Statistica Sinica **23** (2013), 271-303
doi:http://dx.doi.org/10.5705/ss.2010.202

EFFICIENT ESTIMATION IN PANEL DATA PARTIALLY ADDITIVE LINEAR MODEL WITH SERIALLY CORRELATED ERRORS

Jinhong You and Xian Zhou

Shanghai University of Finance and Economics and Macquarie University

Abstract: The partially linear additive model arises in many scientific endeavors. In this paper, we look at inference given panel data and a serially correlated error component structure. By combining polynomial spline series approximation with least squares and the estimation of correlation, we propose a weighted semiparametric least squares estimator (WSLSE) for the parametric components, and a weighted polynomial spline series estimator (WPSSE) for the nonparametric components. The WSLSE is shown to be asymptotically normal and more efficient than the unweighted one. In addition, based on the WSLSE and WPSSE, a two-stage local polynomial estimator (TSLLE) of the nonparametric components is proposed that takes both contemporaneous correlation and additive structure into account. The TSLLE has several advantages, including higher asymptotic efficiency and an oracle property that achieves the same asymptotic distribution of each additive component as if the parametric and other nonparametric components were known with certainty. Some simulation studies were conducted to illustrate the finite sample performance of the proposed procedure. An example of application to a set of panel data from a wage study is illustrated.

Key words and phrases: Additive model, efficient estimation, error component, panel data, semiparametric, serial correlation.

1. Introduction

Panel data arise frequently in biological and economic applications in such as surveys of workers, households, countries, firms, patients, etc., over several time periods. Two well-known examples of panel data in the United States are the Panel Study of Income Dynamics (PSID), collected by the Institute for Social Research at the University of Michigan (`http://psidonline.isr.umich.edu`), and the National Longitudinal Surveys (NLS) sponsored by the Bureau of Labor Statistics (`http://www.bls.gov/nld/home.htm`). See Baltagi (2008) for the details. Various parametric models and statistical tools have been developed for panel data analysis; see, for instance, Ahn and Schmidt (2000), Verbeke and Molenberghs (2000), Hsiao (2003), Baltagi (2008) and the references therein. While parametric models are very useful for analyzing panel data and providing parsimonious descriptions of the relationships between the response variables and their covariates, they are often subject to the risk of modeling biases. To relax the assumptions of parametric forms, Ruckstuhl, Welsh, and Carroll (2000) and Wang (2003) proposed nonparametric panel data regression models that allow one to explore possible hidden structures in the data and to reduce modeling biases of the traditional parametric methods. Such nonparametric models, however, have several shortcomings including the *curse of dimensionality*, difficulty of interpretation, lack of extrapolation capability, and so on. To overcome these shortcomings, *semiparametric* panel data regression models, especially the partially linear panel data regression models, have been considered recently that embody a compromise between a general nonparametric approach and a fully parametric specification; see Horowitz and Markatou (1996), Li and Ullah (1998), Ullah and Mundra (2002), You and Zhou (2006, 2009) You, Zhou, and Zhou (2010), to mention only a few. However, when the number of the nonparametric covariates is large, the curse of dimensionality is still a problem. In order to ease it further, we propose the *semiparametric panel data partially linear additive model*

$$Y_{it} = \mathbf{X}_{it}^{\tau}\boldsymbol{\beta} + \alpha_1(U_{it1}) + \cdots + \alpha_q(U_{itq}) + \varepsilon_{it}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \quad (1.1)$$

where Y_{it}'s are responses, $(\mathbf{X}_{it}^{\tau}, \mathbf{U}_{it}^{\tau})^{\tau} = (X_{it1}, \ldots, X_{itp}, U_{it1}, \ldots, U_{itq})^{\tau}$ are explanatory variables, $\boldsymbol{\beta}$ is an unknown p-dimensional parametric vector, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^{\tau}$ is an unknown q-dimensional function vector, ε_{it} are random errors, and the superscript $(^{\tau})$ denotes the transpose of a vector or matrix. Here $((\mathbf{X}_{i1}^{\tau}, \mathbf{U}_{i1}^{\tau})^{\tau}, \ldots, (\mathbf{X}_{iT}^{\tau}, \mathbf{U}_{iT}^{\tau})^{\tau})$ are i.i.d. random vectors over i. We further assume that the random errors ε_{it} in (1.1) follow a serially correlated error component structure:

$$\varepsilon_{it} = \mu_i + \nu_{it}, \quad \nu_{it} = \rho\nu_{i,t-1} + e_{it}, \quad |\rho| < 1, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \quad (1.2)$$

where μ_i and e_{it} are i.i.d. random variables with mean zero and variances σ_{μ}^2 and σ_e^2, respectively, and ρ is an unknown autoregressive coefficient. Obviously, error structure (1.2) is more general than the classical one-way error component structure that assumes the same correlation between e_{it} and $e_{it'}$ no matter how far t' is from t (cf., Baltagi (2008)). Under (1.2), the observations $Y_{i1}, \ldots, Y_{iT}$ from the same individual i are allowed to be dependent. Typically, we choose $E\alpha_s(U_{its}) = 0$ as our identification condition.

We use a data example to demonstrate the need for model (1.1)−(1.2). The National Longitudinal Surveys (NLS) conducted by the US Department of Labor have a database on women aged between 14 and 24 in 1968. This example is based on a subset of the survey data conducted in 1983, 1985 and 1987; Section 15.4.3 of Carter Hill, Griffiths and Lim (2008). In this dataset, observations on five variables were collected from 716 women. They include the year interviewed, log(wage/GNP deflator), total work experience, job tenure and whether getting a college degree. It is well known that total work experience and job tenure affect the log(wage/GNP deflator) nonlinearly. Carter Hill, Griffiths and Lim (2008) used quadratic forms to model the nonlinearity of both factors. By applying model (1.1)−(1.2), we can avoid this kind of parametric assumption and let the data speak for themselves.

It is easy to see that model (1.1)−(1.2) includes many usual parametric, nonparametric and semiparametric models as special cases. For example, when $q = 1$, model (1.1)−(1.2) reduces to the panel data partially linear regression model. Many researchers have explored the panel data partially linear regression model (cf., Zeger and Diggle (1994), Roy (1997), Mundra (1997), Honore (1992), Kniesner and Li (1994)). When $\alpha_1(\cdot) \equiv 0, \ldots, \alpha_q(\cdot) \equiv 0$, the model becomes the well-known panel data linear regression model (cf., Ahn and Schmidt (2000), Baltagi (2008)). When $\boldsymbol{\beta} = 0$, the model reduces to the panel data nonparametric additive regression model (cf., You and Zhou (2007)). When $T \equiv 1$, it is the non-panel structure partially linear additive model that has been studied by Fan, Härdle, and Mammen (1998), Li (2000), and Fan and Li (2003), among others.

For model (1.1)−(1.2), our goal is to develop a satisfactory procedure for estimating the unknown parametric and nonparametric components that meets four criteria: (i) it takes both of the error components and additive structures into account; (ii) it is computationally efficient, (iii) theoretically reliable, and (iv) intuitively appealing. The last three criteria have been used by Wang and Yang (2007) and Liu and Yang (2009) for non-cluster additive models. To meet the criteria we take the following approach. By applying a polynomial spline approximation to the nonparametric components and estimating the error structure, we construct a weighted semiparametric least squares estimator (WSLSE) for the parametric components $\boldsymbol{\beta}$ that achieves $\sqrt{n}$-consistency without under-smoothing and is asymptotically efficient. Then by applying the under-smoothing technique, and taking both the contemporaneous correlation and additive structures into account, we propose a two-stage local linear estimator (TSLLE) for the unknown nonparametric functions $\boldsymbol{\alpha}(\cdot)$. This TSLLE has several advantages, including higher asymptotic efficiency than the one neglecting the contemporaneous correlation and an oracle property that achieves the same asymptotic distribution of each additive component as if the parametric and other nonparametric components were known with certainty.

The rest of this paper is organized as follows. Section 2 introduces several initial estimators. In Section 3, we develop a class of weighted semiparametric least squares estimators for the parametric and nonparametric components. A two-stage local polynomial estimator of the nonparametric components is proposed in Section 4. Section 5 reports on some simulation studies. An application of the model and estimation procedure to a set of economic data is illustrated in Section 6. Concluding remarks are given in Section 7. Proofs of the main results are relegated to Appendix.

Throughout this paper we assume large n and small T. This is typical in labor or consumer panel data situation (cf., Baltagi (2008)). In addition, it should be noted that our results can be extended to a higher order autoregressive structure on the ν_{it} although we focus on AR(1) in this paper.

2. Several Initial Estimators

Polynomial splines are piecewise polynomials joined together smoothly at a set of interior points (*knots*). A polynomial spline of degree $d_s \geq 0$ on $\mathcal{U}_s$ with knot sequence $\eta_{s0} < \eta_{s1} < \cdots < \eta_{s,M_s+1}$, where η_{s0} and η_{s,M_s+1} are the two end points of the interval $\mathcal{U}_s$, is a function that is a polynomial of degree d_s on each of the intervals $[\eta_{sm}, \eta_{s,m+1})$, $0 \leq m \leq M_s - 1$, and $[\eta_{sM_s}, \eta_{s,M_s+1}]$, and globally has continuous $d_s - 1$ continuous derivatives for $d_s \geq 1$. A piecewise constant function, linear spline, quadratic spline and cubic spline corresponds to $d_s = 0, 1, 2, 3$, respectively. The collection of spline functions of a particular degree and knot sequence form a linear space; books by de Boor (1978) and Schumaker (1981) are good references. Without loss of generality, we take all intervals $\mathcal{U}_s = [0, 1]$, $s = 1, \ldots, q$.

We approximate each $\alpha_s(u)$ by some spline function:

$$\alpha_s(u) \approx \sum_{l=1}^{\kappa_s} \theta_{sl}\zeta_{sl}(u), \quad s = 1, \ldots, q,$$

where $\{\zeta_{sl}(\cdot)\}_{l=1}^{\kappa_s}$ is a basis for a linear space $\mathcal{G}_s$ of spline function on $[0, 1]$ with a fixed degree, and knot sequence, $\boldsymbol{\theta}_s = (\theta_{s1}, \ldots, \theta_{s\kappa_s})^\tau$ is an unknown κ_s vector; the κ_s play the role of smoothing parameters. Thus, model (1.1) can be approximated by

$$Y_{it} \approx \mathbf{X}_{it}^\tau \boldsymbol{\beta} + \sum_{s=1}^{q}\sum_{l=1}^{\kappa_s} \theta_{sl}\zeta_{sl}(U_{its}) + \varepsilon_{it}, \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T. \quad (2.1)$$

If we denote

$$\begin{aligned}
\boldsymbol{\zeta}(\mathbf{U}_{it\cdot}) &= (\zeta_{11}(U_{it1}), \ldots, \zeta_{1\kappa_1}(U_{it1}), \ldots, \zeta_{q1}(U_{itq}), \ldots, \zeta_{q\kappa_q}(U_{itq}))^\tau, \\
\mathbf{D}_i &= (\boldsymbol{\zeta}^\tau(\mathbf{U}_{i1\cdot}), \ldots, \boldsymbol{\zeta}^\tau(\mathbf{U}_{iT\cdot}))^\tau, \quad \mathbf{D} = (\mathbf{D}_1, \ldots, \mathbf{D}_n)^\tau, \\
\mathbf{Y}_i &= (Y_{i1}, \ldots, Y_{iT})^\tau, \quad \mathbf{Y} = (\mathbf{Y}_1^\tau, \ldots, \mathbf{Y}_n^\tau)^\tau, \quad \boldsymbol{\theta} = (\boldsymbol{\theta}_1^\tau, \ldots, \boldsymbol{\theta}_q^\tau)^\tau, \\
\boldsymbol{\varepsilon}_i &= (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\tau \quad \text{and} \quad \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\tau, \ldots, \boldsymbol{\varepsilon}_n^\tau)^\tau,
\end{aligned}$$

(2.1) can be written as

$$\mathbf{Y} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{2.2}$$

Let $\mathbf{M}_{\mathbf{D}} = \mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}$. Then $\mathbf{M}_{\mathbf{D}}\mathbf{D}\boldsymbol{\theta} = \mathbf{D}\boldsymbol{\theta} - \mathbf{D}\boldsymbol{\theta} = \mathbf{0}$, and (2.2) leads to

$$\mathbf{M}_{\mathbf{D}}\mathbf{Y} \approx \mathbf{M}_{\mathbf{D}}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}_{\mathbf{D}}\boldsymbol{\varepsilon}. \tag{2.3}$$

If we write $\mathbf{M}_{\mathbf{D}}\boldsymbol{\varepsilon}$ for the residuals, then model (2.3) is a version of the usual linear regression. An estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_n = (\mathbf{X}^{\tau}\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}^{\tau}\mathbf{M}_{\mathbf{D}}\mathbf{Y}.$$

Substitute $\widehat{\boldsymbol{\beta}}_n$ into (2.2) we get an estimator of $\boldsymbol{\theta}$ as $\widehat{\boldsymbol{\theta}}_n = (\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_n)$. This implies that the polynomial spline series estimator of $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^{\tau}$ is

$$\widehat{\boldsymbol{\alpha}}_n(u) = (\widehat{\alpha}_{1n}(u), \ldots, \widehat{\alpha}_{qn}(u))^{\tau} = \boldsymbol{\zeta}(u)\widehat{\boldsymbol{\theta}}_n,$$

a nonparametric projection estimator, where $\boldsymbol{\zeta}(\cdot) = (\zeta_{11}(\cdot), \ldots, \zeta_{1\kappa_1}(\cdot), \ldots, \zeta_{q1}(\cdot), \ldots, \zeta_{q\kappa_q}(\cdot))^{\tau}$. In addition, it is not difficult to calculate that

$$\boldsymbol{\Sigma} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\tau}) = \sigma_{\mu}^2(\mathbf{I}_n \otimes \boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\tau}) + \sigma_e^2(\mathbf{I}_n \otimes \mathbf{A}),$$

where $\boldsymbol{\iota}_T$ is a vector of ones of dimension T and

$$\mathbf{A} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}.$$

Based on $\widehat{\boldsymbol{\beta}}_n$ and $(\widehat{\alpha}_{1n}(\cdot), \ldots, \widehat{\alpha}_{qn}(\cdot))^{\tau}$, we obtain the estimated residuals

$$\widehat{\varepsilon}_{it} = Y_{it} - \mathbf{X}_{it}^{\tau}\widehat{\boldsymbol{\beta}}_n - \widehat{\alpha}_{1n}(U_{it1}) - \cdots - \widehat{\alpha}_{qn}(U_{itq}), \quad i = 1, \ldots, n,\ t = 1, \ldots, T.$$

As in Baltagi and Li (1991), we can consistently estimate ρ, σ_e^2 and σ_{μ}^2 by

$$\widehat{\rho}_n = \frac{\widehat{Q}_0 - \widehat{Q}_2}{\widehat{Q}_0 - \widehat{Q}_1} - 1 = \frac{\widehat{Q}_1 - \widehat{Q}_2}{\widehat{Q}_0 - \widehat{Q}_1},$$

$$\widehat{\sigma}_{en}^2 = \frac{1}{n(T-1)}\widehat{\boldsymbol{\varepsilon}}^{\tau}(\mathbf{I}_n \otimes \widehat{\mathbf{C}}^{\tau})(\mathbf{I}_n \otimes \mathbf{E}_T^{\widehat{\rho}_n})(\mathbf{I}_n \otimes \widehat{\mathbf{C}})\widehat{\boldsymbol{\varepsilon}},$$

$$\widehat{\sigma}_{\mu n}^2 = \frac{1}{\{\widehat{\rho}_n(2-T) + T\}(1 - \widehat{\rho}_n)}\left\{\frac{1}{n}\widehat{\boldsymbol{\varepsilon}}^{\tau}(\mathbf{I}_n \otimes \widehat{\mathbf{C}}^{\tau})(\mathbf{I}_n \otimes \bar{\mathbf{J}}_T^{\widehat{\rho}_n})(\mathbf{I}_n \otimes \widehat{\mathbf{C}})\widehat{\boldsymbol{\varepsilon}} - \widehat{\sigma}_{en}^2\right\},$$

where

$$\widehat{Q}_j=\sum_{i=1}^{n}\sum_{t=1}^{T-2}\frac{\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{i,t+j}}{n(T-2)},\quad \widehat{\mathbf{C}}=\begin{bmatrix}\sqrt{1-\widehat{\rho}_n^2} & 0 & 0 & \cdots & 0 & 0\\ -\widehat{\rho}_n & 1 & 0 & \cdots & 0 & 0\\ 0 & -\widehat{\rho}_n & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & -\widehat{\rho}_n & 1 & 0\\ 0 & 0 & 0 & 0 & -\widehat{\rho}_n & 1\end{bmatrix}, \tag{2.4}$$

$\bar{\mathbf{J}}_T^{\widehat{\rho}_n}=\dfrac{1}{(T-1)+(1+\widehat{\rho}_n)/(1-\widehat{\rho}_n)}\boldsymbol{\iota}^{\widehat{\rho}_n}\boldsymbol{\iota}^{\widehat{\rho}_n\tau}$ with

$\boldsymbol{\iota}^{\widehat{\rho}_n}=(\sqrt{(1+\widehat{\rho}_n)/(1-\widehat{\rho}_n)},\boldsymbol{\iota}_{T-1}^{\tau})^{\tau}$ and $\mathbf{E}_T^{\widehat{\rho}_n}=\mathbf{I}_T-\bar{\mathbf{J}}_T^{\widehat{\rho}_n}$.

In order to present the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n,(\widehat{\alpha}_{1n}(\cdot),\ldots,\widehat{\alpha}_{qn}(\cdot))^{\tau},\widehat{\rho}_n$, $\widehat{\sigma}_{\mu n}^2$ and $\widehat{\sigma}_{en}^2$, and other estimators proposed in the following sections, we first introduce some notation and technical assumptions.

Definition 1. A function $\xi(u_1,\ldots,u_q)$ belongs to an additive class of functions $\mathcal{G}$ ($\xi\in\mathcal{G}$) if (i) $\xi(u_1,\ldots,u_q)=\sum_{s=1}^{q}\xi_s(u_s)$, $\xi_s(u_s)$ is continuous in its support $\mathcal{U}_s$, where $\mathcal{U}_s$ is a compact subset of $\mathcal{R}$ for $s=1,\ldots,q$, (ii) $\sum_{t=1}^{T}\sum_{s=1}^{q}E\{\xi_s(U_{its})\}^2<\infty$, and (iii) $E\{\xi_s(U_{its})\}=0$ for $s=1,\ldots,q$ and $t=1,\ldots,T$.

Take $\psi_{tj}(u_1,\ldots,u_q)=E(X_{itj}|U_{it1}=u_1,\ldots,U_{itq}=u_q)$ and let $h_{tj}(u_1,\ldots,u_q)$ be the projection of $\psi_{tj}(u_1,\ldots,u_q)$ onto $\mathcal{G}$. According to Li (2000), $h_{tj}(u_1,\ldots,u_q)=\sum_{s=1}^{q}h_{tjs}(u_s)\in\mathcal{G}$ is the solution of the minimization problem

$$\begin{aligned}&E\left\{[\psi_{tj}(U_{it1},\ldots,U_{itq})-h_{tj}(U_{it1},\ldots,U_{itq})]^2\right\}\\&=\inf_{\xi_{tj}=\sum_{s=1}^{q}\xi_{tjs}}E\Big\{\Big[\psi_{tj}(U_{it1},\ldots,U_{itq})-\sum_{s=1}^{q}\xi_{tjs}(U_{its})\Big]^2\Big\}.\end{aligned}$$

Assumption 1. For fixed t and s, the U_{its}'s have a distribution which has bounded support $\mathcal{U}_{ts}$ and a Lipschitz continuous density function $p_{ts}(\cdot)$ such that

$$0<\inf_{\mathcal{U}_{ts}}p_{ts}(\cdot)\le\sup_{\mathcal{U}_{ts}}p_{ts}(\cdot)<\infty.$$

Assumption 2. For fixed t and j, $\mathbf{X}_{itj}$ satisfies that $E\|\mathbf{X}_{itj}\|^{2s}<\infty$ for some $s>2$. In addition, $E|\mu_i|^s<\infty$ and $E|\mu_{it}|^s<\infty$.

Assumption 3. For fixed t, j, and s, $\alpha_s(\cdot)$ and $h_{tjs}(\cdot)$ have continuous second derivatives on $\mathcal{U}_{ts}$.

Assumption 4. (i) $\limsup_n\dfrac{\max(\kappa_1,\ldots,\kappa_q)}{\min(\kappa_1,\ldots,\kappa_q)}<\infty$;
(ii) $\max(\kappa_1,\ldots,\kappa_q)=o(n^{1/2})$ and $n^{1/2}[\min(\kappa_1,\ldots,\kappa_q)]^{-4}=o(1)$.

Assumption 5. There exist positive constants c_1 and c_2 such that $c_1\mathbf{I}_T \leq E\{\mathbf{\Pi}_{i\cdot}\mathbf{\Pi}_{i\cdot}^\tau\} \leq c_2\mathbf{I}_T$, where $\mathbf{\Pi}_{i\cdot} = (\mathbf{\Pi}_{i1}, \ldots, \mathbf{\Pi}_{iT})$ with $\mathbf{\Pi}_{it} = \mathbf{X}_{it} - \sum_{s=1}^q \mathbf{H}_{ts}(U_{its})$, and $\mathbf{H}_{ts} = (h_{t1s}(U_{its}), \ldots, h_{tps}(U_{its}))^\tau$.

Let $\|a_s\|_{L_2}$ denote the L_2 norm of a square integrable function $a_s(u)$ on $\bigcup_{t=1}^T \mathcal{U}_{ts}$, and φ_s be the L_∞ distance between $\alpha_s(\cdot)$ and $\mathcal{G}$,

$$\varphi_s = \operatorname{dist}(\alpha_s, \mathcal{G}) = \inf_{a\in\mathcal{G}} \sup_{u\in\bigcup_{t=1}^T \mathcal{U}_{ts}} |\alpha_s(u) - a(u)|.$$

Theorem 1. *Suppose that Assumptions* 1 *to* 5 *hold. Then*

(i) $\sqrt{nT}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to_D N(0, \mathbf{\Omega}_1^{-1}\mathbf{\Omega}_2\mathbf{\Omega}_1^{-1})$ as $n \to \infty$, where $\mathbf{\Omega}_1 = T^{-1}E\{\mathbf{\Pi}_{i\cdot}\mathbf{\Pi}_{i\cdot}^\tau\}$, $\mathbf{\Pi}_{i\cdot} = (\mathbf{\Pi}_{i1}, \ldots, \mathbf{\Pi}_{iT})$, $\mathbf{\Omega}_2 = T^{-1}E\{\mathbf{\Pi}_{i\cdot}\mathbf{\Sigma}_0\mathbf{\Pi}_{i\cdot}^\tau\}$ and $\mathbf{\Sigma}_0 = E(\boldsymbol{\varepsilon}_{i\cdot}\boldsymbol{\varepsilon}_{i\cdot}^\tau) = \sigma_\mu^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T^\tau + \sigma_e^2\mathbf{A}$.

(ii) $\max_{1\leq s\leq q} \|\widehat{\alpha}_{sn} - \alpha_s\|_{L_2}^2 = O_p\left(\max_{1\leq s\leq q} \kappa_s n^{-1} + \max_{1\leq s\leq q} \varphi_s^2\right)$.

Theorem 2. *Under Assumptions* 1 *to* 5, *as* $n \to \infty$,

(i) $\sqrt{nT}(\widehat{\rho}_n - \rho) \to_D N(0, \omega_\rho)$, where

$$\omega_\rho = \frac{2T}{T-2}\left(\frac{\sigma_e^2}{1+\rho}\right)^{-2} \frac{(1-\rho^2)\sigma_\mu^2\sigma_e^2 + \sigma_e^4}{(1-\rho^2)}.$$

(ii) $\sqrt{nT}(\widehat{\sigma}_{en}^2 - \sigma_e^2) \to_D N(0, \omega_e)$, where

$$\omega_e = \frac{T}{(T-1)^2}\operatorname{Var}\left\{(\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})\mathbf{E}_T^\rho(\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})^\tau\right\}$$

and $\mathbf{E}_T^\rho$ has the form of $\mathbf{E}_T^{\widehat{\rho}_n}$ except that $\widehat{\rho}_n$ is replaced by ρ.

(iii) $\sqrt{nT}(\widehat{\sigma}_{\mu n}^2 - \sigma_\mu^2) \to_D N(0, \omega_\mu)$, where

$$\begin{aligned}\omega_\mu = &\frac{T}{\{\rho(2-T) + T(1-\rho)\}^2} \\ &\cdot\operatorname{Var}\Big\{\Big(\sqrt{1-\rho^2}(\mu_i + \nu_{i1}), (1-\rho)\mu_i + e_{i2}, \ldots, (1-\rho)\mu_i + e_{iT}\Big) \\ &\cdot\left(\bar{\mathbf{J}}_T^\rho - \frac{1}{T-1}\mathbf{E}_T^\rho\right) \\ &\cdot\Big(\sqrt{1-\rho^2}(\mu_i + \nu_{i1}), (1-\rho)\mu_i + e_{i2}, \ldots, (1-\rho)\mu_i + e_{iT}\Big)^\tau\Big\}\end{aligned}$$

and $\bar{\mathbf{J}}_T^\rho$ has the form of $\bar{\mathbf{J}}_T^{\widehat{\rho}_n}$ except that $\widehat{\rho}_n$ is replaced by ρ.

Remark 1. Baltagi and Li (1991) established the consistency of $\widehat{\rho}_n, \widehat{\sigma}_{en}^2$, and $\widehat{\sigma}_{\mu n}^2$ under the setting of panel data linear regression. We here show that they are root-n consistent and asymptotically normal under a more general setting.

In order to obtain $\widehat{\boldsymbol{\beta}}_n$, $(\widehat{\alpha}_{1n}(\cdot),\ldots,\widehat{\alpha}_{qn}(\cdot))^\tau$, $\widehat{\rho}_n$, $\widehat{\sigma}^2_{\mu n}$, and $\widehat{\sigma}^2_{en}$, we need to select the degrees of splines and the numbers and locations of knots. Due to computational complexity, it is often impractical to automatically select all three components. Similar to Rice and Wu (2001), we use splines with equally spaced knots and fixed degrees and select κ_s, the numbers of knots, by a data-driven cross-validation method. Here $\kappa_s = M_s + 1 + d_s$ for $1 \le s \le q$.

$\widehat{\boldsymbol{\beta}}_n$ and $(\widehat{\alpha}_{1n}(\cdot),\ldots,\widehat{\alpha}_{qn}(\cdot))^\tau$ do not take contemporaneous correlation into account, hence may not be asymptotically efficient. We construct more efficient estimators by implementing the estimated error variances and contemporaneous correlation in the following sections.

3. Weighted Semiparametric Least Squares Estimation

Let

$$\widehat{\boldsymbol{\Sigma}}^{-1} = \mathbf{I}_n \otimes \left[\frac{1}{\widehat{\sigma}^2_{en}}\left(\widehat{\mathbf{A}}^{-1} - \frac{\widehat{\sigma}^2_{\mu n}}{(\widehat{\sigma}^2_{en} + \boldsymbol{\iota}_T^\tau \widehat{\mathbf{A}}^{-1}\boldsymbol{\iota}_T^\tau)}\widehat{\mathbf{A}}^{-1}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^\tau\widehat{\mathbf{A}}^{-1}\right)\right],$$

where $\widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{C}}\widehat{\mathbf{C}}^\tau$ and $\widehat{\mathbf{C}}$ is given by (2.4). Pre-multiplying (2.2) by $\widehat{\boldsymbol{\Sigma}}^{-1/2}$ leads to

$$\widehat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Y} \approx \widehat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \widehat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{D}\boldsymbol{\theta} + \widehat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\varepsilon}. \tag{3.1}$$

Based on (3.1) we can obtain the weighted semiparametric least squares estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as

$$\widehat{\boldsymbol{\beta}}^w_n = (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X})^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}}\mathbf{Y} \quad \text{and} \quad \widehat{\boldsymbol{\theta}}^w_n = (\mathbf{D}^\tau\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^w_n),$$

respectively, where $\mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}} = \widehat{\boldsymbol{\Sigma}}^{-1} - \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{D}(\mathbf{D}^\tau\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\widehat{\boldsymbol{\Sigma}}^{-1}$. This gives the weighted polynomial spline series estimator of $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot),\ldots,\alpha_q(\cdot))^\tau$ as

$$\widehat{\boldsymbol{\alpha}}^w_n(u) = (\widehat{\alpha}^w_{1n}(u),\ldots,\widehat{\alpha}^w_{qn}(u))^\tau = \boldsymbol{\zeta}(u)\widehat{\boldsymbol{\theta}}^w_n.$$

For $\widehat{\boldsymbol{\beta}}^w_n$ and $(\widehat{\alpha}_{1n}(\cdot),\ldots,\widehat{\alpha}_{qn}(\cdot))^\tau$ we have an asymptotic result.

Theorem 3. *Under Assumptions* 1 *to* 5*,*

(i) $\sqrt{nT}(\widehat{\boldsymbol{\beta}}^w_n - \boldsymbol{\beta}) \to_D N(0, \boldsymbol{\Omega}_3^{-1})$ as $n \to \infty$, where $\boldsymbol{\Omega}_3 = T^{-1}E\left\{\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\right\}$ with $\boldsymbol{\Pi}_{i\cdot} = (\boldsymbol{\Pi}_{i1},\ldots,\boldsymbol{\Pi}_{iT})$;

(ii) $\max\limits_{1\le s\le q}\|\widehat{\alpha}^w_{sn} - \alpha_s\|^2_{L_2} = O_p\left(\max\limits_{1\le s\le q}\kappa_s n^{-1} + \max\limits_{1\le s\le q}\varphi_s^2\right).$

Remark 2. Let $\mathbf{O} = \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Sigma}_0^{1/2} - \boldsymbol{\Omega}_3^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Sigma}_0^{-1/2}$. Then

$$\mathbf{O}\mathbf{O}^\tau = \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Sigma}_0\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Omega}_1^{-1} - \boldsymbol{\Omega}_3^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Omega}_1^{-1} - \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Omega}_3^{-1} + \boldsymbol{\Omega}_3^{-1}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Omega}_3^{-1}.$$

Since $\mathbf{O}\mathbf{O}^\tau$ is nonnegative definite, we have

$$\begin{aligned}0 \le T^{-1}E(\mathbf{O}\mathbf{O}^\tau) &= \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1} - \boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_1^{-1} - \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_3^{-1} + \boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3\boldsymbol{\Omega}_3^{-1}\\ &= \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1} - \boldsymbol{\Omega}_3^{-1}.\end{aligned}$$

Thus $\boldsymbol{\Omega}_3^{-1} \le \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1}$. This implies that $\widehat{\boldsymbol{\beta}}_n^w$ has smaller asymptotic covariance matrix, hence is asymptotically more efficient, than $\widehat{\boldsymbol{\beta}}_n$.

In order to apply Theorem 3, a consistent estimator of $\boldsymbol{\Omega}_3$ is needed. This is given by $\widehat{\boldsymbol{\Omega}}_3 = (nT)^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X}$ via the following theorem.

Theorem 4. *Suppose that Assumptions* 1 *to* 5 *hold. Then we have that* $\widehat{\boldsymbol{\Omega}}_3 \to_p \boldsymbol{\Omega}_3$ *as* $n \to \infty$.

Combining Theorems 3 and 4, we can construct the asymptotic confidence intervals for $\boldsymbol{\beta}$ or check whether $\mathbf{C}\boldsymbol{\beta} = 0$, where $\mathbf{C}$ is a known $d \times p$ constant matrix with $d \le p$.

4. Two-Stage Estimation of Nonparametric Components

Our weighted polynomial spline estimator $(\widehat{\alpha}_{1n}^w(\cdot), \ldots, \widehat{\alpha}_{qn}^w(\cdot))^\tau$ takes contemporaneous correlation into account, is fast to compute, and is intuitive. However, its distribution theory is not available. In this section, we propose a two-stage local linear estimator of $(\alpha_s(u), \alpha_s'(u))^\tau$, with $\alpha_s'(u) = \partial\alpha_s(u)/\partial u$, and establish its asymptotic distribution. The estimator of $\alpha_s'(u)$ is of interest in some situations; see, for example, Mundra (2005). The two-stage local linear estimation extends that of You and Zhou (2007), who assumed the contemporaneous correlation totally known, did not involve any parametric component, and established only the asymptotic distribution of the estimator of $\alpha_s(u)$. As in Fan and Zhang (2000), the established asymptotic distribution is useful for constructing the simultaneous confidence bands for the underlying additive functions, which can then be used to check if an estimated additive function is significantly different from zero, or if the estimated additive function is really varying.

Let $\mathbf{Y}_{i\cdot} = (Y_{i1}, \ldots, Y_{iT})^\tau$, $\boldsymbol{\varepsilon}_{i\cdot} = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\tau$, $\boldsymbol{\Sigma}_0 = E(\boldsymbol{\varepsilon}_{i\cdot}\boldsymbol{\varepsilon}_{i\cdot}^\tau) = \sigma_\mu^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T^\tau + \sigma_e^2\mathbf{A} = (\sigma_{t_1t_2}^2)_{t_1,t_2=1}^T$, and $\boldsymbol{\Sigma}_0^{-1} = (\sigma^{t_1t_2})_{t_1,t_2=1}^T$. The t-th element of $\boldsymbol{\Sigma}_0^{-1}\mathbf{Y}_{i\cdot}$ is

$$\widetilde{Y}_{it} = \sum_{t_1=1}^T \sigma^{tt_1}Y_{it_1} = \sigma^{tt}Y_{it} + \sum_{t_1\ne t}^T \sigma^{tt_1}Y_{it_1}.$$

Write $a_{it} = \mathbf{X}_{it}^\tau\boldsymbol{\beta} + \alpha_1(U_{it1}) + \cdots + \alpha_q(U_{itq})$ and

$$a_{it}^{-s} = \mathbf{X}_{it}^\tau\boldsymbol{\beta} + \alpha_1(U_{it1}) + \cdots + \alpha_{s-1}(U_{it,s-1}) + \alpha_{s+1}(U_{it,s+1}) + \cdots + \alpha_q(U_{itq}).$$

Then we have

$$(\sigma^{tt})^{-1}\Big(\widetilde{Y}_{it}-\sigma^{tt}a_{it}^{-s}-\sum_{t_1\neq t}^{T}\sigma^{tt_1}a_{it_1}\Big)=\alpha_s(U_{its})+(\sigma^{tt})^{-1}\sum_{t_1=1}^{T}\sigma^{tt_1}\varepsilon_{it_1}.$$

Let $\widetilde{Y}_{it}^{*}=(\sigma^{tt})^{-1}\left(\widetilde{Y}_{it}-\sigma^{tt}a_{it}^{-s}-\sum_{t_1\neq t}^{T}\sigma^{tt_1}a_{it_1}\right)$. Since

$$E(\widetilde{Y}_{it}^{*}|U_{its})=\alpha_s(U_{its})\ \text{and}\ \ \mathrm{Var}\Big\{(\sigma^{tt})^{-1}\sum_{t_1=1}^{T}\sigma^{tt_1}\varepsilon_{it_1}\Big\}=(\sigma^{tt})^{-1}\leq\sigma_{tt}^{2},$$

applying the local polynomial estimation to $\widetilde{Y}_{it}^{*}$ results in a more efficient estimator of the unknown function $\alpha_s(\cdot)$ in model (1.1)−(1.2).

For U_{its} in a small neighborhood of u, $\alpha_s(u)$ can be approximated by

$$\alpha_s(U_{its})\approx\alpha_s(u)+\alpha_s'(u)(U_{its}-u)\equiv a_s+b_s(U_{its}-u).$$

This leads to the local least squares problem: find $\{(a_s,b_s)\}$ to minimize

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\big[\widetilde{Y}_{it}^{*}-\{a_s+b_s(U_{its}-u)\}\big]^2K_{h_s}(U_{its}-u),\tag{4.1}$$

where $K(\cdot)$ is a kernel function, h_s is a bandwidth, and $K_{h_s}(\cdot)=h_s^{-1}K(\cdot/h_s)$. Simple algebra leads to the solution to (4.1):

$$(\widetilde{a}_s,\widetilde{b}_s)^{\tau}=(\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\mathbf{D}_{su}\mathbf{W}_{su}\widetilde{\mathbf{Y}}^{*},$$

where $\widetilde{\mathbf{Y}}^{*}=(\widetilde{Y}_{11}^{*},\ldots,\widetilde{Y}_{1T}^{*},\ldots,\widetilde{Y}_{nT}^{*})^{\tau}$, and

$$\mathbf{W}_{su}=\mathrm{diag}(K_{h_s}(U_{11s}-u),\ldots,K_{h_s}(U_{1Ts}-u),,\ldots,K_{h_s}(U_{nTs}-u)),$$
$$\mathbf{D}_{su}=\begin{pmatrix}1&\cdots&1&\cdots&1\\(U_{11s}-u)&\cdots&(U_{1Ts}-u)&\cdots&(U_{nTs}-u)\end{pmatrix}^{\tau}.$$

In applications, a_{it}, $\boldsymbol{\beta}$, $\alpha_1(\cdot),\ldots,\alpha_{s-1}(\cdot)$, $\alpha_{s+1}(\cdot),\ldots,\alpha_q(\cdot)$, and $\sigma^{t_1t_2}$ are unknown. For $\widetilde{Y}_{it}^{*}$, we replace them by

$$\widehat{a}_{it}=\mathbf{X}_{it}^{\tau}\widehat{\boldsymbol{\beta}}_n+\widehat{\alpha}_{1n}(U_{it1})+\cdots+\widehat{\alpha}_{qn}(U_{itq}),$$

$\widehat{\boldsymbol{\beta}}_n$, $\widehat{\alpha}_{1n}(\cdot)$, $\widehat{\alpha}_{s-1,n}(\cdot)$, …,$\widehat{\alpha}_{s+1,n}(\cdot)$, …, $\widehat{\alpha}_{qn}(\cdot)$, and $\widehat{\sigma}^{t_1t_2}$. Then

$$\widehat{\mathbf{\Sigma}}_0=\frac{1}{\widehat{\sigma}_{en}^2}\Big(\widehat{\mathbf{A}}^{-1}-\frac{\widehat{\sigma}_{\mu n}^2}{(\widehat{\sigma}_{en}^2+\boldsymbol{\iota}_T^{\tau}\widehat{\mathbf{A}}^{-1}\boldsymbol{\iota}_T^{\tau})}\widehat{\mathbf{A}}^{-1}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\tau}\widehat{\mathbf{A}}^{-1}\Big)=(\widehat{\sigma}^{t_1t_2})_{t_1,t_2=1}^{T}.$$

Thus, a feasible two-stage estimator of $(\alpha_s(u), \alpha_s'(u))$ is

$$(\widehat{\alpha}_{sn}^{TS}(u), \widehat{\alpha}_{sn}^{'TS}(u))^\tau = (\mathbf{D}_{su}^\tau \mathbf{W}_{su} \mathbf{D}_{su})^{-1} \mathbf{D}_{su} \mathbf{W}_{su} \widehat{\mathbf{Y}}^*,$$

where $\widehat{\mathbf{Y}}^* = (\widehat{Y}_{11}^*, \ldots, \widehat{Y}_{1T}^*, \ldots, \widehat{Y}_{nT}^*)^\tau$ with

$$\widehat{Y}_{it}^* = (\widehat{\sigma}^{tt})^{-1}\Big(\widehat{Y}_{it} - \widehat{\sigma}^{tt}\widehat{a}_{it}^{-s} - \sum_{t_1 \neq t}^{T} \widehat{\sigma}^{tt_1}\widehat{a}_{it_1}\Big), \quad \widehat{Y}_{it} = \sum_{t_1=1}^{T} \widehat{\sigma}^{tt_1} Y_{it_1}$$

and $\widehat{a}_{it}^{-s} = \mathbf{X}_{it}^\tau \widehat{\boldsymbol{\beta}}_n + \widehat{\alpha}_{1n}(U_{it1}) + \cdots + \widehat{\alpha}_{s-1,n}(U_{it,s-1}) + \widehat{\alpha}_{s+1,n}(U_{it,s+1}) + \cdots + \widehat{\alpha}_{qn}(U_{itq})$.

To achieve the asymptotic properties of $(\widehat{\alpha}_{sn}^{TS}(u), \widehat{\alpha}_{sn}^{'TS}(u))^\tau$, assumptions are needed.

Assumption 6. $\kappa_s = c_s n^{-1/5} \log n$ for some constants c_s satisfying $0 < c_s < \infty$ and $s = 1, \ldots, q$.

Assumption 7. The function $K(\cdot)$ is a density function with a compact support.

Assumption 8. The bandwidth $h_s = c_{sh} n^{-1/5}$ for some constant c_{sh} satisfying $0 < c_{sh} < \infty$ and $s = 1, \ldots, q$.

Under the above assumptions and those in Section 2, we can state the asymptotic properties of $(\widehat{\alpha}_{sn}^{TS}(u), \widehat{\alpha}_{sn}^{'TS}(u))^\tau$. Let

$$\varrho_j = \int_{-\infty}^{\infty} u^j K(u) du, \quad \varsigma_j = \int_{-\infty}^{\infty} u^j K^2(u) du, \quad j = 1, 2, 3.$$

Theorem 5. *Under Assumptions* 1 *to* 3, 5, 6 *to* 8, *as* $n \to \infty$,

$$\sqrt{nTh_s}\left[\mathbf{H}_s\left\{\begin{pmatrix}\widehat{\alpha}_{sn}^{TS}(u)\\ \widehat{\alpha}_{sn}^{'TS}(u)\end{pmatrix} - \begin{pmatrix}\alpha_s(u)\\ \alpha_s'(u)\end{pmatrix}\right\} - \frac{h_s^2}{2}\frac{1}{\varrho_2 - \varrho_1^2}\begin{pmatrix}(\varrho_2 - \varrho_1\varrho_3)\alpha_s''(u)\\ (\varrho_3 - \varrho_1\varrho_2)\alpha_s''(u)\end{pmatrix}\right] \xrightarrow{D} N(0, \mathbf{\Omega}_s^{TS}),$$

where $\mathbf{H}_s = diag(1, h_s)$, $\alpha_s''(u) = \partial^2 \alpha_s(u)/\partial u^2$, *and*

$$\mathbf{\Omega}_s^{TS} = \frac{\sum_{t=1}^T (\sigma^{tt})^{-1} p_{ts}(u)}{\left(\sum_{t=1}^T p_{ts}(u)\right)^2 (\varrho_2 - \varrho_1^2)^2} \cdot \begin{pmatrix} \varrho_2^2\varsigma_0 - 2\varrho_1\varrho_2\varsigma_1 + \varrho_1^2\varsigma_2 & (\varrho_1^2 + \varrho_2)\varsigma_1 - \varrho_1\varrho_2\varsigma_0 - \varrho_1\varsigma_2 \\ (\varrho_1^2 + \varrho_2)\varsigma_1 - \varrho_1\varrho_2\varsigma_0 - \varrho_1\varsigma_2 & \varsigma_2 - \varrho_1(2\varsigma_1 + \varrho_1\varsigma_0) \end{pmatrix}.$$

Corollary 1. *Under Assumptions* 1 *to* 3, 5, 6 *to* 8,

$$\sqrt{nTh_s}\left\{\widehat{\alpha}_{sn}^{TS}(u)-\alpha_s(u)-\frac{h_s^2}{2}\frac{\varrho_2^2-\varrho_1\varrho_3}{\varrho_2-\varrho_1^2}\alpha_s^{''}(u)\right\}\xrightarrow{D}N(0,\sigma_{\alpha_s}^{TS})\quad as\ n\to\infty,$$

where

$$\sigma_{\alpha_s}^{TS}=(\varrho_2^2\varsigma_0-2\varrho_1\varrho_2\varsigma_1+\varrho_1^2\varsigma_2)\left\{\left(\sum_{t=1}^{T}p_{ts}(u)\right)^2(\varrho_2-\varrho_1^2)^2\right\}^{-1}\sum_{t=1}^{T}(\sigma^{tt})^{-1}p_{ts}(u).$$

Remark 3. If the correlation within the response is ignored, we can apply the same method to construct a two-stage estimator for $(\alpha_s(\cdot),\alpha_s^{'}(\cdot))^\tau$ as $(\check{\alpha}_{sn}^{TS}(u),\check{\alpha}_{sn}^{'TS}(u))^\tau=(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\mathbf{D}_{su}\mathbf{W}_{su}\check{\mathbf{Y}}$, where $\check{\mathbf{Y}}=(\check{Y}_{11},\ldots,\check{Y}_{1T},\ldots,\check{Y}_{nT})^\tau$ with $\check{Y}_{it}=Y_{it}-\widehat{a}_{it}^{-s}$. For this $(\check{\alpha}_{sn}^{TS}(u),\check{\alpha}_{sn}^{'TS}(u))^\tau$,

$$\sqrt{nTh_s}\left[\mathbf{H}_s\left\{\begin{pmatrix}\check{\alpha}_{sn}^{TS}(u)\\ \check{\alpha}_{sn}^{'TS}(u)\end{pmatrix}-\begin{pmatrix}\alpha_s(u)\\ \alpha_s^{'}(u)\end{pmatrix}\right\}-\frac{h_s^2}{2}\frac{1}{\varrho_2-\varrho_1^2}\begin{pmatrix}(\varrho_2-\varrho_1\varrho_3)\alpha_s^{''}(u)\\ (\varrho_3-\varrho_1\varrho_2)\alpha_s^{''}(u)\end{pmatrix}\right]$$
$$\xrightarrow{D}N(0,\check{\mathbf{\Omega}}_s^{TS})$$

as $n\to\infty$, where

$$\check{\mathbf{\Omega}}_s^{TS}=\frac{\sum_{t=1}^{T}\sigma_{tt}^2p_{ts}(u)}{\left(\sum_{t=1}^{T}p_{ts}(u)\right)^2(\varrho_2-\varrho_1^2)^2}\cdot\begin{pmatrix}\varrho_2^2\varsigma_0-2\varrho_1\varrho_2\varsigma_1+\varrho_1^2\varsigma_2 & (\varrho_1^2+\varrho_2)\varsigma_1-\varrho_1\varrho_2\varsigma_0-\varrho_1\varsigma_2\\ (\varrho_1^2+\varrho_2)\varsigma_1-\varrho_1\varrho_2\varsigma_0-\varrho_1\varsigma_2 & \varsigma_2-\varrho_1(2\varsigma_1+\varrho_1\varsigma_0)\end{pmatrix}.$$

Since $(\sigma^{tt})^{-1}\geq\sigma_{tt}^2$, the estimator $(\widehat{\alpha}_{sn}^{TS}(\cdot),\widehat{\alpha}_{sn}^{'TS}(\cdot))^\tau$ in Theorem 5 is asymptotically more efficient than $(\check{\alpha}_{sn}^{TS}(\cdot),\check{\alpha}_{sn}^{'TS}(\cdot))^\tau$, which ignores the correlation within the response.

Remark 4. In order to apply Theorem 5 or Corollary 1 to make statistical inference for $\alpha_s(\cdot)$ or $(\alpha_s(\cdot),\alpha_s'(\cdot))^\tau$, a consistent estimator of $\sigma_{\alpha_s^{TS}}$ or $\widehat{\mathbf{\Omega}}_s^{TS}$ is needed. As $\varsigma_0,\varsigma_1,\varsigma_2$, ϱ_1, ϱ_2, and ϱ_3 are known constants, we just need to estimate σ^{tt} and $p_{ts}(\cdot)$ for $s=1,\ldots,q$ and $t=1,\ldots,T$. According to Theorem 2, $\widehat{\sigma}^{tt}$ is a consistent estimator of σ^{tt} where $\widehat{\mathbf{\Sigma}}_0^{-1}=(\widehat{\sigma}^{t_1t_2})_{t_1,t_2=1}^T$ and $\mathbf{\Sigma}_0^{-1}=(\sigma^{t_1t_2})_{t_1,t_2=1}^T$. As for $p_{ts}(\cdot)$, we can use the usual kernel density method to estimate it:

$$\widehat{p}_{ts}(\cdot)=\frac{1}{h_s}\sum_{i=1}^{n}K_{h_s}(U_{its}-\cdot).$$

Remark 5. It should be noted that $\widehat{\alpha}_{sn}^{TS}(\cdot)$ involves the smoothing parameters h_s and $\kappa_1, \ldots, \kappa_q$. The asymptotic result of Theorem 5 shows that the smoothing parameter h_s should be of standard order. However, the smoothing parameters $\kappa_1, \ldots, \kappa_q$ at the initial estimators $\widehat{\alpha}_{1n}(\cdot), \ldots, \widehat{\alpha}_{qn}(\cdot)$ should be of bigger order than the standard one, $O(h_s^{-1})$. That is, undersmoothing is needed, which means that more knots are used than needed to achieve the optimal rate of convergence. This requirement controls the bias in the preliminary step of the estimation. In practice, standard smoothing parameter selection in the second step can be utilized. Simulation experiments show that the final results are not very sensitive to the choice of the smoothing parameters $\kappa_1, \ldots, \kappa_S$. In practice, the usual optimal smoothing parameters multiplied by a constant, say 1.5 or 2, can be used. Undersmoothing is widely used in two-stage estimation. See, for example, Horowitz and Mammen (2004), Wang and Yang (2007), Liu, Cheng, and Yao (2009) and so on.

Remark 6. This paper has assumed an AR(1) model for the noise. Our results can be extended to higher order autoregressive structure of ν_{it}. For example, if

$$\varepsilon_{it} = \mu_i + \nu_{it}, \quad \nu_{it} = \rho_1 \nu_{i,t-1} + \rho_2 \nu_{i,t-2} + e_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T,$$

where ρ_1 and ρ_2 satisfy the stationary condition: $1 - \rho_1 z - \rho_2 z^2 \neq 0$ for $|z| \leq 1$, we can estimate $(\rho_1, \rho_2)^\tau$ by $(\widehat{\rho}_{1n}, \widehat{\rho}_{2n})^\tau = (\widehat{\mathbf{Q}}_0 - \widehat{\mathbf{Q}}_1)^{-1}(\widehat{\mathbf{Q}}_2 - \widehat{\mathbf{Q}}_3)$, where

$$\widehat{\mathbf{Q}}_0 = \frac{1}{n(T-3)} \sum_{i=1}^{n} \sum_{t=1}^{T-3} (\widehat{\varepsilon}_{i,t+1}, \widehat{\varepsilon}_{it})^\tau (\widehat{\varepsilon}_{i,t+1}, \widehat{\varepsilon}_{it}),$$

$$\widehat{\mathbf{Q}}_1 = \frac{1}{n(T-3)} \sum_{i=1}^{n} \sum_{t=1}^{T-3} (\widehat{\varepsilon}_{i,t+1}, \widehat{\varepsilon}_{it})^\tau (\widehat{\varepsilon}_{i,t+2}, \widehat{\varepsilon}_{i,t+1}),$$

$$\widehat{\mathbf{Q}}_2 = \frac{1}{n(T-3)} \sum_{i=1}^{n} \sum_{t=1}^{T-3} (\widehat{\varepsilon}_{i,t+1}, \widehat{\varepsilon}_{it})^\tau \widehat{\varepsilon}_{i,t+2},$$

$$\widehat{\mathbf{Q}}_3 = \frac{1}{n(T-3)} \sum_{i=1}^{n} \sum_{t=1}^{T-3} (\widehat{\varepsilon}_{i,t+1}, \widehat{\varepsilon}_{it})^\tau \widehat{\varepsilon}_{i,t+3},$$

and $\widehat{\varepsilon}_{it}$ are the estimated residuals.

5. Simulation Studies

In this section, we report on simulation studies of the finite sample performance of the proposed procedures.

The data were generated from the partially linear panel data additive model

$$Y_{it} = X_{it1}\beta_1 + X_{it2}\beta_2 + X_{it3}\beta_3 + \alpha_1(U_{it1}) + \alpha_2(U_{it2}) + \varepsilon_{it},$$
$$\varepsilon_{it} = \mu_i + \nu_{it}, \quad \nu_{it} = \rho \nu_{i,t-1} + e_{it},$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $X_{it1} = \xi_{it1} + \xi_{i1} \sim N(1, 9) + N(0, 1)$, $X_{it2} = \xi_{it2} + \xi_{i2} \sim N(0, 6.25) + N(0, 1)$, $X_{it3} = \xi_{it3} + \xi_{i3} \sim N(0, 2.25) + N(0, 1)$, $U_{it1} \sim U(0, 2)$, $U_{it2} \sim U(0, 2)$, $\beta_1 = 0$, $\beta_2 = 1.5$, $\beta_2 = 2$, $\alpha_1(U_{it1}) = 2\cos(2\pi U_{it1}) - 2E\left[\cos(2\pi U_{it1})\right]$, $\alpha_2(U_{it2}) = 2U_{it2}^2 + \cos(2\pi U_{it2}) - E\left[2U_{it2}^2 + \cos(2\pi U_{it2})\right]$, $\mu_i \sim N(0, 1)$, and $\nu_{it} \sim N(0, 1)$. Moreover, we took $n = 50, 100, 200$, $T = 3, 4, 5, 10$, and $\rho = 0.1, 0.3, 0.6$.

In each case the number of simulated realizations was $1,000$. We used the univariate cubic B-spline basis and uniform knots. The number of knots was selected by cross validation. For the WSLSE $(\widehat{\beta}_{1n}^w, \widehat{\beta}_{2n}^w, \widehat{\beta}_{3n}^w)^\tau$ of the parametric components $(\beta_1,\, \beta_2,\, \beta_3)^\tau$, given a sample size, the sample mean (sm), standard error (se), which is the sample standard deviation from the simulated estimates of the parametric components, mean of the estimate of the standard deviation (mstd) based on the asymptotic covariance matrix, and coverage percentage of the 95% confidence intervals (cp) are summarized in Tables 1−3. In Tables 1−3, we also present the sm, se, mstd, and cp of the unweighted SLSE $(\widehat{\beta}_{1n}, \widehat{\beta}_{2n}, \widehat{\beta}_{3n})^\tau$ that neglects the serially correlated error component structure, as well the **ratio** of the se of the weighted estimator over the se of the unweighted estimator.

From Tables 1−3 we make the following observations:

1. Under the model studied, both WSLSE and SLSE of the parametric components are unbiased.
2. The WSLSE has smaller se than the SLSE.
3. The WSLSE can substantially improve the estimation of the parametric components over the SLSE, especially when the contemporaneous correlation is strong.
4. The mstd approximates the se very well.
5. The coverage of the confidence interval is very close to the 95% nominal level.

For the estimators of the nonparametric components, we computed a measure of estimation accuracy referred to as the *root average squared error* (RASE):

$$\text{RASE}_s = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\widetilde{\alpha}_s(U_{its}) - \alpha_s(U_{its})\}^2\right]^{1/2}, \quad s = 1, 2,$$

where $\widetilde{\alpha}_s(u)$ is either $\widehat{\alpha}_{sn}(u)$, or $\widehat{\alpha}_{sn}^w(u)$, or $\widehat{\alpha}_{sn}^{TS}(u)$. The results are summarized in Table 4, which shows that the two-stage local linear estimator $\widehat{\alpha}_{sn}^{TS}(u)$ or the weighted polynomial spline estimator $\widehat{\alpha}_{sn}^w(u)$ of the nonparametric components $\alpha_s(u)$ outperforms the $\widehat{\alpha}_{sn}(u)$ that ignores the contemporaneous correlation and is constructed by the polynomial spline approximation. Furthermore, $\widehat{\alpha}_{sn}^{TS}(u)$ outperforms the $\widehat{\alpha}_{sn}^w(u)$ that uses the contemporaneous correlation but has larger mean in terms of RASE.

Table 1. The finite sample performance of the parametric component estimators with $\rho = 0.1$.

		$n=50$				$n=100$				$n=200$			
		$T=3$	$T=4$	$T=5$	$T=10$	$T=3$	$T=4$	$T=5$	$T=10$	$T=3$	$T=4$	$T=5$	$T=10$
$\widehat{\beta}_{1n}$	sm	0.0030	0.0019	0.0027	0.0044	0.0058	0.0047	0.0041	0.0043	0.0044	0.0048	0.0040	0.0029
	se	0.0413	0.0358	0.0319	0.0250	0.0280	0.0258	0.0231	0.0179	0.0192	0.0177	0.0158	0.0118
	mstd	0.0403	0.0350	0.0317	0.0241	0.0277	0.0244	0.0222	0.0170	0.0194	0.0171	0.0156	0.0120
	cp	0.9430	0.9390	0.9440	0.9320	0.9410	0.9280	0.9380	0.9310	0.9450	0.9310	0.9390	0.9560
$\widehat{\beta}_{2n}$	sm	1.5021	1.5005	1.4999	1.5005	1.4978	1.4990	1.4996	1.4995	1.4999	1.4996	1.4989	1.5001
	se	0.0527	0.0451	0.0395	0.0304	0.0336	0.0298	0.0273	0.0215	0.0241	0.0205	0.0200	0.0150
	mstd	0.0482	0.0422	0.0385	0.0299	0.0333	0.0294	0.0268	0.0212	0.0231	0.0206	0.0189	0.0150
	cp	0.9240	0.9290	0.9470	0.9470	0.9470	0.9410	0.9430	0.9430	0.9370	0.9510	0.9340	0.9400
$\widehat{\beta}_{3n}$	sm	2.0011	2.0048	1.9999	2.0011	1.9980	2.0009	1.9976	2.0001	2.0006	1.9991	2.0016	1.9989
	se	0.0776	0.0732	0.0674	0.0555	0.0551	0.0506	0.0455	0.0401	0.0386	0.0359	0.0315	0.0275
	mstd	0.0767	0.0690	0.0641	0.0543	0.0529	0.0482	0.0450	0.0383	0.0370	0.0338	0.0316	0.0271
	cp	0.9490	0.9380	0.9540	0.9390	0.9400	0.9300	0.9460	0.9350	0.9380	0.9420	0.9470	0.9570
$\widehat{\beta}^w_{1n}$	sm	0.0042	0.0022	0.0024	0.0045	0.0053	0.0044	0.0042	0.0043	0.0044	0.0047	0.0041	0.0032
	se	0.0344	0.0297	0.0253	0.0167	0.0229	0.0202	0.0173	0.0120	0.0164	0.0140	0.0122	0.0079
	mstd	0.0331	0.0273	0.0238	0.0158	0.0225	0.0188	0.0165	0.0111	0.0157	0.0131	0.0115	0.0078
	cp	0.9390	0.9350	0.9320	0.9270	0.9390	0.9210	0.9260	0.9130	0.9210	0.9130	0.9240	0.9310
	ratio	**0.8329**	**0.8296**	**0.7931**	**0.6680**	**0.8179**	**0.7829**	**0.7489**	**0.6704**	**0.8542**	**0.7910**	**0.7722**	**0.6695**
$\widehat{\beta}^w_{2n}$	sm	1.5012	1.4994	1.4992	1.5008	1.4976	1.4980	1.4995	1.4998	1.4997	1.4996	1.4993	1.5002
	se	0.0438	0.0360	0.0301	0.0203	0.0290	0.0235	0.0208	0.0142	0.0200	0.0163	0.0148	0.0093
	mstd	0.0395	0.0327	0.0284	0.0189	0.0270	0.0225	0.0197	0.0133	0.0187	0.0157	0.0138	0.0093
	cp	0.9190	0.9240	0.9410	0.9290	0.9250	0.9310	0.9210	0.9310	0.9350	0.9560	0.9330	0.9550
	ratio	**0.8311**	**0.7982**	**0.7620**	**0.6678**	**0.8631**	**0.7886**	**0.7619**	**0.6605**	**0.8299**	**0.7951**	**0.7400**	**0.6200**
$\widehat{\beta}^w_{3n}$	sm	2.0020	2.0008	1.9999	2.0017	1.9974	2.0015	1.9994	2.0002	1.9991	1.9992	2.0008	1.9992
	se	0.0652	0.0590	0.0481	0.0331	0.0444	0.0397	0.0345	0.0236	0.0325	0.0277	0.0242	0.0162
	mstd	0.0627	0.0525	0.0459	0.0312	0.0428	0.0363	0.0319	0.0218	0.0298	0.0253	0.0223	0.0154
	cp	0.9410	0.9270	0.9330	0.9430	0.9380	0.9280	0.9430	0.9280	0.9230	0.9230	0.9260	0.9420
	ratio	**0.8402**	**0.8060**	**0.7136**	**0.5964**	**0.8058**	**0.7846**	**0.7582**	**0.5885**	**0.8420**	**0.7716**	**0.7683**	**0.5891**

6. An Application

We now demonstrate the application of the proposed procedures to a data example from The National Longitudinal Surveys (NLS), as described in Section 1.

It is well known that total work experience and job tenure affect the log(wage/GNP deflator) nonlinearly (see Carter Hill, Griffiths and Lim (2008)). Therefore, we consider a panel data partially linear additive model. For $i = 1, \ldots, 716$, $t = 1, 2$ and 3,

$$Y_{it} = \beta_1 X_{it} + \alpha_1(U_{it1}) + \alpha_2(U_{it2}) + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it}, \quad \nu_{it} = \rho\nu_{i,t-1} + e_{it},$$

where Y_{it} is the log(wage/GNP deflator), X_{it} is an indicator of college degree, U_{it1} is total work experience and U_{it2} is the job tenure. The estimates of $(\rho, \sigma_\mu, \sigma_e)^\tau$ are found to be $(0.4279, 0.0864, 0.0483)^\tau$. The SLSE of β is 0.3828 with std 0.0300

Table 2. The finite sample performance of the parametric component estimators with $\rho = 0.3$.

		$n = 50$				$n = 100$				$n = 200$			
		$T=3$	$T=4$	$T=5$	$T=10$	$T=3$	$T=4$	$T=5$	$T=10$	$T=3$	$T=4$	$T=5$	$T=10$
$\widehat{\beta}_{1n}$	sm	-0.0007	0.0050	0.0049	0.0033	0.0037	0.0042	0.0038	0.0042	0.0045	0.0037	0.0037	0.0044
	se	0.0431	0.0380	0.0352	0.0246	0.0294	0.0257	0.0231	0.0175	0.0207	0.0178	0.0172	0.0128
	mstd	0.0415	0.0364	0.0328	0.0249	0.0286	0.0251	0.0229	0.0176	0.0200	0.0177	0.0161	0.0124
	cp	0.9340	0.9450	0.9220	0.9590	0.9540	0.9480	0.9430	0.9460	0.9410	0.9350	0.9280	0.9240
$\widehat{\beta}_{2n}$	sm	1.5010	1.5027	1.4970	1.4986	1.4990	1.4995	1.5007	1.4999	1.5002	1.5008	1.4999	1.4992
	se	0.0503	0.0470	0.0412	0.0318	0.0358	0.0310	0.0282	0.0227	0.0247	0.0221	0.0197	0.0161
	mstd	0.0498	0.0438	0.0397	0.0309	0.0342	0.0303	0.0278	0.0218	0.0240	0.0213	0.0195	0.0154
	cp	0.9530	0.9310	0.9450	0.9460	0.9400	0.9410	0.9400	0.9340	0.9420	0.9410	0.9480	0.9370
$\widehat{\beta}_{3n}$	sm	2.0032	2.0001	2.0013	2.0031	1.9977	1.9983	1.9990	2.0005	2.0015	1.9999	1.9998	2.0013
	se	0.0782	0.0787	0.0681	0.0561	0.0595	0.0511	0.0504	0.0406	0.0397	0.0361	0.0327	0.0289
	mstd	0.0799	0.0724	0.0671	0.0561	0.0554	0.0504	0.0470	0.0396	0.0387	0.0354	0.0331	0.0281
	cp	0.9570	0.9200	0.9460	0.9470	0.9300	0.9540	0.9340	0.9450	0.9360	0.9450	0.9550	0.9490
$\widehat{\beta}^w_{1n}$	sm	0.0022	0.0037	0.0065	0.0028	0.0050	0.0042	0.0038	0.0041	0.0046	0.0035	0.0035	0.0041
	se	0.0332	0.0284	0.0245	0.0164	0.0214	0.0189	0.0164	0.0111	0.0162	0.0127	0.0118	0.0079
	mstd	0.0307	0.0256	0.0222	0.0150	0.0209	0.0175	0.0153	0.0105	0.0145	0.0122	0.0108	0.0074
	cp	0.9340	0.9220	0.9140	0.9180	0.9430	0.9220	0.9210	0.9160	0.9060	0.9370	0.9130	0.8970
	ratio	**0.7703**	**0.7474**	**0.6960**	**0.6667**	**0.7279**	**0.7354**	**0.7100**	**0.6343**	**0.7826**	**0.7135**	**0.6860**	**0.6172**
$\widehat{\beta}^w_{2n}$	sm	1.5010	1.5013	1.4982	1.4984	1.4998	1.5001	1.4997	1.4998	1.5003	1.5005	1.5002	1.4998
	se	0.0388	0.0348	0.0283	0.0195	0.0276	0.0225	0.0193	0.0136	0.0186	0.0156	0.0141	0.0095
	mstd	0.0367	0.0305	0.0265	0.0179	0.0249	0.0209	0.0184	0.0125	0.0173	0.0146	0.0128	0.0088
	cp	0.9350	0.9240	0.9270	0.9280	0.9190	0.9340	0.9330	0.9370	0.9270	0.9270	0.9300	0.9300
	ratio	**0.7714**	**0.7404**	**0.6869**	**0.6132**	**0.7709**	**0.7258**	**0.6844**	**0.5991**	**0.7530**	**0.7059**	**0.7157**	**0.5901**
$\widehat{\beta}^w_{3n}$	sm	2.0011	1.9998	2.0010	2.0015	1.9985	1.9989	1.9993	1.9995	2.0000	1.9991	1.9992	2.0003
	se	0.0599	0.0536	0.0494	0.0304	0.0437	0.0365	0.0328	0.0219	0.0295	0.0258	0.0223	0.0155
	mstd	0.0588	0.0495	0.0432	0.0295	0.0401	0.0340	0.0299	0.0206	0.0279	0.0237	0.0210	0.0145
	cp	0.9570	0.9270	0.9080	0.9440	0.9310	0.9300	0.9300	0.9330	0.9230	0.9280	0.9360	0.9360
	ratio	**0.7660**	**0.6811**	**0.7254**	**0.5419**	**0.7345**	**0.7143**	**0.6508**	**0.5394**	**0.7431**	**0.7147**	**0.6820**	**0.5363**

and the WSLSE of β is 0.3875 with std 0.0299, where the std is the standard deviation from the estimated asymptotic covariance matrix. This implies that the education has a positive effect on wage in the sense that a college degree is associated with higher wage. In addition, the estimates of $\alpha_1(\cdot)$, $\alpha_2(\cdot)$, and the corresponding pointwise 95% confidence bands are shown in Figure 1.

From Figure 1 we can see that the relationship between the response log(wage/GNP deflator) and total work experience is concave. This has the accumulation of work experience with a greater potential to increase the wage in earlier career than in later working years, and the wage tends to be static or even decreasing a bit approaching retirement. In addition, log(wage/GNP deflator) always increases with the accumulation of job tenure. More importantly, log(wage/GNP deflator) does not change much when the job tenure is below 10 years. Beyond 10 years, however, log(wage/GNP deflator) accelerates with longer tenure. These observations are consistent with classic labor economics.

Table 3. The finite sample performance of the parametric component estimators with $\rho = 0.6$.

		$n = 50$				$n = 100$				$n = 200$			
		$T = 3$	$T = 4$	$T = 5$	$T = 10$	$T = 3$	$T = 4$	$T = 5$	$T = 10$	$T = 3$	$T = 4$	$T = 5$	$T = 10$
$\widehat{\beta}_{1n}$	sm	0.0046	0.0025	0.0046	0.0048	0.0029	0.0035	0.0050	0.0043	0.0038	0.0037	0.0035	0.0042
	se	0.0492	0.0410	0.0375	0.0276	0.0339	0.0288	0.0269	0.0195	0.0233	0.0204	0.0188	0.0139
	mstd	0.0463	0.0403	0.0366	0.0277	0.0319	0.0281	0.0256	0.0196	0.0223	0.0197	0.0180	0.0138
	cp	0.9360	0.9400	0.9450	0.9490	0.9230	0.9450	0.9400	0.9480	0.9370	0.9400	0.9360	0.9330
$\widehat{\beta}_{2n}$	sm	1.5015	1.5006	1.4996	1.4999	1.5001	1.5010	1.4998	1.4997	1.4999	1.5007	1.4993	1.5002
	se	0.0568	0.0517	0.0452	0.0368	0.0398	0.0340	0.0313	0.0252	0.0282	0.0246	0.0224	0.0171
	mstd	0.0555	0.0489	0.0445	0.0344	0.0383	0.0340	0.0311	0.0243	0.0268	0.0239	0.0220	0.0172
	cp	0.9460	0.9390	0.9420	0.9330	0.9330	0.9480	0.9460	0.9390	0.9360	0.9480	0.9450	0.9510
$\widehat{\beta}_{3n}$	sm	2.0004	2.0043	2.0010	1.9995	2.0021	2.0017	2.0027	2.0003	1.9997	1.9993	1.9994	1.9991
	se	0.0949	0.0847	0.0791	0.0661	0.0641	0.0557	0.0517	0.0453	0.0454	0.0405	0.0377	0.0318
	mstd	0.0905	0.0817	0.0762	0.0629	0.0626	0.0572	0.0535	0.0444	0.0439	0.0401	0.0375	0.0314
	cp	0.9360	0.9480	0.9410	0.9320	0.9440	0.9550	0.9590	0.9510	0.9420	0.9460	0.9450	0.9490
$\widehat{\beta}^w_{1n}$	sm	0.0041	0.0029	0.0052	0.0041	0.0030	0.0031	0.0036	0.0040	0.0038	0.0033	0.0035	0.0035
	se	0.0321	0.0265	0.0228	0.0145	0.0213	0.0174	0.0164	0.0102	0.0141	0.0119	0.0111	0.0071
	mstd	0.0286	0.0234	0.0203	0.0135	0.0194	0.0160	0.0140	0.0094	0.0134	0.0112	0.0098	0.0066
	cp	0.9170	0.9140	0.9090	0.9160	0.9240	0.9170	0.9020	0.9070	0.9320	0.9270	0.8960	0.8990
	ratio	**0.6524**	**0.6463**	**0.6080**	**0.5254**	**0.6283**	**0.6042**	**0.6097**	**0.5231**	**0.6052**	**0.5833**	**0.5904**	**0.5108**
$\widehat{\beta}^w_{2n}$	sm	1.5013	1.4988	1.5009	1.5004	1.4987	1.5009	1.5000	1.4999	1.4999	1.5004	1.4994	1.4996
	se	0.0384	0.0317	0.0272	0.0180	0.0256	0.0205	0.0182	0.0126	0.0171	0.0145	0.0124	0.0088
	mstd	0.0342	0.0281	0.0243	0.0162	0.0231	0.0191	0.0167	0.0113	0.0160	0.0133	0.0117	0.0080
	cp	0.9250	0.9210	0.9100	0.9200	0.9300	0.9310	0.9370	0.9250	0.9340	0.9280	0.9380	0.9250
	ratio	**0.6761**	**0.6132**	**0.6018**	**0.4891**	**0.6432**	**0.6029**	**0.5815**	**0.5000**	**0.6064**	**0.5894**	**0.5536**	**0.5146**
$\widehat{\beta}^w_{3n}$	sm	1.9999	1.9991	2.0001	1.9988	2.0009	1.9998	2.0005	1.9996	1.9994	1.9989	1.9998	1.9995
	se	0.0618	0.0515	0.0431	0.0291	0.0420	0.0330	0.0303	0.0205	0.0275	0.0244	0.0207	0.0142
	mstd	0.0556	0.0458	0.0399	0.0268	0.0377	0.0314	0.0276	0.0187	0.0261	0.0219	0.0192	0.0131
	cp	0.9190	0.9110	0.9290	0.9260	0.9280	0.9330	0.9300	0.9230	0.9330	0.9190	0.9300	0.9360
	ratio	**0.6512**	**0.6080**	**0.5449**	**0.4402**	**0.6552**	**0.5925**	**0.5861**	**0.4525**	**0.6057**	**0.6025**	**0.5491**	**0.4465**

In addition, we also use the quadratic model in Carter Hill, Griffiths and Lim (2008) to fit this data set and the corresponding fitting error is 0.4427. In comparison, our model has a smaller fitting error 0.4374, suggesting it may be more suitable.

7. Concluding Remarks

In this paper, we have investigated the statistical inference of the panel data partially linear additive regression with serially correlated error component structure. We have proposed a weighted semiparametric least squares estimator for the parametric components and a weighted polynomial spline series estimator for the nonparametric components. The weighted semiparametric least squares estimator is shown to be asymptotically normal and more efficient than the unweighted one. Based on these estimators a two-stage local polynomial estimator

Table 4. The finite sample performance of the nonparametric component estimators.

	$n=50$				$n=100$				$n=300$			
	$T=2$	$T=3$	$T=5$	$T=10$	$T=2$	$T=3$	$T=5$	$T=10$	$T=2$	$T=3$	$T=5$	$T=10$
$\rho=0.1$												
RASE($\widehat{\alpha}_{1n}$) sm	0.4552	0.3555	0.3084	0.3818	0.3258	0.3886	0.3824	0.3111	0.3618	0.3335	0.3120	0.2992
RASE($\widehat{\alpha}_{2n}$) sm	0.4800	0.4377	0.3927	0.3191	0.4219	0.3443	0.3186	0.3379	0.3287	0.3367	0.3383	0.3185
RASE($\widehat{\alpha}^{w}_{1n}$) sm	0.3681	0.3037	0.2550	0.3293	0.3065	0.3574	0.3501	0.2707	0.3385	0.3150	0.2881	0.2568
ratio	**0.8087**	**0.8543**	**0.8268**	**0.8625**	**0.9408**	**0.9197**	**0.9155**	**0.8701**	**0.9356**	**0.9445**	**0.9234**	**0.8583**
RASE($\widehat{\alpha}^{w}_{2n}$) sm	0.4664	0.3935	0.3494	0.2656	0.3728	0.3103	0.2838	0.2869	0.3055	0.3088	0.3063	0.2800
ratio	**0.9717**	**0.8990**	**0.8897**	**0.8323**	**0.8836**	**0.9012**	**0.8908**	**0.8491**	**0.9294**	**0.9171**	**0.9054**	**0.8791**
RASE($\widehat{\alpha}^{TS}_{1n}$) sm	0.3357	0.2719	0.2397	0.1438	0.2777	0.2430	0.2059	0.1720	0.2774	0.2664	0.1867	0.1351
ratio	**0.7375**	**0.7648**	**0.7772**	**0.3766**	**0.8524**	**0.6253**	**0.5384**	**0.5529**	**0.7667**	**0.7988**	**0.5984**	**0.4515**
RASE($\widehat{\alpha}^{TS}_{2n}$) sm	0.3017	0.3069	0.2534	0.1789	0.3039	0.2906	0.2567	0.1779	0.2902	0.2452	0.2040	0.1320
ratio	**0.6285**	**0.7012**	**0.6453**	**0.5606**	**0.7203**	**0.8440**	**0.8057**	**0.5265**	**0.8829**	**0.7282**	**0.6030**	**0.4144**
$\rho=0.3$												
RASE($\widehat{\alpha}_{1n}$) sm	0.3775	0.4468	0.4169	0.3587	0.3950	0.3451	0.3595	0.3143	0.3411	0.2657	0.3154	0.2988
RASE($\widehat{\alpha}_{2n}$) sm	0.4412	0.3967	0.4122	0.3042	0.3425	0.3498	0.3024	0.3169	0.3448	0.3361	0.3149	0.2994
RASE($\widehat{\alpha}^{w}_{1n}$) sm	0.3065	0.3896	0.3407	0.2894	0.3568	0.3007	0.3115	0.2687	0.2962	0.2383	0.2856	0.2582
ratio	**0.8119**	**0.8720**	**0.8172**	**0.8068**	**0.9033**	**0.8713**	**0.8665**	**0.8549**	**0.8684**	**0.8969**	**0.9055**	**0.8641**
RASE($\widehat{\alpha}^{w}_{2n}$) sm	0.3866	0.3362	0.3517	0.2448	0.2941	0.3134	0.2595	0.2631	0.3274	0.3057	0.2822	0.2534
ratio	**0.8762**	**0.8475**	**0.8532**	**0.8047**	**0.8587**	**0.8959**	**0.8581**	**0.8302**	**0.9495**	**0.9096**	**0.8962**	**0.8464**
RASE($\widehat{\alpha}^{TS}_{1n}$) sm	0.2637	0.2382	0.1716	0.1294	0.2094	0.1549	0.1221	0.1059	0.1724	0.1058	0.0931	0.0835
ratio	**0.6985**	**0.5331**	**0.4116**	**0.3607**	**0.5301**	**0.4489**	**0.3396**	**0.3369**	**0.5054**	**0.3982**	**0.2952**	**0.2795**
RASE($\widehat{\alpha}^{TS}_{2n}$) sm	0.2802	0.2313	0.1775	0.1366	0.2162	0.1565	0.1213	0.1142	0.1659	0.1124	0.0933	0.0934
ratio	**0.6351**	**0.5831**	**0.4306**	**0.4490**	**0.6312**	**0.4474**	**0.4011**	**0.3604**	**0.4811**	**0.3344**	**0.2963**	**0.3120**
$\rho=0.6$												
RASE($\widehat{\alpha}_{1n}$) sm	0.5095	0.4084	0.3800	0.3583	0.3491	0.3989	0.2981	0.3271	0.3234	0.3254	0.3277	0.3088
RASE($\widehat{\alpha}_{2n}$) sm	0.4834	0.4306	0.4110	0.3705	0.3465	0.3595	0.3560	0.2970	0.3428	0.3373	0.2999	0.3122
RASE($\widehat{\alpha}^{w}_{1n}$) sm	0.3956	0.3030	0.2897	0.2771	0.2844	0.2833	0.2463	0.2758	0.2647	0.2644	0.2886	0.2511
ratio	**0.7764**	**0.7419**	**0.7624**	**0.7734**	**0.8147**	**0.7102**	**0.8262**	**0.8432**	**0.8185**	**0.8125**	**0.8807**	**0.8131**
RASE($\widehat{\alpha}^{w}_{2n}$) sm	0.3494	0.3326	0.3230	0.2888	0.2605	0.3018	0.2949	0.2340	0.3000	0.2912	0.2451	0.2575
ratio	**0.7228**	**0.7724**	**0.7859**	**0.7795**	**0.7518**	**0.8395**	**0.8284**	**0.7879**	**0.8751**	**0.8633**	**0.8173**	**0.8248**
RASE($\widehat{\alpha}^{TS}_{1n}$) sm	0.3495	0.1907	0.2185	0.1751	0.2546	0.2459	0.2176	0.1460	0.2523	0.2622	0.2067	0.1818
ratio	**0.6860**	**0.4669**	**0.5750**	**0.4887**	**0.7293**	**0.6164**	**0.7300**	**0.4463**	**0.7801**	**0.8058**	**0.6308**	**0.5887**
RASE($\widehat{\alpha}^{TS}_{2n}$) sm	0.3357	0.3119	0.2262	0.1722	0.2482	0.2912	0.2044	0.1672	0.2327	0.2782	0.2160	0.1783
ratio	**0.6945**	**0.7243**	**0.5504**	**0.4648**	**0.7163**	**0.8100**	**0.5742**	**0.5630**	**0.6788**	**0.8248**	**0.7202**	**0.5711**

ratio = (the sm of weighted nonparametric component estimator's RASE) divided by (the sm of unweighted nonparametric component estimator's RASE).

of the nonparametric components was proposed that has the advantages of higher asymptotic efficiency and an oracle property.

Parametric regression models, if specified correctly, can provide a more parsimonious description of the relationship between the response variable and its covariates than semiparametric or nonparametric regression models. Therefore, it is of interest to check whether the nonlinear function $\alpha_s(\cdot)$ can be described by a parametric structure. This amounts to testing if $\alpha_s(\cdot)$ is of a certain parametric

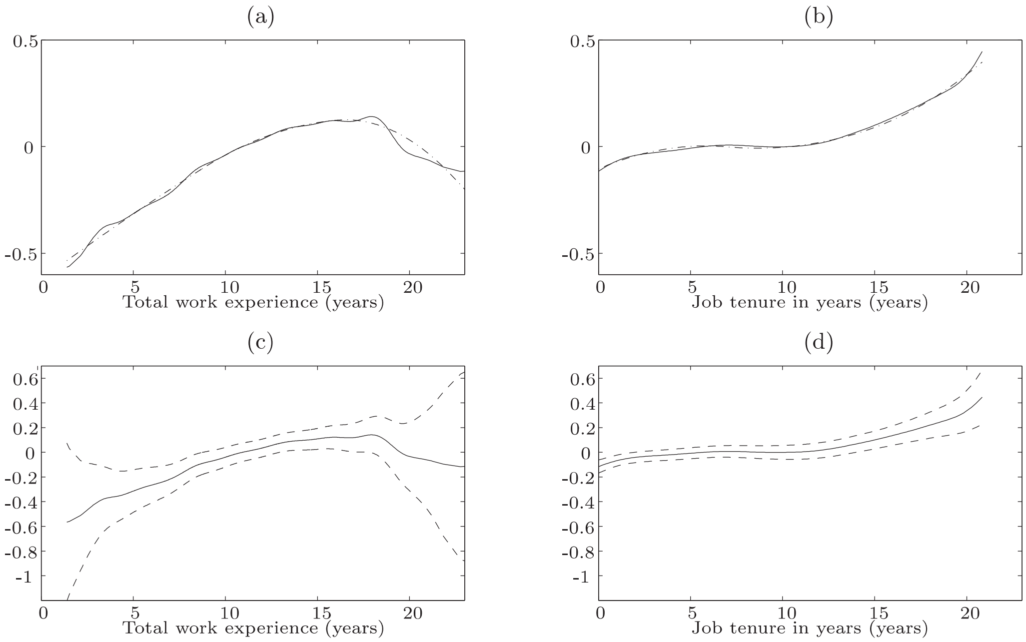

Figure 1. Plots for the estimators of the nonparametric components. (a) Relation between log(wage/GNP deflator) and total work experience. (b) Relation between log(wage/GNP deflator) and job tenure. In (a) and (b), the solid line is the proposed two-stage estimate and the dash-dotted line is the weighted polynomial spline approximation estimate. (c) Pointwise 95% confidence band of the relation between log(wage/GNP deflator) and total work experience. (d) Pointwise 95% confidence band of the relation between log(wage/GNP deflator) and job tenure. The relations in (c) and (d) are estimated by the proposed two-stage estimation.

form. Fan, Zhang, and Zhang (2001) proposed a generalized likelihood-ratio test statistic to check whether an unknown function has a certain parametric form in classic nonparametric regression models. Extending this kind of method to model (1.1)−(1.2) calls for further research efforts.

Throughout, we did not include the lag response $Y_{i,t-d}$ on the right side of model (1.1). If this lag response is incorporated into the model, it becomes the indigenous covariate and a new estimation method (for example, the instrumental variable method) is needed to deal with it. We will investigate this problem in future research work.

In addition, we have focused on large n and fixed T. In some situations, the observed time points may increase with the increase in the number of individuals. In other cases, the observed time points may be greater than the number of individuals. How to extend our methods to these cases is still an open problem.

Acknowledgement

You's research was supported by grants from the National Natural Science Foundation of China (NSFC) (No. 11071154). The work is also partiallysupported by the Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics.

Appendix: Proofs of Main Results

We first prove several lemmas.

Lemma A.1. Suppose that Assumptions 1, 7 and 8 hold. Then, as $n \to \infty$,

$$\sup_{u \in \bigcup_{t=1}^T \mathcal{U}_{ts}} \left| \frac{1}{nTh_s} \sum_{i=1}^n \sum_{t=1}^T K\left(\frac{U_{its}-u}{h_s}\right)\left(\frac{U_{its}-u}{h_s}\right)^k - \frac{1}{T}\sum_{t=1}^T p_{ts}(u)\varrho_s \right| = O_p\left\{h_s^2 + \left(\frac{\log n}{nh_s}\right)^{1/2}\right\},$$

$$\sup_{u \in \bigcup_{t=1}^T \mathcal{U}_j} \frac{1}{nTh_s} \sum_{i=1}^n \sum_{t=1}^T K\left(\frac{U_{its}-u}{h_s}\right)\left(\frac{U_{its}-u}{h_s}\right)^k \varepsilon_{it} = O_p\left\{\left(\frac{\log n}{nh_s}\right)^{1/2}\right\},$$

where $s = 1, \ldots, q$ and $k = 0, 1, 2, 4$.

Proof. Lemma A.1 follows immediately from Mack and Silverman (1982).

Lemma A.2. Let $\widetilde{\boldsymbol{\beta}}_n^w = (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}} \mathbf{X})^{-1} \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}} \mathbf{Y}$, $\widetilde{\boldsymbol{\theta}}_n^w = (\mathbf{D}^\tau \boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}$ $(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}_n^w)$, and $\widetilde{\boldsymbol{\alpha}}_n^w(u) = (\widetilde{\alpha}_{1n}^w(u), \ldots, \widetilde{\alpha}_{qn}^w(u))^\tau = \boldsymbol{\zeta}(u)\widetilde{\boldsymbol{\theta}}_n^w$, where $\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}$. Then under Assumptions 1 to 5,

(i) $\sqrt{nT}(\widetilde{\boldsymbol{\beta}}_n^w - \boldsymbol{\beta}) \to_D N(0, \boldsymbol{\Omega}_3^{-1})$ as $n \to \infty$, where $\boldsymbol{\Omega}_3 = T^{-1}E\left\{\boldsymbol{\Pi}_{i\cdot}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Pi}_{i\cdot}\right\}$,

(ii) $\max_{1\le s\le q} \|\widetilde{\alpha}_{sn}^w - \alpha_s\|_{L_2}^2 = O_p\left(\max_{1\le s\le q}\kappa_s n^{-1} + \max_{1\le s\le q}\varphi_s^2\right)$.

Proof. It is easy to see that the j-th element of $\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}$ can be decomposed as

$$\begin{aligned}\mathbf{X}_j^{*\tau}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} &= \boldsymbol{\Pi}_j^{*\tau}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - \boldsymbol{\Pi}_j^{*\tau}\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} + \mathbf{H}_j^{*\tau}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\\ &= J_1 - J_2 + J_3, \quad \text{say},\end{aligned}$$

where $\mathbf{X}_j^* = (X_{11j}, \ldots, X_{1Tj}, \ldots, X_{nTj})^\tau$, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{1T}, \ldots, \varepsilon_{nT})^\tau$, $\boldsymbol{\Pi}_j^* = (\Pi_{11j}, \ldots, \Pi_{nTj})^\tau$, and $\mathbf{H}_s^* = (\sum_{s=1}^q h_{1js}(U_{11s}), \ldots, \sum_{s=1}^q h_{Tjs}(U_{1Ts}), \ldots, \sum_{s=1}^q h_{Tjs}(U_{nTs}))^\tau$. We have $E(J_2) = 0$ and

$$\begin{aligned}\mathrm{Var}(J_2) &= E\left\{\boldsymbol{\Pi}_j^{*\tau}\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}_j^*\right\}\\ &= O(1)\cdot\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\tau\boldsymbol{\Sigma}^{-1}\right\}\\ &= O(\kappa) = o(n),\end{aligned}$$

so that $J_2 = o_p(n^{1/2})$. Moreover, there exits a real vector $\boldsymbol{\gamma}$ of dimension $\kappa_1 + \cdots + \kappa_q$ such that

$$\max_{1\le i\le n, 1\le t\le T}\left|\sum_{s=1}^{q} h_{tjs}(U_{its}) - \boldsymbol{\gamma}^\tau \boldsymbol{\zeta}(\mathbf{U}_{it\cdot})\right| = O(\max_{1\le s\le q}\varphi_s).$$

Hence

$$\begin{aligned}\mathrm{Var}(J_3) &\le O(1)\cdot E\left\{(\mathbf{H}_j^* - \boldsymbol{\gamma}^\tau\mathbf{D})^\tau \boldsymbol{\Sigma}^{-1}(\mathbf{H}_j^* - \boldsymbol{\gamma}^\tau\mathbf{D})\right\}\\ &\le O(1)\cdot \sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{s=1}^{q} h_{tjs}(U_{its}) - \boldsymbol{\gamma}^\tau\boldsymbol{\zeta}(\mathbf{U}_{it\cdot})\right\}^2 = O\left(n\max_{1\le s\le q}\varphi_s^2\right).\end{aligned}$$

Thus $J_3 = o_p(n^{1/2})$ and $n^{-1}\mathbf{X}_j^{*\tau}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\Pi}_j^{*\tau}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} + o_p(n^{-1/2})$. The same argument leads to the same results for other elements of $\mathbf{X}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}$, giving $n^{-1}\mathbf{X}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\Pi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} + o_p(n^{-1/2})$. Now let

$$\boldsymbol{\alpha} = \left(\sum_{s=1}^{q}\alpha_s(U_{11s}),\ldots,\sum_{s=1}^{q}\alpha_s(U_{1Ts}),\ldots,\sum_{s=1}^{q}\alpha_s(U_{nTs})\right)^\tau.$$

There exists a real vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\tau,\ldots,\boldsymbol{\theta}_q^\tau)^\tau$ such that $\max_{1\le i\le n,1\le t\le T}|\alpha_s(U_{its}) - \boldsymbol{\theta}_s^\tau\boldsymbol{\zeta}_s(U_{its})| = O(\varphi_s)$. Thus the j-th element of $\mathbf{X}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha}$ can be written as

$$\begin{aligned}\mathbf{X}_j^*\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha} &= \boldsymbol{\Pi}_j^*\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha} + \mathbf{H}_j^*\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha}\\ &\le \left\{\boldsymbol{\Pi}_j^{*\tau}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}-\mathbf{D}\boldsymbol{\theta})(\boldsymbol{\alpha}-\mathbf{D}\boldsymbol{\theta})^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\Pi}_j^{*\tau}\right\}^{1/2}\\ &\quad+\frac{1}{2}\left\{(\mathbf{H}_j^*-\boldsymbol{\gamma}^\tau\mathbf{D})^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\mathbf{H}_j^*-\boldsymbol{\gamma}^\tau\mathbf{D}) + (\boldsymbol{\alpha}-\mathbf{D}\boldsymbol{\theta})^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}-\mathbf{D}\boldsymbol{\theta})\right\}\\ &= O_p\left(\max_{1\le s\le q}\kappa_s\right) + O_p\left(n\max_{1\le s\le q}\varphi_s^2\right) = o_p(n^{1/2}).\end{aligned}$$

Following the same line, we can show the same results for other elements of $\mathbf{X}\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha}$. Thus $n^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha} = o_p(n^{-1/2})$ and, similarly, $n^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X} = n^{-1}\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} + o_p(1)$. It follows that

$$\begin{aligned}&\sqrt{nT}(\widetilde{\boldsymbol{\beta}}_n^w - \boldsymbol{\beta})\\ &= \sqrt{nT}\left(\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X}\right)^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} + \sqrt{nT}\left(\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X}\right)^{-1}\mathbf{X}^\tau\mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\alpha}\\ &= \sqrt{nT}(\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} + o_p(1).\end{aligned}$$

Therefore, to complete the proof of part (i) it suffices to show

$$\frac{1}{\sqrt{nT}}\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \to_D N(0,\boldsymbol{\Omega}_3) \quad \text{as} \quad n\to\infty.$$

For any nonzero p-vector $\boldsymbol{\lambda}$ we have

$$\frac{1}{\sqrt{nT}}\boldsymbol{\lambda}\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \frac{1}{\sqrt{nT}}\sum_{i=1}^n \boldsymbol{\lambda}\boldsymbol{\Pi}_{i\cdot}^\tau\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varepsilon}_{i\cdot} = \frac{1}{\sqrt{nT}}\sum_{i=1}^n \psi_i,$$

where the ψ_i are independent random variables with mean zero; it is easy to check that $\{\psi_i\}$ satisfy the Lindeberg condition. Moreover, as Var $\left((nT)^{-1/2}\boldsymbol{\lambda}\boldsymbol{\Pi}^\tau\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\right)$ $= \boldsymbol{\lambda}^\tau\boldsymbol{\Omega}_3\boldsymbol{\lambda}$, part (i) is proved. Part (ii) follows from the root-n consistency of $\widetilde{\boldsymbol{\beta}}_n^w$ and the standard method.

Proof of Theorem 1. The proof of Theorem 1 is the same as that of Lemma A.2. We omit the details.

Proof of Theorem 2. (i) Let $Q_j = [n(T-2)]^{-1}\sum_{i=1}^n\sum_{t=1}^{T-2}\varepsilon_{it}\varepsilon_{i,t+j}$ and

$$\Delta_{it} = \mathbf{X}_{it}^\tau(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n) - (\alpha_1(U_{it1}) - \widehat{\alpha}_{1n}(U_{it1})) - \cdots - (\alpha_q(U_{itq}) - \widehat{\alpha}_{qn}(U_{itq})).$$

Since

$$\widehat{\varepsilon}_{it} = Y_{it} - \mathbf{X}_{it}^\tau\widehat{\boldsymbol{\beta}}_n - \widehat{\alpha}_{1n}(U_{it1}) - \cdots - \widehat{\alpha}_{qn}(U_{itq}), \quad i=1,\ldots,n,\ t=1,\ldots,T, \quad \text{(A.1)}$$

we have

$$\begin{aligned}\widehat{Q}_j &= Q_j + \frac{1}{n(T-2)}\sum_{i=1}^n\sum_{t=1}^{T-2}\Delta_{it}\Delta_{i,t+j} + \frac{1}{n(T-2)}\sum_{i=1}^n\sum_{t=1}^{T-2}\varepsilon_{it}\Delta_{i,t+j}\\ &\quad + \frac{1}{n(T-2)}\sum_{i=1}^n\sum_{t=1}^{T-2}\Delta_{it}\varepsilon_{i,t+j}\\ &= Q_j + J_1 + J_2 + J_3, \text{ say.}\end{aligned}$$

According to Theorem 1, $\Delta_{it} = O_p\left(n^{-1/2}\right) + O_p\left(\sqrt{\max_{1\le s\le q}\kappa_s n^{-1} + \max_{1\le s\le q}\varphi_s^2}\right)$, so that

$$J_1 = O_p\left(n^{-1}\right) + O_p\left(\max_{\{1\le s\le q\}}\kappa_s n^{-1} + \max_{\{1\le s\le q\}}\varphi_s^2\right) = o_p\left(n^{-1/2}\right).$$

In addition,

$$\begin{aligned}J_2 &= \frac{1}{n(T-2)}\sum_{i=1}^n\sum_{t=1}^{T-2}\varepsilon_{it}\mathbf{X}_{i,t+j}^\tau(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\\ &\quad + \frac{1}{n(T-2)}\sum_{i=1}^n\sum_{t=1}^{T-2}\varepsilon_{it}\sum_{s=1}^q(\alpha_s(U_{i,t+j,s}) - \widehat{\alpha}_{sn}(U_{i,t+j,s}))\\ &= J_{21} + J_{22}, \quad \text{say.}\end{aligned}$$

According to the proof of Theorem 1, $[n(T-2)]^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\varepsilon_{it}\mathbf{X}_{i,t+j}^{\tau}=O_p(n^{-1/2})$. Therefore Theorem 1 (i) leads to $J_{21}=O_p(n^{-1})=o_p(n^{-1/2})$. Based on the definition of $\widehat{\alpha}_{sn}(U_{i,t+j,s})$,

$$\begin{aligned}\sum_{s=1}^{q}(\alpha_s(U_{i,t+j,s})-\widehat{\alpha}_{sn}(U_{i,t+j,s}))&=\sum_{s=1}^{q}\alpha_s(U_{i,t+j,s})-(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}\widehat{\boldsymbol{\theta}}_n\\ &=\left\{\sum_{s=1}^{q}\alpha_s(U_{i,t+j,s})-(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\boldsymbol{\alpha}\right\}\\ &\quad-(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\boldsymbol{\varepsilon}-(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n).\end{aligned}$$

Hence, J_{22} can be decomposed as

$$\begin{aligned}J_{22}&=\frac{1}{n(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\varepsilon_{it}\left\{\sum_{s=1}^{q}\alpha_s(U_{i,t+j,s})-(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\boldsymbol{\alpha}\right\}\\ &\quad-\frac{1}{n(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\varepsilon_{it}(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\boldsymbol{\varepsilon}\\ &\quad-\frac{1}{n(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\varepsilon_{it}(\boldsymbol{\zeta}(\mathbf{U}_{i,t+j,\cdot}))^{\tau}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\mathbf{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)\\ &=J_{221}+J_{222}+J_{223},\quad\text{say.}\end{aligned}$$

By the properties of polynomial splines we can show that

$$J_{221}=O_p(n^{-1/2})\cdot O_p\left(\sqrt{\max_{\{1\le s\le q\}}\kappa_s n^{-1}+\max_{\{1\le s\le q\}}\varphi_s^2}\right)=o_p(n^{-1/2}).$$

Further,

$$J_{222}=O_p(n^{-1})+\frac{1}{n}\boldsymbol{\varepsilon}^{\tau}\mathbf{D}(\mathbf{D}^{\tau}\mathbf{D})^{-1}\mathbf{D}^{\tau}\boldsymbol{\varepsilon}=O_p(n^{-1}\max_{1\le s\le q}\kappa_s)=o_p(n^{-1/2}).$$

The $\sqrt{n}$ consistency of $\widehat{\boldsymbol{\beta}}_n$, and the same argument for J_{222}, lead to $J_{223}=O(n^{-1}\max_{1\le s\le q}\kappa_s)$ as well. Together we have $J_{22}=o_p(n^{-1/2})$. As a result, $J_2=o_p(n^{-1/2})$. Following the same line, we can show that $J_3=O_p(n^{-1/2})$. This implies that $\widehat{Q}_j=Q_j+o_p(n^{-1/2})$ and, for $\widehat{\rho}_n$,

$$\begin{aligned}&\sqrt{nT}(\widehat{\rho}_n-\rho)=\sqrt{nT}\left(\frac{\widehat{Q}_1-\widehat{Q}_2}{\widehat{Q}_0-\widehat{Q}_1}-\rho\right)\\ &=\sqrt{nT}\left(\frac{Q_1-Q_2}{Q_0-Q_1}-\rho\right)+o_p(1)\\ &=\sqrt{nT}\left[\left\{\sum_{i=1}^{n}\sum_{t=1}^{T-2}(\varepsilon_{it}^2-\varepsilon_{it}\varepsilon_{i,t+1})\right\}^{-1}\left\{\sum_{i=1}^{n}\sum_{t=1}^{T-2}(\varepsilon_{it}\varepsilon_{i,t+1}-\varepsilon_{it}\varepsilon_{i,t+2})\right\}-\rho\right]+o_p(1).\end{aligned}$$

According to (1.2) we have $\varepsilon_{i,t+1} = \mu_i + \nu_{i,t+1} = \mu_i + \rho\nu_{it} + e_{i,t+1} = \rho\varepsilon_{it} + (1-\rho)\mu_i + e_{i,t+1}$ and $\varepsilon_{i,t+2} = \mu_i + \nu_{i,t+2} = \mu_i + \rho\nu_{i,t+1} + e_{i,t+2} = \rho\varepsilon_{i,t+1} + (1-\rho)\mu_i + e_{i,t+2}$. Therefore,

$$\begin{aligned}
&\sqrt{nT}(\widehat{\rho}_n - \rho) \\
&\quad = \sqrt{nT}\frac{\sum_{i=1}^{n}\sum_{t=1}^{T-2}\left[\varepsilon_{it}\left\{(1-\rho)\mu_i + e_{i,t+1}\right\} - \varepsilon_{it}\left\{(1-\rho)\mu_i + e_{i,t+2}\right\}\right]}{\sum_{i=1}^{n}\sum_{t=1}^{T-2}(\varepsilon_{it}^2 - \varepsilon_{it}\varepsilon_{i,t+1})} + o_p(1).
\end{aligned}$$

Write

$$\begin{aligned}
&\frac{1}{\sqrt{n}(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\left[\varepsilon_{it}\left\{(1-\rho)\mu_i + e_{i,t+1}\right\} - \varepsilon_{it}\left\{(1-\rho)\mu_i + e_{i,t+2}\right\}\right] \\
&\quad = \frac{1}{\sqrt{n}(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}\left\{(\mu_i + \nu_{it})e_{i,t+1} - (\mu_i + \nu_{it})e_{i,t+2}\right\} \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{T-2}\sum_{t=1}^{T-2}\left\{(\mu_i + \nu_{it})e_{i,t+1} - (\mu_i + \nu_{it})e_{i,t+2}\right\}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_i, \quad \text{say,}
\end{aligned}$$

where $\{\chi_i\}_{i=1}^n$ is an i.i.d. random variable sequence with mean 0 and variance $\mathrm{Var}(\chi_i) = 2\sigma_e^2/(T-2)\left(\sigma_\mu^2 + \sigma_e^2/(1-\rho^2)\right)$. It is easy to see that, as $n \to \infty$,

$$\frac{1}{n(T-2)}\sum_{i=1}^{n}\sum_{t=1}^{T-2}(\varepsilon_{it}^2 - \varepsilon_{it}\varepsilon_{i,t+1}) \to_p E(\varepsilon_{it}^2 - \varepsilon_{it}\varepsilon_{i,t+1}) = \sigma_e^2(1+\rho)^{-1}.$$

Therefore the result of $\widehat{\rho}_n$ holds. Based on the definition of $\widehat{\sigma}_{en}^2$, we have

$$\begin{aligned}
\widehat{\sigma}_{en}^2 &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\widehat{\boldsymbol{\varepsilon}}_{i\cdot}^\tau\mathbf{C}^\tau\mathbf{E}_T^\rho\mathbf{C}\widehat{\boldsymbol{\varepsilon}}_{i\cdot} + \frac{1}{n(T-1)}\sum_{i=1}^{n}\widehat{\boldsymbol{\varepsilon}}_{i\cdot}^\tau(\widehat{\mathbf{C}}^\tau\mathbf{E}_T^{\widehat{\rho}_n}\widehat{\mathbf{C}} - \mathbf{C}^\tau\mathbf{E}_T^\rho\mathbf{C})\widehat{\boldsymbol{\varepsilon}}_{i\cdot} \\
&= J_4 + J_5, \quad \text{say}
\end{aligned}$$

where $\widehat{\boldsymbol{\varepsilon}}_{i\cdot} = (\widehat{\varepsilon}_{i1}, \ldots, \widehat{\varepsilon}_{iT})^\tau$. Combining this with (A.1) we get

$$\begin{aligned}
J_4 &= \frac{1}{n(T-1)} \sum_{i=1}^{n} (\mu_i \mathbf{1}_T + \boldsymbol{\nu}_{i\cdot})^\tau \mathbf{C}^\tau \mathbf{E}_T^\rho \mathbf{C} (\mu_i \mathbf{1}_T + \boldsymbol{\nu}_{i\cdot}) + \frac{1}{n(T-1)} \sum_{i=1}^{n} \boldsymbol{\varepsilon}_{i\cdot}^\tau \mathbf{C}^\tau \mathbf{E}_T^\rho \mathbf{C} \boldsymbol{\varepsilon}_{i\cdot} \\
&+ \frac{1}{n(T-1)} \sum_{i=1}^{n} \left\{ \mathbf{X}_{i\cdot}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) + \sum_{s=1}^{q} (\boldsymbol{\alpha}_s(\mathbf{U}_{i\cdot s}) - \widehat{\boldsymbol{\alpha}}_s(\mathbf{U}_{i\cdot s})) \right\}^\tau \\
&\cdot \mathbf{C}^\tau \mathbf{E}_T^\rho \mathbf{C} \left\{ \mathbf{X}_{i\cdot}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) + \sum_{s=1}^{q} (\boldsymbol{\alpha}_s(\mathbf{U}_{i\cdot s}) - \widehat{\boldsymbol{\alpha}}_s(\mathbf{U}_{i\cdot s})) \right\} \\
&+ \frac{2}{n(T-1)} \sum_{i=1}^{n} \left\{ \mathbf{X}_{i\cdot}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) + \sum_{s=1}^{q} (\boldsymbol{\alpha}_s(\mathbf{U}_{i\cdot s}) - \widehat{\boldsymbol{\alpha}}_s(\mathbf{U}_{i\cdot s})) \right\}^\tau \mathbf{C}^\tau \mathbf{E}_T^\rho \mathbf{C} \boldsymbol{\varepsilon}_{i\cdot} \\
&= J_{41} + J_{42} + J_{43}, \quad \text{say,}
\end{aligned}$$

where $\boldsymbol{\varepsilon}_{i\cdot} = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\tau$, $\mathbf{X}_{i\cdot} = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})^\tau$, $\boldsymbol{\alpha}_s(\mathbf{U}_{i\cdot s}) = (\alpha_s(U_{i1s}), \ldots, \alpha_s(U_{iTs}))^\tau$, and $\widehat{\boldsymbol{\alpha}}_s(\mathbf{U}_{i\cdot s}) = (\widehat{\alpha}_s(U_{i1s}), \ldots, \widehat{\alpha}_s(U_{iTs}))^\tau$. By Theorem 1,

$$J_{42} = O_p(n^{-1}) + O_p\left(\max_{\{1 \le s \le q\}} \kappa_s n^{-1} + \max_{\{1 \le s \le q\}} \varphi_s^2 \right) = o_p(n^{-1/2}).$$

Same as for J_2, we can show that $J_{43} = o_p(n^{-1/2})$. In addition,

$$\begin{aligned}
J_{41} &= \frac{1}{n(T-1)} \sum_{i=1}^{n} (\mu_i \mathbf{C} \mathbf{1}_T + \mathbf{C}\boldsymbol{\nu}_{i\cdot})^\tau \mathbf{E}_T^\rho (\mu_i \mathbf{C} \mathbf{1}_T + \mathbf{C}\boldsymbol{\nu}_{i\cdot}) \\
&= \frac{1}{n(T-1)} \sum_{i=1}^{n} \mathbf{C}\boldsymbol{\nu}_{i\cdot}^\tau \mathbf{E}_T^\rho \mathbf{C}\boldsymbol{\nu}_{i\cdot} \\
&= \frac{1}{n(T-1)} \sum_{i=1}^{n} (\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT}) \mathbf{E}_T^\rho (\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})^\tau \\
&= \frac{1}{n(T-1)} \sum_{i=1}^{n} \varpi_i \quad \text{say,} \quad \text{where } \; \boldsymbol{\nu}_{i\cdot} = (\nu_{i1}, \ldots, \nu_{iT})^\tau.
\end{aligned}$$

It is easy to see that the ϖ_i's are i.i.d. random variables with mean

$$\begin{aligned}
E(\varpi_i) &= \mathrm{tr}\left\{ \mathbf{E}_T^\rho E((\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})^\tau (\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})) \right\} \\
&= \sigma_e^2 \mathrm{tr}(\mathbf{E}_T^\rho) = (T-1)\sigma_e^2
\end{aligned}$$

and variance $\mathrm{Var}((\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT}) \mathbf{E}_T^\rho (\sqrt{1-\rho^2}\nu_{i1}, e_{i2}, \ldots, e_{iT})^\tau)$. This proves (ii).

Part (iii) follows from the same arguments as for $\widehat{\sigma}_{en}^2$. We omit the details.

Proof of Theorem 3. According to the definitions of $\widehat{\boldsymbol{\beta}}_n^w$ and $\widetilde{\boldsymbol{\beta}}_n^w$ and the equality $a_1b_1 - a_2b_2 = (a_1-b_1)(a_2-b_2) + (a_1-b_1)b_2 + b_1(a_2-b_2)$, we have

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_n^w - \boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}_n^w - \boldsymbol{\beta} &+ \left\{(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X})^{-1} - (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1}\right\} \\
&\cdot \left\{\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q) - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q)\right\} \\
&+ \left\{(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X})^{-1} - (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1}\right\} \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q) \\
&+ (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1} \left\{\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q) - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q)\right\} \\
&+ \left\{(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X})^{-1} - (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1}\right\} \left(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\boldsymbol{\varepsilon} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}\right) \\
&+ \left\{(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X})^{-1} - (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1}\right\} \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} \\
&+ (\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X})^{-1} \left(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\boldsymbol{\varepsilon} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}\right),
\end{aligned}
$$

where $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{1T}, \ldots, \varepsilon_{nT})^\tau$. Consequently, by combining Lemma A.2 with the fact that $(\mathbf{A} + a\mathbf{B})^{-1} = \mathbf{A}^{-1} - a\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} + O(a^2)$ as $a \to 0$, it suffices to show

$$n^{-1}(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X}) = O_p(n^{-1/2}), \quad \text{(A.2)}$$

$$n^{-1}\{\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q) - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q)\} = o_p(n^{-1/2}), \quad \text{(A.3)}$$

$$n^{-1}(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\boldsymbol{\varepsilon} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}) = o_p(n^{-1/2}), \quad \text{(A.4)}$$

$$n^{-1}\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X} = O_p(1), \quad \text{(A.5)}$$

$$n^{-1}\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q) = o_p(n^{-1/2}) \text{ and } n^{-1}\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon} = O_p(n^{-1/2}). \quad \text{(A.6)}$$

According to the proof of Lemma A.2 and the $\sqrt{n}$ consistency of $\widehat{\boldsymbol{\Sigma}}_0$ we have

$$n^{-1}(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\mathbf{X} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\mathbf{X}) = n^{-1}(\boldsymbol{\Pi}^\tau \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Pi} - \boldsymbol{\Pi}^\tau \boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}) + o_p(n^{-1/2}),$$

where $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_{11}, \ldots, \boldsymbol{\Pi}_{1T}, \ldots, \boldsymbol{\Pi}_{nT})^\tau$. Further, $\boldsymbol{\Pi}^\tau \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Pi}$ can be written as $\sum_{i=1}^n \boldsymbol{\Pi}_{i\cdot} \widehat{\boldsymbol{\Sigma}}_0^{-1} \boldsymbol{\Pi}_{i\cdot})^\tau$. Correspondingly, $\boldsymbol{\Pi}^\tau \boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} = \sum_{i=1}^n \boldsymbol{\Pi}_{i\cdot} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Pi}_{i\cdot}^\tau$. By the $\sqrt{n}$ consistency of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$,

$$n^{-1}\boldsymbol{\Pi}_{i\cdot} \widehat{\boldsymbol{\Sigma}}_0^{-1} \boldsymbol{\Pi}_{i\cdot}^\tau - n^{-1}\boldsymbol{\Pi}_{i\cdot} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Pi}_{i\cdot}^\tau = O_p(n^{-1/2}).$$

This implies (A.2). Since the left side of (A.3) is

$$n^{-1}\{\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q - \mathbf{D}\boldsymbol{\theta}) - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_q - \mathbf{D}\boldsymbol{\theta})\} = o_p(n^{-1/2})$$

we can prove (A.3) by the same argument. Moreover, (A.5) and (A.6) follow from the proof of Lemma A.2. It remains to prove (A.4). By the proof of Lemma A.2 and the $\sqrt{n}$ consistency of $\widehat{\boldsymbol{\Sigma}}_0^{-1}$,

$$n^{-1}(\mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\widehat{\boldsymbol{\Sigma}}^{-1}}\boldsymbol{\varepsilon} - \mathbf{X}^\tau \mathbf{M}_{\mathbf{D}}^{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\varepsilon}) = n^{-1}(\boldsymbol{\Pi}^\tau \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\varepsilon} - \boldsymbol{\Pi}^\tau \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}) + o_p(n^{-1/2}).$$

Further, $\mathbf{\Pi}^\tau \widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\varepsilon} = \sum_{i=1}^n \mathbf{\Pi}_{i\cdot}\widehat{\mathbf{\Sigma}}_0^{-1}\mathbf{\Pi}_{i\cdot}^\tau$ and $\mathbf{\Pi}^\tau \mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon} = \sum_{i=1}^n \mathbf{\Pi}_{i\cdot}\mathbf{\Sigma}_0^{-1}\mathbf{\Pi}_{i\cdot}^\tau$. Hence the $\sqrt{n}$ consistency of $\widehat{\mathbf{\Sigma}}_0^{-1}$ gives $\mathbf{\Pi}_{i\cdot}\widehat{\mathbf{\Sigma}}_0^{-1}(\varepsilon_{i1},\ldots,\varepsilon_{iT})^\tau - \mathbf{\Pi}_{i\cdot}\mathbf{\Sigma}_0^{-1}(\varepsilon_{i1},\ldots,\varepsilon_{iT})^\tau = O_p(n^{-1/2})$. This implies (A.4) and completes the proof.

Proof of Theorem 4. The proof of Theorem 4 is straightforward. We omit the details.

Proof of Theorem 5. Let

$$\widehat{Y}_{it}^{**} = (\sigma^{tt})^{-1}\Big(\widehat{Y}_{it} - \sigma^{tt}\widehat{a}_{it}^{-s} - \sum_{t_1\neq t}^{T}\sigma^{tt_1}\widehat{a}_{it_1}\Big)$$
$$= (\sigma^{tt})^{-1}\Big(\sum_{t_1=1}^{T}\sigma^{tt_1}Y_{it_1} - \sigma^{tt}\widehat{a}_{it}^{-s} - \sum_{t_1\neq t}^{T}\sigma^{tt_1}\widehat{a}_{it_1}\Big)$$

and $(\widehat{\alpha}_{sn}^{TS*}(u), \widehat{\alpha}_{sn}^{'TS*}(u))^\tau = (\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\mathbf{D}_{su}\mathbf{W}_{su}\widehat{\mathbf{Y}}^{**}$, where $\widehat{\mathbf{Y}}^{**} = (\widehat{Y}_{11}^{**},\ldots, \widehat{Y}_{1T}^{**},\ldots,\widehat{Y}_{nT}^{**})^\tau$. Then we have

$$(\widehat{\alpha}_{sn}^{TS*}(u), \widehat{\alpha}_{sn}^{'TS*}(u))^\tau - (\alpha_s(u), \alpha_s^{'}(u))^\tau = J_1 + J_2 + J_3 - J_4,$$

where

$$J_1 = (\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T \begin{pmatrix}1\\ U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\alpha_s(U_{its}) - (\alpha_s(u),\alpha_s^{'}(u))^\tau,$$
$$J_2 = (\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T \begin{pmatrix}1\\ U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\sum_{t_1=1}^{T}(\sigma^{tt})^{-1}\sigma^{tt_1}\varepsilon_{it_1},$$
$$J_3 = (\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T \begin{pmatrix}1\\ U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)(\sigma^{tt})^{-1}$$
$$\cdot\Big\{\sigma^{tt}\sum_{s_1\neq s}^{q}[\alpha_{s_1}(U_{its_1}) - \widehat{\alpha}_{s_1 n}(U_{its_1})] + \sum_{t_1\neq t}^{T}\sigma^{tt_1}\sum_{s_1=1}^{q}[\alpha_{s_1}(U_{it_1s_1}) - \widehat{\alpha}_{s_1 n}(U_{it_1s_1})]\Big\},$$
$$J_4 = (\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T \begin{pmatrix}1\\ U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)$$
$$\sum_{t_1=1}^{T}\sigma^{tt_1}\sum_{s_1=1}^{q}[\alpha_{s_1}(U_{iTs_1}) - \widehat{\alpha}_{s_1 n}(U_{iTs_1})].$$

We first show that $J_s = o_p(n^{-2/5}),\ s = 3, 4$. To do so it suffices to prove

$$\left\|(\mathbf{D}_{su}^\tau \mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T \begin{pmatrix}1\\ U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)(\alpha_{s_1}(U_{its_1}) - \widehat{\alpha}_{s_1 n}(U_{its_1}))\right\|$$
$$= o_p(n^{-2/5}). \tag{A.7}$$

Write $\boldsymbol{\zeta}_s(u) = (\zeta_{s1}(u), \ldots, \zeta_{s\kappa_s}(u))^\tau$, $\boldsymbol{\alpha}_j = (\alpha_j(X_{11j}), \ldots, \alpha_j(X_{1Tj}), \ldots, \alpha_j(X_{nTj}))^\tau$, $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{1T}, \ldots, \nu_{nT})^\tau$, and $\boldsymbol{\nabla} = (\mathbf{0}_{\kappa_{s_1}\times(\kappa_1+\cdots+\kappa_{s_1-1})}, \mathbf{I}_{\kappa_{s_1}\times\kappa_{s_1}}, \mathbf{0}_{\kappa_{s_1}\times(\kappa_{s_1+1}+\cdots+\kappa_q)})(\mathbf{D}^\tau\mathbf{D})^{-1}\mathbf{D}^\tau$. Then

$$\begin{aligned}
&\alpha_{s_1}(U_{its_1}) - \widehat{\alpha}_{s_1n}(U_{its_1}) = \alpha_{s_1}(U_{its_1}) - \boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\mathbf{Y}\\
&\quad = \left\{\alpha_{s_1}(U_{its_1}) - \boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\theta}_{s_1}\right\} - \boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\left(\sum_{s=1}^p \boldsymbol{\alpha}_s - \mathbf{D}\boldsymbol{\theta}\right) - \boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\boldsymbol{\varepsilon},
\end{aligned}$$

where $n^{-1}\left(\sum_{s=1}^q \boldsymbol{\alpha}_s - \mathbf{D}\boldsymbol{\theta}\right)^\tau\left(\sum_{s=1}^q \boldsymbol{\alpha}_s - \mathbf{D}\boldsymbol{\theta}\right) = O_p\left(\max_{1\le s\le q}\kappa_s^{-2}\right)$. It is easy to see that

$$\begin{aligned}
&\left\|(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T\begin{pmatrix}1\\U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\left\{\alpha_{s_1}(U_{its_1})-\boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\theta}_{s_1}\right\}\right\|\\
&\le \left\|(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T\begin{pmatrix}1\\U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\right\|\\
&\quad\cdot\max_{1\le i\le n,1\le t\le T}\left|\alpha_{s_1}(U_{its_1}) - \boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\theta}_{s_1}\right|\\
&= O_p(\kappa_{s_1}^{-2}) = O_p\left\{n^{-2/5}(\log n)^{-2}\right\} = o_p(n^{-2/5}),
\end{aligned}$$

and

$$\begin{aligned}
&\left\|(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T\begin{pmatrix}1\\U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\left(\sum_{s=1}^q\boldsymbol{\alpha}_s-\mathbf{D}\boldsymbol{\theta}\right)\right\|\\
&\le \left\|(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T\begin{pmatrix}1\\U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\right\|\\
&\quad\cdot\max_{1\le i\le n,1\le t\le T}\left|\boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\left(\sum_{s=1}^q\boldsymbol{\alpha}_s - \mathbf{D}\boldsymbol{\theta}\right)\right|\\
&\le O_p(1)\cdot\max_{1\le i\le n,1\le t\le T}\sqrt{\left(\sum_{s=1}^q\boldsymbol{\alpha}_s-\mathbf{D}\boldsymbol{\theta}\right)^\tau\boldsymbol{\nabla}^\tau\boldsymbol{\zeta}_{s_1}(U_{its_1})\boldsymbol{\zeta}_{s_1}^\tau(U_{its_1})\boldsymbol{\nabla}\left(\sum_{s=1}^q\boldsymbol{\alpha}_s-\mathbf{D}\boldsymbol{\theta}\right)}\\
&= O_p(\max_{1\le s\le q}\kappa_s^{-2}) = O_p\left\{n^{-2/5}(\log n)^{-2}\right\} = o_p(n^{-2/5}).
\end{aligned}$$

Following the same line as at Lemma 5.1 in Wang and Yang (2007), we can show that

$$\begin{aligned}
&\left\|(\mathbf{D}_{su}^\tau\mathbf{W}_{su}\mathbf{D}_{su})^{-1}\sum_{i=1}^n\sum_{t=1}^T\begin{pmatrix}1\\U_{its}\end{pmatrix}\frac{1}{h_s}K\left(\frac{U_{its}-u}{h_s}\right)\boldsymbol{\zeta}_{s_1}(U_{its_1})\boldsymbol{\nabla}\boldsymbol{\varepsilon}\right\|\\
&\quad = O_p\left\{\max_{1\le s\le q}\kappa_s(\log n)^2n^{-1}\right\} = o_p(n^{-2/5}).
\end{aligned}$$

Together, we have established (A.7) and hence $J_s = o_p(n^{-2/5})$, $s = 3, 4$. In addition, note that

$$\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}\mathbf{D}_{su} = \begin{pmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} K_{h_s}(U_{its}-u) & \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{U_{its}-u}{h_s}\right)K_{h_s}(U_{its}-u) \\ \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{U_{its}-u}{h_s}\right)K_{h_s}(U_{its}-u) & \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{U_{its}-u}{h_s}\right)^2 K_{h_s}(U_{its}-u) \end{pmatrix}.$$

Each element of the above matrix is in the form of kernel regression. By Lemma A.1,

$$\frac{1}{n}\mathbf{D}_{su}^{\tau}\mathbf{W}_{su}\mathbf{D}_{su} = \sum_{t=1}^{T} p_{ts}(u) \otimes \mathbf{H}_s \begin{pmatrix} 1 & \varsigma_1 \\ \varsigma_1 & \varsigma_2 \end{pmatrix} \mathbf{H}_s \cdot O_p\left(1 + \left\{\frac{\log n}{nh_s}\right\}^{1/2}\right).$$

Therefore, by the usual nonparametric regression result, we have

$$\sqrt{nTh_s}\left[\mathbf{H}_s^{-1}\left\{J_1 - \begin{pmatrix}\alpha_s(u)\\ \alpha_s'(u)\end{pmatrix}\right\} - \frac{h_s^2}{2}\begin{pmatrix}\Im_1\alpha_s''(u)\\ \Im_2\alpha_s''(u)\end{pmatrix} + o(h_s^2)\right] = o_p(1).$$

Next we show that

$$\sqrt{nTh_s}\mathbf{H}_s^{-1}J_2 \xrightarrow{D} N(0, \mathbf{\Omega}_s^{TS}) \quad \text{as} \quad n \to \infty. \tag{A.8}$$

Let

$$Q_s = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{d_1 + d_2\left(\frac{U_{its}-u}{h_s}\right)\right\}K_{h_s}(U_{its}-u)\sum_{t_1=1}^{T}(\sigma^{tt})^{-1}\sigma^{tt_1}\varepsilon_{it_1},$$

where d_1 and d_2 are any non-zero constants. It is easy to see that $E(Q_s) = 0$ and

$$\begin{aligned}
&\mathrm{Var}(\sqrt{nTh_s}Q_s)\\
&= \frac{h_s}{nT}\sum_{i=1}^{n}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} E\Big\{d_1 + d_2\Big(\frac{U_{it_1s}-u}{h_s}\Big)\Big\}\Big\{d_1 + d_2\Big(\frac{U_{it_2s}-u}{h_s}\Big)\Big\}\\
&\quad \cdot K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)\Big\{\sum_{t_3=1}^{T}(\sigma^{t_1t_1})^{-1}\sigma^{t_1t_3}\varepsilon_{it_3}\Big\}\Big\{\sum_{t_3=1}^{T}(\sigma^{t_2t_2})^{-1}\sigma^{t_2t_3}\varepsilon_{it_3}\Big\}\\
&= \frac{h_s}{nT}\sum_{i=1}^{n}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} E\Big[\Big\{d_1 + d_2\Big(\frac{U_{it_1s}-u}{h_s}\Big)\Big\}\Big\{d_1 + d_2\Big(\frac{U_{it_2s}-u}{h_s}\Big)\Big\}\\
&\quad \cdot K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)\Big]\\
&\quad E\Big\{\Big(\sum_{t_3=1}^{T}(\sigma^{t_1t_1})^{-1}\sigma^{t_1t_3}\varepsilon_{it_3}\Big)\Big(\sum_{t_3=1}^{T}(\sigma^{t_2t_2})^{-1}\sigma^{t_2t_3}\varepsilon_{it_3}\Big)\Big\}.
\end{aligned}$$

Note that for $t_1 \neq t_2$,

$$E\left\{K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)\right\}=O(1),$$

$$E\left\{\left(\frac{U_{it_2s}-u}{h_s}\right)K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)\right\}=O(1),$$

$$E\left\{\left(\frac{U_{it_1s}-u}{h_s}\right)K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)\right\}=O(1),$$

$$E\left(\frac{U_{it_1s}-u}{h_s}\right)\left(\frac{U_{it_2s}-u}{h_s}\right)K_{h_s}(U_{it_1s}-u)K_{h_s}(U_{it_2s}-u)=O(1).$$

In addition,

$$E\Big\{\sum_{t_2=1}^{T}(\sigma^{t_1t_1})^{-1}\sigma^{t_1t_2}\varepsilon_{it_2}\Big\}^2=(\sigma^{t_1t_1})^{-2}\mathbf{e}_{t_1T}\mathbf{\Sigma}_0^{-1}\mathbf{e}_{t_1T}=(\sigma^{t_1t_1})^{-1},$$

where $\mathbf{e}_{t_1T}$ is a T-vector with 1 in the t_1-th position and zeros elsewhere. It follows that

$$\begin{aligned}
&\mathrm{Var}(\sqrt{nTh_s}Q_s)\\
&=o(1)+\frac{h_s}{nT}\sum_{i=1}^{n}\sum_{t_1=1}^{T}E\left[\left\{d_1+d_2\left(\frac{U_{it_1s}-u}{h_s}\right)\right\}^2K_{h_x}^2(U_{it_1s}-u)\right]\\
&\quad E\left(\sum_{t_2=1}^{T}(\sigma^{t_1t_1})^{-1}\sigma^{t_1t_2}\varepsilon_{it_2}\right)^2\\
&=o(1)+\frac{h_s}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\sigma^{tt})^{-1}E\left[\left\{d_1+d_2\left(\frac{U_{it_1s}-u}{h_s}\right)\right\}^2K_{h_s}^2(U_{its}-u)\right]\\
&=o(1)+\frac{h_sd_1d_2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\sigma^{tt})^{-1}EK_{h_s}^2(U_{its}-u))\\
&\quad+\frac{h_sd_1d_2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\sigma^{tt})^{-1}E\left\{\left(\frac{U_{its}-u}{h_s}\right)^2K_{h_s}^2(U_{its}-u)\right\}\\
&\quad+\frac{2h_sd_1d_2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\sigma^{tt})^{-1}E\left\{\left(\frac{U_{its}-u}{h_s}\right)K_{h_s}^2(U_{its}-u)\right\}\\
&=o(1)+J_5+J_6+J_7,\ \text{say}.
\end{aligned}$$

Direct computations show that $J_5\to_p d_1d_2\sum_{t=1}^{T}(\sigma^{tt})^{-1}p_{ts}(u)\varrho_0$, $J_6\to_p d_1d_2\sum_{t=1}^{T}(\sigma^{tt})^{-1}p_{ts}(u)\varrho_2$ and $J_7\to_p 2d_1d_2\sum_{t=1}^{T}p_{ts}(u)\varrho_1$ as $n\to\infty$. Therefore,

$$\mathrm{Var}(\sqrt{nTh_s}Q_s)=d_1d_2\sum_{t=1}^{T}(\sigma^{tt})^{-1}p_{ts}(u)(\varrho_0+2\varrho_1+\varrho_2)+o(1).$$

Let

$$b_{is}=\sqrt{h_s}\sum_{t=1}^{T}\left\{d_1+d_2\left(\frac{U_{its}-u}{h_s}\right)\right\}K_{h_s}(U_{its}-u)\Big\{\sum_{t_1=1}^{T}(\sigma^{tt})^{-1}\sigma^{tt_1}\varepsilon_{it_1}\Big\}$$

and $B_{ns}^2=\sum_{i=1}^n E(b_{is}^2)$. Then

$$B_{ns}^2=nd_1d_2\sum_{t=1}^{T}(\sigma^{tt})^{-1}p_{ts}(u)(\varrho_0+2\varrho_1+\varrho_2)+o(n).$$

Simple calculation shows that

$$\begin{aligned}\sum_{i=1}^{n}E|b_{is}|^3&\le O(1)\cdot\sum_{i=1}^{n}\sum_{t=1}^{T}h_s^{3/2}E\left\{|d_1|+|d_2|\cdot\left|\frac{U_{its}-u}{h_s}\right|\right\}^3K_{h_s}^3(U_{its}-u)\\&=O(h_s^{-1/2}).\end{aligned}$$

It follows that the Lindeberg condition $\lim_{n\to\infty}B_{ns}^{-3}\sum_{i=1}^n E|b_{is}^3|=0$ is satisfied. Hence (A.8) follows from the Central Limit Theorem. Together with the $\sqrt{n}$ consistency of $\widehat{\sigma}^{tt_1}$, the proof of Theorem 5 is complete.

References

Ahn, S. C. and Schmidt, P. (2000). Estimation of linear panel data models using GMM. In *Generalized Method of Moments Estimation* (Edited by L. Mátyás). Cambridge University Press.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. Wiley.

Baltagi, B. H. and Li, Q. (1991). A transformation that will circumvent the problem of autocorrelation in an error-component model. *J. Econometrics* **48**, 385-393.

Carter Hill, R., Griffiths, W. E. and Lim, G. C. (2008). *Principles of Econometrics*. Wiley.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York-Berlin.

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of additive and linear components for high dimensional data. *Ann. Statist.* **26**, 943-971.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715-731.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Fan, Y. and Li, Q. (2003). A kernel-based method for estimating additive partially linear models, *Statist. Sinica* **13**, 739-762.

Honore, B. E. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* **60**, 533-565.

Horowitz, J. L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63** 145-168.

Horowitz, J. L. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *Ann. Statist.* **32**, 2412-2443.

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

Kniesner, T. and Li, Q. (1994). Semiparametric panel data models with dynamic adjustment: Theoretical considerations and an application to labor supply. Manuscript, University of Guelph.

Li, Q. (2000). Efficient estimation of additive partially linear models. *Internat. Economic Rev.* **41**, 1073-1098.

Li, Q. and Ullah, A. (1998). Estimating partially linear panel data models with one-way error components. *Economic Rev.* **17**, 145-166.

Liu, M., Cheng, R. and Yao, Q. (2009). Nonparametric transfer function models. Accepted by *J. Econometrics.*

Liu, R. and Yang, L. (2009). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theo.* **25**, 1-30.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61**, 405-415.

Mundra, K. (1997). Immigration and U.S. bilateral trade flows: A semiparametric empirical analysis. Manuscript, University of California, Riverside.

Mundra, K. (2005). Nonparametric slope estimators for fixed-effect panel data. Manuscript, San Diego State University.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.

Roy, N. (1997). Nonparametric and semiparametric analysis of panel data models: An application to Calorie-Income relation for rural south India. Ph.D. Thesis, University of California, Riverside.

Ruckstuhl, A. F., Welsh, A. H. and Carroll, R. J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statist. Sinica* **10**, 51-71.

Schumaker, L. L. (1981). *Spline functions: Basic Theory. Pure and Applied Mathematics.* Wiley, New York.

Ullah, A. and Mundra, K. (2002). Semiparametric panel data estimation: an application to immigrants' homelink effect on U.S. producer trade flows. In *Handbook of Applied Econometrics and Statistical Inference*, 591-608. Dekker, New York.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Springer-Verlag, New York.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35**, 2474-2503.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90**, 43-52.

You, J. H. and Zhou, H. (2007). Two-stage efficient estimation of longitudinal nonparametric additive models. *Statist. Prob. Lett.* **77**, 1666-1675.

You, J. H. and Zhou, X. (2006). Statistical inference in a panel data semiparametric model with serially correlated errors. *J. Multivariate Anal.* **97**, 844-873.

You, J. H. and Zhou, X. (2009). Partially linear models and polynomial spline approximations for the analysis of unbalanced panel data. *J. Statist. Plann. Inference* **139**, 679-695.

You, J. H., Zhou, X. and Zhou, Y. (2010). Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors. *J. Multivariate Anal.* **101**,1079-1101.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric model for longitudinal data with application to CD4 cell numbers in HIV seroconvertiers. *Biometrics* **50**, 689-699.

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, P.R. China;
Key Laboratory of Mathematical Economics (SUFE), Ministry of Education of China.

E-mail: johnyou07@gmail.com

Department of Applied Finance and Actuarial Studies, Macquarie University, North Ryde, NSW 2109, Australia.

E-mail: xian.zhou@mq.edu.au

(Received September 2010; accepted November 2011)