quad \alpha(t) \le C_\alpha t^{-\gamma_0},$$

where $C_\alpha > 0$ and $\gamma_0 > \frac{(2+\delta_*)(2+\delta)}{2(\delta-\delta_*)}$, in which δ and δ_* are positive constants satisfying $\delta > \delta_*$.

A2 The kernel function $H(\cdot)$: $\mathbb{R}^d \to \mathbb{R}^+$ is a bounded and Lipschitz continuous probability density function with a compact support. Furthermore, $H(\mathbf{x})$ is symmetric and $\int \mathbf{x}\mathbf{x}^\top H(\mathbf{x})d\mathbf{x}$ is positive definite.

A3 The density function $f_{\mathbf{X}}(\cdot)$ of $\mathbf{X}_{it}$ is second–order continuous and has gradient $f'_{\mathbf{X}}(\cdot)$. Moreover, $f_{\mathbf{X}}(\cdot)$ is positive and bounded in $\mathcal{X} := \left\{\mathbf{x} : \|\mathbf{x}\| \le C(nT)^{\frac{1}{2+\delta}}\right\}$ for any $C > 0$ and $\mathsf{E}\|X_{it}\|^{2+\delta} < \infty$, where $\|\cdot\|$ is the L_2–distance and δ was defined in **A1**.

A4 Let $g_1(\mathbf{x}) := \mathsf{E}\left[\mathbf{Z}_{it}|\mathbf{X}_{it} = \mathbf{x}\right]$ and $g_2(\mathbf{x}) := \mathsf{E}\left[\mathbf{Z}_{it}\mathbf{Z}_{it}^\top|\mathbf{X}_{it} = \mathbf{x}\right]$. Both $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ have bounded and continuous derivatives. In addition, $\mathsf{E}\|\mathbf{Z}_{it}\|^{2+\delta} < \infty$ and

$$\mathsf{E}\left\{\left(\mathbf{Z}_{it} - \mathsf{E}(\mathbf{Z}_{it}|\mathbf{X}_{it})\right)\left(\mathbf{Z}_{it} - \mathsf{E}(\mathbf{Z}_{it}|\mathbf{X}_{it})\right)^\top\right\}$$

is a positive definite matrix, where δ was defined in **A1**.

A5 $\{v_{it}\}$ is independent of $\{(\mathbf{Z}_{it}, \mathbf{X}_{it})\}$ with $\mathsf{E}[v_{it}] = 0$, $0 < \sigma^2 := \mathsf{E}[v_{it}^2] < \infty$ and $\mathsf{E}[|v_{it}|^{2+\delta_*}] < \infty$ for $\delta_* > 0$ defined in **A1**.

A6 The link function $\eta(\cdot)$ has continuous derivatives up to the second order.

A7 The bandwidth h_1 involved in the multivariate weights satisfies

$$h_1 \to 0, \quad \frac{\log T}{Th_1^{p+2}} = O(1), \quad \frac{(nT)^{2\gamma_0 - 4p(1+\frac{1}{2+\delta})-3} h_1^{2p\gamma_0+4p^2+9p+2}}{\log^{2\gamma_0-4p+1}(nT)} \to \infty,$$

where p is the dimension of $\mathbf{X}_{it}$, and γ_0 and δ were defined in **A1**.

To establish asymptotic distribution for the final parametric estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, we further need the following set of regularity conditions.

B1 The kernel function $K(\cdot)$: $\mathbb{R} \to \mathbb{R}^+$ is a bounded and symmetric probability density function. Furthermore, $K(\cdot)$ is Lipschitz continuous and has a compact support.

B2 The density function $f_{\boldsymbol{\theta}}(\cdot)$ of $\mathbf{X}_{it}^\top \boldsymbol{\theta}$ is positive and second–order continuous for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. Moreover, $f_{\boldsymbol{\theta}_0}(\cdot)$ is positive and bounded in

$$\mathcal{U} := \left\{ u = \mathbf{x}^\top \boldsymbol{\theta}_0 : \|\mathbf{x}\| \leq C(nT)^{\frac{1}{2+\delta}} \right\}$$

for any $C > 0$ and δ were defined in **A1**.

B3 The conditional expectation $g_3(u) := \mathsf{E}[\mathbf{Z}_{it} | \mathbf{X}_{it}^\top \boldsymbol{\theta} = u]$ has a bounded and continuous derivative for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

B4 The bandwidth h_2 involved in the single–index weights satisfies

$$0 < \lim_{n,T\to\infty} (nT)\ h_2^5 < \infty.$$

Furthermore, there exists a relationship between n and T,

$$\frac{T^{\delta_*\delta+2\delta} \log^{5(2+\delta)(2+\delta_*)}(nT)}{n^{4\delta\delta_*+10\delta_*-2\delta}} = o(1).$$

In **A1**, we assume that $(\mathbf{Z}_i, \mathbf{X}_i, V_i)$, $1 \leq i \leq n$, are cross–sectional independent (see, for example, Su & Ullah 2006, Sun *et al* 2009) and each component time series is α–mixing dependent, which can be satisfied by many linear and nonlinear time series (see, for example, the discussion in Section 4). Assumption **A2** involves some mild conditions on the multivariate kernel function $H(\cdot)$. **A3** and **A4** are similar to the corresponding conditions in Xia & Härdle (2006). Since α_i are allowed to be correlated with $(\mathbf{X}_{it}, \mathbf{Z}_{it})$, $u_{it} = \alpha_i + v_{it}$ thus may be correlated with $(\mathbf{X}_{it}, \mathbf{Z}_{it})$ even though v_{it} are independent of $(\mathbf{X}_{it}, \mathbf{Z}_{it})$. Assumption **A4** is needed to ensure that both $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ and $\eta(\cdot)$ are identifiable and estimable. Meanwhile, the independence between $\{(\mathbf{Z}_{it}, \mathbf{X}_{it})\}$ and $\{v_{it}\}$ in **A5** is imposed to simplify our proofs and it can be removed at the expense of more tedious proofs. **A6** is a common condition for local linear estimators (see, for example, Fan & Gijbels 1996, Fan & Yao 2003). We next show that the bandwidth restrictions in **A7** are satisfied under mild conditions if we take $h_1 \sim (nT)^{-\vartheta}$, $0 < \vartheta < 1/(p+2)$. It is easy to check that $h_1 \sim (nT)^{-\vartheta} = o(1)$ and the second condition in **A7** is also satisfied when $n = O\left(T^{\frac{1}{\vartheta(p+2)}-1} / \log^{\frac{1}{\vartheta(p+2)}} T\right)$. If we let

$p_1 = 2\gamma_0 - \frac{4p(3+\delta)}{2+\delta} - 3$, $p_2 = 2p\gamma_0 + 4p^2 + 9p + 2$ and $p_3 = 2\gamma_0 - 4p + 1$, the left hand side of the last term in **A7** becomes

$$\frac{(nT)^{p_1} h_1^{p_2}}{\log^{p_3}(nT)} = \frac{(nT)^{p_1 - p_2\vartheta}}{\log^{p_3}(nT)}$$

which tends to ∞ when $p_1 > p_2\vartheta$. As $\vartheta < 1/(p+2)$, $2 - 2p\vartheta > 0$. By some elementary calculation, it is easy to show that if

$$\gamma_0 > \frac{(4p^2 + 9p + 2)\vartheta}{2 - 2p\vartheta} + \frac{(4p+3)(2+\delta) + 4p}{(2+\delta)(2 - 2p\vartheta)},$$

then $p_1 > p_2\vartheta$ and thus the third condition in **A7** holds.

Assumptions **B1**–**B3** are natural extensions of conditions C2, C4 and C5 in Xia & Härdle (2006). The rate of the bandwidth h_2 in **B4** is optimal for pooled local linear estimators. In particular, if we take $\delta_* = 1$ and $\delta = 2$,

$$\delta_*\delta + 2\delta = 6, \quad 5(2+\delta)(2+\delta_*) = 60, \quad 4\delta\delta_* + 10\delta_* - 2\delta = 14.$$

Then, the condition on the relationship between n and T in **B4** would become

$$\frac{T^3 \log^{30}(nT)}{n^7} = o(1),$$

which includes two cases: (i) the time series length T is larger than the cross–sectional dimension n, and (ii) the cross–sectional dimension n is larger than the time series length T.

Appendix B: Proof of Theorem 3.1

Define $a_{\mathbf{x}} = \eta(\mathbf{x}^\top \boldsymbol{\theta}_0)$, $a_{it} = \eta(\mathbf{X}_{it}^\top \boldsymbol{\theta}_0)$, $b_{\mathbf{x}} = \eta'(\mathbf{x}^\top \boldsymbol{\theta}_0)$ and $b_{it} = \eta'(\mathbf{X}_{it}^\top \boldsymbol{\theta}_0)$. Let $\widetilde{a}_{\mathbf{x}}$, $\widetilde{a}_{it}$, $\widetilde{b}_{\mathbf{x}}$, and $\widetilde{b}_{it}$ be the local linear estimators obtained from (2.6) using the set of multivariate weights in (2.8). Let $e_{\mathbf{x},*}$, $\mathbb{X}_{\mathbf{x},*}$, $\overline{\mathbb{X}}_{\mathbf{x},*}$, $\mathbb{W}_{\mathbf{x}}$ and $\mathbb{Z}_{\mathbf{x},*}$ be the counterparts of $e_{it,*}$, $\mathbb{X}_{it,*}$, $\overline{\mathbb{X}}_{it,*}$, $\mathbb{W}_{it}$ and $\mathbb{Z}_{it,*}$ when $\mathbf{X}_{it}$ are replaced by $\mathbf{x}$. Furthermore, define

$$D_{\mathbf{x},*} = D - D\left(D^\top \mathbb{W}_{\mathbf{x}} D\right)^{-1} D^\top \mathbb{W}_{\mathbf{x}} D,$$
$$\mathbb{V}_{\mathbf{x},*} = \mathbb{V} - D\left(D^\top \mathbb{W}_{\mathbf{x}} D\right)^{-1} D^\top \mathbb{W}_{\mathbf{x}} \mathbb{V},$$
$$\mathbb{V}_{it,*} = \mathbb{V} - D\left(D^\top \mathbb{W}_{it} D\right)^{-1} D^\top \mathbb{W}_{it} \mathbb{V}.$$

For simplicity, define $\tau(T) = \sqrt{\frac{\log T}{T h_1^p}}$, $\tau_{nT}(1) = \sqrt{\frac{\log nT}{nT h_1^p}}$, $\tau_{nT}(2) = \sqrt{\frac{\log nT}{nT h_2}}$, $\zeta_{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ and $\zeta_{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$.

To prove the weak consistency of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ in Theorem 3.1, we need to establish the asymptotic uniform expansions of $\widetilde{a}_{\mathbf{x}}$ and $\widetilde{b}_{\mathbf{x}}$ in $\{\mathbf{x} : \|\mathbf{x}\| \leq C_{nT}\}$, where $C_{nT} = C_0\ (nT)^{1/(2+\delta)}$ and $0 < C_0 < \infty$.

Lemma B.1. *Let Assumptions* **A1**–**A7** *in Appendix A hold. Then, we have*

$$\widetilde{a}_{\mathbf{x}} = a_{\mathbf{x}} + g_1^\top(\mathbf{x})(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + O_P\left(h_1^2 + \tau(T)\right), \tag{B.1}$$

and

$$\widetilde{b}_{\mathbf{x}} = \boldsymbol{\theta}^\top \boldsymbol{\theta}_0 b_{\mathbf{x}} + O_P\left(\|\zeta_{\boldsymbol{\beta}}\| + h_1^2 + \tau(T)\right) \tag{B.2}$$

uniformly in $\{\mathbf{x} : \|\mathbf{x}\| \leq C_{nT}\}$*, where* $g_1^\top(\mathbf{x})$ *was defined in* **A4**.

Proof. By the definition of $\widetilde{a}_{\mathbf{x}}$ and $\widetilde{b}_{\mathbf{x}}$, we have

$$\begin{aligned}
\left(\widetilde{a}_{\mathbf{x}},\ \widetilde{b}_{\mathbf{x}}\right)^\top &= \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{Z}_{\mathbf{x},*}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \\
&\quad + \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}D_{\mathbf{x},*}\boldsymbol{\alpha} \\
&\quad + \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) \\
&\quad + \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{V}_{\mathbf{x},*} \\
&= \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{Z}_{\mathbf{x},*}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \\
&\quad + \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) \\
&\quad + \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{V}_{\mathbf{x},*},
\end{aligned} \tag{B.3}$$

where $\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) = \eta(\mathbb{X},\boldsymbol{\theta}_0) - D\left(D^\top\mathbb{W}_{\mathbf{x}}D\right)^{-1}D^\top\mathbb{W}_{\mathbf{x}}\eta(\mathbb{X},\boldsymbol{\theta}_0)$.

By Taylor expansion of $\eta(\mathbf{X}_{it}^\top\boldsymbol{\theta}_0)$, we have

$$\eta(\mathbf{X}_{it}^\top\boldsymbol{\theta}_0) = \eta(\mathbf{x}^\top\boldsymbol{\theta}_0) + \eta'(\mathbf{x}^\top\boldsymbol{\theta}_0)d_{it}^\top(\mathbf{x})\boldsymbol{\theta}_0 + \eta''(\mathbf{x}^\top\boldsymbol{\theta}_0)\left(d_{it}^\top(\mathbf{x})\boldsymbol{\theta}_0\right)^2 + O\left(\left(d_{it}^\top(\mathbf{x})\boldsymbol{\theta}_0\right)^3\right), \tag{B.4}$$

where $d_{it}(\mathbf{x}) = \mathbf{X}_{it} - \mathbf{x}$. By (B.4), the definition of $\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0)$ and following the proof of Lemma D.2 in Appendix D of the supplemental document, we have

$$\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) = \left(a_{\mathbf{x}},\ \boldsymbol{\theta}^\top\boldsymbol{\theta}_0 b_{\mathbf{x}}\right)^\top + O_P(h_1^2 + \tau(T)). \tag{B.5}$$

By Lemmas D.4 and D.5 in Appendix D of the supplemental document, we have

$$(1,0)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{Z}_{\mathbf{x},*} = g_1(\mathbf{x}) + O_P\left(h_1^2 + \tau(T)\right) \tag{B.6}$$

and

$$\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{V}_{\mathbf{x},*} = O_P(\tau_{nT}(1)) = o_P(\tau(T)) \tag{B.7}$$

uniformly in $\|\mathbf{x}\| \leq C_{nT}$.

By (B.3), (B.5)–(B.7), we have proved that (B.1) holds.

On the other hand, by Lemma D.4, we have, uniformly in $\|\mathbf{x}\| \leq C_{nT}$,

$$(0,1)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^{\top}(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^{\top}(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}\mathbb{Z}_{\mathbf{x},*} = O_P\left(1 + h_1^{-1}\tau(T)\right). \tag{B.8}$$

With (B.3), (B.5), (B.7) and (B.8), we have shown that (B.2) holds. ■

We next give the proof of Theorem 3.1 by making use of Lemma B.1.

Proof of Theorem 3.1. Note that for any small $\varepsilon > 0$,

$$\begin{aligned}
& \mathsf{P}\left(\max_{1\le i\le n}\max_{1\le t\le T}\|\mathbf{X}_{it}\| > C(nT)^{1/(2+\delta)}\right) \\
\le\ & \sum_{i=1}^{n}\sum_{t=1}^{T}\mathsf{P}\left(\|\mathbf{X}_{it}\| > C(nT)^{1/(2+\delta)}\right) \\
\le\ & \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\mathsf{E}\|\mathbf{X}_{it}\|^{2+\delta}}{(C^{2+\delta}\ nT)} < \varepsilon
\end{aligned}$$

if $C > \left(\mathsf{E}\|\mathbf{X}_{it}\|^{2+\delta}/\varepsilon\right)^{\frac{1}{(2+\delta)}}$. Hence, we need only to consider the case of $\max_{1\le i\le n}\max_{1\le t\le T}\|\mathbf{X}_{it}\| \leq C\ (nT)^{\frac{1}{2+\delta}}$.

By (2.7) and (B.1), we have

$$\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \left(\mathsf{E}\left[\mathbf{Z}_{it}\mathbf{Z}_{it}^{\top}\right]\right)^{-1}\mathsf{E}\left[g_1(\mathbf{X}_{it})g_1^{\top}(\mathbf{X}_{it})\right](\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_P(1). \tag{B.9}$$

Since we use the multivariate kernel $H(\cdot)$ for producing initial estimates of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$, (B.9) does not involve $\boldsymbol{\theta}$. From (B.9), we have

$$\widetilde{\boldsymbol{\beta}}_{k+1} - \boldsymbol{\beta}_0 = \left(\mathsf{E}\left[\mathbf{Z}_{it}\mathbf{Z}_{it}^{\top}\right]\right)^{-1}\mathsf{E}\left[g_1(\mathbf{X}_{it})g_1^{\top}(\mathbf{X}_{it})\right](\widetilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_0) + o_P(1), \tag{B.10}$$

where $\widetilde{\boldsymbol{\beta}}_k$ is the estimate of $\boldsymbol{\beta}_0$ from the k-th iteration in the process of producing initial estimates.

By Assumption **A4** in Appendix A, it can be shown that the matrix $\mathsf{E}\left[\mathbf{Z}_{it}\mathbf{Z}_{it}^{\top}\right] - \mathsf{E}\left[g_1(\mathbf{X}_{it})g_1^{\top}(\mathbf{X}_{it})\right]$ is positive definite. Similarly to the proofs of Lemma 1 and Theorem 1 in Xia & Härdle (2006), the eigenvalues of the matrix $\left(\mathsf{E}\left[\mathbf{Z}_{it}\mathbf{Z}_{it}^{\top}\right]\right)^{-1}\mathsf{E}\left[g_1(\mathbf{X}_{it})g_1^{\top}(\mathbf{X}_{it})\right]$ are all less than 1. Hence, after a sufficiently large number of iterations,

$$\widetilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_0 = o_P(1),$$

which implies that the first result in (3.1) holds.

By (2.7) and (B.2), we have

$$\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}^\top \boldsymbol{\theta}_0\right)^{-1} \left(1 - \boldsymbol{\theta}^\top \boldsymbol{\theta}_0\right) \boldsymbol{\theta}_0 + O(\|\zeta_{\boldsymbol{\beta}}\|) + o_P(1), \tag{B.11}$$

which implies that

$$\widetilde{\boldsymbol{\theta}} = \left(\boldsymbol{\theta}^\top \boldsymbol{\theta}_0\right)^{-1} \boldsymbol{\theta}_0 + O(\|\zeta_{\boldsymbol{\beta}}\|) + o_P(1). \tag{B.12}$$

Following the proof of Lemma 1 in Xia & Härdle (2006), we can also show that the second result in (3.1) holds. ■

Appendix C: Proofs of Theorems 3.2 and 3.3

For simplicity, let $\mathbb{W}_{it}(\boldsymbol{\theta})$ be defined as $\mathbb{W}_{it}$ with the weights in (2.7) replaced by those in (2.8), and $e_{it,*}$, $\mathbb{X}_{it,*}$, $\overline{\mathbb{X}}_{it,*}$, $\mathbb{V}_{it}$ and $\mathbb{Z}_{it,*}$ be defined in the same way as in Appendix B. Throughout this section, $\widehat{a}_{\mathbf{x}}$, $\widehat{a}_{it}$, $\widehat{b}_{\mathbf{x}}$, and $\widehat{b}_{it}$ are the local linear estimators obtained from (2.6) using the single–index weights defined in (2.9). As in Appendix B, $e_{\mathbf{x},*}$, $\mathbb{X}_{\mathbf{x},*}$, $\overline{\mathbb{X}}_{\mathbf{x},*}$, $\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})$, $\mathbb{V}_{\mathbf{x},*}$ and $\mathbb{Z}_{\mathbf{x},*}$ are defined similarly to $e_{it,*}$, $\mathbb{X}_{it,*}$, $\overline{\mathbb{X}}_{it,*}$, $\mathbb{W}_{it}(\boldsymbol{\theta})$, $\mathbb{V}_{it,*}$ and $\mathbb{Z}_{it,*}$ with $\mathbf{X}_{it}$ replaced by $\mathbf{x}$. Furthermore, define

$$d_{\mathbf{x}}(\boldsymbol{\theta}) = \left(\left(d_{11}^\top(\mathbf{x})\boldsymbol{\theta}\right)^2, \cdots, \left(d_{1T}^\top(\mathbf{x})\boldsymbol{\theta}\right)^2, \left(d_{21}^\top(\mathbf{x})\boldsymbol{\theta}\right)^2, \cdots, \left(d_{nT}^\top(\mathbf{x})\boldsymbol{\theta}\right)^2\right)^\top,$$

$$d_{\mathbf{x},*}(\boldsymbol{\theta}) = d_{\mathbf{x}}(\boldsymbol{\theta}) - D\left(D^\top \mathbb{W}(\boldsymbol{\theta}) D\right)^{-1} D^\top \mathbb{W}(\boldsymbol{\theta}) d_{\mathbf{x}}(\boldsymbol{\theta}),$$

where $d_{it}(\mathbf{x})$ was defined in the proof of Lemma B.1.

To prove the asymptotic distributions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ given in Theorem 3.2, we need the following asymptotic uniform expansions of $\widehat{a}_{\mathbf{x}}$ and $\widehat{b}_{\mathbf{x}}$ in $\{\mathbf{x} : \ \|\mathbf{x}\| \le C_{nT}\}$.

Lemma C.1. *Let Assumptions* **A1**–**A7** *and* **B1**–**B4** *in Appendix A hold. Then, uniformly in* $\{\mathbf{x} : \ \|\mathbf{x}\| \le C_{nT}\}$,

$$\widehat{a}_{\mathbf{x}} = a_{\mathbf{x}} + b_{\mathbf{x}} U_{\mathbf{x}}^\top(1)\zeta_{\boldsymbol{\theta}} + U_{\mathbf{x}}^\top(2)\zeta_{\boldsymbol{\beta}} + R_{\mathbf{x}}(1) + h_2^2 \eta''(\mathbf{x}^\top \boldsymbol{\theta}_0) U_{\mathbf{x}}(3) + O_P(h_2^3) \tag{C.1}$$

and

$$\widehat{b}_{\mathbf{x}} = b_{\mathbf{x}} + R_{\mathbf{x}}(2) + O_P\left(h_2^2 + \|\zeta_{\boldsymbol{\beta}}\| + \|\zeta_{\boldsymbol{\theta}}\|\right), \tag{C.2}$$

where

$$U_{\mathbf{x}}^\top(1) = (1, 0)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1} \overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \mathbb{X}_{\mathbf{x},*},$$

$$U_{\mathbf{x}}^\top(2) = (1, 0)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1} \overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \mathbb{Z}_{\mathbf{x},*},$$

$$U_{\mathbf{x}}(3) = (1, 0)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1} \overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) d_{\mathbf{x},*}(\boldsymbol{\theta}_0),$$

$$(R_{\mathbf{x}}(1), R_{\mathbf{x}}(2))^\top = \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1} \overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta}) \mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta}) \mathbb{V}_{\mathbf{x},*}.$$

Proof. By the definition of $\widehat{a}_{\mathbf{x}}$ and $\widehat{b}_{\mathbf{x}}$, we have

$$\begin{aligned}
\left(\widehat{a}_{\mathbf{x}},\ \widehat{b}_{\mathbf{x}}\right)^\top &= \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\mathbb{Z}_{\mathbf{x},*}(\boldsymbol{\beta}_0-\boldsymbol{\beta}) \\
&\quad+\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})D_{\mathbf{x},*}\boldsymbol{\alpha} \\
&\quad+\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) \\
&\quad+\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\mathbb{V}_{\mathbf{x},*} \\
&= \left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\mathbb{Z}_{\mathbf{x},*}(\boldsymbol{\beta}_0-\boldsymbol{\beta}) \\
&\quad+\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) \\
&\quad+\left(\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\overline{\mathbb{X}}_{\mathbf{x},*}(\boldsymbol{\theta})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^\top(\boldsymbol{\theta})\mathbb{W}_{\mathbf{x}}(\boldsymbol{\theta})\mathbb{V}_{\mathbf{x},*}, \qquad \text{(C.3)}
\end{aligned}$$

where $\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0)$ is defined in the same way as in Appendix B with $\mathbb{W}_{it}$ replaced by $\mathbb{W}_{it}(\boldsymbol{\theta})$.

By (C.3), Lemma D.3 in the supplementary document and the same Taylor expansion for $\eta(\mathbf{X}_{it}^\top\boldsymbol{\theta}_0)$ as in the proof of Lemma B.1, we complete the proofs of (C.1) and (C.2). ■

Before giving the proof of Theorem 3.2, we introduce the following notations. Let

$$\begin{aligned}
&\mathbb{U}_*(j) = \left(\mathbb{U}_{11,*}^\top(j),\cdots,\mathbb{U}_{1T,*}^\top(j),\mathbb{U}_{21,*}^\top(j),\cdots,\mathbb{U}_{nT,*}^\top(j)\right)^\top, \\
&\mathbb{U}_{it,*}^\top(j) = e_{nT}U_{it}^\top(j) - D(D^\top\mathbb{W}_{it}(\boldsymbol{\theta})D)^{-1}D^\top\mathbb{W}_{it}(\boldsymbol{\theta})e_{nT}U_{it}^\top(j), \quad j=1,2, \\
&\widetilde{\mathbb{V}}_* = \left(\widetilde{\mathbb{V}}_{11,*}^\top,\cdots,\widetilde{\mathbb{V}}_{1T,*}^\top,\widetilde{\mathbb{V}}_{21,*}^\top,\cdots,\widetilde{\mathbb{V}}_{nT,*}^\top\right)^\top,
\end{aligned}$$

where $U_{it}(j)$ is defined in the same way as $U_{\mathbf{x}}(j)$ with $\mathbf{x}$ replaced by $\mathbf{X}_{it}$, $\widetilde{\mathbb{V}}_{it,*}$ is defined as $\mathbb{V}_{it,*}$ with $\mathbb{V}$ replaced by $\mathbb{V}-R_{it}(1)e_{nT}$, and $R_{it}(1)$ is defined as $R_{\mathbf{x}}(1)$ with $\mathbf{x}$ replaced by $\mathbf{X}_{it}$.

Proof of Theorem 3.2. Since the main idea of the proof is a non–trivial extension of the proof of Theorem 1 in Xia & Härdle (2006), we still need to provide the following details.

By Lemma C.1 and following the proof of Lemma 6.3 in Xia & Härdle (2006), we have

$$\begin{aligned}
\begin{pmatrix} \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0 \end{pmatrix} &= \mathbb{J}_{nT}^{-1}\mathbb{M}_{nT} + \mathbb{J}_{nT}^{-1}\mathbb{U}_{nT}\begin{pmatrix} \boldsymbol{\beta}-\boldsymbol{\beta}_0 \\ \boldsymbol{\theta}-\boldsymbol{\theta}_0 \end{pmatrix} \\
&+ O_P\left(\|\zeta_r\|^2 + \left(h_2+h_2^{-1}\tau_{nT}(2)\right)\|\zeta_r\| + h_2^3\right), \qquad \text{(C.4)}
\end{aligned}$$

where $\zeta_r = \left(\zeta_{\boldsymbol\beta}^\top, \zeta_{\boldsymbol\theta}^\top\right)^\top$, $\zeta_{\boldsymbol\beta}$ and $\zeta_{\boldsymbol\theta}$ were defined in Appendix B,

$$\begin{aligned}
\mathbb{J}_{nT} &= \begin{pmatrix} \frac{1}{nT}\mathbb{Z}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{Z}_* & \frac{1}{nT}\mathbb{Z}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{X}_*, \\ \frac{1}{nT}\mathbb{X}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{Z}_* & \frac{1}{nT}\mathbb{X}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{X}_* \end{pmatrix}, \quad \mathbb{M}_{nT} = \frac{1}{nT}\begin{pmatrix} \mathbb{Z}_*^\top \\ \mathbb{X}_*^\top \end{pmatrix} \mathbb{W}(\boldsymbol\theta)\widetilde{\mathbb{V}}_*, \\
\mathbb{U}_{nT} &= \mathsf{diag}\left(\frac{1}{nT}\mathbb{Z}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{U}_*(2), \frac{1}{nT}\mathbb{X}_*^\top \mathbb{W}(\boldsymbol\theta)\mathbb{U}_*(1)\right).
\end{aligned}$$

Following the proof of Lemma D.4 in Appendix D of the supplemental document, we have

$$\mathbb{J}_{nT} \xrightarrow{P} \begin{pmatrix} \mathbb{J}_{11} & \mathbb{J}_{12} \\ \mathbb{J}_{12}^\top & \mathbb{J}_{22} \end{pmatrix} =: \mathbb{J} \tag{C.5}$$

and

$$\mathbb{U}_{nT} \xrightarrow{P} \mathsf{diag}\left(\mathsf{E}\left[v_{\boldsymbol\theta_0}(\mathbf{X}_{11})v_{\boldsymbol\theta_0}^\top(\mathbf{X}_{11})\right], \frac{1}{2}\mathbb{J}_{22}\right) =: \mathbb{U}, \tag{C.6}$$

where

$$\begin{aligned}
&\mathbb{J}_{11} = \mathsf{E}\left(\mathbf{Z}_{11}\mathbf{Z}_{11}^\top\right), \quad \mathbb{J}_{12} = \mathsf{E}\left[\mathbf{Z}_{11}\eta'(\mathbf{X}_{11}^\top\boldsymbol\theta_0)\left(\mu_{\boldsymbol\theta_0}(\mathbf{X}_{11}) - \mathbf{X}_{11}\right)^\top\right], \\
&\mathbb{J}_{22} = 2\mathsf{E}\left[\left(\eta'(\mathbf{X}_{11}^\top\boldsymbol\theta_0)\right)^2\left(\mathbf{X}_{11} - \mu_{\boldsymbol\theta_0}(\mathbf{X}_{11})\right)\left(\mathbf{X}_{11} - \mu_{\boldsymbol\theta_0}(\mathbf{X}_{11})\right)^\top\right], \\
&\mu_{\boldsymbol\theta}(\mathbf{x}) = \mathsf{E}\left(\mathbf{X}_{11}|\mathbf{X}_{11}^\top\boldsymbol\theta = \mathbf{x}^\top\boldsymbol\theta\right), \quad v_{\boldsymbol\theta}(\mathbf{x}) = \mathsf{E}\left(\mathbf{Z}_{11}|\mathbf{X}_{11}^\top\boldsymbol\theta = \mathbf{x}^\top\boldsymbol\theta\right).
\end{aligned}$$

Following the proof of Theorem 1 in Xia & Härdle (2006), it can be shown that $\widetilde{\mathbb{N}} := \left(\mathbb{J}^{-1}\right)^{1/2}\mathbb{U}\left(\mathbb{J}^{-1}\right)^{1/2}$ is a semi–positive definite matrix with rank $d+p-1$ and all eigenvalues being less than 1. Let $1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d+p-1} > 0$ be the eigenvalues of $\widetilde{\mathbb{N}}$.

Let $\mathbb{J}_{nT}(k)$ and $\mathbb{U}_{nT}(k)$ be the corresponding versions of $\mathbb{J}_{nT}$ and $\mathbb{U}_{nT}$ at the k–th iteration. Then, by (C.5) and (C.6), the eigenvalues of

$$\widetilde{\mathbb{N}}(k) := \left(\mathbb{J}_{nT}^{-1}(k)\right)^{1/2}\mathbb{U}_{nT}(k)\left(\mathbb{J}_{nT}^{-1}(k)\right)^{1/2}$$

satisfy $1 > \lambda_1(k) \geq \lambda_2(k) \geq \cdots \geq \lambda_{d+p-1}(k) > 0$, $\lambda_1(k) = \lambda_1 + o_P(1)$ for all $k \geq 1$.

Define $\mathbf{r}_k = \mathbb{J}_{nT}^{1/2}(k)\left(\left(\widehat{\boldsymbol\beta}_k - \boldsymbol\beta_0\right)^\top, \left(\widehat{\boldsymbol\theta}_k - \boldsymbol\theta_0\right)^\top\right)^\top$. By (C.4), we have

$$\mathbf{r}_{k+1} = \left(\mathbb{J}_{nT}^{-1}(k)\right)^{1/2}\mathbb{M}_{nT}(k) + \widetilde{\mathbb{N}}(k)\mathbf{r}_k + O_P\left(\left(\|\zeta_{\mathbf{r}_k}\| + h_2 + h_2^{-1}\tau_{nT}(2)\right)\|\zeta_{\mathbf{r}_k}\| + h_2^3\right), \tag{C.7}$$

which, together with the proof of Lemma D.5, implies

$$\left\|\zeta_{\mathbf{r}_{k+1}}\right\| \leq \tau_{nT}(2) + \lambda_1(k)\|\zeta_{\mathbf{r}_k}\| + c_0\|\zeta_{\mathbf{r}_k}\|^2 + c_0\left(h_2 + h_2^{-1}\tau_{nT}(2)\right)\|\zeta_{\mathbf{r}_k}\| + c_0 h_2^3, \tag{C.8}$$

where $c_0 > 1$ is a constant.

By Theorem 3.1, we have

$$\left\|\zeta_{\mathbf{r}_1}\right\| \le \frac{(1-\lambda_1)}{3c_0}. \tag{C.9}$$

By the definition of $\tau_{nT}(2)$ and **B4**, we have

$$\tau_{nT}(2) + c_0 h_2^3 \le \frac{(1-\lambda)^2}{9c_0} \quad \text{and} \quad c_0\left(h_2 + h_2^{-1}\tau_{nT}(2)\right) \le \frac{(1-\lambda_1)}{3}. \tag{C.10}$$

By (C.8)–(C.10) and the fact that $\lambda_1(k) \sim \lambda_1$, we have

$$\left\|\zeta_{\mathbf{r}_k}\right\| \le \frac{(1-\lambda_1)}{3c_0}. \tag{C.11}$$

for all $k \ge 1$. Then, following the proof of Theorem 1 in Xia & Härdle (2006), we have, for sufficiently large k,

$$\left\|\zeta_{\mathbf{r}_{k+1}}\right\| = O_P\left(\tau_{nT}(2) + h_2^3\right). \tag{C.12}$$

Note that

$$nTh_2^6 \to 0, \quad \sqrt{nT}\left(\tau_{nT}^2(2) + h_2^6\right) \to 0, \quad \sqrt{nT}\left(h_2 + h_2^{-1}\tau_{nT}(2)\right)\left(h_2^3 + \tau_{nT}(2)\right) \to 0. \tag{C.13}$$

By (C.5) and (C.6), we have

$$\mathbb{J}_{nT} - \mathbb{U}_{nT} \xrightarrow{P} \mathbf{\Sigma}_0. \tag{C.14}$$

By some standard arguments, it can be shown that the leading term of $\mathbb{M}_{nT}$ is

$$\mathbb{M}_{nT}^* = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\begin{pmatrix} \widetilde{\mathbf{Z}}_{it} \\ \eta'(\mathbf{X}_{it}^\top\boldsymbol{\theta}_0)\widetilde{\mathbf{X}}_{it}\end{pmatrix} v_{it}$$

as $nT \to \infty$. Applying the central limit theorem for α–mixing processes (see, for example, Theorem 2.21 of Fan and Yao 2003), we have

$$\sqrt{nT}\mathbb{M}_{nT}^* \xrightarrow{d} \mathsf{N}\left(0, \mathbf{\Sigma}_1\right). \tag{C.15}$$

By (C.4) and (C.13)–(C.15), we have shown that Theorem 3.2 holds. ■

Proof of Theorem 3.3. By the definition of local linear estimators, it is easy to show that

$$\begin{aligned}
& \widehat{\eta}(\mathbf{x}^\top\widehat{\boldsymbol{\theta}}) - \eta(\mathbf{x}^\top\boldsymbol{\theta}_0) \\
= \; & (1,0)\left(\mathbb{X}_{\mathbf{x},*}^\top(\widehat{\boldsymbol{\theta}})\mathbb{W}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}})\mathbb{X}_{\mathbf{x},*}(\widehat{\boldsymbol{\theta}})\right)^{-1}\mathbb{X}_{\mathbf{x},*}^\top(\widehat{\boldsymbol{\theta}})\mathbb{W}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}})\left(\mathbb{Y}_{\mathbf{x},*} - \mathbb{Z}_{\mathbf{x},*}\widehat{\boldsymbol{\beta}}\right) - \eta(\mathbf{x}^\top\boldsymbol{\theta}_0) \\
= \; & \left(S_{\mathbf{x}}^\top(\widehat{\boldsymbol{\theta}})\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) - \eta(\mathbf{x}^\top\boldsymbol{\theta}_0)\right) + S_{\mathbf{x}}^\top(\widehat{\boldsymbol{\theta}})\mathbb{V}_{\mathbf{x},*} + S_{\mathbf{x}}^\top(\widehat{\boldsymbol{\theta}})\mathbf{Z}_{\mathbf{x},*}\left(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}\right) \\
=: \; & \Pi_{nT}(1) + \Pi_{nT}(2) + \Pi_{nT}(3),
\end{aligned}$$

where $S_{\mathbf{x}}^{\top}(\widehat{\boldsymbol{\theta}}) = (1,0)\left(\overline{\mathbb{X}}_{\mathbf{x},*}^{\top}(\widehat{\boldsymbol{\theta}})\mathbb{W}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}})\overline{\mathbb{X}}_{\mathbf{x},*}(\widehat{\boldsymbol{\theta}})\right)^{-1}\overline{\mathbb{X}}_{\mathbf{x},*}^{\top}(\widehat{\boldsymbol{\theta}})\mathbb{W}_{\mathbf{x}}(\widehat{\boldsymbol{\theta}})$.

By Theorem 3.2, we have

$$\Pi_{nT}(3) = O_P\left((nT)^{-1/2}\right). \tag{C.16}$$

Let us now consider $\Pi_{nT}(1)$ and $\Pi_{nT}(2)$. Note that

$$\begin{aligned}
\Pi_{nT}(1) &= \left(S_{\mathbf{x}}^{\top}(\widehat{\boldsymbol{\theta}})\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) - \eta(\mathbf{x}^{\top}\boldsymbol{\theta}_0)\right) \\
&= \left(S_{\mathbf{x}}^{\top}(\boldsymbol{\theta}_0)\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) - \eta(\mathbf{x}^{\top}\boldsymbol{\theta}_0)\right) + \left(S_{\mathbf{x}}^{\top}(\widehat{\boldsymbol{\theta}}) - S_{\mathbf{x}}^{\top}(\boldsymbol{\theta}_0)\right)\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) \\
&=: \Pi_{nT}(1,1) + \Pi_{nT}(1,2).
\end{aligned}$$

Noticing that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_P\left((nT)^{-1/2}\right)$ by Theorem 3.2, we have

$$\Pi_{nT}(1,2) = O_P\left((nT)^{-1/2}\right).$$

Meanwhile, by the property of local linear smoothing, we have

$$\begin{aligned}
S_{\mathbf{x}}^{\top}(\boldsymbol{\theta}_0)\eta_{\mathbf{x},*}(\mathbb{X},\boldsymbol{\theta}_0) - \eta(\mathbf{x}^{\top}\boldsymbol{\theta}_0) &= \frac{1}{2}h_2^2\eta''(\mathbf{x}^{\top}\boldsymbol{\theta}_0)\int u^2K(u)du + o_P(h_2^2) \\
&= b_\eta(\mathbf{x}^{\top}\boldsymbol{\theta}_0) + o_P(h_2^2).
\end{aligned}$$

Hence, we have

$$\Pi_{nT}(1) = b_\eta(\mathbf{x}^{\top}\boldsymbol{\theta}_0) + o_P(h_2^2). \tag{C.17}$$

We next turn to the asymptotic distribution of $\Pi_{nT}(2)$. By **B1**, we have

$$K\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\widehat{\boldsymbol{\theta}}}{h_2}\right) = K\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\boldsymbol{\theta}_0}{h_2}\right) + K'\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\boldsymbol{\theta}^*}{h_2}\right)\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)}{h_2},$$

where $K'(\cdot)$ is the first–order derivative of $K(\cdot)$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 + \lambda^*(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ for some $0 < \lambda^* < 1$. Hence,

$$\begin{aligned}
&\frac{1}{\sqrt{nTh_2}}\sum_{i=1}^{n}\sum_{t=1}^{T}K\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\widehat{\boldsymbol{\theta}}}{h_2}\right)v_{it} = \frac{1}{\sqrt{nTh_2}}\sum_{i=1}^{n}\sum_{t=1}^{T}K\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\boldsymbol{\theta}_0}{h_2}\right)v_{it} \\
&+ \frac{1}{\sqrt{nTh_2}}\sum_{i=1}^{n}\sum_{t=1}^{T}K'\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}\boldsymbol{\theta}^*}{h_2}\right)\frac{(\mathbf{X}_{it}-\mathbf{x})^{\top}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)}{h_2}v_{it} \\
=:\ & \Pi_{nT}(2,1) + \Pi_{nT}(2,2).
\end{aligned}$$

By Theorem 3.2 and following the same argument as in the proof of Lemma D.5 of the supplemental document, we have

$$\Pi_{nT}(2,2) = o_P(1), \tag{C.18}$$

which implies that the leading term of $\frac{1}{\sqrt{nTh_2}} \sum_{i=1}^{n} \sum_{t=1}^{T} K\left(\frac{(\mathbf{X}_{it}-\mathbf{x})^\top \widehat{\boldsymbol{\theta}}}{h_2}\right) v_{it}$ is $\Pi_{nT}(2,1)$.

In a similar way to the proof of Theorem 2.21 of Fan and Yao (2003), applying Doob's large–block and small–block argument in the proof of asymptotic distribution for nonparametric kernel estimator under α–mixing dependence, we can show that

$$\Pi_{nT}(2,1) \xrightarrow{d} \mathsf{N}\left(0, \sigma_*^2\right), \tag{C.19}$$

where $\sigma_*^2 = \sigma_{\boldsymbol{\theta}_0}^2(\mathbf{x}^\top \boldsymbol{\theta}_0) f_{\boldsymbol{\theta}_0}(\mathbf{x}^\top \boldsymbol{\theta}_0)\nu_0$. By (C.18), (C.19) and the uniform convergence results in Appendix D of the supplemental document, we have

$$\Pi_{nT}(2) \xrightarrow{d} \mathsf{N}\left(0, \sigma_\eta^2(\mathbf{x}^\top \boldsymbol{\theta}_0)\right), \tag{C.20}$$

By (C.16), (C.17) and (C.20), Theorem 3.3 holds. ■

References

An, H. Z. & 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: Oxford.

Bai, Y., Fung, W. & Zhu, Z. (2008). Penalized quadratic inference functions for single–index models with longitudinal data. *Journal of Multivariate Analysis* **100**, 152-161.

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

Baltagi, B. & Li, D. (2002) Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance* **3**, 103-116.

Cai, Z. & Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory* **24**, 1321-1342.

Carroll, R. J., Fan, J. Gijbels, I. & Wand, M. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* **92**, 477-489.

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

Fan, J. & 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.

Henderson, D., Carroll, R. & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* **144**, 257-275.

Henderson, D. & Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters* **88**, 403-407.

Hjellvik, V., Chen, R. & Tjøstheim, D. (2004). Nonparametric estimation and testing in panels of intercorrelated time series. *Journal of Time Series Analysis* **25**, 831-872.

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

Li, Q. & Racine, J. (2007). *Nonparametric Econometrics.* Princeton University Press: New Jersey.

Li, Q. & Stengos, T. (1996). Semiparametric estimation of partially linear regression models. *Journal of Econometrics* **71**, 389-397.

Liang, H., Liu, X., Li, R. & Tsai, C. (2010). Estimation and testing for partially linear single–index models. *Annals of Statistics* **38**, 3811-3836.

Mammen, E., Støve, B. & Tjøstheim, D. (2009). Nonparametric additive models for panels of time series. *Econometric Theory* **25**, 442-481.

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

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

Ruppert, D., Wand, M. & Carroll, R. J. (2003). Semiparametric Regression. Cambridge University Press.

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

Sun, Y., Carroll, R. J. & Li, D. (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics* **25**, 101-129.

Ullah, A. and Roy, N. (1998). Nonparametric and semiparametric econometrics of panel data. *Handbook of Applied Economics Statistics*, Ullah, A., Giles, D.E.A. (Eds.). Marcel Dekker, New York, pp. 579C604.

Wang, J., Xue, L., Zhu, L. & Chong, Y. (2010). Estimation for a partially–linear single–index model. *Annals of Statistics* **38**, 246-274.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single–index model. *Econometric Theory* **22** 1112-1137.

Xia, Y. & Härdle, W. (2006). Semi–parametric estimation of partially linear single–index models. *Journal of Multivariate Analysis* **97**, 1162-1184.

Xia, Y., Tong, H. & Li, W. K. (1999). On extended partially linear single–index models. *Biometrika* **86**, 831–842.

Xia, Y., Tong, H., Li, W. K. & Zhu L. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society Series B* **64**, 363-410.

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