

Article

Statistical Inference for Partially Linear Varying Coefficient Spatial Autoregressive Panel Data Model

Sanying Feng ¹, Tiejun Tong ² and Sung Nok Chiu ^{2,*}¹ School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China; fsy5801@zzu.edu.cn² Department of Mathematics, Hong Kong Baptist University, Hong Kong; tongt@hkbu.edu.hk

* Correspondence: snchiu@hkbu.edu.hk

Abstract: This paper studies the estimation and inference of a partially linear varying coefficient spatial autoregressive panel data model with fixed effects. By means of the basis function approximations and the instrumental variable methods, we propose a two-stage least squares estimation procedure to estimate the unknown parametric and nonparametric components, and meanwhile study the asymptotic properties of the proposed estimators. Together with an empirical log-likelihood ratio function for the regression parameters, which follows an asymptotic chi-square distribution under some regularity conditions, we can further construct accurate confidence regions for the unknown parameters. Simulation studies show that the finite sample performance of the proposed methods are satisfactory in a wide range of settings. Lastly, when applied to the public capital data, our proposed model can also better reflect the changing characteristics of the US economy compared to the parametric panel data models.

Keywords: partially linear varying coefficient model; panel data; spatial autoregressive model; instrumental variable; two-stage least squares; empirical likelihood

MSC: 62G08; 62G20



Citation: Feng, S.; Tong, T.; Chiu, S.N. Statistical Inference for Partially Linear Varying Coefficient Spatial Autoregressive Panel Data Model. *Mathematics* **2023**, *11*, 4606. <https://doi.org/10.3390/math11224606>

Academic Editor: Jin-Ting Zhang

Received: 18 October 2023

Revised: 8 November 2023

Accepted: 8 November 2023

Published: 10 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spatial econometrics is a branch of econometrics that mainly deals with the interactions of economic units in space, where the space can be in both physical and economic dimensions. Early work on spatial econometrics dates back to [1], when the spatial autoregressive model was first introduced, and since then, it has become an active area thanks to its simplicity in estimation and interpretation. For more details on the traditional spatial autoregressive model, one may refer to [2–4], and the references therein. On the other side, however, the parametric structure of the spatial autoregressive model is often subject to the risk of model mis-specification, resulting in modeling bias and even inconsistent estimates. To overcome this shortcoming, nonparametric and semiparametric spatial autoregressive models have also been introduced in the last decade. To name a few, Su and Jin [5] proposed a profile quasi-maximum likelihood estimation approach for a partially linear spatial autoregressive model; Su [6] considered a nonparametric spatial autoregressive model; Malikov and Sun [7] proposed a flexible semiparametric varying coefficient spatial autoregressive model; Sun [8] studied a spatial varying coefficient model with nonparametric spatial weights; and Du et al. [9] developed a partially linear additive spatial autoregressive model and studied the asymptotic properties of the proposed estimators.

Panel data track individual units over time, enabling the estimation of complex models and an extraction of information not possible with cross-sectional or time series data. This two-dimensional information also allows for more comprehensive analysis and inference of panel data ([10,11]). Specifically in spatial econometrics, panel data models are also popular since they take into account the spatial dependence and control of the unobservable

heterogeneity. For instance, Lee and Yu ([12]) focused on a spatial autoregressive panel data model with individual fixed effects. Zhang and Shen [13] considered a partially linear spatial autoregressive panel data model with functional coefficients and random effects. Ai and Zhang [14] considered the estimation of a partially specified spatial panel data model with fixed effects. Li [15] proposed a quasi-maximum-likelihood estimation method for a dynamic spatial panel data model. Sun and Malikov [16] studied a varying coefficient spatial autoregressive panel data model with fixed effects.

In this paper, we are interested in the partially linear varying coefficient spatial panel data model with fixed effects, namely,

$$y_{it} = \lambda_0 \sum_{j=1}^N w_{ij} y_{jt} + \mathbf{x}_{it}^{\tau} \beta_0 + \mathbf{z}_{it}^{\tau} \gamma_0(u_{it}) + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where y_{it} is the response variable, and $\mathbf{x}_{it} = (x_{it,1}, \dots, x_{it,p})^{\tau}$, $\mathbf{z}_{it} = (z_{it,1}, \dots, z_{it,q})^{\tau}$, and $u_{it} \in [a, b]$ are the associated covariates. In addition, $\beta_0 = (\beta_{01}, \dots, \beta_{0p})^{\tau}$ is a p -dimensional vector of unknown parameters, $\gamma_0(\cdot) = (\gamma_{01}(\cdot), \dots, \gamma_{0q}(\cdot))^{\tau}$ is a q -dimensional vector of unknown functions, and ε_{it} is a random error with zero mean and finite variance σ^2 . The unobserved individual-specific effect α_i is time-invariant to account for the individual's unobserved ability, which is also allowed to be correlated with covariates $\mathbf{x}_{it}$, $\mathbf{z}_{it}$, and u_{it} with an unknown correlation structure; and w_{ij} describes the spatial weight of observation j to i , which can be a decreasing function of the spatial distance between i and j . Lastly, we note that the scalar parameter λ_0 measures the strength of spatial dependence.

Model (1) is a unified and flexible model that includes a variety of existing models as special cases. If $\gamma_0(u)$ does not vary over u , it reduces to a vector of constants so that model (1) becomes the traditional spatial autoregressive panel data model [12]. If $q = 1$ and $z_{it} = 1$, the model reduces to the partially linear spatial autoregressive model studied by [14]. If $\mathbf{x}_{it} = 0$, the model is given as a varying coefficient spatial autoregressive model. If $\lambda_0 = 0$, model (1) becomes the partially linear varying coefficient panel data model considered in [17]. Moreover, if $\lambda_0 = 0$ and $\mathbf{x}_{it} = 0$, model (1) reduces to the classical varying coefficient panel data model with fixed effects. For further development of this model, one may refer to, for example, [18–20].

This paper considers the estimation and empirical likelihood inference for model (1). For panel data models with fixed effects, the individual effects are often viewed as the nuisance parameters. We first tackle the fixed effects issue by applying differencing techniques. We then, from the perspective of computational costs, use B-spline to approximate the nonparametric functions and propose a two-stage least squares estimation method to consistently estimate the unknown parameters. The consistency and asymptotic normality properties of the resulting estimators are established under some mild conditions. Moreover, to construct confidence regions of β_0 in model (1), we also propose an empirical log-likelihood ratio function for the regression parameter and show that it follows, asymptotically, a standard chi-square distribution.

The rest of the paper is organized as follows. Section 2 introduces the two-stage least squares estimation and the empirical likelihood inference for the model. Section 3 provides the regularity conditions and then derives the asymptotic properties of the estimators. Section 4 reports the simulation results for assessing the finite sample performance of the proposed methods. Section 5 demonstrates the usefulness of the proposed methods via a real data analysis. Finally, Section 6 concludes the paper with some future directions, and Section 7 presents the technical results.

2. Model Estimation

Throughout the paper, we assume that N is large and T is finite. We first deal with the fixed effects issue by applying the first difference:

$$\begin{aligned} y_{it} - y_{i,t-1} &= \lambda_0 \sum_{j=1}^N w_{ij}(y_{jt} - y_{j,t-1}) + (\mathbf{x}_{it}^\tau - \mathbf{x}_{i,t-1}^\tau)\beta_0 \\ &\quad + \mathbf{z}_{it}^\tau \gamma_0(u_{it}) - \mathbf{z}_{i,t-1}^\tau \gamma_0(u_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1}. \end{aligned} \quad (2)$$

For the additive structure in $\mathbf{z}_{it}^\tau \gamma_0(u_{it}) - \mathbf{z}_{i,t-1}^\tau \gamma_0(u_{i,t-1})$, we further assume $E(\gamma_0(u_{it})) = 0$ for the model identification. Let also $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i,t-1}$, $\zeta = (1, -1)^\tau$, $\tilde{\mathbf{z}}_{it} = \text{diag}\{\mathbf{z}_{it}, \mathbf{z}_{i,t-1}\}$, $\tilde{\gamma}_0(u_{it}) = (\gamma_0(u_{it})^\tau, \gamma_0(u_{i,t-1})^\tau)^\tau$, and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$. Model (2) can then be written as

$$\Delta y_{it} = \lambda_0 \sum_{j=1}^N w_{ij} \Delta y_{jt} + \Delta \mathbf{x}_{it}^\tau \beta_0 + \zeta^\tau \tilde{\mathbf{z}}_{it}^\tau \tilde{\gamma}_0(u_{it}) + \Delta \varepsilon_{it}. \quad (3)$$

Next, we apply the centralized B-spline method to approximate $\gamma_0(u)$. Specifically, let $p^K(u) = (p_1^{K_1}(u)^\tau, \dots, p_q^{K_q}(u)^\tau)^\tau$ denote a sequence of known basis functions, where $p_l^{K_l}(u) = (p_{l1}^{K_l}(u), \dots, p_{lK_l}^{K_l}(u))^\tau$. For each $l = 1, \dots, q$, there exists some constant θ_{0l} such that $\gamma_{0l}(u) = (p_l^{K_l}(u))^\tau \theta_{0l} + \vartheta_{0l}(u)$ with $\vartheta_{0l}(u)$ as the approximation error. For ease of notation, we also define $\vartheta_0(u) = (\vartheta_{01}(u), \dots, \vartheta_{0q}(u))^\tau$ and

$$p(u) = \begin{pmatrix} p_{11}^{K_1}(u) & \cdots & p_{1K_1}^{K_1}(u) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & p_{q1}^{K_q}(u) & \cdots & p_{qK_q}^{K_q}(u) \end{pmatrix},$$

yielding that

$$\gamma_0(u) = p(u)\theta_0 + \vartheta_0(u),$$

where $\theta_0 = (\theta_{01}^\tau, \dots, \theta_{0q}^\tau)^\tau$. With the above notations, model (3) can then be expressed as

$$\begin{aligned} \Delta y_{it} &= \lambda_0 \sum_{j=1}^N w_{ij} \Delta y_{jt} + \Delta \mathbf{x}_{it}^\tau \beta_0 + \zeta^\tau \tilde{\mathbf{z}}_{it}^\tau \{p_0(u_{it})\theta_0 + (\vartheta_0(u_{it})^\tau, \vartheta_0(u_{i,t-1})^\tau)^\tau\} + \Delta \varepsilon_{it} \\ &= \lambda_0 \sum_{j=1}^N w_{ij} \Delta y_{jt} + \Delta \mathbf{x}_{it}^\tau \beta_0 + q_{it}^\tau \theta_0 + v_{0it} + \Delta \varepsilon_{it}, \end{aligned} \quad (4)$$

where $p_0(u_{it}) = (p(u_{it})^\tau, p(u_{i,t-1})^\tau)^\tau$ is a $2q \times \sum_{l=1}^q K_l$ matrix, $q_{it} = (\zeta^\tau \tilde{\mathbf{z}}_{it}^\tau p_0(u_{it}))^\tau$ is a $(\sum_{l=1}^q K_l)$ -dimensional vector, and $v_{0it} = \zeta^\tau \tilde{\mathbf{z}}_{it}^\tau (\vartheta_0(u_{it})^\tau, \vartheta_0(u_{i,t-1})^\tau)^\tau$.

For the sake of descriptive convenience, we denote $\Delta Y_i = (\Delta y_{i2}, \Delta y_{i3}, \dots, \Delta y_{iT})^\tau$ and $\Delta Y = (\Delta Y_1^\tau, \Delta Y_2^\tau, \dots, \Delta Y_N^\tau)^\tau$; analogously, $\Delta X_i, \Delta X, Q_i, Q, V_{0i}, V_0, \Delta \varepsilon_i$ and $\Delta \varepsilon$ can be defined as well. Also let $w_i = (w_{i1}, w_{i2}, \dots, w_{iN})$, $W_i = w_i \otimes I_{T-1}$, and $W = (W_1^\tau, W_2^\tau, \dots, W_N^\tau)^\tau$. We can hence rewrite model (4) as

$$\Delta Y_i = \lambda_0 W_i \Delta Y + \Delta X_i \beta_0 + Q_i \theta_0 + V_{0i} + \Delta \varepsilon_i$$

and

$$\begin{aligned} \Delta Y &= \lambda_0 W \Delta Y + \Delta X \beta_0 + Q \theta_0 + V_0 + \Delta \varepsilon \\ &= D \delta_0 + Q \theta_0 + V_0 + \Delta \varepsilon, \end{aligned} \quad (5)$$

where $D = (W \Delta Y, \Delta X)$ and $\delta_0 = (\lambda_0, \beta_0^\tau)^\tau$.

Let $S = Q(Q^\tau Q)^{-1}Q^\tau$ denote the projection matrix onto the space spanned by Q . Partialling out the B-spline approximation, we obtain

$$(I - S)\Delta Y = (I - S)D\delta_0 + (I - S)V_0 + (I - S)\Delta \varepsilon. \quad (6)$$

To further tackle the endogeneity issue, we apply the two-stage least squares (2SLS) procedure proposed by [21] and obtain the 2SLS estimator of δ_0 with

$$\hat{\delta}_0 = \{D^\tau(I - S)M(I - S)D\}^{-1}D^\tau(I - S)M(I - S)\Delta Y, \quad (7)$$

where $M = H(H^\tau H)^{-1}H^\tau$ and H is a matrix of instrumental variables. This leads to an estimator of θ_0 as

$$\hat{\theta}_0 = (Q^\tau Q)^{-1}Q^\tau(\Delta Y - D\hat{\delta}_0),$$

and, accordingly, we can also apply $\hat{\gamma}_{0l}(u) = (p_l^{K_l}(u))^\tau \hat{\theta}_{0l}$ as the 2SLS estimator of $\gamma_{0l}(u)$, $l = 1, \dots, q$.

For the instrumental matrix H , we construct it in a similar way to [13]. Specifically, in the first step, we select the instrumental variables

$$\tilde{H} = (W(I - \tilde{\lambda}W)^{-1}(Q\tilde{\theta}, \Delta X), \Delta X),$$

where $\tilde{\theta} = (Q^\tau Q)^{-1}Q^\tau(\Delta Y - D\tilde{\delta})$ and $\tilde{\delta} = (\tilde{\lambda}, \tilde{\beta}^\tau)^\tau = \{D^\tau(I - S)D\}^{-1}D^\tau(I - S)\Delta Y$. In the second step, we use the instrumental variable $\tilde{H}$ to obtain the initial consistent estimators $\tilde{\delta}$ and $\tilde{\theta}$, and then use them to construct the instrumental variables

$$H = (W(I - \tilde{\lambda}W)^{-1}(Q\tilde{\theta} + \Delta X\tilde{\beta}), \Delta X).$$

Finally, we use the instrumental variable H to obtain the final estimators $\hat{\delta}_0$ and $\hat{\theta}_0$.

In what follows, we apply the empirical likelihood method to construct the confidence regions of β_0 and λ_0 in model (1). The empirical likelihood method was first introduced by [22], and has now been applied to various regression models ([23,24]). Compared with the two-stage least squares method, an advantage of the empirical likelihood method is that it uses only the data to determine the shape and orientation of confidence regions of β_0 and λ_0 . Another advantage is that the empirical likelihood method can construct confidence regions without estimating the asymptotic covariance, which can be rather complicated for the partially linear varying coefficient spatial autoregressive panel data model with fixed effects.

Specifically for model (6), let $\Delta \tilde{Y} = (I - S)\Delta Y = (\Delta \tilde{Y}_1^\tau, \dots, \Delta \tilde{Y}_N^\tau)^\tau$ and $\tilde{D} = (I - S)D = (\tilde{D}_1^\tau, \dots, \tilde{D}_N^\tau)^\tau$. If the covariate D is exogenous, then the estimating equation for the parametric components δ_0 can be defined as

$$\sum_{i=1}^N \tilde{D}_i^\tau (\Delta \tilde{Y}_i - \tilde{D}_i \delta_0) = 0. \quad (8)$$

In practice, however, D is often an endogenous covariate. In this case, the estimating equation defined by (8) cannot obtain a consistent estimator of δ_0 . To overcome this problem, we propose an adjustment for (8) based on the instrumental variable H , where the key idea is to obtain a linear projection of $\tilde{D}$. From the model $\tilde{D} = H\zeta + e$, the estimator of ζ is known as

$$\hat{\zeta} = (H^\tau H)^{-1}H^\tau \tilde{D}.$$

Also letting $\hat{D} = H\hat{\zeta} = (\hat{D}_1^\tau, \dots, \hat{D}_N^\tau)^\tau$, then our adjustment for (8) is given by

$$\sum_{i=1}^N \hat{D}_i^\tau (\Delta \tilde{Y}_i - \tilde{D}_i \delta_0) = 0.$$

To define the empirical likelihood ratio, we first treat $\eta_i(\delta_0) = \hat{D}_i^\tau(\Delta\tilde{Y}_i - \tilde{D}_i\delta_0)$ as the auxiliary random vector. Then, by [22], an empirical log-likelihood ratio function for δ_0 can be defined as

$$L(\delta_0) = -2 \max \left\{ \sum_{i=1}^N \log(Np_i) \mid p_i \geq 0, \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \eta_i(\delta_0) = 0 \right\},$$

where p_i , $i = 1, \dots, N$, are non-negative real numbers. Finally, through the Lagrange multiplier method, we can show that

$$L(\delta_0) = 2 \sum_{i=1}^N \log[1 + \phi^\tau \eta_i(\delta_0)], \quad (9)$$

where ϕ is a $(p+1)$ -dimensional vector that satisfies the equation of

$$\sum_{i=1}^N \frac{\eta_i(\delta_0)}{1 + \phi^\tau \eta_i(\delta_0)} = 0. \quad (10)$$

3. Asymptotic Properties

Let $G_{0it} = z_{it}^\tau \gamma_0(u_{it})$, $\Delta G_{0it} = G_{0it} - G_{0i,t-1}$, $\Delta G_{0i} = (\Delta G_{0i2}, \Delta G_{0i3}, \dots, \Delta G_{0iT})^\tau$, and $\Delta G_0 = (\Delta G_{01}^\tau, \Delta G_{02}^\tau, \dots, \Delta G_{0N}^\tau)^\tau$. Define $X_i = (x_{i1}, x_{i2}, \dots, x_{iT})^\tau$, and define Z_i , u_i , and ε_i analogously. To derive the asymptotic properties of the proposed estimators, we need the following regularity conditions.

(C1) $\{(X_i, Z_i, u_i, \varepsilon_i), i = 1, \dots, N\}$ are independent and identically distributed, and for all (i, t) , $E[\varepsilon_{it} | X_i, Z_i, u_i, H_i, W] = 0$, $E(\varepsilon_{it}^2 | X_i, Z_i, u_i, H_i, W) = \sigma^2 > 0$. Moreover, there exists a positive constant ϵ such that $E(|\varepsilon_{it}|^{2+\epsilon}) < \infty$, $\sup_{i,t} E(\|z_{it}\|^{2+\epsilon}) < \infty$, and

$$\sup_{i,t} E(\|x_{it}\|^{2+\epsilon}) < \infty.$$

(C2) The matrix $I - \lambda_0 W$ is nonsingular with $|\lambda_0| < 1$, and the row and column sums of the matrices W and $(I - \lambda_0 W)^{-1}$ are bounded uniformly in absolute value for any $|\lambda_0| < 1$. Moreover, for the matrix $C_0 = W(I - \lambda_0 W)^{-1}$, there exists a constant λ_c such that $\lambda_c I - C_0 C_0'$ is positive semidefinite.

(C3) Let the internal knots of the spline be s_j , $j = 1, \dots, K_l$. Also, letting $d_j = s_j - s_{j-1}$ and $d = \max_{1 \leq j \leq K_l} d_j$, there exists a constant M_0 such that

$$\max_{1 \leq j \leq K_l} |d_j - d_{j-1}| = o(K_l^{-1}) \quad \text{and} \quad \frac{d}{\min_{1 \leq j \leq K_l} d_j} \leq M_0.$$

(C4) $\gamma_{0l}(u) \in C^m[a, b]$, $l = 1, \dots, q$, $m \geq 2$, where $C^m[a, b]$ denotes the set of functions with the m th bounded continuous derivatives on the interval $[a, b]$.

(C5) For the knot number K_l , it is assumed that $K_1 \asymp \dots \asymp K_q \asymp K$, and $K = O(N^{1/(2m+1)})$, where $a_0 \asymp b_0$ means that the ratio a_0/b_0 is bounded away from zero and infinity.

(C6) $\Gamma^\tau \Gamma / N$ converges in probability to a positive definite matrix Π , where $\Gamma^\tau = D^{*\tau}(I - S)M$ and $D^* = (W(I - \lambda_0 W)^{-1}(\Delta X \beta_0 + \Delta G_0), \Delta X)$.

(C7) For the matrix D^* , there exists a constant λ_{c^*} such that $\lambda_{c^*} I - D^* D^{*\tau}$ is positive semidefinite.

(C8) The density function of u_{it} , $f(u)$, is bounded away from zero and infinity on $[a, b]$. Furthermore, we assume that $f(u)$ is continuously differentiable on (a, b) .

(C9) Denote $\bar{H} = (I - S)H = (\bar{H}_1^\tau, \dots, \bar{H}_N^\tau)^\tau$ and $\Sigma_i = E(\Delta \varepsilon_i \Delta \varepsilon_i^\tau)$. We assume that $\Psi = \lim_{N \rightarrow \infty} E\left(N^{-1} \sum_{i=1}^N \bar{\xi}^\tau \bar{H}_i^\tau \Sigma_i \bar{H}_i \bar{\xi}\right)$ is a nonsingular constant matrix.

Condition (C1) or its variant is commonly assumed in the spatial panel data models. It requires the explanatory variables (x_{it}, z_{it}, u_{it}) , the instrumental variables H , and the spatial weighting matrix W to be exogenous. Condition (C2) imposes restrictions on the spatial

weighting matrix. These restrictions are required in the setting of a spatial autoregressive model ([3,4]). Condition (C3) is a standard condition on the polynomial spline function method ([25]). Condition (C4) ensures that the functions $\gamma_{0l}(u)$ are sufficiently smooth. Condition (C5) is required to achieve the optimal convergence rate of $\gamma_{0l}(u)$. Condition (C6) is required to establish the asymptotic results. Condition (C7) is required to ensure the identifiability of parameters. Condition (C8) is commonly used in the nonparametric literature. And lastly, Condition (C9) is also routinely used in the empirical likelihood inference ([23,26]).

Let $\xrightarrow{D}$ represent the convergence in distribution. The following two theorems derive the asymptotic distribution and the convergence rate of the 2SLS estimators $\hat{\delta}_0$ and $\hat{\gamma}(u)$, respectively.

Theorem 1. *Under conditions (C1)–(C8), we have*

$$\sqrt{N}(\hat{\delta}_0 - \delta_0) \xrightarrow{D} \mathcal{N}(0, \Omega),$$

where $\Omega = (\Gamma^\tau \Gamma)^{-1} \Gamma^\tau (I - S) \Sigma (I - S) \Gamma (\Gamma^\tau \Gamma)^{-1}$ and $\Sigma = \text{diag}\{\Sigma_1, \dots, \Sigma_N\}$.

Theorem 2. *Under conditions (C1)–(C8), we have*

$$\|\hat{\gamma}(u) - \gamma_0(u)\|^2 = O_p(N^{-2m/(2m+1)}).$$

Theorem 1 shows that the 2SLS estimator of the parametric component $\hat{\delta}_0$ is $\sqrt{N}$ -consistent. Theorem 2 indicates that the 2SLS estimator $\hat{\gamma}(u)$ achieves the optimal convergence rate for nonparametric regression with independent and identically distributed data in [27]. In addition, the above two theorems allow us to construct the confidence region for δ_0 provided a consistent estimator of the asymptotic covariance Ω is obtained. Letting $\Delta \hat{\varepsilon}_i = \Delta Y_i - \hat{\lambda}_0 W_i \Delta Y - \Delta X_i \hat{\beta}_0 - Q_i \hat{\theta}_0$, $\hat{\Sigma}_i = \Delta \hat{\varepsilon}_i \Delta \hat{\varepsilon}_i^\tau$, and $\hat{\Sigma} = \text{diag}\{\hat{\Sigma}_1, \dots, \hat{\Sigma}_N\}$, we then propose to estimate Ω with

$$\hat{\Omega} = (\hat{\Gamma}^\tau \hat{\Gamma})^{-1} \hat{\Gamma}^\tau (I - S) \hat{\Sigma} (I - S) \hat{\Gamma} (\hat{\Gamma}^\tau \hat{\Gamma})^{-1},$$

where $\hat{\Gamma}^\tau = D^\tau (I - S) M$.

Theorem 3. *Under conditions (C1)–(C8), we have*

$$N(\hat{\Omega} - \Omega) = o_p(1).$$

Theorem 3 shows that $\hat{\Omega}$ is given as a consistent estimator. Moreover, by Theorem 1 and Slutsky's theorem, it can be shown that $\hat{\Omega}^{-1/2} \sqrt{N}(\hat{\delta}_0 - \delta_0) \xrightarrow{D} \mathcal{N}(0, I)$. Hence, the asymptotic $100(1 - \alpha)\%$ confidence intervals for δ_{0k} can be constructed as

$$\hat{\delta}_{0k} \pm z_{1-\alpha/2} N^{-1/2} \hat{\Omega}_{kk}^{1/2}, \quad k = 1, \dots, p+1,$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, and $\hat{\Omega}_{kk}$ is the k th diagonal element of $\hat{\Omega}$.

Next, the following theorem establishes the asymptotic distribution of the empirical log-likelihood ratio function $L(\delta_0)$ in (9).

Theorem 4. *Under conditions (C1)–(C9), if δ_0 is the true value of the parameter, then*

$$L(\delta_0) \xrightarrow{D} \chi_{p+1}^2,$$

where χ_{p+1}^2 is a standard chi-square distribution with $p+1$ degrees of freedom.

Theorem 4 can be used to construct the empirical likelihood confidence regions for δ_0 . For any $0 < \alpha < 1$, an approximate $1 - \alpha$ confidence region for δ_0 is given by

$$C_\alpha(\delta_0) = \{\delta_0 \mid L(\delta_0) \leq \chi_{p+1}^2(1 - \alpha)\},$$

where $\chi_{p+1}^2(1 - \alpha)$ is the $1 - \alpha$ quantile of the standard chi-square distribution with $p + 1$ degrees of freedom.

4. Simulation Study

In this section, we investigate the finite sample performance of the proposed estimation and inference methods with a simulation study. The data are generated from the model

$$y_{it} = \lambda_0 \sum_{j=1}^N w_{ij} y_{jt} + x_{it,1} \beta_1 + x_{it,2} \beta_2 + z_{it} \gamma_0(u_{it}) + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

where $x_{it,1} \sim \mathcal{N}(0, 2.25)$, $x_{it,2} \sim \mathcal{N}(0, 1)$, $z_{it} \sim \mathcal{N}(0, 1.69)$, $u_{it} \sim \text{Uniform}(0, 1)$, $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, $\beta_1 = 5$, $\beta_2 = 2$, $\gamma_0(u_{it}) = 0.5 \sin(2\pi u_{it})$, $\alpha_i = \bar{x}_{i,1} + \varepsilon_i$ with $\bar{x}_{i,1} = T^{-1} \sum_{t=1}^T x_{it,1}$ and $\varepsilon_i \sim N(0, 1)$, and $\alpha_1 = -\sum_{i=2}^N \alpha_i$.

Throughout the simulation, we use the centered cubic B-splines as the basis functions. The smoothing parameter K is selected using the generalized cross-validation (GCV) criterion. Similar to [28], we focus on the spatial scenario with a total of R districts, where in each district, there are l members with each neighbor of a member giving equal weight such that $W_0 = (w_{ij}) = I_R \otimes (\frac{e_l e_l^T - I_l}{l-1})$, where e_l is an l -dimensional column vector with all elements being 1 and $\otimes$ is the Kronecker product. In our simulation, the sample sizes are set to be $T = 4$ and 6 , $N = R \times l$ where $R = 30$ and 50 , $l = 4$ and 8 . For comparison, three different values, $\lambda_0 = 0.2, 0.5$, and 0.8 , are also considered, where $\lambda_0 = 0.2$ represents weak spatial dependence, $\lambda_0 = 0.5$ represents mild spatial dependence, and $\lambda_0 = 0.8$ represents strong spatial dependence.

We assess the performance of the two-stage least squares estimation by checking the average bias (Bias) and the sample standard deviation (SD) of the parametric components, and assess the varying coefficient function $\gamma_0(\cdot)$ by checking the square root of the average squared error (RASE), which is defined as

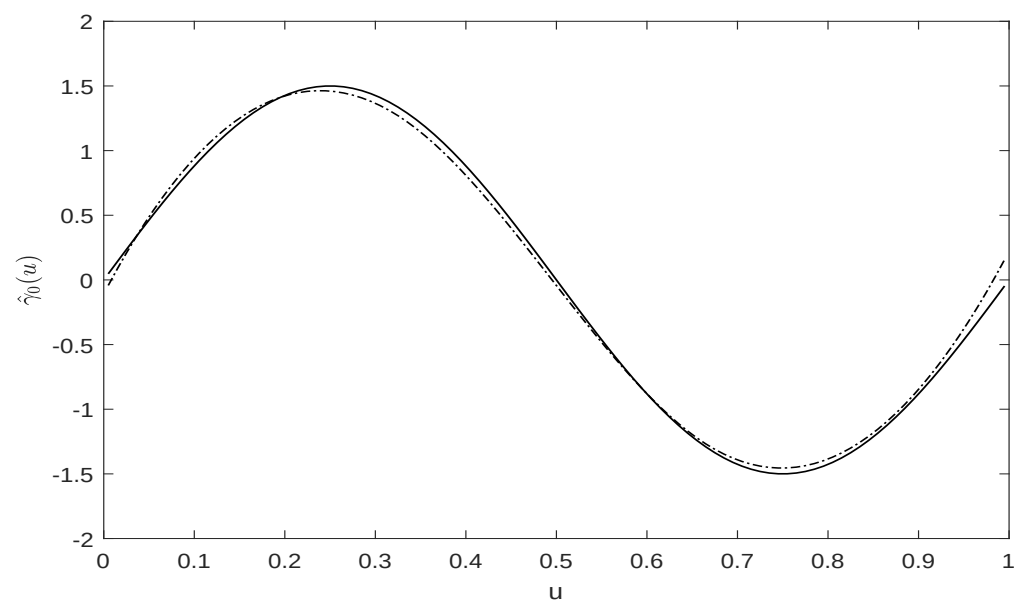
$$\text{RASE} = \left\{ \frac{1}{N_0} \sum_{j=1}^{N_0} [\hat{\gamma}_0(u_j) - \gamma_0(u_j)]^2 \right\}^{1/2},$$

where $\{u_j, j = 1, \dots, N_0\}$ are the regular grid points at which the function $\hat{\gamma}_0(u)$ is evaluated. In our simulation, $N_0 = 100$ is used. We carry out 1000 simulations for each setting and then summarize the results in Table 1 and Figure 1. Table 1 lists the average biases and standard deviations of the estimators of λ_0 , β_1 , and β_2 , and the average RASEs of the estimator of $\gamma_0(\cdot)$. Figure 1 presents the estimator of $\gamma_0(\cdot)$ in a typical sample, which is selected in such a way that its RASE is equal to the median in the 1000 replications.

From Table 1 and Figure 1, we can make a few interesting observations: (i) All the estimators of parameters are close to the true value. (ii) The standard deviations of $\hat{\lambda}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ decrease as the sample size increases. (iii) The RASEs of $\hat{\gamma}_0(u)$ are small for all cases and decrease as the sample size increases, and it can be concluded that the estimate curves fit well to the corresponding true curve, which also coincides with what was discovered from Figure 1. To conclude, the simulation results verify the validity and effectiveness of the proposed estimation procedure.

Table 1. The finite sample performance of the two-stage least squares estimators.

λ_0	(R, l)	T	$\hat{\lambda}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\gamma}_0(u)$
			Bias	SD	Bias	SD	Bias	SD	RASE
0.2	(30, 4)	4	0.0003	0.0077	0.0010	0.0211	−0.0023	0.0264	0.0861
		6	0.0004	0.0067	0.0015	0.0132	0.0031	0.0189	0.0656
	(30, 8)	4	−0.0001	0.0082	0.0011	0.0113	0.0033	0.0184	0.0615
		6	−0.0002	0.0062	−0.0012	0.0095	0.0034	0.0141	0.0465
	(50, 4)	4	0.0005	0.0064	−0.0005	0.0132	0.0035	0.0092	0.0682
		6	−0.0001	0.0056	0.0035	0.0099	0.0019	0.0089	0.0523
	(50, 8)	4	0.0002	0.0060	0.0007	0.0095	0.0022	0.0086	0.0485
		6	−0.0003	0.0047	0.0007	0.0082	−0.0010	0.0075	0.0372
	(30, 4)	4	−0.0013	0.0050	0.0020	0.0225	0.0004	0.0265	0.0824
		6	0.0002	0.0046	0.0013	0.0141	0.0030	0.0191	0.0656
	(30, 8)	4	−0.0001	0.0053	0.0011	0.0113	0.0031	0.0181	0.0615
		6	−0.0001	0.0041	−0.0012	0.0096	0.0033	0.0145	0.0466
0.5	(50, 4)	4	0.0003	0.0044	−0.0008	0.0139	0.0024	0.0094	0.0688
		6	0.0001	0.0032	0.0045	0.0104	0.0002	0.0091	0.0522
	(50, 8)	4	0.0001	0.0039	0.0007	0.0097	0.0020	0.0086	0.0481
		6	−0.0002	0.0031	0.0005	0.0082	0.0001	0.0079	0.0372
0.8	(30, 4)	4	−0.0006	0.0022	0.0032	0.0241	0.0008	0.0266	0.0872
		6	0.0001	0.0020	0.0011	0.0145	0.0029	0.0194	0.0665
	(30, 8)	4	0.0000	0.0021	0.0012	0.0113	0.0033	0.0181	0.0616
		6	0.0002	0.0017	−0.0012	0.0099	0.0013	0.0144	0.0465
	(50, 4)	4	0.0001	0.0019	−0.0011	0.0151	0.0031	0.0098	0.0682
		6	0.0003	0.0014	0.0046	0.0112	0.0020	0.0094	0.0522
	(50, 8)	4	0.0001	0.0016	0.0006	0.0103	0.0012	0.0088	0.0485
		6	−0.0001	0.0013	−0.0009	0.0089	0.0005	0.0081	0.0372

**Figure 1.** The simulation result of $\hat{\gamma}_0(u)$ when $\lambda_0 = 0.5$, $R = 30$, $l = 8$, $T = 4$. The solid curve denotes the true curve, the dash curve denotes its estimate.

The second aim of this simulation study is to construct the confidence intervals for the parameters λ_0 , β_1 , and β_2 , respectively. We consider two approaches for comparison, including the empirical likelihood (EL) approach and the normal approximation using the two-stage least squares estimator (2SLS). The average lengths of the confidence intervals and their corresponding empirical coverage probabilities, with a confidence level of $1 - \alpha = 95\%$, are computed with 1000 simulation runs. The simulation results are pre-

sented in Table 2. It is evident that EL has shorter interval lengths and higher coverage probabilities. This implies that EL performs better than 2SLS in terms of coverage accuracy of the confidence intervals. Lastly, we note that most of the interval lengths decrease and the empirical coverage probabilities increase as the sample size increases.

Table 2. The coverage probabilities and average lengths (in parentheses) of the 95% confidence intervals for δ_0 using different methods.

λ_0	(R, l)	T	$\hat{\lambda}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$	
			EL	2SLS	EL	2SLS	EL	2SLS
0.2	(30, 4)	4	0.9320 (0.0219)	0.9240 (0.0264)	0.9310 (0.0727)	0.9270 (0.0753)	0.9350 (0.1185)	0.9280 (0.1296)
		6	0.9360 (0.0211)	0.9270 (0.0263)	0.9340 (0.0624)	0.9310 (0.0652)	0.9390 (0.0814)	0.9310 (0.0908)
	(30, 8)	4	0.9370 (0.0197)	0.9310 (0.0189)	0.9350 (0.0613)	0.9320 (0.0667)	0.9390 (0.0765)	0.9320 (0.0764)
		6	0.9410 (0.0158)	0.9340 (0.0171)	0.9420 (0.0508)	0.9360 (0.0516)	0.9410 (0.0546)	0.9380 (0.0529)
	(50, 4)	4	0.9390 (0.0206)	0.9320 (0.0217)	0.9360 (0.0677)	0.9320 (0.0717)	0.9380 (0.0783)	0.9320 (0.0822)
		6	0.9410 (0.0160)	0.9330 (0.0152)	0.9410 (0.0551)	0.9370 (0.0572)	0.9420 (0.0638)	0.9360 (0.0637)
	(50, 8)	4	0.9450 (0.0145)	0.9390 (0.0151)	0.9440 (0.0438)	0.9410 (0.0496)	0.9450 (0.0588)	0.9400 (0.0607)
		6	0.9480 (0.0139)	0.9440 (0.0148)	0.9520 (0.0332)	0.9450 (0.0375)	0.9510 (0.0579)	0.9450 (0.0594)
0.5	(30, 4)	4	0.9320 (0.0175)	0.9220 (0.0165)	0.9310 (0.0754)	0.9260 (0.0818)	0.9330 (0.1257)	0.9290 (0.1359)
		6	0.9350 (0.0111)	0.9260 (0.0139)	0.9330 (0.0642)	0.9320 (0.0710)	0.9380 (0.0899)	0.9330 (0.1084)
	(30, 8)	4	0.9350 (0.0129)	0.9290 (0.0137)	0.9360 (0.0643)	0.9330 (0.0695)	0.9390 (0.0856)	0.9320 (0.0976)
		6	0.9390 (0.0109)	0.9330 (0.0123)	0.9410 (0.0537)	0.9370 (0.0641)	0.9420 (0.0751)	0.9370 (0.0825)
	(50, 4)	4	0.9390 (0.0116)	0.9340 (0.0136)	0.9350 (0.0706)	0.9320 (0.0742)	0.9370 (0.0875)	0.9310 (0.0922)
		6	0.9420 (0.0096)	0.9370 (0.0105)	0.9410 (0.0551)	0.9360 (0.0612)	0.9420 (0.0779)	0.9350 (0.0793)
	(50, 8)	4	0.9440 (0.0101)	0.9410 (0.0120)	0.9450 (0.0509)	0.9410 (0.0528)	0.9440 (0.0636)	0.9410 (0.0737)
		6	0.9470 (0.0089)	0.9430 (0.0108)	0.9490 (0.0388)	0.9460 (0.0439)	0.9480 (0.0680)	0.9440 (0.0687)
0.8	(30, 4)	4	0.9270 (0.0097)	0.9210 (0.0095)	0.9330 (0.0928)	0.9290 (0.1087)	0.9320 (0.1286)	0.9310 (0.1421)
		6	0.9310 (0.0085)	0.9240 (0.0088)	0.9350 (0.0723)	0.9310 (0.0862)	0.9380 (0.1014)	0.9320 (0.1017)
	(30, 8)	4	0.9330 (0.0072)	0.9280 (0.0093)	0.9350 (0.0653)	0.9320 (0.0729)	0.9380 (0.0896)	0.9330 (0.0934)
		6	0.9360 (0.0066)	0.9310 (0.0078)	0.9390 (0.0508)	0.9360 (0.0571)	0.9410 (0.0813)	0.9360 (0.0869)
	(50, 4)	4	0.9350 (0.0066)	0.9310 (0.0086)	0.9340 (0.0716)	0.9320 (0.0838)	0.9360 (0.0908)	0.9330 (0.0992)
		6	0.9390 (0.0053)	0.9350 (0.0067)	0.9390 (0.0551)	0.9370 (0.0657)	0.9410 (0.0845)	0.9360 (0.0911)
	(50, 8)	4	0.9420 (0.0057)	0.9380 (0.0062)	0.9440 (0.0514)	0.9420 (0.0554)	0.9430 (0.0816)	0.9390 (0.0838)
		6	0.9450 (0.0043)	0.9410 (0.0059)	0.9470 (0.0388)	0.9470 (0.0513)	0.9470 (0.0774)	0.9420 (0.0825)

5. A Real Data Example

In this section, we apply the proposed estimation methods for model (1) to investigate the productivity of public capital in private production based on data for 48 US states observed over 17 years (1970–1986). The public capital data had been considered in [11,29–31], and can be downloaded from <http://www.mysmu.edu/faculty/zlyang/> (accessed on 1 March 2022). We also note that the previous works were all conducted within

the parametric framework, assuming constant elasticities of the specified models across all the states and all the years. Nevertheless, due to changes in policies as well as the change in the economic environment, including the 1973 oil crisis and the 1979 energy crisis, the constant elasticity assumption can be questionable. In addition, the spatial spillover effects are also discussed in the literature. For example, Xu and Yang [32] employed spatial panel data models to capture the possible spatial spillover effects, and they further pointed out that a temporal heterogeneity pattern is observed in the parameter estimation. In view of this, we propose the following partially linear varying coefficient spatial autoregressive panel data model:

$$\begin{aligned} \log(y_{it}) = & \alpha_i + \lambda_0 \sum_{j=1}^N w_{ij} \log(y_{jt}) + \log(Pc_{it})\beta_1 + \log(L_{it})\beta_2 + Unemp_{it}\beta_3 \\ & + \log(Ps_{it})\gamma_0(t/T) + \varepsilon_{it}, \quad i = 1, \dots, 48, t = 1, \dots, 17, \end{aligned} \quad (11)$$

where y_{it} denotes the gross state product of state i in year t ; α_i reflects the unobserved individual fixed effect; Pc_{it} denotes the public capital including highways and streets, water and sewer facilities, and other public buildings; L_{it} denotes the labor input measured as employment in non-agricultural payrolls; Ps_{it} is the stock of private capital; and $Unemp_{it}$ is the state unemployment rate included to capture business cycle effects. The spatial weight matrix W is specified using a contiguity form, where the (i, j) th element is indicated as 1 if the states i and j share a common border, otherwise it is 0. Note that the final W is also row-normalized.

The fitted results are reported in Table 3 including the estimates (EST) of the parameters and the 95% confidence intervals (CI). The results in the left panel of Table 3 show that Pc_{it} does not have a significant effect on the states' private economic growth. This conclusion is consistent with the finding in [30]. This leads to a reconstructed model as follows:

$$\begin{aligned} \log(y_{it}) = & \alpha_i + \lambda_0 \sum_{j=1}^N w_{ij} \log(y_{jt}) + \log(L_{it})\beta_2 + Unemp_{it}\beta_3 \\ & + \log(Ps_{it})\gamma_0(t/T) + \varepsilon_{it}. \end{aligned} \quad (12)$$

Table 3. The estimates of the parameters and their 95% confidence intervals.

	Model (11)			Model (12)		
	EST	CI _{2SLS}	CI _{EL}	EST	CI _{2SLS}	CI _{EL}
λ_0	0.0908	[0.0036, 0.1780]	[0.0419, 0.1386]	0.0838	[0.0108, 0.1568]	[0.0346, 0.1318]
β_1	−0.0122	[−0.1110, 0.0866]	[−0.1312, 0.1067]	—	—	—
β_2	0.8633	[0.7773, 0.9493]	[0.8104, 0.9169]	0.8657	[0.7797, 0.9517]	[0.8122, 0.9197]
β_3	−0.0039	[−0.0055, −0.0023]	[−0.0048, −0.0030]	−0.0040	[−0.0056, −0.0025]	[−0.0049, −0.0031]

From the right panel of Table 3, we can see that the significance of the spatial coefficient estimate reflects the spatial dependence and confirms the existences of spillover effects between states. Moreover, L_{it} affects the states' private economic growth positively, and $Unemp_{it}$ affects the states' private economic growth negatively. Further, the fitted varying coefficient function curve is presented in Figure 2. The estimated curve has two inflection points, which approximately correspond to the 1973 oil crisis and the 1979 energy crisis. Figure 2 indicates the fluctuating effects of Ps_{it} on the states' private economic growth. In the mid-1970s, the effect of Ps_{it} on the states' private economic growth was approximately unchanged, while in the early 1970s and also the mid-1980s, the negative effect of Ps_{it} on the states' private economic growth increased rapidly. This demonstrates that the standard applications of parametric panel models may not be valid.

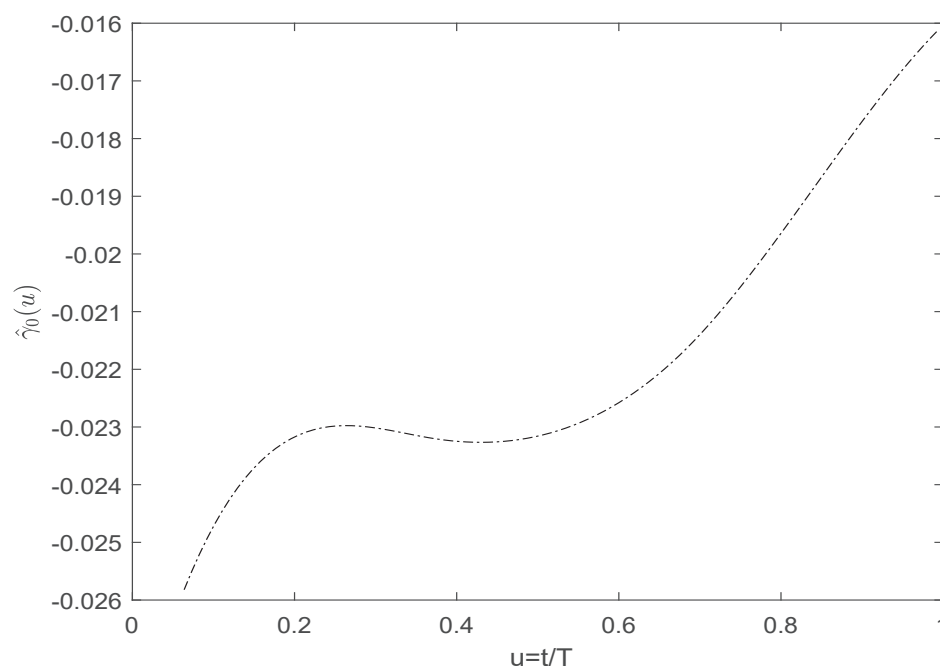


Figure 2. The estimated curve of the varying coefficient function $\gamma_0(\cdot)$.

6. Conclusions and Discussion

In this paper, we studied the statistical inference for a partially linear varying coefficient spatial autoregressive panel data model with fixed effects. By means of the basis function approximations and instrumental variable methods, we proposed a two-stage least squares estimation procedure to estimate the unknown parametric and nonparametric components, and meanwhile derived the convergence rate and asymptotic distributions of the estimators under some regularity conditions. We further constructed an empirical log-likelihood ratio function to derive the empirical likelihood confidence regions for the parametric component, which is shown to have an asymptotically correct coverage probability. Simulation studies and real data analysis also demonstrated that the proposed method performs well in the finite sample settings.

Lastly, we note that there are some interesting directions for future research. First, extending the model to a case with spatial errors would be useful yet challenging work. Second, the present paper assumes the spatial matrix W to be predetermined and time-invariant. In practice, however, the spatial structure W may change along with T , especially when it is large. In addition, the spatial coefficient λ_0 may also change with time. These circumstances are outside the scope of the present paper and are left for future research.

7. Proof of the Main Results

To prove the theorems obtained in Section 3, we first present several lemmas. Note that the first three are essentially the same as Corollary 6.21 in [25], Lemma 4.5 in [33], and Lemma A.2 in [34], respectively. For convenience and simplicity, we also express C as a positive constant that may be different at each appearance throughout this section.

Lemma 1. Under conditions (C4) and (C5), there exists a constant $C > 0$ such that

$$\sup_{u \in [a, b]} |\gamma_{0l}(u) - (p_l^{K_l}(u))^{\tau} \theta_{0l}| \leq CK^{-m}.$$

Lemma 2. Under condition (C1), there exist two constants $0 < c_1 < c_2 < \infty$ such that

$$(c_1 + o_p(1)) \frac{N}{K} \leq \lambda_{\min}^{Q^{\tau}Q} \leq \lambda_{\max}^{Q^{\tau}Q} \leq (c_2 + o_p(1)) \frac{N}{K}.$$

Lemma 3. We denote $\varsigma_i, i = 1, \dots, N$, as independent random variables satisfying $E(\varsigma_i) = 0$ and $E(\varsigma_i^2) < C < \infty$. Then, we have

$$\max_{1 \leq k \leq N} \left| \sum_{i=1}^k \varsigma_i \right| = O_p(\sqrt{N} \log N).$$

Lemma 4. Under conditions (C1)–(C9), if δ_0 is the true value of the parameter, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i(\delta_0) \xrightarrow{D} \mathcal{N}(0, \Psi).$$

Proof. We have

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i(\delta_0) &= \frac{1}{\sqrt{N}} \hat{D}^\tau (\Delta \tilde{Y} - \tilde{D} \delta_0) = \frac{1}{\sqrt{N}} \hat{D}^\tau (I - S) (\Delta Y - D \delta_0) \\ &= \frac{1}{\sqrt{N}} \hat{D}^\tau (I - S) V_0 + \frac{1}{\sqrt{N}} (\hat{\xi} - \xi)^\tau H^\tau (I - S) \Delta \varepsilon \\ &\quad + \frac{1}{\sqrt{N}} \xi^\tau H^\tau (I - S) \Delta \varepsilon. \end{aligned} \quad (13)$$

By Lemma 1, we have $\|V_{0i}\| = O(K^{-m})$. Also, by [35], we have $\hat{\xi} = \xi + O_p(N^{-1/2})$. Letting $\tilde{D} = (I - S)\hat{D} = (\tilde{D}_1, \dots, \tilde{D}_N)$, and invoking condition (C1) and Lemma 3, we can obtain

$$\frac{1}{\sqrt{N}} \hat{D}^\tau (I - S) V_0 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{D}_i^\tau V_{0i} = O_p\left(\frac{1}{\sqrt{N}} \cdot \sqrt{N} \log N \cdot K^{-m}\right) = o_p(1) \quad (14)$$

and

$$\frac{1}{\sqrt{N}} (\hat{\xi} - \xi)^\tau H^\tau (I - S) \Delta \varepsilon = o_p(1). \quad (15)$$

Also let $\omega_i = \xi^\tau \tilde{H}_i^\tau \Delta \varepsilon_i$. It is easy to verify that $E(\omega_i | H_i, Z_i) = 0$ and $\text{Cov}(\omega_i | H_i, Z_i) = \xi^\tau \tilde{H}_i^\tau \Sigma_i \tilde{H}_i \xi$. Further, by invoking condition (C9) and by the law of large numbers, we can derive that

$$\frac{1}{N} \sum_{i=1}^N \xi^\tau \tilde{H}_i^\tau \Sigma_i \tilde{H}_i \xi \xrightarrow{P} \Psi. \quad (16)$$

For any given vector $a \in R^{p+1}$ with $a^\tau a = 1$, invoking condition (C1), it is easy to show that $E(a^\tau \omega_i) = 0$ and

$$\sup_i E\|a^\tau \omega_i\|^{2+\delta} \leq \|a\|^{2+\delta} \sup_i E\|\omega_i\|^{2+\delta} \leq C \sup_i E\|\omega_i\|^{2+\delta} < \infty.$$

Hence, $a^\tau \omega_i$ satisfies the Lyapunov condition for the central limit theorem, yielding that

$$\frac{\sum_{i=1}^N a^\tau \omega_i}{(a^\tau \sum_{i=1}^N \text{Cov}(\omega_i) a)^{1/2}} \xrightarrow{D} \mathcal{N}(0, 1). \quad (17)$$

Finally, by (13)–(17), we have

$$\frac{1}{\sqrt{N}} \xi^\tau H^\tau (I - S) \Delta \varepsilon = \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi^\tau \tilde{H}_i^\tau \Delta \varepsilon_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N \omega_i \xrightarrow{D} \mathcal{N}(0, \Psi)$$

and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i(\delta_0) = \frac{1}{N} \sum_{i=1}^N \omega_i + o_p(1) \xrightarrow{D} \mathcal{N}(0, \Psi).$$

This proves the lemma. $\square$

Lemma 5. Under conditions (C1)–(C9), if δ_0 is the true value of the parameter, we have

$$\frac{1}{N} \sum_{i=1}^N \eta_i(\delta_0) \eta_i(\delta_0)^\tau \xrightarrow{P} \Psi.$$

Proof. Following the same notation as in Lemma 4, we can derive that

$$\frac{1}{N} \sum_{i=1}^N \eta_i(\delta_0) \eta_i(\delta_0)^\tau = \frac{1}{N} \sum_{i=1}^N \omega_i \omega_i^\tau + o_p(1).$$

Further, by a similar argument as that for (16), we have

$$\frac{1}{N} \sum_{i=1}^N \omega_i \omega_i^\tau \xrightarrow{P} \Psi.$$

This thus proves the lemma. $\square$

Proof of Theorem 1. By (5), a simple calculation yields that

$$\begin{aligned} \hat{\delta}_0 - \delta_0 &= \{D^\tau(I - S)M(I - S)D\}^{-1} D^\tau(I - S)M(I - S)(D\delta_0 + Q\theta_0 + V_0 + \Delta\epsilon) - \delta_0 \\ &= \{D^\tau(I - S)M(I - S)D\}^{-1} D^\tau(I - S)M(I - S)V_0 \\ &\quad + \{D^\tau(I - S)M(I - S)D\}^{-1} D^\tau(I - S)M(I - S)\Delta\epsilon. \end{aligned}$$

Note also that $D^* = (C_0(\Delta X\beta_0 + \Delta G_0), \Delta X)$ and

$$D = (W(I - \lambda_0 W)^{-1}(\Delta X\beta_0 + \Delta G_0 + \Delta\epsilon), \Delta X) = (C_0(\Delta X\beta_0 + \Delta G_0 + \Delta\epsilon), \Delta X).$$

Thus, $D = D^* + \Lambda$, where $\Lambda = (C_0\Delta\epsilon, 0)$. Furthermore, we have

$$\begin{aligned} D^\tau(I - S)M(I - S)D &= D^{*\tau}(I - S)M(I - S)D^* + \Lambda^\tau(I - S)M(I - S)\Lambda \\ &\quad + D^{*\tau}(I - S)M(I - S)\Lambda + \Lambda^\tau(I - S)M(I - S)D^* \\ &=: R_{11} + R_{12} + R_{13} + R_{14}. \end{aligned}$$

It is easy to show that $E(\Delta\epsilon\Delta\epsilon^\tau) = \sigma^2 I_N \otimes A$, where $A = 2I_{T-1} - J_{T-1}(0) - J_{T-1}(0)^\tau$ is a $(T-1) \times (T-1)$ matrix, and

$$J_{T-1}(0) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{(T-1) \times (T-1)}.$$

Based on Gerschgorin's disk theorem ([36]), if λ is the eigenvalue of A , then $0 \leq \lambda \leq 4$. Thus, there exists a constant $\lambda_{\max}$, such that $E(\Delta\epsilon\Delta\epsilon^\tau) < \lambda_{\max}I$. Further, by condition (C2), we have

$$\begin{aligned} &E[\Delta\epsilon^\tau C_0^\tau(I - S)M(I - S)C_0\Delta\epsilon] \\ &= \text{tr}\{E[(H^\tau H)^{-1/2}H^\tau(I - S)C_0\Delta\epsilon\Delta\epsilon^\tau C_0^\tau(I - S)H(H^\tau H)^{-1/2}]\} \\ &\leq \lambda_{\max}\lambda_c \text{tr}\{E[(H^\tau H)^{-1/2}H^\tau(I - S)H(H^\tau H)^{-1/2}]\} \\ &\leq \lambda_{\max}\lambda_c \text{tr}\{E[(H^\tau H)^{-1/2}H^\tau H(H^\tau H)^{-1/2}]\} \\ &= O(1). \end{aligned}$$

This shows that

$$\Delta \varepsilon^\tau C_0^\tau (I - S) M (I - S) C_0 \Delta \varepsilon = O_p(1), \quad (18)$$

or, equivalently, $R_{12} = O_p(1)$. Invoking conditions (C2) and (C7), we have

$$\begin{aligned} & E \| D^{*\tau} (I - S) M (I - S) C_0 \Delta \varepsilon \|^2 \\ &= \text{tr} \{ E [D^{*\tau} (I - S) M (I - S) C_0 \Delta \varepsilon \Delta \varepsilon^\tau C_0^\tau (I - S) M (I - S) D^*] \} \\ &\leq \lambda_{\max} \text{tr} \{ E [D^{*\tau} (I - S) M (I - S) C_0 C_0^\tau (I - S) M (I - S) D^*] \} \\ &\leq \lambda_{\max} \lambda_c \text{tr} \{ E [D^{*\tau} (I - S) M (I - S) M (I - S) D^*] \} \\ &\leq \lambda_{\max} \lambda_c \text{tr} \{ E [(H^\tau H)^{-1/2} H^\tau D^* D^{*\tau} H (H^\tau H)^{-1/2}] \} \\ &\leq \lambda_{\max} \lambda_c \lambda_{c^*} \text{tr} \{ E [(H^\tau H)^{-1/2} H^\tau H (H^\tau H)^{-1/2}] \} \\ &= O(1). \end{aligned} \quad (19)$$

This implies that $R_{13} = O_p(1)$. Similarly, it can also be shown that $R_{14} = O_p(1)$. Taken together, we have

$$D^\tau (I - S) M (I - S) D = D^{*\tau} (I - S) M (I - S) D^* + O_p(1). \quad (20)$$

Next, for the term $D^\tau (I - S) M (I - S) \Delta \varepsilon$, we note that

$$\begin{aligned} D^\tau (I - S) M (I - S) \Delta \varepsilon &= D^{*\tau} (I - S) M (I - S) \Delta \varepsilon + \Lambda^\tau (I - S) M (I - S) \Delta \varepsilon \\ &=: R_{21} + R_{22}. \end{aligned}$$

By a simple calculation, we have

$$\begin{aligned} & E [\Delta \varepsilon^\tau (I - S) M (I - S) \Delta \varepsilon] \\ &= \text{tr} \{ E [(H^\tau H)^{-1/2} H^\tau (I - S) \Delta \varepsilon \Delta \varepsilon^\tau (I - S) H (H^\tau H)^{-1/2}] \} \\ &\leq \lambda_{\max} \text{tr} \{ E [(H^\tau H)^{-1/2} H^\tau (I - S) H (H^\tau H)^{-1/2}] \} \\ &\leq \lambda_{\max} \text{tr} \{ E [(H^\tau H)^{-1/2} H^\tau H (H^\tau H)^{-1/2}] \} \\ &= O(1). \end{aligned}$$

Hence

$$\Delta \varepsilon^\tau (I - S) M (I - S) \Delta \varepsilon = O_p(1). \quad (21)$$

Applying the triangular inequality and invoking (18) and (21), we obtain

$$\Delta \varepsilon^\tau C_0^\tau (I - S) M (I - S) \Delta \varepsilon = O_p(1).$$

That is, $R_{22} = O_p(1)$. Thus, we have

$$D^\tau (I - S) M (I - S) \Delta \varepsilon = D^{*\tau} (I - S) M (I - S) \Delta \varepsilon + O_p(1).$$

Lastly, for the term $D^\tau (I - S) M (I - S) V_0$, we can represent it as

$$\begin{aligned} D^\tau (I - S) M (I - S) V_0 &= D^{*\tau} (I - S) M (I - S) V_0 + \Lambda^\tau (I - S) M (I - S) V_0 \\ &=: R_{31} + R_{32}. \end{aligned}$$

Invoking condition (C7) and Lemma 1, we have

$$\begin{aligned} E (\| R_{31} \|^2) &= E [\text{tr} (V_0^\tau (I - S) M (I - S) D^* D^{*\tau} (I - S) M (I - S) V_0)] \\ &\leq \lambda_{c^*} E [\text{tr} (V_0^\tau (I - S) M (I - S) M (I - S) V_0)] \\ &\leq \lambda_{c^*} E [\text{tr} (V_0^\tau (I - S) M (I - S) V_0)] \\ &\leq \lambda_{c^*} E [\text{tr} (V_0^\tau V_0)] \\ &= O(NK^{-2m}). \end{aligned}$$

This shows that $R_{31} = O_p(\sqrt{N}K^{-m})$. Similarly, we can have $R_{32} = O_p(\sqrt{N}K^{-m})$. That is, $D'(I - S)M(I - S)V_0 = O_p(\sqrt{N}K^{-m})$.

Combining the above results, we can obtain that

$$\begin{aligned}\sqrt{N}(\hat{\delta}_0 - \delta_0) &= \left\{ \frac{1}{N} D^{*\tau}(I - S)M(I - S)D^* + O_p(N^{-1}) \right\}^{-1} \\ &\quad \times \frac{1}{\sqrt{N}} D^{*\tau}(I - S)M(I - S)\Delta\varepsilon + O_p(K^{-m}) + O_p(N^{-1/2}).\end{aligned}$$

Finally, invoking conditions (C1) and (C6), and using the central limit theorem and Slutsky's theorem, we have

$$\sqrt{N}(\hat{\delta}_0 - \delta_0) \xrightarrow{D} \mathcal{N}(0, \Omega).$$

This completes the proof of Theorem 1. $\square$

Proof of Theorem 2. By a simple calculation, we have

$$\begin{aligned}\hat{\theta}_0 &= (Q^\tau Q)^{-1} Q^\tau (\Delta Y - D\delta_0) + (Q^\tau Q)^{-1} Q^\tau D(\delta_0 - \hat{\delta}) \\ &= \theta_0 + (Q^\tau Q)^{-1} Q^\tau \Delta\varepsilon + (Q^\tau Q)^{-1} Q^\tau V_0 + (Q^\tau Q)^{-1} Q^\tau D(\delta_0 - \hat{\delta}) \\ &=: \theta_0 + R_{41} + R_{42} + R_{43}.\end{aligned}$$

For R_{41} , by Lemma 2, we have

$$\begin{aligned}E(\|R_{41}\|^2) &= E(\Delta\varepsilon^\tau Q(Q^\tau Q)^{-1}(Q^\tau Q)^{-1} Q^\tau \Delta\varepsilon) \\ &= O(N^{-1}K)E[\text{tr}\{(Q^\tau Q)^{-1/2} Q^\tau \Delta\varepsilon \Delta\varepsilon^\tau Q(Q^\tau Q)^{-1/2}\}] \\ &= O(N^{-1}K)\text{tr}\{E[(Q^\tau Q)^{-1/2} Q^\tau \Delta\varepsilon \Delta\varepsilon^\tau Q(Q^\tau Q)^{-1/2}]\} \\ &= O(1)O(N^{-1}K)\text{tr}\{E[(Q^\tau Q)^{-1/2} Q^\tau Q(Q^\tau Q)^{-1/2}]\} \\ &= O(N^{-1}K)E[\text{tr}\{(Q^\tau Q)^{-1/2} Q^\tau Q(Q^\tau Q)^{-1/2}\}] \\ &= O(N^{-1}K^2).\end{aligned}$$

This shows that $R_{41} = O_p(N^{-1/2}K)$. Similarly, by Lemmas 1 and 2, it is easy to show that $R_{42} = O_p(K^{-m+1/2})$. Moreover, by Theorem 1 and Lemma 2,

$$\|R_{43}\|^2 = (\hat{\delta} - \delta_0)^\tau D^\tau Q(Q^\tau Q)^{-1}(Q^\tau Q)^{-1} Q^\tau D(\hat{\delta} - \delta_0) = O_p(N^{-1}).$$

or, equivalently, $R_{43} = O_p(N^{-1/2})$. Taken together, we have

$$\|\hat{\theta}_0 - \theta_0\|^2 = O_p(N^{-1}K^2 + K^{-2m+1} + N^{-1}) = O_p(N^{-1}K^2). \quad (22)$$

Furthermore, it is easy to show that

$$\begin{aligned}\|\hat{\gamma}(u) - \gamma_0(u)\|^2 &= \|p(u)(\hat{\theta}_0 - \theta_0) + p(u)\theta_0 - \gamma_0(u)\|^2 \\ &\leq 2\|p(u)(\hat{\theta}_0 - \theta_0)\|^2 + 2\|p(u)\theta_0 - \gamma_0(u)\|^2 \\ &= 2(\hat{\theta}_0 - \theta_0)^\tau p(u)^\tau p(u)(\hat{\theta}_0 - \theta_0) + O_p(K^{-2m}) \\ &= O_p(K^{-1})(\hat{\theta}_0 - \theta_0)^\tau (\hat{\theta}_0 - \theta_0) + O_p(K^{-2m}) \\ &= O_p(N^{-1}K + K^{-2m}) \\ &= O_p(N^{-2m/(2m+1)}).\end{aligned}$$

This completes the proof of Theorem 2. $\square$

Proof of Theorem 3. Note that

$$N\hat{\Omega} = \left(\frac{\hat{\Gamma}^{\tau}\hat{\Gamma}}{N} \right)^{-1} \frac{\hat{\Gamma}^{\tau}(I-S)\hat{\Sigma}(I-S)\hat{\Gamma}}{N} \left(\frac{\hat{\Gamma}^{\tau}\hat{\Gamma}}{N} \right)^{-1},$$

$$N\Omega = \left(\frac{\Gamma^{\tau}\Gamma}{N} \right)^{-1} \frac{\Gamma^{\tau}(I-S)\Sigma(I-S)\Gamma}{N} \left(\frac{\Gamma^{\tau}\Gamma}{N} \right)^{-1}.$$

Hence, to prove the theorem, we only need to show that

$$\frac{\hat{\Gamma}^{\tau}\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}\Gamma}{N} = o_p(1), \quad (23)$$

$$\frac{\hat{\Gamma}^{\tau}(I-S)\hat{\Sigma}(I-S)\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}(I-S)\Sigma(I-S)\Gamma}{N} = o_p(1). \quad (24)$$

For (23), by $D = D^* + \Lambda$ and $\hat{\Gamma}^{\tau} - \Gamma^{\tau} = \Lambda^{\tau}(I-S)M$, we have

$$\begin{aligned} \frac{\hat{\Gamma}^{\tau}\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}\Gamma}{N} &= \frac{(\hat{\Gamma} - \Gamma)^{\tau}(\hat{\Gamma} - \Gamma)}{N} + \frac{\Gamma^{\tau}(\hat{\Gamma} - \Gamma)}{N} + \frac{(\hat{\Gamma} - \Gamma)^{\tau}\Gamma}{N} \\ &= \frac{\Lambda^{\tau}(I-S)M(I-S)\Lambda}{N} + \frac{D^{*\tau}(I-S)M(I-S)\Lambda}{N} \\ &\quad + \frac{\Lambda^{\tau}(I-S)M(I-S)D^*}{N} \\ &=: B_{21} + B_{22} + B_{23}. \end{aligned}$$

Further, from the proof of Theorem 1, it is easy to verify that $B_{21} = O_p(N^{-1})$, $B_{22} = O_p(N^{-1})$, and $B_{23} = O_p(N^{-1})$. Therefore,

$$\frac{\hat{\Gamma}^{\tau}\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}\Gamma}{N} = O_p(N^{-1}) + O_p(N^{-1}) + O_p(N^{-1}) = o_p(1).$$

For (24), a simple calculation yields that

$$\begin{aligned} &\frac{\hat{\Gamma}^{\tau}(I-S)\hat{\Sigma}(I-S)\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}(I-S)\Sigma(I-S)\Gamma}{N} \\ &= \frac{\hat{\Gamma}^{\tau}(I-S)(\hat{\Sigma} - \Sigma)(I-S)\hat{\Gamma}}{N} + \frac{\hat{\Gamma}^{\tau}(I-S)\Sigma(I-S)\hat{\Gamma}}{N} - \frac{\Gamma^{\tau}(I-S)\Sigma(I-S)\Gamma}{N}. \end{aligned}$$

Note also that

$$\Delta\hat{\varepsilon}_i\Delta\hat{\varepsilon}_i^{\tau} - \Delta\varepsilon_i\Delta\varepsilon_i^{\tau} = (\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)(\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)^{\tau} + \Delta\varepsilon_i(\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)^{\tau} + (\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)\Delta\varepsilon_i^{\tau}$$

and

$$\begin{aligned} \Delta\hat{\varepsilon}_i - \Delta\varepsilon_i &= \Delta Y_i - D_i\hat{\delta}_0 - Q_i\hat{\theta}_0 - (\Delta Y_i - D_i\delta_0 - Q_i\theta_0 - V_{0i}) \\ &= D_i(\delta_0 - \hat{\delta}_0) + Q_i(\theta_0 - \hat{\theta}_0) + V_{0i}. \end{aligned}$$

Thus, invoking Lemma 1, Theorem 1, and (22), we have

$$\begin{aligned} &(\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)(\Delta\hat{\varepsilon}_i - \Delta\varepsilon_i)^{\tau} \\ &\leq 4D_i(\delta_0 - \hat{\delta})(\delta_0 - \hat{\delta})^{\tau}D_i^{\tau} + 4Q_i(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})^{\tau}Q_i^{\tau} + 4V_{0i}V_{0i}^{\tau} \\ &= O_p(N^{-1}) + O_p(N^{-1}K^2) + O_p(K^{-2m}) = o_p(1) \end{aligned}$$

and

$$\begin{aligned}
\Delta \varepsilon_i (\Delta \hat{\varepsilon}_i - \Delta \varepsilon_i)^\tau &= \Delta \varepsilon_i (\delta_0 - \hat{\delta}_0)^\tau D_i^\tau + \Delta \varepsilon_i (\theta_0 - \hat{\theta}_0)^\tau Q_i^\tau + \Delta \varepsilon_i V_{0i}^\tau \\
&= O_p(N^{-1/2}) + O_p(N^{-1/2}K) + O_p(K^{-m}) \\
&= o_p(1).
\end{aligned}$$

Hence, we obtain

$$\Delta \hat{\varepsilon}_i \Delta \hat{\varepsilon}_i^\tau - \Delta \varepsilon_i \Delta \varepsilon_i^\tau = o_p(1) \quad \text{and} \quad \hat{\Sigma} - \Sigma = o_p(1).$$

Using condition (C7), similar to the proof of (20), we can obtain that

$$\begin{aligned}
\left\| \frac{\hat{\Gamma}^\tau (I - S)(\hat{\Sigma} - \Sigma)(I - S)\hat{\Gamma}}{N} \right\| &= \left\| \frac{D^\tau (I - S)M(I - S)(\hat{\Sigma} - \Sigma)(I - S)M(I - S)D}{N} \right\| \\
&= o_p(1) \left\| \frac{D^\tau (I - S)M(I - S)M(I - S)D}{N} \right\| \\
&\leq o_p(1) \left\| \frac{D^\tau (I - S)M(I - S)D}{N} \right\| \\
&= o_p(1).
\end{aligned}$$

Moreover, similar to the proof of (19), it is easy to show that

$$\begin{aligned}
&\frac{\hat{\Gamma}^\tau (I - S)\Sigma(I - S)\hat{\Gamma}}{N} - \frac{\Gamma^\tau (I - S)\Sigma(I - S)\Gamma}{N} \\
&= \frac{\hat{\Gamma}^\tau (I - S)\Sigma(I - S)(\hat{\Gamma} - \Gamma)}{N} - \frac{(\hat{\Gamma} - \Gamma)^\tau (I - S)\Sigma(I - S)\Gamma}{N} \\
&= \frac{D^\tau (I - S)M(I - S)\Sigma(I - S)M(I - S)\Lambda}{N} \\
&\quad + \frac{\Lambda^\tau (I - S)M(I - S)\Sigma(I - S)M(I - S)D^*}{N} \\
&= O_p(N^{-1}) + O_p(N^{-1}) \\
&= o_p(1).
\end{aligned}$$

Taken together, (24) also holds. This completes the proof of Theorem 3. $\square$

Proof of Theorem 4. Using a similar argument as in [22], we can obtain

$$\|\phi\| = O_p(N^{-1/2})$$

and

$$\max_{1 \leq i \leq N} \|\eta_i(\delta_0)\| = o_p(N^{1/2}).$$

Further using the Taylor expansion to $L(\delta_0)$, it yields that

$$L(\delta_0) = 2 \sum_{i=1}^N \log[1 + \phi^\tau \eta_i(\delta_0)] = 2 \sum_{i=1}^N [\phi^\tau \eta_i(\delta_0) - \frac{1}{2}(\phi^\tau \eta_i(\delta_0))^2] + o_p(1). \quad (25)$$

In addition, from (10), we have

$$\begin{aligned}
0 &= \frac{1}{N} \sum_{i=1}^N \frac{\eta_i(\delta_0)}{1 + \phi^\tau \eta_i(\delta_0)} \\
&= \frac{1}{N} \sum_{i=1}^N \eta_i(\delta_0) - \frac{1}{N} \sum_{i=1}^N \eta_i(\delta_0) \eta_i(\delta_0)^\tau \phi + \frac{1}{N} \sum_{i=1}^N \frac{\eta_i(\delta_0) [\phi^\tau \eta_i(\delta_0)]^2}{1 + \phi^\tau \eta_i(\delta_0)}. \quad (26)
\end{aligned}$$

Note also that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{\|\eta_i(\delta_0)\|^3 \|\phi\|^2}{|1 + \phi^\tau \eta_i(\delta_0)|} &\leq \|\phi\|^2 \max_{1 \leq i \leq N} \|\eta_i(\delta_0)\| \frac{1}{N} \sum_{i=1}^N \|\eta_i(\delta_0)\|^2 \\ &= O_p(N^{-1}) o_p(N^{1/2}) O_p(1) \\ &= o_p(N^{-1/2}). \end{aligned}$$

Then, from (26), it leads to

$$\phi = \left[\sum_{i=1}^N \eta_i(\delta_0) \eta_i(\delta_0)^\tau \right]^{-1} \sum_{i=1}^N \eta_i(\delta_0) + o_p(N^{-1/2}). \quad (27)$$

Moreover, from (25) and (27), we have

$$L(\delta_0) = \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i(\delta_0) \right]^\tau \left[\frac{1}{N} \sum_{i=1}^N \eta_i(\delta_0) \eta_i(\delta_0)^\tau \right]^{-1} \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i(\delta_0) \right] + o_p(1).$$

Finally, invoking Lemmas 4 and 5, the proof of Theorem 4 is complete. $\square$

Author Contributions: Methodology, S.F. and T.T.; software, S.F.; formal analysis, S.F.; investigation, S.N.C.; writing—original draft, S.F.; writing—review and editing, T.T. and S.N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Social Science Foundation of China (23BTJ061), the Humanities and Social Science Project of the Ministry of Education of China (21YJC910003), the Foundation of Henan Educational Committee (21A910004), the General Research Fund of Hong Kong (HKBU12300123, HKBU12303421), the National Natural Science Foundation of China (1207010822), and the Research Matching Grant Scheme (RMGS-2022-11-08) from the Research Grants Council of Hong Kong.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors are grateful to the three reviewers for the constructive comments and suggestions that led to significant improvements to the original manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cliff, A.; Ord, J.K. *Spatial Autocorrelation*; Pion: London, UK, 1973.
2. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1988.
3. Lee, L.F. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econom. Rev.* **2003**, *22*, 307–335. [\[CrossRef\]](#)
4. Lee, L.F.; Yu, J.H. A spatial dynamic panel data model with both time and individual fixed effects. *Econom. Theory* **2007**, *26*, 564–597. [\[CrossRef\]](#)
5. Su, L.J.; Jin, S.N. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *J. Econom.* **2010**, *157*, 18–33. [\[CrossRef\]](#)
6. Su, L.J. Semiparametric GMM estimation of spatial autoregressive models. *J. Econom.* **2012**, *167*, 543–560. [\[CrossRef\]](#)
7. Malikov, E.; Sun, Y.G. Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. In *Working Paper*; University of Guelph: Guelph, ON, Canada, 2015.
8. Sun, Y.G. Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *J. Econom.* **2016**, *195*, 134–153. [\[CrossRef\]](#)
9. Du, J.; Sun, X.Q.; Cao, R.Y.; Zhang, Z.Z. Statistical inference for partially linear additive spatial autoregressive models. *Spat. Stat.* **2018**, *25*, 52–67. [\[CrossRef\]](#)
10. Arellano, M. *Panel Data Econometrics*; Oxford University Press: New York, NY, USA, 2003.
11. Baltagi, B.H. *Econometric Analysis of Panel Data*, 5th ed.; Wiley: New York, NY, USA, 2013.
12. Lee, L.F.; Yu, J.H. Estimation of spatial autoregressive panel data models with fixed effects. *J. Econom.* **2010**, *154*, 165–185. [\[CrossRef\]](#)

13. Zhang, Y.Q.; Shen, D.M. Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects. *J. Stat. Plan. Inference* **2015**, *159*, 64–80. [\[CrossRef\]](#)
14. Ai, C.R.; Zhang, Y.Q. Estimation of partially specified spatial panel data models with fixed-effects. *Econom. Rev.* **2017**, *36*, 6–22. [\[CrossRef\]](#)
15. Li, K.P. Fixed-effects dynamic spatial panel data models and impulse response analysis. *J. Econom.* **2017**, *198*, 102–121. [\[CrossRef\]](#)
16. Sun, Y.G.; Malikov, E. Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects. *J. Econom.* **2018**, *203*, 359–378. [\[CrossRef\]](#)
17. He, B.Q.; Hong, X.J.; Fan, G.L. Empirical likelihood for semi-varying coefficient models for panel data with fixed effects. *J. Korean Stat. Soc.* **2016**, *45*, 395–408. [\[CrossRef\]](#)
18. Li, D.G.; Chen, J.; Gao, J.T. Nonparametric time-varying coefficient panel data models with fixed effects. *Econom. J.* **2011**, *14*, 387–408. [\[CrossRef\]](#)
19. Rodriguez-Poo, J.M.; Soberon, A. Nonparametric estimation of fixed effects panel data varying coefficient models. *J. Multivar. Anal.* **2015**, *133*, 95–122. [\[CrossRef\]](#)
20. Feng, S.Y.; He, W.Q.; Li, F. Model detection and estimation for varying coefficient panel data models with fixed effects. *Comput. Stat. Data Anal.* **2020**, *152*, 107054. [\[CrossRef\]](#)
21. Kelejian, H.H.; Prucha, I.R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Financ. Econ.* **1998**, *17*, 99–121. [\[CrossRef\]](#)
22. Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249. [\[CrossRef\]](#)
23. Xue, L.G.; Zhu, L.X. Empirical likelihood-based inference in a partially linear model for longitudinal data. *Sci. China Ser. A Math.* **2008**, *51*, 115–130. [\[CrossRef\]](#)
24. Li, G.R.; Lin, L.; Zhu, L.X. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. *J. Multivar. Anal.* **2012**, *105*, 85–111. [\[CrossRef\]](#)
25. Schumaker, L.L. *Spline Functions*; Wiley: New York, NY, USA, 1981.
26. Allen, J.; Gregory, A.W.; Shimotsu, K. Empirical likelihood block bootstrapping. *J. Econom.* **2011**, *161*, 110–121. [\[CrossRef\]](#)
27. Stone, C.J. Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* **1982**, *10*, 1348–1360. [\[CrossRef\]](#)
28. Case, A.C. Spatial patterns in household demand. *Econometrica* **1991**, *59*, 953–965. [\[CrossRef\]](#)
29. Munnell, A.H. Why has productivity growth declined? Productivity and public investment. *N. Engl. Econ. Rev.* **1990**, *30*, 3–22.
30. Baltagi, B.H.; Pinnoi, N. Public capital stock and state productivity growth: Further evidence from an error components model. *Empir. Econ.* **1995**, *20*, 351–359. [\[CrossRef\]](#)
31. Feng, S.Y.; Li, G.R.; Tong, T.J.; Luo, S.H. Testing for heteroskedasticity in two-way fixed effects panel data models. *J. Appl. Stat.* **2020**, *47*, 91–116. [\[CrossRef\]](#)
32. Xu, Y.H.; Yang, Z.L. Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Reg. Sci. Urban Econ.* **2020**, *81*, 103488. [\[CrossRef\]](#)
33. Zhang, R.Q.; Lu, Y.Q. *Varying-Coefficient Model*; Science Press: Beijing, China, 2004.
34. Zhao, P.X.; Xue, L.G. Empirical likelihood inferences for semiparametric varying-coefficient partially linear models with longitudinal data. *Commun. Stat. Theory Methods* **2010**, *39*, 1898–1914. [\[CrossRef\]](#)
35. Cai, Z.W.; Xiong, H.Y. Partially varying coefficient instrumental variables models. *Stat. Neerl.* **2012**, *66*, 85–110. [\[CrossRef\]](#)
36. Wilkinson, J.H. *The Algebraic Eigenvalue Problem*; Clarendon: Oxford, UK, 1965.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.