Statistica Sinica **25** (2015), 275-294
doi:http://dx.doi.org/10.5705/ss.2013.190w

PANEL DATA PARTIALLY LINEAR VARYING-COEFFICIENT MODEL WITH ERRORS CORRELATED IN SPACE AND TIME

Yang Bai[1,2], Jianhua Hu[1,2] and Jinhong You[1,2]

[1]*Shanghai University of Finance and Economics and*
[2]*Key Laboratory of Mathematical Economics (SUFE), Ministry of Education*

Abstract: In this paper, we consider a panel data varying-coefficient partially linear model errors correlated in space and time. A serially correlated error structure is adopted for the correlation in time, and we propose an estimating procedure for the autoregressive coefficients in our set-up by combining a polynomial spline series approximation with least squares. The resulted estimators are shown to enjoy asymptotic properties. We construct a weighted semiparametric least squares estimator (WSLSE) and a weighted polynomial spline series estimator (WPSSE) for the parametric and nonparametric components of the mean model, respectively. The WSLSE is shown to be asymptotically normal and more efficient than the unweighted one, and the WPSSE is shown to achieve the optimal nonparametric convergence rate. Some simulation studies are reported to illustrate the finite sample performance of the proposed procedure. An application to Indonesian rice farming data is given.

Key words and phrases: Asymptotic normality, panel data, partially linear varying coefficient model, spatial, time-wise correlation.

1. Introduction

In Statistics and Econometrics, panel data refers to observations on a cross-section of countries, households, firms, individuals, or patients, etc., over multiple time periods. Compared with traditional time series or cross-sectional data, there are several benefits of its use: panel data can control individual heterogeneity; panel data give more data points, therefore are more informative and less collinear among the explanatory variables; panel data are better able to study the dynamics of adjustment. See Baltagi (2008) for details. Various parametric, nonparametric, and semiparametric models and corresponding statistical methods have been developed for analyzing panel data with assuming cross-sectional independence in the last decades. See, for instance, Diggle, Liang, and Zeger (1994) and Baltagi (2008) for the parametric modeling; Ullah and Roy (1998) and Ruckstuhl, Welsh, and Carroll (2000) for the nonparametric modeling; and

Horowitz and Markatou (1996), Wang, Carroll, and Lin (2005) Fan, Huang, and Li (2007) and Henderson, Carroll, and Li (2008) for the semiparametric modeling.

When sample data are randomly drawn from a population, the cross-sectional independence may not be a worry. However, if competitions between cross-sectional units, copy-cat policies, net work issues, spill-overs, externalities, regional issues, etc., are involved, the cross-sectional independence may not hold (e.g., see Kapoor, Kelejian, and Prucha (2007)). An attractive way of allowing for interdependence between cross-sectional units in an empirical modeling is by means of a spatial method. In spatial models, interactions between cross sectional units are typically modeled in terms of some measure of distance between them. So far, the most widely used spatial models are variants of the one considered by Cliff and Ord (1981). The statistical inferences of spatial cross-sectional data models are discussed in Kelejian and Robinson (1992), Kelejian and Prucha (2001, 2004, 2010), Lee (2002, 2005), Su and Jin (2010), Su (2012), and so on. Recent theoretical contributions and applications on panel data spatial models include Baltagi, Song, and Koh (2003), Druska and Horrace (2004), Egger, Pfaffermayr, and Winner (2005), Baltagi, Egger, and Pfaffermayr (2007), Lee and Yu (2010), and Korniotis (2010). Kapoor, Kelejian, and Prucha (2007) generalized the procedure of generalized method of moments (GMM) in Kelejian and Prucha (1999) to panel data models involving a first-order spatially autoregressive disturbance term, whose innovations have an error component structure; Badinger and Egger (2013) developed an estimation for higher-order spatial autoregressive panel data error component models with spatial autoregressive disturbances, derived the moment conditions and optimal weighting matrix without distributional assumptions for a generalized moments estimation procedure of the spatial autoregressive parameters of the disturbance process and defined a generalized two-stages least squares estimator for the mean regression parameters of the model; Mutl and Pfaffermayr (2011) discussed an instrumental variable estimation under both the fixed and the random effects specifications, proposed a spatial Hausman test which compared these two models accounting for spatial autocorrelation in the disturbances and showed that the test statistic follows, asymptotically, a chi-squared distribution.

Almost all works on spatial panel data mentioned above focus on parametric models, although some scholars have investigated the estimating problem of spatial cross-sectional data semiparametric models (e.g., see Su and Jin (2010) and Su (2012)). Parametric models are useful for analyzing panel data and providing a parsimonious description of the relationship between the response variable and its covariates. They are, however, often subject to the risk of increasing modeling biases. Semiparametric and structural nonparametric panel data regression models are good alternatives that keep a balance between a general nonparametric framework and a fully parametric specification. In this paper, we propose a semiparametric spatial panel data model, namely spatial panel data varying-coefficient partially linear model in which the errors are allowed to be correlated in space and time, and the correlation in time is described by a serially correlated error component structure. Obviously, a serially correlated error component structure is more general than the classical one-way error component structure that assumes the same correlation between errors no matter how large the distance of the two observed time points (e.g., see Baltagi (2008)).

We consider the panel data varying-coefficient partially linear model

$$\mathbf{Y}_N(t) = \mathbf{X}_N(t)\boldsymbol{\beta} + \mathbf{Z}_N(t) \odot \mathbf{M}_N(\mathbf{U}_N(t)) + \boldsymbol{\varepsilon}_N(t),\ t = 1, \ldots, T, \tag{1.1}$$

where $\mathbf{Y}_N(t) = (Y_{1t}, \ldots, Y_{Nt})^\tau$ are the response vectors, $\mathbf{X}_N(t) = (\mathbf{X}_{1t}, \ldots, \mathbf{X}_{Nt})^\tau_{p\times N}$, $\mathbf{Z}_N(t) = (\mathbf{Z}_{1t}, \ldots, \mathbf{Z}_{Nt})^\tau_{q\times N}$, and $\mathbf{U}_N(t) = (U_{1t}, \ldots, U_{Nt})^\tau$ are realizations of explanatory variables $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$, respectively. In addition, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a p-dimensional unknown coefficient vector, $\mathbf{M}_N(\mathbf{U}_N(t)) = (\mathbf{m}(U_{1t}), \ldots, \mathbf{m}(U_{Nt}))^\tau_{q\times N}$ with $\mathbf{m}(\cdot) = (m_1(\cdot), \ldots, m_q(\cdot))^\tau$ a q-dimensional unknown function vector, $\boldsymbol{\varepsilon}_N(t) = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\tau$ are random noises and "$\odot$" denotes an operator of two matrices such that $\mathbf{G_1} \odot \mathbf{G_2} = (\sum_{j=1}^{p_2} g_{11j}g_{21j}, \ldots, \sum_{j=1}^{p_2} g_{1p_1j}g_{2p_1j})^\tau$ for two $p_1 \times p_2$ matrices $\mathbf{G}_1 = (g_{1ij})$ and $\mathbf{G}_2 = (g_{2ij})$.

We further assume that the random errors $\boldsymbol{\varepsilon}_N(t)$ in the model (1.1) follow the spatially and serially correlated error component structure

$$\boldsymbol{\varepsilon}_N(t) = \lambda\mathbf{W}_N\boldsymbol{\varepsilon}_N(t) + \boldsymbol{\eta}_N(t) \quad \text{and} \quad \boldsymbol{\eta}_N(t) = \boldsymbol{\mu}_N + \boldsymbol{\nu}_N(t), \tag{1.2}$$

where λ is a scalar cross-sectional autoregressive parameter, and $\mathbf{W}_N$ is an $N \times N$ weighting matrix of known constants which does not depend on t. More specifically, the serially time-wisely correlated part $\boldsymbol{\eta}_N(t)$ has been partitioned into time independent error $\boldsymbol{\mu}_N = (\mu_1, \ldots, \mu_N)^\tau$ and correlated error $\boldsymbol{\nu}_N(t) = (\nu_{1t}, \ldots, \nu_{Nt})^\tau$ which satisfies

$$\nu_{it} = \rho_1\nu_{i(t-1)} + \cdots + \rho_s\nu_{i(t-s)} + e_{it},\quad 1 - \rho_1 z - \cdots - \rho_s z^s \neq 0 \text{ for } |z| \leq 1. \tag{1.3}$$

Here μ_i and e_{it} are i.i.d. random variables with zero mean and variances σ_μ^2 and σ_e^2, respectively, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)^\tau$ is an s-dimensional unknown autoregressive coefficient vector.

The model (1.1)−(1.3) is a generalization of many usual parametric, nonparametric and semiparametric models. When $\lambda = 0$ and $T = 1$, the model is the cross-sectional data varying-coefficient partially linear model studied by Fan, Peng, and Huang (2005) who proposed a profile least squares estimation and a quasi likelihood ratio test. When $\lambda = 0$, the model is the panel data varying-coefficient partially linear model without cross-sectional interdependence studied by Fan, Huang, and Li (2007); they proposed an estimation procedure for regression coefficients by combining a profile weighted least squares approach and estimating parameters in the correlation structure. When $m_1(\cdot) = \cdots = m_q(\cdot) \equiv 0$ and $\rho_1 = \cdots = \rho_s = 0$, the model is the panel data linear model with spatially correlated error components. See Kapoor, Kelejian, and Prucha (2007) for details.

To the best of our knowledge, this is the first work in which the estimating problem of modeling panel data with both spatially and time-wise correlated errors is investigated. Specifically, our main contributions include constructing a new generalized moment estimator for the autoregressive parameter in a spatial model by combining the polynomial spline series approximation with semiparametric least squares; investigating the fitting of a time correlated structure; based on the estimated spatially and time-wise correlated error structure, constructing, respectively, a weighted semiparametric least squares estimator (WSLSE) and a weighted polynomial spline series estimator (WPSSE) for the parametric and nonparametric components of the mean model; showing that the WSLSE is asymptotically normal and more efficient than the unweighted one, and that the WPSSE achieves the optimal nonparametric convergence rate.

The layout of the rest of the paper is as follows. In Section 2, we describe a semiparametric least squares estimation for the parametric and nonparametric components. In Section 3, the fitting of the spatial and time-wise correlation structures are investigated. In Section 4, we construct the WSLSE and WPSSE procedures for the parametric and nonparametric components, respectively, and establish their asymptotic properties. Section 5 illustrates results from simulation studies. An application is analyzed in Section 6. Concluding remarks are presented in Section 7. All proofs of main results are summarized in a online supplementary document due to space limitation.

2. Semiparametric Least Squares Estimation

As in Kelejian and Prucha (2001), we use a_{ij} to denote the ijth element of some matrix $\mathbf{A}$, $a_{i\cdot}$ and $a_{\cdot j}$ to denote ith row and jth column, respectively, while a_i denotes the ith element of a vector $\mathbf{a}$. Let $\mathbf{A}_N$ denote a sequence of $d_N \times d_N$ matrices with some fixed integer d_N. We say that the row and column sums of the sequence of matrices $\mathbf{A}_N$ are bounded uniformly in absolute value if there exist a constant $c_A < \infty$, not depending on N, such that

$$\max_{1\le i\le d_N} \sum_{j=1}^{d_N} |a_{ij,N}| \le c_A \quad \text{and} \quad \max_{1\le j\le d_N} \sum_{i=1}^{d_N} |a_{ij,N}| \le c_A \text{ for any } N \ge 1.$$

We choose the Euclidean norm, $||\mathbf{A}_N|| = \{\mathrm{tr}(\mathbf{A}_N^{\tau}\mathbf{A}_N)\}^{1/2}$, for the sequence $\mathbf{A}_N = (a_{ij,N})$.

Polynomial splines are piecewise polynomials jointed together smoothly at a set of interior points (*knots*). A polynomial spline of degree $r \geq 0$ on an interval $\mathcal{U}$ (without loss of generality, let $\mathcal{U} = [0,1]$) with inner knots $0 < \eta_1 < \cdots < \eta_M < 1$ is made of piecewise r-degree polynomial functions on each subinterval $[\eta_k, \eta_{k+1}), 0 \leq k \leq M-1$, and with continuous $r-1$ derivatives for $r > 2$. A piecewise constant function, linear spline, quadratic spline and cubic spline correspond to $r = 0, 1, 2, 3$ respectively. See de Boor (1978) for details.

In our case, let $\{\zeta_l(\cdot)\}_{l=1}^{\kappa_N}$ be a set of r-degree B-spline bases with $\kappa_N = r + 1 + M$. We approximate each unknown smoothing function $m_j(u)$ as

$$m_j(u) \approx \sum_{l=1}^{\kappa_N} \theta_{jl}\zeta_l(u),\ j = 1, \ldots, q,$$

where κ_N plays a role of smoothing parameter and $\boldsymbol{\theta}_j = (\theta_{j1}, \ldots, \theta_{j\kappa_N})^\tau$ is an unknown κ_N-dimensional pseudo coefficient vector. Thus, the model (1.1) can be approximated by

$$Y_{it} \approx \mathbf{X}_{it}^\tau \boldsymbol{\beta} + \sum_{j=1}^{q} Z_{itj}\Big\{\sum_{l=1}^{\kappa_N} \theta_{jl}\zeta_l(U_{it})\Big\} + \varepsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T. \quad (2.1)$$

Write $\mathbf{Y}_N = (\mathbf{Y}_N^\tau(1), \ldots, \mathbf{Y}_N^\tau(T))^\tau, \mathbf{X}_N = (\mathbf{X}_N^\tau(1), \ldots, \mathbf{X}_N^\tau(T))^\tau$, $\boldsymbol{\varepsilon}_N = (\boldsymbol{\varepsilon}_N^\tau(1), \ldots, \boldsymbol{\varepsilon}_N^\tau(T))^\tau$, $\mathbf{Z}_N^* = (\mathbf{Z}_N^{*\tau}(1), \ldots, \mathbf{Z}_N^{*\tau}(T))^\tau$ and $\mathbf{Z}_N^*(t) = (\mathbf{Z}_{1t}^*, \ldots, \mathbf{Z}_{Nt}^*)^\tau$ with $\mathbf{Z}_{it}^* = (Z_{it1}(\zeta_1(U_{it}), \cdots, \zeta_{\kappa_N}(U_{it})), \cdots, Z_{itq}(\zeta_1(U_{it}), \cdots, \zeta_{\kappa_N}(U_{it})))^\tau$. In matrix form, the model (2.1) is

$$\mathbf{Y}_N \approx \mathbf{X}_N\boldsymbol{\beta} + \mathbf{Z}_N^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}_N \quad (2.2)$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\tau, \ldots, \boldsymbol{\theta}_q^\tau)^\tau$. If we define $\mathbf{M}_{\mathbf{Z}_N^*} = \mathbf{I}_{NT} - \mathbf{Z}_N^*(\mathbf{Z}_N^{*\tau}\mathbf{Z}_N^*)^{-1}\mathbf{Z}_N^{*\tau}$, then $\mathbf{M}_{\mathbf{Z}_N^*}\mathbf{Z}_N^*\boldsymbol{\theta}_N = \mathbf{Z}_N^*\boldsymbol{\theta}_N - \mathbf{Z}_N^*\boldsymbol{\theta}_N = \mathbf{0}$, the model (2.2) leads to

$$\mathbf{M}_{\mathbf{Z}_N^*}\mathbf{Y}_N \approx \mathbf{M}_{\mathbf{Z}_N^*}\mathbf{X}_N\boldsymbol{\beta} + \mathbf{M}_{\mathbf{Z}_N^*}\boldsymbol{\varepsilon}_N. \quad (2.3)$$

If we take $\mathbf{M}_{\mathbf{Z}_N^*}\boldsymbol{\varepsilon}_N$ as the residuals, the model (2.3) can result in the conventional least squared estimator of $\boldsymbol{\beta}$,

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

Substituting $\widehat{\boldsymbol{\beta}}_N$ into (2.2), we get a profiled least squared estimator of $\boldsymbol{\theta}_N$, written as

$$\widehat{\boldsymbol{\theta}}_N = (\mathbf{Z}_N^{*\tau}\mathbf{Z}_N^*)^{-1}\mathbf{Z}_N^{*\tau}(\mathbf{Y}_N - \mathbf{X}_N\widehat{\boldsymbol{\beta}}_N).$$

This helps us attain the polynomial spline series estimator of $\mathbf{m}(u) = (m_1(u), \ldots, m_q(u))^\tau$, denoted by

$$\widehat{\mathbf{m}}_N(u) = (\widehat{m}_{1,N}(u), \ldots, \widehat{m}_{q,N}(u))^\tau = \boldsymbol{\zeta}_N^*(u)\widehat{\boldsymbol{\theta}}_N$$

with $\boldsymbol{\zeta}_N^*(\cdot) = \mathbf{1}_q \otimes (\zeta_1(\cdot), \ldots, \zeta_{\kappa_N}(\cdot))$; that is a nonparametric projecting estimator.

In order to present asymptotic properties of $\widehat{\boldsymbol{\beta}}_N, \widehat{\mathbf{m}}_N(u)$ and other estimators proposed in the following sections, some notations and technical assumptions are useful.

Assumption A.

(A1) $\{U_{it}\}, i = 1, \ldots, N, t = 1, \ldots, T$, form a sequence of designs generated by a "design density" $f_U(u)$ which is bounded away from zero and infinity on $\mathcal{U}$. $(\mathbf{X}_{it}^\tau, \mathbf{Z}_{it}^\tau, U_{it})^\tau, i = 1, \ldots, N, t = 1, \ldots, T$ are nonstochastic regressors. $(\mathbf{X}_{it}^\tau, \mathbf{Z}_{it}^\tau)^\tau$ are uniformly bounded on the space $\mathbb{X} \times \mathbb{Z}$.

(A2) There exist some functions $\varphi_{j1}(u), \ldots, \varphi_{jq}(u)$ such that

$$X_{itj} = \mathbf{Z}_{it}^\tau (\varphi_{j1}(U_{it}), \ldots, \varphi_{jq}(U_{it}))^\tau + \Pi_{itj},$$

for $i = 1, \ldots, N, t = 1, \ldots, T$, and $j = 1, \ldots, p$, and the real sequences $\{\Pi_{itj}\}$ satisfy

$$\lim_{N \to \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\Pi}_{it} \boldsymbol{\Pi}_{it}^\tau = \boldsymbol{\Omega}$$

with $\boldsymbol{\Pi}_{it} = (\Pi_{it1}, \ldots, \Pi_{itp})^\tau$.

(A3) $m_1(\cdot), \ldots, m_q(\cdot)$, $\varphi_1(\cdot), \ldots, \varphi_p(\cdot)$ are 2-times continuously differentiable and their 2th derivatives are Lipschitz continuous of order one.

Assumption B.

(B1) For each $N \geq 1$, the individual effects $\{\mu_i, 1 \leq i \leq N\}$ are independently distributed with zero mean and variance σ_μ^2, where $0 < \sigma_\mu^2 < c_\mu$ with $c_\mu < \infty$. Further, $\sup_{1 \leq i \leq N} E(|\mu_i|^{4+\delta_\mu}) < \infty$ for some $\delta_\mu > 0$.

(B2) For each $N \geq 1$, the individual effects $\{e_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$ are independently distributed with zero mean and variance σ_e^2, where $0 < \sigma_e^2 < c_e$ with $c_e < \infty$. Further, $\sup_{1 \leq i \leq N, 1 \leq t \leq T} E(|e_{it}|^{4+\delta_e}) < \infty$ for some $\delta_e > 0$.

(B3) The processes $\{\mu_i\}$ and $\{e_{it}\}$ are independent.

(B4) $1 - \rho_1 z - \cdots - \rho_s z^s \neq 0$ for $|z| \leq 1$.

Assumption C.

(C1) $\lambda \in (-\underline{a}^\lambda, \overline{a}^\lambda)$ with $0 < \underline{a}^\lambda < \overline{a}^\lambda < c^\lambda < \infty$.

(C2) All diagonal elements of $\mathbf{W}_N$ are zero.

(C3) The row and column sums of $\mathbf{W}_N$ and $(\mathbf{I}_N - \lambda \mathbf{W}_N)^{-1}$ are bounded uniformly in absolute value.

(C4) The matrix $\mathbf{I}_N - \lambda \mathbf{W}_N$ is nonsingular for all $\lambda \in (-\underline{a}^\lambda, \overline{a}^\lambda)$.

Assumption D. $\lim_{N\to\infty}(NT)^{-1}\mathbf{\Pi}_N^\tau\mathbf{\Pi}_N=\mathbf{\Omega}>\mathbf{0}$ and $\lim_{N\to\infty}(NT)^{-1}\mathbf{\Pi}_N^\tau\mathbf{\Sigma}^{-1}\mathbf{\Pi}_N=\mathbf{\Omega}^w>\mathbf{0}$, where $\mathbf{\Sigma}=(\sigma_\mu^2\mathbf{1}_T\mathbf{1}_T^\tau+\mathbf{\Gamma})\otimes\left\{(\mathbf{I}_N-\lambda\mathbf{W}_N)^{-1}(\mathbf{I}_N-\lambda\mathbf{W}_N^\tau)^{-1}\right\}$ with $\mathbf{\Gamma}=\left(\gamma_\nu(j_1-j2)\right)_{T\times T}$ and $\mathbf{\Pi}_N=(\mathbf{\Pi}_{11},\ldots,\mathbf{\Pi}_{N1},\ldots,\mathbf{\Pi}_{NT})^\tau$.

Assumption E. $\kappa_N=o(N^{1/2})$ and $N^{1/2}\kappa_N^{-4}=o(1)$.

Assumption F. $\max_{1\le j\le p}(NT)^{-1}\left|\left|\sum_{i=1}^N\sum_{t=1}^T\boldsymbol{\zeta}(U_{it})\Pi_{itj}\right|\right|=O(\kappa_N^{1/2}N^{-1/2})$. For any smooth function $h(\cdot)$, there exists some vector $\boldsymbol{\pi}$ such that

$$\max_{1\le j\le p}\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(h(U_{it})-\boldsymbol{\pi}^\tau\boldsymbol{\zeta}(U_{it}))\Pi_{itj}\right|=O\left(\kappa_N^{-2}N^{-1/2}\right)$$

with $\boldsymbol{\zeta}(u)=(\zeta_1(u),\ldots,\zeta_{\kappa_N}(u))^\tau$.

Remark 1. These assumptions are quite mild and are easily satisfied. The fixed and bounded design assumption in Assumption A1 is usually made in the literature on spatially correlated data. See, for example, Kapoor, Kelejian, and Prucha (2007), Kelejian and Prucha (2010), Su (2012), and so on. Assumptions A2 and F parallel those in the literature on semiparametric modeling (e.g., Ahmad, Leelahanon, and Li (2005)), and do not preclude $\{\mathbf{X}_{it},\mathbf{Z}_{it},U_{it}\}$ from being generated by some random mechanism. We focus on the fixed regressor case, but our analysis holds with probability one if $\{\mathbf{X}_{it},\mathbf{Z}_{it},U_{it}\}$ are generated randomly. And, in this case, we can interpret our analysis as being conditional on $\{\mathbf{X}_{it},\mathbf{Z}_{it},U_{it}\}$. Assumption A3 is a standard smoothing condition in the nonparametric and semiparametric regression literature. Assumption C part concerns the essential features of the spatial weights matrix, of which, C2 implies that each unit is not a neighbour of itself. Assumptions C1 and C4 imply that the dependent variable $\mathbf{Y}_N(t)$ is uniquely determined in terms of the disturbances conditional on the regressors. The uniform boundedness condition on $\mathbf{W}_N$ and $(\mathbf{I}_N-\lambda\mathbf{W}_N)^{-1}$ in C3 originated in papers Kelejian and Prucha (1998), in order to limit the spatial dependence across units to a manageable degree. Assumption D is necessary to establish the asymptotic normality of the parametric component; they are easy to verify if $\{\mathbf{X}_{it},\mathbf{Z}_{it},U_{it}\}$ are random designs. Assumption E is standard in the literature on polynomial spline approximation.

Let $\|a_s\|_{L_2}$ denote the L_2 norm of a square integrable function $m(u)$ on $\mathcal{U}$, and φ_N be the L_∞ distance between $m_N(\cdot)$ and $\mathcal{G}$ given by $\varphi_N=\mathrm{dist}(m,\mathcal{G})=\inf_{a\in\mathcal{G}}\sup_{u\in\mathcal{U}}|m(u)-a(u)|$.

Theorem 1. *Under Assumptions A–F,*

(i) $\sqrt{NT}(\widehat{\boldsymbol{\beta}}_N-\boldsymbol{\beta})\to_D N(0,\mathbf{\Omega}^{-1}\mathbf{\Omega}_1\mathbf{\Omega}^{-1})$ *as* $N\to\infty$, where $\mathbf{\Omega}_1=\lim_{N\to\infty}(1/NT)\,\mathbf{\Pi}_N^\tau\mathbf{\Sigma}\mathbf{\Pi}_N$,

(ii) $\max_{1\le j\le q}\|\widehat{m}_{j,N}-m_j\|_{L_2}^2=O_p(\max_{1\le j\le q}\kappa_N N^{-1}+\varphi_N^2)=O_p\left(\kappa_N N^{-1}+\kappa_N^{-4}\right).$

Here, $\widehat{\boldsymbol{\beta}}_N$ and $(\widehat{m}_{1,N}(\cdot),\ldots,\widehat{m}_{j,N}(\cdot))^\tau$ do not take the spatial and time-wise correlations into account, hence may not be asymptotically efficient. We will construct more efficient estimators by implementing estimated correlations in the following sections.

3. Estimation of the Spatial and Time-wise Error Structure

We investigate the estimation of the spatial and time-wise error structure. Based on $\widehat{\boldsymbol{\beta}}_N$ and $(\widehat{m}_{1,N}(\cdot),\ldots,\widehat{m}_{q,N}(\cdot))^\tau$, we can obtain the estimated residuals as

$$\widehat{\varepsilon}_{it,N}=Y_{it}-\mathbf{X}_{it}^\tau\widehat{\boldsymbol{\beta}}_N-Z_{it1}\widehat{m}_{1,N}(U_{it,N})-\cdots-Z_{itq}\widehat{\alpha}_{q,N}(U_{it}),\ i=1,\ldots,N,\ t=1,\ldots,T.$$

As the generalized moments estimation proposed by Kapoor, Kelejian, and Prucha (2007) does not apply to our scenario, we propose a new generalized moments method based on the temporally averaged disturbances.

Let $\bar{\boldsymbol{\eta}}_N=\sum_{t=1}^T\boldsymbol{\eta}_N(t)/T$ and $\bar{\bar{\boldsymbol{\eta}}}_N=\mathbf{W}_N\bar{\boldsymbol{\eta}}_N$. Then $E(\bar{\bar{\boldsymbol{\eta}}}_N^\tau\bar{\bar{\boldsymbol{\eta}}}_N)=(\sigma_\mu^2+\sigma_\nu^2)\mathrm{tr}(\mathbf{W}_N\mathbf{W}_N^\tau)$ and $E(\bar{\bar{\boldsymbol{\eta}}}_N^\tau\bar{\boldsymbol{\eta}}_N)=0$. Like Kelejian and Prucha (2010), it is convenient to rewrite the above moment conditions as

$$\frac{1}{N}\mathrm{E}\begin{pmatrix}c\bar{\boldsymbol{\eta}}_N^\tau\mathbf{A}_{1N}\bar{\boldsymbol{\eta}}_N\\ \bar{\boldsymbol{\eta}}_N^\tau\mathbf{A}_{2N}\bar{\boldsymbol{\eta}}_N\end{pmatrix}=\mathbf{0},\tag{3.1}$$

where $\mathbf{A}_{1,N}=\mathbf{W}_N^\tau\mathbf{W}_N-\mathrm{diag}(\mathbf{W}_N^\tau\mathbf{W}_N)$ and $\mathbf{A}_{2,N}=(\mathbf{W}_N^\tau+\mathbf{W}_N)/2$. We estimate λ based on the moment conditions in (3.1).

Noting that $\bar{\boldsymbol{\eta}}_N=(\mathbf{I}_N-\lambda\mathbf{W}_N)\bar{\boldsymbol{\varepsilon}}_N$ with $\bar{\boldsymbol{\varepsilon}}_N=\sum_{t=1}^T\boldsymbol{\varepsilon}_N(t)/T$, we can substitute this expression into (3.1) to yield $\boldsymbol{\psi}_N-\boldsymbol{\Psi}_N\boldsymbol{\theta}_N=\mathbf{0}$, where $\boldsymbol{\theta}=(\lambda,\lambda^2)^\tau$,

$$\boldsymbol{\psi}_N=\begin{pmatrix}\psi_{1,N}\\ \psi_{2,N}\end{pmatrix}=\begin{pmatrix}N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{A}_{1,N}\bar{\boldsymbol{\varepsilon}}_N)\\ N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{A}_{2,N}\bar{\boldsymbol{\varepsilon}}_N)\end{pmatrix},$$

$$\boldsymbol{\Psi}_N=\begin{pmatrix}\psi_{11,N}\ \psi_{12,N}\\ \psi_{21,N}\ \psi_{22,N}\end{pmatrix}=\begin{pmatrix}2N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{1,N}\bar{\boldsymbol{\varepsilon}}_N)\ N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{1,N}\mathbf{W}_N\bar{\boldsymbol{\varepsilon}}_N)\\ 2N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{2,N}\bar{\boldsymbol{\varepsilon}}_N)\ N^{-1}\mathrm{E}(\bar{\boldsymbol{\varepsilon}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{2,N}\mathbf{W}_N\bar{\boldsymbol{\varepsilon}}_N)\end{pmatrix}.$$

We then obtain the corresponding estimators $\widehat{\boldsymbol{\Psi}}_N=\left(\widehat{\psi}_{s_1s_2,N}\right)_{2\times2}$ and $\widehat{\boldsymbol{\psi}}_N=\left(\widehat{\psi}_{1,N},\widehat{\psi}_{2,N}\right)^\tau$ of $\boldsymbol{\Psi}_N=(\psi_{s_1s_2,N})_{2\times2}$ and $\boldsymbol{\psi}_N=\left(\widehat{\psi}_{1,N},\psi_{2,N}\right)^\tau$ in which

$$\begin{aligned}\widehat{\psi}_{11,N}&=2N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{1,N}\widehat{\bar{\boldsymbol{\varepsilon}}}_N, &\widehat{\psi}_{12,N}&=2N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{1,N}\mathbf{W}_N\widehat{\bar{\boldsymbol{\varepsilon}}}_N,\\ \widehat{\psi}_{21,N}&=2N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{2,N}\widehat{\bar{\boldsymbol{\varepsilon}}}_N, &\widehat{\psi}_{22,N}&=2N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{W}_N^\tau\mathbf{A}_{2,N}\mathbf{W}_N\widehat{\bar{\boldsymbol{\varepsilon}}}_N,\\ \widehat{\psi}_{1,N}&=N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{A}_{1,N}\widehat{\bar{\boldsymbol{\varepsilon}}}_N, &\widehat{\psi}_{2,N}&=N^{-1}\widehat{\bar{\boldsymbol{\varepsilon}}}_N^\tau\mathbf{A}_{2,N}\widehat{\bar{\boldsymbol{\varepsilon}}}_N,\end{aligned}$$

where $\widehat{\bar{\boldsymbol{\varepsilon}}}_N = \sum_{t=1}^{T} \widehat{\boldsymbol{\varepsilon}}_N(t)/T$ and $\widehat{\boldsymbol{\varepsilon}}_N(t) = (\widehat{\varepsilon}_{1t,N}, \ldots, \widehat{\varepsilon}_{Nt,N})^\tau$.

Let $\boldsymbol{\hbar}(\lambda) = \widehat{\boldsymbol{\psi}}_N - \widehat{\boldsymbol{\Psi}}_N \boldsymbol{\theta}_N$. Then we obtain the generalized moments estimator $\widehat{\lambda}_N \equiv \widehat{\lambda}_N(\boldsymbol{\Upsilon}_N)$ for λ by minimizing the objective function $Q_N = (\boldsymbol{\hbar}(\lambda))^\tau \boldsymbol{\Upsilon}_N \boldsymbol{\hbar}(\lambda)$, where $\boldsymbol{\Upsilon}_N$ is a 2×2 symmetric positive semidefinite matrix.

Theorem 2. *Assume that* $\lambda_{\min}(\boldsymbol{\Psi}_N^\tau \boldsymbol{\Psi}_N) \geq c_1 > 0$, $\lambda_{\min}(\boldsymbol{\Upsilon}_N) \geq c_2 > 0$ *and* $\lambda_{\min}(\boldsymbol{\Phi}_N) \geq c_3 > 0$. *Under Assumptions A-F,* $(\widehat{\lambda}_N - \lambda) = O_p(N^{-1/2})$.

Due to the cross-sectional correlation and the individual effect, the conventional Yule-Walker equations-based method could not be used to estimate the temporal autoregressive parameter vector $\boldsymbol{\rho}$. We here propose a new method to estimate $\boldsymbol{\rho}$ that can be taken as an extension of the method of Baltagi and Li (1991) who focused on the scenario of AR(1) and without cross-sectional correlation.

Set $T_s = T - (s+1)$ for brevity, and let

$$\widehat{\eta}_{it,N} = \widehat{\varepsilon}_{it,N} - \widehat{\lambda}_N \sum_{i_1=1}^{N} W_{ii_1} \widehat{\varepsilon}_{i_1 t,N},$$

$$\widehat{\mathbf{Q}}_{0,N} = \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} (\widehat{\eta}_{i(t+(s-1)),N}, \ldots, \widehat{\eta}_{it,N})^\tau (\widehat{\eta}_{i(t+(s-1)),N}, \ldots, \widehat{\eta}_{it,N}),$$

$$\widehat{\mathbf{Q}}_{1,N} = \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} (\widehat{\eta}_{i(t+(s-1)),N}, \ldots, \widehat{\eta}_{it,N})^\tau (\widehat{\eta}_{i(t+s),N}, \ldots, \widehat{\eta}_{i(t+1),N}),$$

$$\widehat{\mathbf{Q}}_{2,N} = \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} (\widehat{\eta}_{i(t+(s-1)),N}, \ldots, \widehat{\eta}_{it,N})^\tau \widehat{\eta}_{i(t+s),N},$$

$$\widehat{\mathbf{Q}}_{3,N} = \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} (\widehat{\eta}_{i(t+(s-1)),N}, \ldots, \widehat{\eta}_{it,N})^\tau \widehat{\eta}_{i(t+s+1),N}.$$

We can estimate $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)^\tau$ by

$$\widehat{\boldsymbol{\rho}}_N = (\widehat{\rho}_{1,N}, \ldots, \widehat{\rho}_{s,N})^\tau = (\widehat{\mathbf{Q}}_{0,N} - \widehat{\mathbf{Q}}_{1,N})^{-1} (\widehat{\mathbf{Q}}_{2,N} - \widehat{\mathbf{Q}}_{3,N}).$$

Noting that $\eta_{i(t+s)} = \mu_i + \nu_{i(t+s)} = \mu_i + \rho_1 \nu_{i(t+(s-1))} + \cdots + \rho_s \nu_{it} + e_{it}$, we have

$$\eta_{i(t+s)} - \rho_1 \eta_{i(t+(s-1))} - \cdots - \rho_s \eta_{it} = (\mu_i - \rho_1 \mu_i - \cdots - \rho_s \mu_i) + e_{it}.$$

Set $\ell_{it} = \eta_{i(t+s)} - \rho_1 \eta_{i(t+(s-1))} - \cdots - \rho_s \eta_{it}$ for $t = 1, \ldots, T-s$. Since $\sigma_e^2 = \mathrm{E}(\ell_{it,N}^2) - \mathrm{E}(\ell_{it,N} \ell_{i(t+1),N})$, we can estimate σ_e^2 by

$$\widehat{\sigma}_{e,N}^2 = \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} \widehat{\ell}_{it,N}^2 - \frac{1}{NT_s} \sum_{i=1}^{N} \sum_{t=1}^{T_s} \widehat{\ell}_{it,N} \widehat{\ell}_{i(t+1),N},$$

where $\widehat{\ell}_{it,N} = \widehat{\eta}_{i(t+s),N} - \widehat{\rho}_{1,N} \widehat{\eta}_{i(t+(s-1)),N} - \cdots - \widehat{\rho}_{s,N} \widehat{\eta}_{it,N}$ for $t = 1, \ldots, T_s$.

Theorem 3. *Under Assumptions B-F, we have*

(i) $\widehat{\rho}_N - \rho = O_p(N^{-1/2}T^{-1/2})$.
(ii) $\widehat{\sigma}^2_{e,N} - \sigma^2_e = O_p(N^{-1/2}T^{-1/2})$.

In the following section, we investigate how to apply the estimated spatial and time-wise correlations to improve the estimations of the parametric and nonparametric components in model (1.1).

4. Weighted Semiparametric Least Squares Estimation

Define $\widehat{\mathbf{\Sigma}}_N = (\widehat{\sigma}^2_{\mu,N}\mathbf{1}_T\mathbf{1}_T^\tau + \widehat{\mathbf{\Gamma}}_N) \otimes \Big\{(\mathbf{I}_N - \widehat{\lambda}_N\mathbf{W}_N)^{-1}(\mathbf{I}_N - \widehat{\lambda}_N\mathbf{W}_N^\tau)^{-1}\Big\}$,

$$\widehat{\mathbf{\Gamma}}_N = \begin{pmatrix} \widehat{\gamma}_{\nu,N}(0) & \widehat{\gamma}_{\nu,N}(1) & \cdots & \widehat{\gamma}_{\nu,N}(T-1) \\ \widehat{\gamma}_{\nu,N}(1) & \widehat{\gamma}_{\nu,N}(0) & \cdots & \widehat{\gamma}_{\nu,N}(T-2) \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\gamma}_{\nu,N}(T-1) & \widehat{\gamma}_{\nu,N}(T-2) & \cdots & \widehat{\gamma}_{\nu,N}(0) \end{pmatrix},$$

and

$$\mathbf{M}_{\mathbf{Z}^*}^{\widehat{\mathbf{\Sigma}}^{-1}} = \widehat{\mathbf{\Sigma}}_N^{-1} - \widehat{\mathbf{\Sigma}}_N^{-1}\mathbf{Z}_N^*(\mathbf{Z}_N^{*\tau}\widehat{\mathbf{\Sigma}}_N^{-1}\mathbf{Z}_N^*)^{-1}\mathbf{Z}_N^{*\tau}\widehat{\mathbf{\Sigma}}_N^{-1}.$$

Pre-multiplying (2.2) by $\widehat{\mathbf{\Sigma}}_N^{-1/2}$ leads to

$$\widehat{\mathbf{\Sigma}}_N^{-1/2}\mathbf{Y}_N \approx \widehat{\mathbf{\Sigma}}_N^{-1/2}\mathbf{X}_N\boldsymbol{\beta}_N + \widehat{\mathbf{\Sigma}}_N^{-1/2}\mathbf{Z}_N^*\boldsymbol{\theta}_N + \widehat{\mathbf{\Sigma}}_N^{-1/2}\boldsymbol{\varepsilon}_N. \tag{4.1}$$

Due to the fact that $\mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\mathbf{Z}_N^*\boldsymbol{\theta}_N = \mathbf{Z}_N^*\boldsymbol{\theta}_N - \mathbf{Z}_N^*\boldsymbol{\theta}_N = \mathbf{0}$, (4.1) leads to

$$\mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\mathbf{Y}_N \approx \mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\mathbf{X}_N\boldsymbol{\beta}_N + \mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\boldsymbol{\varepsilon}_N. \tag{4.2}$$

If we take $\mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\boldsymbol{\varepsilon}_N$ as the residuals, then (4.2) is also a version of the usual linear regression. A weighted semiparametric least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_N^w = (\mathbf{X}_N^\tau\mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\mathbf{X}_N)^{-1}\mathbf{X}_N^\tau\mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\mathbf{\Sigma}}_N^{-1}}\mathbf{Y}_N.$$

Substituting $\widehat{\boldsymbol{\beta}}_N^w$ into (2.2) gets an estimator of $\boldsymbol{\theta}$,

$$\widehat{\boldsymbol{\theta}}_N^w = (\mathbf{Z}_N^{*\tau}\widehat{\mathbf{\Sigma}}_N^{-1}\mathbf{Z}_N^*)^{-1}\mathbf{Z}_N^{*\tau}(\mathbf{Y}_N - \mathbf{X}_N\widehat{\boldsymbol{\beta}}_N^w).$$

A weighted polynomial spline series estimator of $\mathbf{m}(\cdot)$ is then

$$\widehat{\mathbf{m}}_N^w(u) = (\widehat{m}_{1,N}^w(u), \ldots, \widehat{m}_{q,N}^w(u))^\tau = \boldsymbol{\zeta}_N^*(u)\widehat{\boldsymbol{\theta}}_N^w,$$

where $\boldsymbol{\zeta}_N^*(u)$ is defined in Section 2.

Theorem 4. *Under Assumptions A-F, the following hold.*

(i) $\sqrt{NT}(\widehat{\boldsymbol{\beta}}_N^w - \boldsymbol{\beta}) \to_D N(0, \boldsymbol{\Omega}^{w-1})$ *as* $N \to \infty$ *where* $\boldsymbol{\Omega}^w$ *is defined in Assumption E.*

(ii) $\max\limits_{1 \le j \le q} \|\widehat{m}_{j,N}^w - m_j\|_{L_2}^2 = O_p\left(\kappa_N N^{-1} + \varphi_N^2\right) = O_p(\max\limits_{1\le j\le q} \kappa_N N^{-1} + \kappa_N^{-4})$.

(iii) $\boldsymbol{\Omega}^{w-1} \le \boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}^{-1}$.

In order to use (i) of Theorem 4 for inference about $\boldsymbol{\beta}$, a consistent estimator of $\boldsymbol{\Omega}^w$ is needed.

Theorem 5. *Under Assumptions A-F,* $\widehat{\boldsymbol{\Omega}}_N^w \equiv (NT)^{-1}\mathbf{X}_N^\tau \mathbf{M}_{\mathbf{Z}_N^*}^{\widehat{\boldsymbol{\Sigma}}^{-1}} \mathbf{X}_N \to_p \boldsymbol{\Omega}^w$ *as* $N \to \infty$.

With Theorems 4 and 5, we can construct 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 $d \le p$.

5. Simulation Studies

We report here on some simulation studies of the finite sample performance of the proposed estimators.

The data were generated from the panel data varying-coefficient partially linear model

$$Y_i(t) = X_{1i}(t)\beta_1 + X_{2i}(t)\beta_2 + Z_{1i}(t)\cdot m_1(U_i(t)) + Z_{2i}\cdot m_2(U_i(t)) + \varepsilon_i(t),$$
$$i = 1, \ldots, N;\ t = 1, \ldots, T,$$

with the true fixed coefficient values as $\beta = (\beta_1, \beta_2)^\tau = (1, -1.5)^\tau$, and the real varying-coefficient functions as $m_1(u) = 2\sin(2\pi u)$ and $m_2(u) = 1.5\cos(1.5\pi u)^3 -(u-0.5)^3 + 1$. We generated $U_i(t)$ from a uniform distribution on the interval $[0,1]$, the random variables Z_{1i} and Z_{2i} independently from two zero mean normal distributions with different standard deviations 0.5 and 0.6, respectively. We generated $X_{1,i}$ and X_{2i} as $X_{1i}(t) = U_i(t) + 1 + \omega_{1i}(t)$ and $X_{2i}(t) = U_i(t)^2 + 1 + \omega_{2i}(t)^2$, where $\omega_{1i}(t)$ and $\omega_{2i}(t)$ were independently generated as standard normal. We took the error structure $\boldsymbol{\varepsilon}_N(t) = (\varepsilon_1(t), \ldots, \varepsilon_N(t))^\tau$ as

$$\boldsymbol{\varepsilon}_N(t) = \lambda \mathbf{W}_N \boldsymbol{\varepsilon}_N(t) + \boldsymbol{\eta}_N(t),\ \boldsymbol{\eta}_N(t) = \boldsymbol{\mu}_N + \boldsymbol{\nu}_N(t),\ t{=}1, \ldots, T,$$

where $\boldsymbol{\mu}_N = (\mu_1, \ldots, \mu_N)^\tau$ with $\mu_i \sim i.i.d.\ N(0, \sigma_\mu^2)$, $\boldsymbol{\nu}_N(t) = (\nu_{1t}, \ldots, \nu_{Nt})^\tau$ with $\nu_{it} = \rho\nu_{i(t-1)} + e_{it}$ and $e_{it} \sim i.i.d.\ N(0, \sigma_e^2)$. We took $(\sigma_\mu^2, \sigma_e^2) = (1,1)$, and $(\lambda, \rho) = (0.3, 0.3)$ and $(0.6, 0.6)$ to represent different degrees of spatial and time-wise correlations. We considered $\mathbf{W}_N$ with each element of $\boldsymbol{\varepsilon}_N(t)$ directly related to the elements immediately after and immediately before it.

Table 1. Finite sample performance of the proposed estimators $\widehat{\lambda}_N, \widehat{\rho}_N, \widehat{\sigma}^2_{\mu,N}, \widehat{\sigma}^2_{e,N}$ of the error structure parameters λ, ρ, σ_μ, σ_e^2.

			$N=100$		$N=200$		$N=300$	
T		(λ,ρ)	$(0.3,0.3)$	$(0.6,0.6)$	$(0.3,0.3)$	$(0.6,0.6)$	$(0.3,0.3)$	$(0.6,0.6)$
5	$\widehat{\lambda}_N$	est	0.3077	0.6081	0.2995	0.6057	0.2995	0.6047
		std	0.0964	0.1093	0.0750	0.0838	0.0690	0.0707
	$\widehat{\rho}_N$	est	0.2844	0.5151	0.2897	0.5369	0.2903	0.5592
		std	0.0707	0.0855	0.0574	0.0692	0.0502	0.0615
	$\widehat{\sigma}^2_{\mu,N}$	est	0.9509	1.0722	0.9663	1.0692	0.9770	1.0201
		std	0.1640	0.3222	0.1312	0.2579	0.1103	0.2433
	$\widehat{\sigma}^2_{e,N}$	est	1.0186	1.0755	1.0177	1.0576	1.0167	1.0440
		std	0.0767	0.1099	0.0610	0.0892	0.0537	0.0762
10	$\widehat{\lambda}_N$	est	0.3060	0.6142	0.2987	0.6066	0.2998	0.6041
		std	0.0905	0.1089	0.0738	0.0836	0.0686	0.0714
	$\widehat{\rho}_N$	est	0.2902	0.5443	0.2930	0.5587	0.2936	0.5710
		std	0.0427	0.0523	0.0338	0.0434	0.0299	0.0344
	$\widehat{\sigma}^2_{\mu,N}$	est	0.9764	1.0688	0.9813	1.0578	0.9888	1.0332
		std	0.1363	0.2258	0.1069	0.1898	0.0967	0.1538
	$\widehat{\sigma}^2_{e,N}$	est	1.0159	1.0502	1.0170	1.0354	1.0141	1.0325
		std	0.0515	0.0894	0.0426	0.0728	0.0375	0.0658

For the first and last elements of $\boldsymbol{\varepsilon}_N(t)$, we took a circular setting such that, for example, ε_{1t} is directly related to the second and last element of $\boldsymbol{\varepsilon}_N(t)$. There are only two nonzero elements in each row of $\mathbf{W}_N$. We took $(N,T) = (100,5),(200,5),(300,5),(100,10),(200,10)$ and $(300,10)$. Each setting was repeated 1,000 times.

Conventional cubic splines with uniformly distributed inner knots were as our base functions. Similar to Wang and Yang (2007), the number of interior knots M was determined by the total sample size $N\times T$ and a tuning constant c,

$$M = \min\left\{\lfloor c(NT)^{1/5}\rfloor + 1, \left\lfloor \frac{1}{2q}(NT-2p)\right\rfloor\right\},$$

in which $\lfloor a\rfloor$ denotes the integer part of a. In our simulation study, we used $c = 0.5, 1.0,$ and 1.5, found the results are not sensitive to the choice, and reported just the results with $c = 1$. The additional constraint $M \le (NT-2p)/(2q)$ ensures that the number of terms in (1.2) or (3.1) is no greater than $NT/2$, which is necessary when the sample size NT is moderate and the dimension q is high.

For the proposed estimators $\widehat{\lambda}_N, \widehat{\rho}_N, \widehat{\sigma}^2_{\mu,N}, \widehat{\sigma}^2_{e,N}$ of the error structure parameters λ, ρ, σ_μ, σ_e^2, given a sample size, the average of the estimates (**est**), and sample standard deviation(**std**) are summarized in Tables 1. From Table 1, we can see that the proposed estimators of the error structure parameters λ, ρ, σ_μ, σ_e^2 worked well, and that their performance improved with increasing sample size.

For the proposed WSLSE $(\widehat{\beta}_{1,N}^w, \widehat{\beta}_{2,N}^w)^\tau$ of the parametric components $(\beta_1, \beta_2)^\tau$, mean estimates (**est**), sample standard deviation (**std**), estimated standard deviation (**estd**), and 95% confidence interval coverage (**cp**) are reported in Table 2 when $T = 5$. We also present results of the unweighted SLSE $(\widehat{\beta}_{1,N}, \widehat{\beta}_{2,N})^\tau$ that ignores spatial and time-wise correlations, and the benchmark estimator $(\tilde{\beta}_{1,N}^w, \tilde{\beta}_{2,N}^w)^\tau$ which has the same definition as the WSLSE except for known error structure parameters λ, ρ, σ_ν, and σ_e^2. In every situation, the estimated standard deviations, **estd**s, matched the Monte Carlo cases, **std**s, reasonably well, and the coverage probabilities were close to the nominal level. The proposed WSLSE outperformed the unweighted SLSE (much smaller variance), and the proposed WSLSE provided comparable estimators with their corresponding benchmark versions that used the true error structure. These findings then also support the proposed estimators $\widehat{\lambda}_N, \widehat{\rho}_N, \widehat{\sigma}_{\mu,N}^2$, and $\widehat{\sigma}_{e,N}^2$.

For the estimators of the nonparametric components, we set a measure of estimation accuracy as the root average squared error (RASE), with

$$\mathrm{RASE}_j = \left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\bar{m}_{j,N}(U_{it}) - m_j(U_{it})\}^2\right]^{1/2}, \quad j = 1, 2,$$

where $\bar{m}_{j,N}(u)$ is either $\widehat{m}_{j,N}(u)$, or $\widehat{m}_{j,N}^w(u)$ or the benchmark estimator $\tilde{m}_{j,N}^w(u)$. The sample mean and standard deviation of the RASEs over 1,000 replications are summarized at the bottom of Table 2. Corresponding box-plots of these RASEs are presented in Figure 1. As we can see, the proposed WPSSE $\widehat{m}_{j,N}^w(u)$ outperformed the initial $\widehat{m}_{j,N}(u)$, having smaller and stabler RASE. Most importantly, the proposed $\widehat{m}_{j,N}^w(u)$ performed nearly as well as benchmark $\check{m}_{j,N}^w(u)$. See the box-plots in Figure 1. Similar results can be obtained in case $T = 10$; corresponding tables and figures are reported in the supplementary document. To check the impact on the estimators of misspecifying $\mathbf{W}_N$, we conducted some simulations in which the true $\mathbf{W}_N$ was defined as before and the misspecified $\mathbf{W}_N$ was specified as each element of $\boldsymbol{\varepsilon}_N(t)$ being related to the three elements immediately after and the three elements immediately before it. The finite sample performances of the WSLSE $(\widehat{\beta}_{1,N}^*, \widehat{\beta}_{2,N}^*)^\tau$ and WPSSE $(\widehat{m}_{1,N}^*(u), \widehat{m}_{2,N}^*(u))^\tau$ are summarized in Table 2. From Table 2, we can see that the resultant estimators of the parametric and nonparametric components were still consistent although $\mathbf{W}_N$ was misspecified.

6. Application

We further illustrate our methodology by an analysis of Indonesian rice farming data. The data were previously analyzed by, for example, Druska and Horrace (2004), and detailed discussion about the data can refer to their paper. We

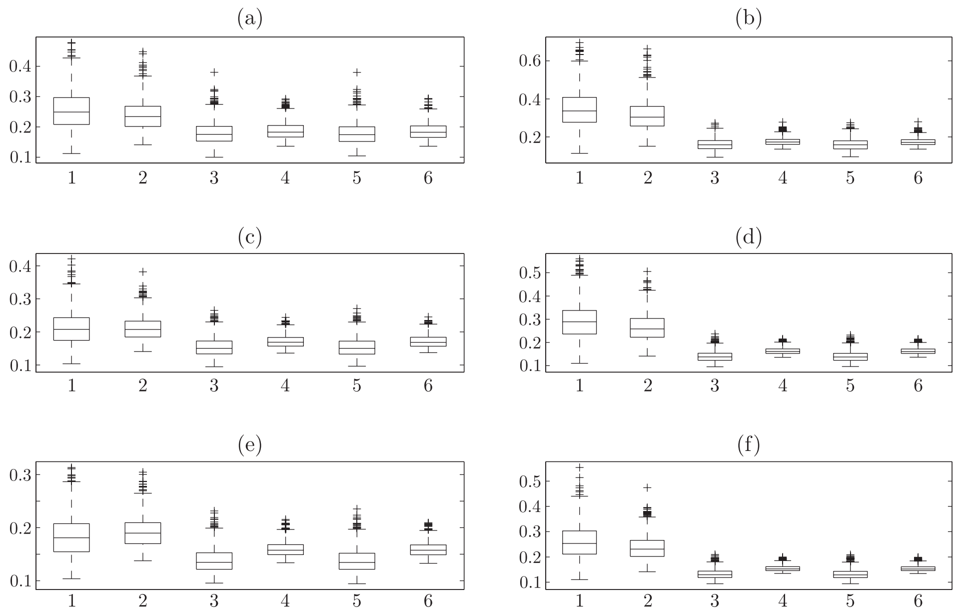

Figure 1. Box plots of the RASE values for the three nonparametric function estimators $(\widehat{m}_{1,N}(u), \widehat{m}_{2,N}(u))^\tau$, $(\widehat{m}^w_{1,N}(u), \widehat{m}^w_{2,N}(u))^\tau$ and $(\tilde{m}^w_{1,N}(u), \tilde{m}^w_{2,N}(u))^\tau$ with $T = 5$. Each boxplot is based on the 1,000 RASE values for a particular combination. Indices 1, 2, 3, 4, 5 and 6 are for $\widehat{m}_{1,N}(u), \widehat{m}_{2,N}(u), \widehat{m}^w_{1,N}(u), \widehat{m}^w_{2,N}(u), \tilde{m}^w_{1,N}(u)$ and $\tilde{m}^w_{2,N}(u)$, respectively. $N = 100, (\lambda, \rho) = (0.3, 0.3)$ in plot (a); $N = 100, (\lambda, \rho) = (0.6, 0.6)$ in plot (b); $N = 200, (\lambda, \rho) = (0.3, 0.3)$ in plot (c); $N = 200, (\lambda, \rho) = (0.6, 0.6)$ in plot (d); $N = 300, (\lambda, \rho) = (0.3, 0.3)$ in plot (e); And $N = 300, (\lambda, \rho) = (0.6, 0.6)$ in plot (f).

have a data set of 171 rice farms over six growing seasons (three wet and three dry seasons). Those farms are located in six different villages. We applied a partially linear model (special case of the proposed partially linear varying-coefficient models) for the data to regress the natural logarithm of output (ln(rice), y) on covariates such as land area (Land, in hectare), seed amount (Sead, in logarithm of kilogram) and whether high yield or mixed varieties. Specifically,

$$y_i(t) = x_{1i}(t) \cdot \beta_1 + x_{2i}(t) \cdot \beta_2 + Seed_i(t) \cdot \beta_3 + m(Land_i(t)) + \varepsilon_i(t),$$

where $x_{1i}(t) = 1$ if the ith form at tth growing season used high yield varieties, otherwise $x_{1i} = 0$; x_2 is a similar indicator for using mixed varieties or not. To describe potential spatial correlations between different villages and timewise correlations between different seasons, we took the error vector $\boldsymbol{\varepsilon}_N(t) = (\varepsilon_1(t), \ldots, \varepsilon_N(t))^\tau$ as

$$\boldsymbol{\varepsilon}_N(t) = \lambda \mathbf{W}_N \boldsymbol{\varepsilon}_N(t) + \boldsymbol{\eta}_N(t) \text{ and } \boldsymbol{\eta}_N(t) = \boldsymbol{\mu}_N + \boldsymbol{\nu}_N(t),\ t = 1, \ldots, T,\ N = 171.$$

Table 2. Finite sample performances of the proposed WSLSE of regression coefficient β and WPSSE of nonprarmetric function $m(\cdot)$ under $T = 5$.

		$N = 100$		$N = 200$		$N = 300$	
	(λ, ρ)	$(0.3, 0.3)$	$(0.6, 0.6)$	$(0.3, 0.3)$	$(0.6, 0.6)$	$(0.3, 0.3)$	$(0.6, 0.6)$
$\widehat{\beta}_{1,N}$	est	1.0002	1.0020	1.0021	0.9995	1.0025	1.0001
	std	0.0497	0.0961	0.0412	0.0758	0.0347	0.0638
	estd	0.0502	0.1157	0.0404	0.0806	0.0351	0.0678
	cp	0.9560	0.9460	0.9410	0.9460	0.9430	0.9460
$\widehat{\beta}_{2,N}$	est	-1.5012	-1.4983	-1.5018	-1.5003	-1.5018	-1.5022
	std	0.0348	0.0509	0.0295	0.0448	0.0236	0.0395
	estd	0.0355	0.0631	0.0288	0.0474	0.0240	0.0401
	cp	0.9510	0.9500	0.9450	0.9580	0.9520	0.9530
$\widehat{\beta}^w_{1,N}$	est	0.9992	1.0053	0.9981	0.9997	1.0027	0.9995
	std	0.0283	0.0279	0.0237	0.0213	0.0203	0.0182
	estd	0.0284	0.0283	0.0236	0.0219	0.0203	0.0183
	cp	0.9530	0.9500	0.9510	0.9580	0.9460	0.9540
$\widehat{\beta}^w_{2,N}$	est	-1.5017	-1.4964	-1.5006	-1.4990	-1.5016	-1.5023
	std	0.0218	0.0167	0.0180	0.0156	0.0145	0.0131
	estd	0.0218	0.0183	0.0180	0.0163	0.0149	0.0136
	cp	0.9510	0.9600	0.9450	0.9600	0.9590	0.9600
$\tilde{\beta}^w_{1,N}$	est	0.9992	1.0054	0.9981	0.9994	1.0027	0.9996
	std	0.0277	0.0274	0.0233	0.0210	0.0201	0.0180
	estd	0.0283	0.0268	0.0235	0.0209	0.0202	0.0177
	cp	0.9590	0.9390	0.9500	0.9520	0.9460	0.9450
$\tilde{\beta}^w_{2,N}$	est	-1.5018	-1.4962	-1.5006	-1.4988	-1.5016	-1.5023
	std	0.0215	0.0166	0.0179	0.0155	0.0144	0.0131
	estd	0.0217	0.0171	0.0178	0.0156	0.0148	0.0132
	cp	0.9560	0.9470	0.9450	0.9530	0.9570	0.9440
$\widehat{\beta}^*_{1,N}$	est	1.0077	0.9936	0.9971	1.0037	0.9977	0.9950
	std	0.0469	0.0489	0.0327	0.0327	0.0268	0.0258
$\widehat{\beta}^*_{2,N}$	est	-1.4977	-1.5022	-1.4981	-1.5007	-1.4993	-1.4991
	std	0.0313	0.0289	0.0239	0.0228	0.0198	0.0201
$\widehat{m}_{1,N}(\cdot)$	sm(RASE)	0.2540	0.3470	0.2111	0.2933	0.1843	0.2592
	std(RASE)	0.0647	0.0971	0.0494	0.0767	0.0393	0.0660
$\widehat{m}_{2,N}(\cdot)$	sm(RASE)	0.2381	0.3117	0.2115	0.2659	0.1921	0.2370
	std(RASE)	0.0495	0.0774	0.0361	0.0592	0.0293	0.0485
$\widehat{m}^w_{1,N}(\cdot)$	sm(RASE)	0.1798	0.1612	0.1541	0.1406	0.1386	0.1319
	std(RASE)	0.0378	0.0312	0.0287	0.0228	0.0226	0.0191
$\widehat{m}^w_{2,N}(\cdot)$	sm(RASE)	0.1870	0.1755	0.1715	0.1634	0.1596	0.1547
	std(RASE)	0.0276	0.0214	0.0193	0.0143	0.0144	0.0116
$\tilde{m}^w_{1,N}(\cdot)$	sm(RASE)	0.1790	0.1601	0.1535	0.1401	0.1382	0.1318
	std(RASE)	0.0371	0.0310	0.0285	0.0227	0.0225	0.0190
$\tilde{m}^w_{2,N}(\cdot)$	sm(RASE)	0.1860	0.1744	0.1710	0.1630	0.1594	0.1544
	std(RASE)	0.0270	0.0208	0.0192	0.0140	0.0142	0.0114
$\widehat{m}^*_{1,N}(\cdot)$	sm(RASE)	0.2568	0.2533	0.1971	0.1870	0.1614	0.1621
	std(RASE)	0.0660	0.0634	0.0438	0.0423	0.0321	0.0311
$\widehat{m}^*_{2,N}(\cdot)$	sm(RASE)	0.2430	0.2311	0.1928	0.1940	0.1768	0.1767
	std(RASE)	0.0507	0.0448	0.0285	0.0287	0.0218	0.0210

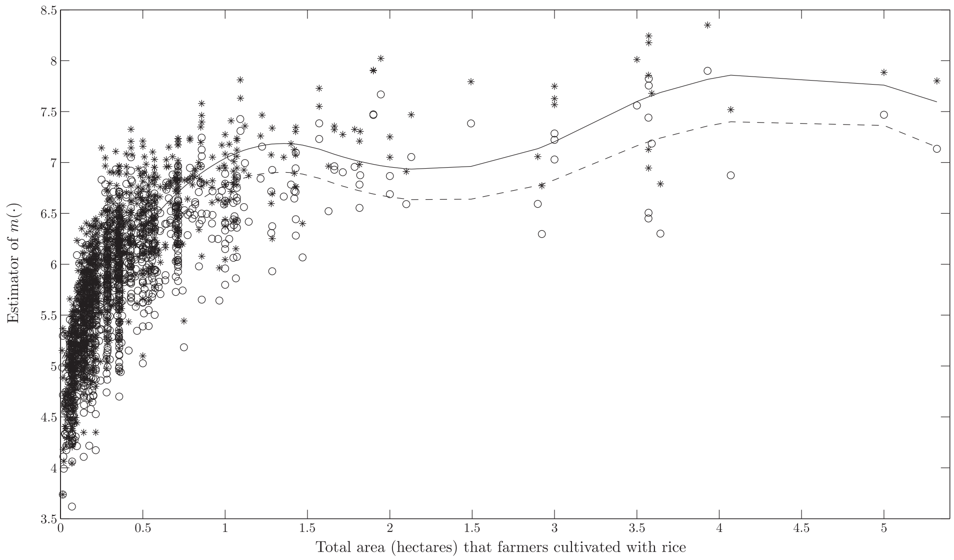

Figure 2. Nonparametric estimation of the effect of the total area that farmers cultivated with rice, measured in hectares on the output (ln(rice)). Dashed curve and "o" is based on the unweighted semiparametric least squares estimation with "o" representing $y_i(t) - x_{1i}(t)\widehat{\beta}_{1,N} - x_{2i}(t)\widehat{\beta}_{2,N} - Seed_i(t)\widehat{\beta}_{3,N}$. Solid curve and "*" is based on our proposed weighted semiparametric least squares estimation with "*" representing $y_i(t) - x_{1i}(t)\widehat{\beta}^w_{1,N} - x_{2i}(t)\widehat{\beta}^w_{2,N} - Seed_i(t)\widehat{\beta}^w_{3,N}$.

The typical element w_{ij} of the spatial weighting matrix $\mathbf{W}_N$ was positive if observations i and j belong to farms located in the same village, and the same growing season. The row sums of $\mathbf{W}_N$ were standardized to one. In addition, $\boldsymbol{\mu}_N = (\mu_1, \ldots, \mu_N)^\tau$ with $\mu_i \sim i.i.d.\ (0, \sigma^2_\mu)$, $\boldsymbol{\nu}_N(t) = (\nu_{1t}, \ldots, \nu_{Nt})^\tau$ with $\nu_{it} = \rho_1 \nu_{i(t-1)} + \cdots + \rho_s \nu_{i(t-s)} + e_{it}$ and $e_{it} \sim i.i.d.\ (0, \sigma^2_e)$.

In applications, the lagged order of the autoregressive process is usually unknown; based on the estimated η_{it}, it can be determined by the classic AIC or BIC criteria. By the AIC criterion, we found the lagged order 1 suitable for this data set. We estimated λ, σ^2_μ, ρ and σ^2_e by $\widehat{\lambda}_N = 0.7553$, $\widehat{\sigma}^2_{\mu,N} = 0.0492$, $\widehat{\rho}_N = 0.0200$, and $\widehat{\sigma}^2_{e,N} = 0.0940$. The results for the estimators of $(\beta_1, \beta_2, \beta_3)^\tau$ and $m(u)$ are shown in Table 3 and Figures 2−3. We see that our proposed weighted estimators $\widehat{\boldsymbol{\beta}}^w_N$ had much smaller standard error than the unweighted estimators $\widehat{\boldsymbol{\beta}}_N$. In particular, in comparison with previous studies where only the logarithm of seed amount was significant because they did not take the spatial and time-wise correlations into account, our proposed estimating procedure has the indicator variables of both high yield and mixed varieties significant. The

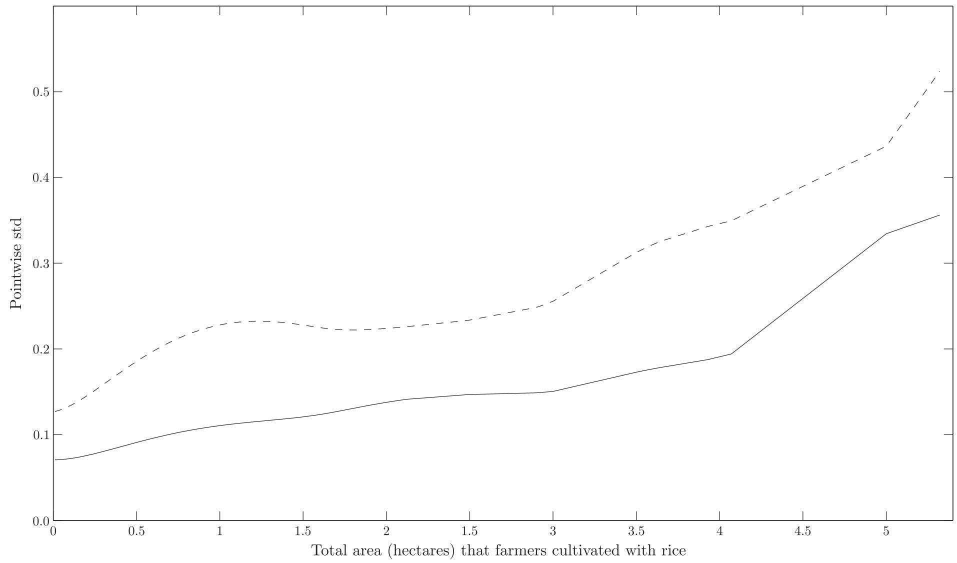

Figure 3. Pointwise std for the estimators of $m(u)$. Dashed curve is based on the unweighted semiparametric least squares estimation. Solid curve is based on the weighted semiparametric least squares estimation.

Table 3. Estimates for the parametric components $(\beta_1, \beta_2, \beta_3)^\tau$ and corresponding 95% confidence intervals.

	Unweighted estimator			Our proposed estimator		
Parameter	Estimate	SE	Confidence Interval	Estimate	SE	Confidence Interval
β_1	0.1165	0.1146	[-0.1081, 0.3411]	0.1806	0.0391	[0.1040, 0.2573]
β_2	0.1898	0.0978	[-0.0019, 0.3814]	0.1197	0.0526	[0.0167, 0.2227]
β_3	0.3925	0.0540	[0.2865, 0.4984]	0.2883	0.0246	[0.2400, 0.3365]

results imply that large amount of seeds from high yield and mixed varieties positively affect the output (ln(rice)).

The effect of the total area that farmers cultivated with rice on the output (ln(rice)) was estimated nonparametrically, with the results given in Figures 2-3. Again we see that the weighted polynomial spline estimator $\widehat{m}_N^w(u)$ has smaller (point wise) standard errors than the unweighted polynomial spline estimator estimator $\widehat{m}_N(u)$. More importantly, we see that the output (ln(rice)) increases very quickly and nonlinearly if the total area that farmers cultivated with rice is less than one hectare, which the output (ln(rice)) changes not much when the total area cultivated with rice is greater than one hectare.

7. Concluding Remarks

There are several ways to generalize our methodology. For example, a referee noticed that it can be easily adapted to the case where T trends to infinite as well. We only considered the balanced panel data. One can also consider unbalanced panel data. Analysis here is more interesting and realistic, particularly when some subjects drop from an experiment or a survey (Baltagi (1998)). For unbalanced panel data, as in Section 2 to 4, we can construct estimators of the parametric and nonparametric components except that T has to be replaced by T_i. The asymptotic properties of the resultant estimators can be derived similarly, but with more complicated notations. Our model does not contain spatially lagged dependent variables. How to extend our proposed method to panel data semiparametric models with both spatially lagged dependent variables and spatially lagged disturbances is an open problem.

Acknowledgement

The authors are grateful to the anonymous referees and Editors for their constructive comments which led to improve the article. Bai's research was supported by National Natural Science Foundation of China (NSFC) Grants 11001162. Hu's research was supported by NSFC Grants 10971126. You's research is supported by grants from the Program for New Century Excellent Talents in University (NCET). The work is also partially supported by Program for Changjiang Scholars and Innovative Research Team in University (IRT13077), Shanghai University of Finance and Economics through Project 211 (Phase IV) and Shanghai Leading Academic Discipline Project(B803).

References

Ahmad, I., Leelahanon, S. and Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Ann. Statist.* **33**, 258-283.

Baltagi, B. H. (1998). Worldwide institutional rankings in econometrics: 1989-1995. *Econometric Theory* **14**, 1-44.

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

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2007). Estimating models of complex FDI: Are there third-country effects? *J. Econometrics* **140**, 260-281.

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

Baltagi, B. H., Song, S. H. and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *J. Econometrics* **117**, 123-150.

Badinger, H. and Egger, P. (2013). Estimation and testing of higher-order spatial autoregressive panel data error component models. *J. Geograph. Systems* **15**, 453-489.

Cliff, A. and Ord, J. (1981). *Spatial Processes, Models and Applications*. Pion, London.

de Boor, C. (1978). *A Practical Guide to Splines*. Applied Mathematical Sciences **27**. Springer-Verlag, New York-Berlin.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press, Oxford.

Druska, V. and Horrace, W. C. (2004). Generalized moments estimation for spatial panel data: Indonesian rice farming. *Amer. J. Agricultural Economics* **86**, 185-198.

Egger, P., Pfaffermayr, M. and Winner, H. (2005). An unbalanced spatial panel data approach to US state tax competition. *Economics Lett.* **88**, 329-335.

Fan, J., Peng, H. and Huang, T. (2005). Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency. *J. Amer. Statist. Assoc.* **100**, 781-813.

Fan, J., Huang, T. and Li, R. Z. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Amer. Statist. Assoc.* **35**, 632-641.

Henderson, D. J., Carroll, R. J. and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *J. Econometrics* **144**, 257-276.

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

Kapoor, M., Kelejian, H. and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *J. Econometrics* **140**, 97-130.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive errors. *J. Real Estate Finance and Economics* **17**, 99-121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Economic Rev.* **40**, 509-533.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *J. Econometrics* **104**, 219-257.

Kelejian, H. H. and Prucha, I. R. (2004) Estimation of systems of spatially interrelated cross sectional equations. *J. Econometrics* **118**, 27-51.

Kelejian, H. H. and Prucha, I. R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econometrics* **157**, 53-67.

Kelejian, H. H. and Robinson, D. P. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics* **22**, 317-331.

Korniotis, G. (2010). Estimating panel models with internal and external habit formation. *J. Business and Economic Statist.* **28**, 145-158.

Lee, L. F. (2002). Consistency and efficiency of least square estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* **18**, 252-277.

Lee, L. F. (2005). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econometrics* **137**, 489-514

Lee, L. F. and Yu, J. H. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *J. Econometrics* **154**, 165-185.

Mutl, J. and Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *Econometrics J.* **14**, 48-76.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models, Asymptotic Theory.* Springer, New York.

Pötscher, B. M. and Prucha, I. R. (2001). Basic elements of asymptotic theory, in *A Companion to Theoretical Econometrics* (Edited by B. H. Baltagi), 201-229. Blackwell, Oxford.

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.

Su, L. (2012). Semi-parametric GMM estimation of spatial autoregressive models. *J. Econometrics* **167**, 543-560.

Su, L. and Jin, S. (2010). Profile quasi-maximum likelihood estimation of spatial autoregressive models. *J. Econometrics* **157**, 18-33.

Ullah, A. and Roy, N. (1998). Parametric and nonparametric panel data models, in *Handbook of Applied Economics and Statistics* (Edited by A. Ullah and David E. A. Giles), 579-604. Marcel Dekker.

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

Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100**, 147-157.

School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China.

Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, China.

E-mail: statbyang@mail.shufe.edu.cn

School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China.

Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, China.

E-mail: frank.jianhuahu@gmail.com

School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China.

Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, China.

E-mail: johnyou07@gmail.com

(Received July 2013; accepted March 2014)