

Article

Profile Maximum Likelihood Estimation of Single-Index Spatial Dynamic Panel Data Model

Mengqi Zhang and Boping Tian * 

Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China; 17b912034@stu.hit.edu.cn

* Correspondence: bopingt361147@hit.edu.cn

Abstract: In this paper, the spatial dynamic panel data (SDPD) model is extended to the single-index spatial dynamic panel data (Si-SDPD) model by introducing a nonlinear connection function to reflect the interaction between explanatory variables. The Si-SDPD model not only retains the advantages of the parametric SDPD model in dealing with spatial and temporal interaction effects and spatio-temporal dependencies, but also solves the limitations of the parametric SDPD model that may lead to missed bias. It reduces the data dimension of non-parametric models and enhances the practicability and explanatory power of parametric models. Since the parts of the model to be estimated contain unknown functions, we propose a new estimation method, a profile maximum likelihood (PML) method, to solve the problem of incidental parameters in the estimation. Under the assumption that the spatial coefficients are known, we preliminarily estimate the unknown function by carrying out local polynomial estimation, so as to transform the model into the parametric form for solving purposes. We then solve the dynamic panel parametric model via quasi-maximum likelihood (QML) estimation. We derive the asymptotic properties of profile maximum likelihood estimators (PMLEs) and find that, under certain regularity conditions, both parametric and non-parametric estimators are consistent. Monte Carlo results show that PMLEs have good finite sample performance.

Keywords: profile maximum likelihood; nonparametric estimation; spatial dynamic panel data; single-index panel model

MSC: 62F12; 62G20; 62H11



Citation: Zhang, M.; Tian, B. Profile Maximum Likelihood Estimation of Single-Index Spatial Dynamic Panel Data Model. *Mathematics* **2023**, *11*, 2947. <https://doi.org/10.3390/math11132947>

Academic Editor: María del Carmen Valls Martínez

Received: 6 June 2023

Revised: 22 June 2023

Accepted: 26 June 2023

Published: 1 July 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

Research into spatial panel data is always a hot topic in econometrics. Scholars have gradually applied the spatial panel data model to analyze social and economic issues, such as the effective allocation of resources, factors influencing economic growth, energy consumption, environmental pollution, foreign direct investments, and so on. Compared with time series or cross-section data, panel data can increase the degree of freedom and reduce collinearity, thereby improving the accuracy of parameter estimation. Spatial panel data models can be divided into parametric and non-parametric models according to different assumptions about the explanatory variables. The parametric panel data model can simply and clearly describe the relationship between dependent and independent variables. Refs. [1–15] and others have made outstanding contributions to parametric panel data modelling and its inference of economic phenomena. However, in many practical problems, the relationship between variables is usually complex, and it is often difficult to accurately describe this relationship using the preset form of the parametric model. In order to solve this problem, some non-parametric panel data models have been widely studied. Because the function form of a non-parametric model can be arbitrary, the model has great adaptability. When there are many explanatory variables in non-parametric models, the so-called “curse of dimensionality” problem often occurs, leading to the reduced reliability

of estimates. Therefore, some new models have been proposed, and one of the effective methods is to build a single-index model.

The basic mathematical form of the single-index model is $Y = g(x'\beta) + \varepsilon$ (Ref. [16]), where β is the unknown parameter and $g(\cdot)$ is the unknown connection function $g(\cdot)$. Its advantages are that it achieves dimension reduction by connecting functions, effectively avoids the problem of “curse of dimensionality”, and (at the same time) better reflects the relationship between variables. In recent years, the research focus on various extended single-index models has been to seek effective estimates of β and $g(\cdot)$. For the single-index model of cross-section data, Ref. [17] proposed a semi-parametric maximum likelihood estimation method, while Ref. [18] proposed non-iterative methods for estimating β and $g(\cdot)$ based on the parameter constraint assumption, which improves the operation speed. Ref. [19] proposed root-n-consistent estimators for the variance components and a local linear smoother for the connection function. Ref. [20] constructed a semi-parametric minimum average variance estimation method in a partially linear single-index panel data model with fixed effects, and demonstrated the asymptotic properties of the estimation. Compared with the cross-sectional data model, there are very few studies on the single-index model for panel data, and even fewer on the single-index spatial panel data model because of its greater estimation complexity.

According to these research trends of spatial econometric models, we design a Si-SDPD model by introducing a nonlinear connection function to reflect the interaction between explanatory variables. The Si-SDPD model not only retains the advantages of the parametric SDPD model in dealing with spatial and temporal interaction effects and spatio-temporal dependencies, but also solves the limitations of the parametric SDPD model that may lead to missed bias. It reduces the data dimension of non-parametric models and enhances the practicability and explanatory power of parametric models. It is worth noting that due to the presence of a spatial lag term and the multi-directionality of the spatial correlation, the traditional estimation of the single-index model may not be directly applied to estimate the Si-SDPD model. In order to overcome the above difficulties, we propose a new estimation method, namely PML, which can first estimate the unknown functions via local polynomial estimation under the assumption that the spatial coefficients are known, and then solves the dynamic panel parametric model by conducting QML estimation. Then, we derive the asymptotic properties of the PMLEs of the Si-SDPD model and find that under certain regularity conditions, both the parametric and non-parametric estimators are consistent. Finally, we present a rigorous theoretical analysis of the asymptotic properties of PMLEs and verify some of their finite-sample properties by carrying out Monte Carlo experiments.

This paper is organized as follows. In Section 2, we introduce the Si-SDPD model and explain our PML estimation method. With the law of large numbers and central limit theorem for our settings developed in the Appendix A, Section 3 establishes the consistency of the parametric and nonparametric parts of the PMLE. We then present a Monte Carlo simulation to verify that the estimators have good finite sample performance. Section 5 concludes the paper. Some useful lemmas and results are provided in the Appendix.

2. The Model and Profile Maximum Likelihood Estimators

2.1. The Model

The model considered in this paper refers to the single-index spatial dynamic panel data (Si-SDPD) model:

$$Y_{nt} = \rho_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \tau_0 W_n Y_{n,t-1} + G_{nt} + V_{nt}, \quad t = 1, 2, \dots, T \quad (1)$$

where $Y_{nt} = (y_{1t}, y_{2t}, \dots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \dots, v_{nt})'$ are $n \times 1$ column vectors; v_{nt} is independent and identically distributed across i and t with a zero mean and variance σ_0^2 ; W_n is the $n \times n$ spatial weight matrix, which is predetermined and generates spatial dependence between cross-sectional units y_{it} ; $G_{nt} = (g(x_1'\beta), g(x_2'\beta), \dots, g(x_n'\beta))', g(\cdot)$

is an unary unknown function, $\|\beta\| = 1$; and the first component of β is positive and is an identifiable condition.

Define $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1})$, $A_n(\rho) = I - \rho W_n$, $\theta = (\beta, \delta', \rho, \sigma^2)'$ and $\zeta = (\beta, \delta', \rho)'$, where $\delta = (\gamma, \tau)'$. At the true value, $A_n = A_n(\rho_0) = I - \rho_0 W_n$, $\theta_0 = (\beta_0, \delta'_0, \rho_0, \sigma_0^2)'$, $\zeta_0 = (\beta_0, \delta'_0, \rho_0)'$, where $\delta_0 = (\gamma_0, \tau_0)'$. Then, presuming A_n is invertible and (1) can be rewritten as $Y_{nt} = A_n^{-1}(Z_{nt}\delta_0 + G_{nt} + V_{nt})$. The likelihood function of (1) is

$$\ln L_{n,T}(\theta, G_{nt}) = -\frac{nT}{2} \ln(2\pi\sigma^2) + T \ln|A_n(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^T V'_{nt}(\zeta) V_{nt}(\zeta) \quad (2)$$

where $V_{nt}(\zeta) = A_n(\rho)Y_{nt} - Z_{nt}\delta - G_{nt}$ and $V_{nt}(\zeta) = (v_1, v_2, \dots, v_n)'$. Thus, $V_{nt} = V_{nt}(\zeta_0)$.

The QMLE $\hat{\theta}_{nT}$ is the extreme estimator derived from the maximization of (2). When the V_{nt} s are normally distributed, $\hat{\theta}_{nT}$ is the MLE; when the V_{nt} s are not normally distributed, $\hat{\theta}_{nT}$ is the QMLE.

2.2. The Profile Maximum Likelihood Estimation

For the likelihood function shown in (2), the parameter estimation method is not feasible because $g(x'_i\beta)$, $i = 1, \dots, n$ is unknown. In order to obtain a feasible estimate, we propose adopting the PML method. First, we consider the parameter $\theta = (\beta, \delta', \rho, \sigma^2)'$ as known, and then (1) becomes a general spatial nonparametric model. The initial estimate $\hat{g}_{IN}(x'_i\beta)$ of $g(\cdot)$ can be obtained using the kernel estimation. Obviously, $\hat{g}_{IN}(x'_i\beta)$ is a function of the parameter θ . By replacing $g(\cdot)$ in (2) with $\hat{g}_{IN}(x'_i\beta)$, we obtain the likelihood function with parameter θ . Then, by maximizing the likelihood function, we obtain the estimator $\hat{\theta} = (\hat{\beta}, \hat{\delta}', \hat{\rho}, \hat{\sigma}^2)'$ of θ . Finally, the final estimate of $g(\cdot)$, $\hat{g}(x'_i\beta)$, is obtained by replacing θ in $\hat{g}_{IN}(x'_i\beta)$ with $\hat{\theta}$.

The specific steps of the profile maximum likelihood method are as follows.

Step 1: Considering $\theta = (\beta, \delta', \rho, \sigma^2)'$ as known, we obtain $\hat{g}_{IN}(x'_i\beta)$ that is an initial estimate of $g(\cdot)$ using local polynomial estimation.

We denote $Y_t = A_n(\rho)Y_{nt} - Z_{nt}\delta = G_{nt} + V_{nt}(\zeta)$ and $Y_t = (y_1, y_2, \dots, y_n)'$, and (1) can be written as $y_i = g(x'_i\beta) + v_i = g(u_i) + v_i$, $i = 1, \dots, n$. At u , the p -order Taylor expansion of $g(u_i)$ is:

$$\begin{aligned} g(u_i) &\approx g(u) + g'(u)(u_i - u) + \dots + \frac{1}{p!}g^{(p)}(u)(u_i - u)^p \\ &\triangleq \lambda_0 + \lambda_1(u_i - u) + \dots + \lambda_p(u_i - u)^p \dots \end{aligned} \quad (3)$$

Therefore, we can use the samples near u to perform weighted regression to estimate $g(u_i)$ and its higher-order derivatives, i.e., solving the following minimization problem:

$$\min_{\{\lambda_0, \lambda_1\}} \frac{1}{n} \sum_{i=1}^n \{y_i - \lambda_0 - \lambda_1(u_i - u)\}^2 k_h(u_i - u) \quad (4)$$

where $k_h(u_i - u) = h^{-p}k\left(\frac{u_i - u}{h}\right)$, $i = 1, \dots, n$ is the multivariate kernel function and h is the bandwidth. In order to simplify the theoretical derivation, all variables in this chapter have the same window width, and the conclusion is also valid under the assumption of different window widths.

For the convenience of the following matrix operations, we denote $p = n - 1$. Then, we denote $M(u, \beta) = \begin{pmatrix} 1 & \cdots & 1 \\ \frac{u_1-u}{h} & \cdots & \frac{u_n-u}{h} \end{pmatrix}'$, $\lambda = (\lambda_0, \lambda_1)'$ and $\kappa(u, \beta) = \text{diag}\left(k\left(\frac{u_1-u}{h}\right), \cdots, k\left(\frac{u_n-u}{h}\right)\right)$, so $M(u, \beta)$ is a $2 \times n$ matrix. We can rewrite the objective function (4) as:

$$\min_{\lambda} \frac{1}{n} (Y_t - M(u, \beta)\lambda)' \kappa(u, \beta) (Y_t - M(u, \beta)\lambda)$$

After minimizing, the estimator of λ is:

$$\hat{\lambda} = \left(M(u, \beta)' \kappa(u, \beta) M(u, \beta) \right)^{-1} M(u, \beta)' \kappa(u, \beta) Y_t$$

The first component of $\hat{\lambda}$ is an estimate of $g(x'_i\beta)$. We denote $\left(M(u, \beta)' \kappa(u, \beta) M(u, \beta) \right)^{-1} M(u, \beta)' \kappa(u, \beta) = S(u, \beta)$, $s(u, \beta) = e_1 S(u, \beta)$, and $s(u, \beta) = e_1 S(u, \beta)$, and the initial estimate of $g(x'_i\beta)$ is

$$\hat{g}_{IN}(u) = e_1 \hat{\lambda} = e_1 S(u, \beta) Y_t = s(u, \beta) Y_t \quad (5)$$

where $e_1 = (1, 0, \cdots, 0)$. Then, the initial estimate of G_{nt} is

$$\begin{aligned} \hat{G}_{IN} &= (\hat{g}_{IN}(u_1), \hat{g}_{IN}(u_2), \cdots, \hat{g}_{IN}(u_n))' \\ &= (s(u_1)Y_t, s(u_2)Y_t, \cdots, s(u_n)Y_t)' \\ &= \left(s(u_1)', s(u_2)', \cdots, s(u_n)' \right)' Y_t \\ &= S(\beta) Y_t \end{aligned} \quad (6)$$

where $S(\beta) = \left(s(x'_1\beta)', s(x'_2\beta)', \cdots, s(x'_n\beta)' \right)'$.

Step 2: By substituting $\hat{G}_{IN}$ for G_{nt} in (1), we obtain the approximate value of the logarithmic likelihood function as:

$$\begin{aligned} \ln L_{n,T}(\theta) &= -\frac{nT}{2} \ln(2\pi\sigma^2) + T \ln|A_n(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^T V'_{nt}(\zeta) V_{nt}(\zeta) \\ &= -\frac{nT}{2} \ln(2\pi\sigma^2) + T \ln|A_n(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^T (Y_t - \hat{G}_{IN})' (Y_t - \hat{G}_{IN}) \\ &= -\frac{nT}{2} \ln(2\pi\sigma^2) + T \ln|A_n(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^T [(I_n - S(\beta))Y_t]' [(I_n - S(\beta))Y_t] \end{aligned} \quad (7)$$

The $\hat{\theta}_{nT}$ that can maximize the above formula is the estimate of θ , i.e.,

$$\hat{\theta}_{nT} = \arg \max_{\theta} \frac{1}{nT} \ln L_{n,T}(\theta)$$

Computationally and analytically, it is convenient to work with the concentrated log-likelihood by concentrating out the σ^2 . Based on the log-likelihood function, the initial estimate of σ^2 is

$$\begin{aligned} \hat{\sigma}_{IN}^2 &= \frac{1}{nT} \sum_{t=1}^T [(I_n - S(\beta))Y_t]' [(I_n - S(\beta))Y_t] \\ &= \frac{1}{nT} \sum_{t=1}^T [(I_n - S(\beta))(A_n(\rho)Y_{nt} - Z_{nt}\delta)]' [(I_n - S(\beta))(A_n(\rho)Y_{nt} - Z_{nt}\delta)] \\ &= \frac{1}{nT} \sum_{t=1}^T [(I_n - S(\beta))(\hat{G}_{IN} + V_{nt}(\zeta))]' [(I_n - S(\beta))(\hat{G}_{IN} + V_{nt}(\zeta))] \end{aligned} \quad (8)$$

The concentrated log-likelihood function of ζ is

$$\ln L_{n,T}(\zeta) = -\frac{nT}{2} \ln(2\pi + 1) - \frac{nT}{2} \ln \hat{\sigma}_{IN}^2 + T \ln|A_n(\rho)| \quad (9)$$

The estimate of ζ can be defined as:

$$\hat{\zeta}_{nT} = \arg \max_{\theta} \frac{1}{nT} \ln L_{n,T}(\zeta) \quad (10)$$

Equation (10) displays a nonlinear optimization problem, which can be solved using an iterative method in the actual estimation. After $\hat{\zeta}_{nT}$ is obtained, the final estimate of σ^2 can be obtained by replacing ζ in Equation (8) with $\hat{\zeta}_{nT}$:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{nT} \sum_{t=1}^T [(I_n - S(\hat{\beta})) Y_t]' [(I_n - S(\hat{\beta})) Y_t] \\ &= \frac{1}{nT} \sum_{t=1}^T [(I_n - S(\hat{\beta})) (\hat{G}_{IN} + V_{nt}(\hat{\zeta}_{nT}))]' [(I_n - S(\hat{\beta})) (\hat{G}_{IN} + V_{nt}(\hat{\zeta}_{nT}))] \end{aligned} \quad (11)$$

Step 3: By using $\hat{\theta} = (\hat{\beta}, \hat{\delta}', \hat{\rho}, \hat{\sigma}^2)'$ obtained in Step 2 to replace the parameters in the model, we describe the final estimate of the non-parametric part $g(x_i'\beta)$ as:

$$\hat{g}(u) = s(u, \hat{\beta}) [A_n(\hat{\rho}) Y_{nt} - Z_{nt} \hat{\delta}] = s(u, \hat{\beta}) \hat{Y}_t$$

3. Profile Likelihood Estimators and Their Asymptotic Properties

To analyze the asymptotic properties of the estimators, we need the following assumptions:

Assumption 1. W_n is a constant spatial weight matrix and its diagonal elements satisfy $w_{n,ii} = 0$ for $i = 1, 2, \dots, n$. In addition, W_n is uniformly bounded in the row and column sums in an absolute value (for short, UB).

Assumption 2. The disturbances $\{v_{it}\}$, $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$, are i.i.d. across i and t with a zero mean, a variance of σ^2 , and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 3. $A_n(\rho)$ is invertible for all $\rho \in \Lambda_\rho$, where Λ_ρ is compact and ρ_0 is in the interior of Λ_ρ . Furthermore, $A_n^{-1}(\rho)$ is UB.

Assumption 4. $\{x_i\}_{i=1}^n$ is an independent and identically distributed random sequence and $E(v_{it}x_i') < \infty$, $i = 1, \dots, n$. $x_i'\beta$ has marginal density functions $f(u_i)$, where $f(u_i)$ is continuously differentiable near $x_i'\beta \in \Lambda_u$ and Λ_u is the support set of $k(\cdot)$.

Assumption 5. $g(\cdot)$ has continuous derivatives $g^{(p)}(\cdot)$, $p = 1, \dots, n$ and $|g(\cdot)| \leq \omega_g$, where ω_g is a positive constant.

Assumption 6. When $n \rightarrow \infty$, $h \rightarrow 0$ and $nh^p \rightarrow \infty$, $p = 1, \dots, n$.

Assumption 7. The kernel function $k(\cdot)$ is a bounded continuous non-negative function whose support set is bounded and closed: $[-\omega_k, \omega_k] \subset \mathbb{R}^p$, where $\omega_k > 0$ is a constant, i.e., $k(m) > 0$ only if $|m| \leq \omega_k$. In addition, $M(u, \beta)$, $\kappa(u, \beta)$, and $(M(u, \beta)' \kappa(u, \beta) M(u, \beta))^{-1}$ are UB.

Assumption 8. $k(\cdot)$ is an even function and $\int k(m) m^l dm = \omega_l$, $\int k^2(m) m^l dm = \varsigma_l$. For any positive odd number l , $\omega_l = \varsigma_l = 0$; also, $\omega_1 = 0$, $\omega_2 \neq 0$.

Assumption 9. β_0 is the inner point of D , where $D \in \mathbb{R}^p$ is a convex compact set. $\|\beta_0 = 1\|$ and the first component of vector quantity β_0 is positive, where $\|\cdot\|$ is the Euclidean norm.

Assumption 1 is a standard normalization assumption in spatial econometrics. Here, the matrix $A = (A_{ij})$ is UB, meaning that there is a non-negative constant ω_A such that $\sum_{i=1}^n A_{ij} \leq \omega_A$ and $\sum_{j=1}^n A_{ij} \leq \omega_A$. The constant ω_A is defined differently in the following sections depending on the matrices. Assumption 2 provides regularity assumptions for v_{it} . The reversibility and compactness of $A_n(\rho)$ in Assumption 3 were derived from Kelejian and Prucha (1998, 2001) and have also been used in many articles about spatial correlation. When exogenous variables u_i are included in the model, it is convenient to assume that the exogenous regressors are uniformly bounded, as in Assumption 4. Assumption 5 is a

necessary condition for (3). Assumptions 6–8 demonstrate the condition of kernel density estimation. The bandwidth of the kernel function, h , is an important parameter that affects the estimation result of the kernel function. Kernel functions that satisfy Assumptions 7–8 exist, such as the product kernel, $k(m) = \prod_{i=1}^p k(m_i)$, where $k(m_i)$ is a symmetric kernel of one variable on the closed interval $[-\omega_k, \omega_k]$. Assumption 9 makes the model (1) recognizable.

For the concentrated log-likelihood Function (4) divided by the sample size nT , the corresponding expected value function is $Q_{n,T}(\theta) = \max_{\sigma^2} E \left[\frac{1}{nT} \ln L_{n,T}(\theta) \right]$, which is

$$Q_{n,T}(\theta) = \frac{1}{nT} E[\ln L_{n,T}(\theta)] \\ = -\frac{1}{2} \ln(2\pi + 1) + \frac{1}{n} \ln |A_n(\rho)| - \frac{1}{2\sigma^2} E \left[\frac{1}{nT} \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) \right] \quad (12)$$

To show the consistency of $\hat{\theta}_{nT}$, we need the following uniform convergence results.

Lemma 1. Under Assumptions 1–9, for an $n \times n$ nonstochastic UB matrix B_n ,

$$\begin{aligned} \frac{1}{nT} \sum_{t=1}^T V_{nt}' B_n V_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T V_{nt}' B_n V_{nt} \right) &= O_p \left(\frac{1}{\sqrt{nT}} \right) \\ \frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n Z_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n Z_{nt} \right) &= O_p \left(\frac{1}{\sqrt{nT}} \right) \\ \frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n V_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n V_{nt} \right) &= O_p \left(\frac{1}{\sqrt{nT}} \right) \end{aligned} \quad (13)$$

where $E \left(\frac{1}{nT} \sum_{t=1}^T V_{nt}' B_n V_{nt} \right) = \frac{1}{n} \sigma_0^2 \text{tr}(B_n) = O(1)$, $E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n Z_{nt} \right) = O(1)$, and $E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' B_n V_{nt} \right) = O \left(\frac{1}{T} \right)$.

Lemma 2. Let Θ be any compact parameter space. Then, under Assumptions 1–9, $\frac{1}{nT} \ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{P} 0$ is uniform in $\theta \in \Theta$.

Lemma 3. Let Θ be any compact parameter space. Then, under Assumptions 1–8, $Q_{n,T}(\zeta)$ is uniformly equicontinuous for $\theta \in \Theta$.

Before obtaining the information matrix, we need to compute the first and second derivatives of the logarithmic likelihood function. The asymptotic distribution of the QMLE $\hat{\theta}_{nT}$ can be derived from the Taylor expansion of $\frac{\partial \ln L_{n,T}(\hat{\theta}_{nT})}{\partial \theta}$ around θ_0 . We define $\theta^\beta = \theta(\hat{\beta}) = (\delta', \rho, \sigma^2)'$ and $\zeta^\beta = \zeta(\hat{\beta}) = (\delta', \rho)'$ at the true value $\theta_0^\beta = \theta_0(\hat{\beta}) = (\delta_0', \rho_0, \sigma_0^2)'$ and $\zeta_0^\beta = \zeta_0(\hat{\beta}) = (\delta_0', \rho_0)'$. The first-order derivative of the concentrated likelihood function involves both linear and quadratic functions of V_{nt} , which are as follows:

$$\begin{aligned} & \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}(\theta_0^\beta)}{\partial \theta^\beta} \\ &= \frac{1}{\sqrt{nT}} \left[\frac{\partial \ln L_{n,T}(\theta_0^\beta)}{\partial \delta} \quad \frac{\partial \ln L_{n,T}(\theta_0^\beta)}{\partial \rho} \quad \frac{\partial \ln L_{n,T}(\theta_0^\beta)}{\partial \sigma^2} \right] \\ &= \frac{1}{\sqrt{nT}} \begin{bmatrix} \frac{1}{\sigma_0^2} \sum_{t=1}^T Z_{nt}' V_{nt} \\ \frac{1}{\sigma_0^2} \sum_{t=1}^T [F_{nt} Z_{nt} \delta_0]' V_{nt} - \sum_{t=1}^T [V_{nt}' F_{nt} V_{nt} - \sigma_0^2 \text{tr}(F_{nt})] \\ \frac{1}{2\sigma_0^4} \left(\sum_{t=1}^T V_{nt}' V_{nt} - nT \sigma_0^2 \right) \end{bmatrix}^T \quad (14) \end{aligned}$$

where $Z_{ns} = [I_n - S(\hat{\beta})]Z_{nt}$ and $Y_{ns} = [I_n - S(\hat{\beta})]W_n Y_{nt}$. Then, the second-order derivatives are:

$$\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}(\theta^\beta)}{\partial \theta^\beta \partial \theta^{\beta'}} = -\frac{1}{nT} \times \begin{bmatrix} \frac{1}{\sigma^2} \sum_{t=1}^T Z'_{ns} Z_{ns} & \frac{1}{\sigma^2} \sum_{t=1}^T Z'_{ns} Y_{ns} & \frac{1}{\sigma^4} \sum_{t=1}^T Z'_{ns} V_{nt}(\zeta^\beta) \\ * & \frac{1}{\sigma^2} \sum_{t=1}^T Y_{ns}' Y_{ns} - T \text{tr}(F_n^2) & \frac{1}{\sigma^4} \sum_{t=1}^T Y_{ns}' V_{nt}(\zeta^\beta) \\ * & * & -\frac{nT}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{t=1}^T V_{nt}(\zeta^\beta)' V_{nt}(\zeta^\beta) \end{bmatrix} \quad (15)$$

The information matrix is as follows:

$$\Sigma_{\theta_0^\beta, nT} = -E \left(\frac{1}{nT} \frac{\partial^2 \ln L_{n,T}(\theta_0^\beta)}{\partial \theta^\beta \partial \theta^{\beta'}} \right) = \begin{bmatrix} \frac{1}{nT} \frac{1}{\sigma_0^2} \sum_{t=1}^T Z'_{ns} Z_{ns} & \frac{1}{nT} \frac{1}{\sigma_0^2} \sum_{t=1}^T Z'_{ns} R_{ns} & 0 \\ * & \frac{1}{nT} \frac{1}{\sigma_0^2} E \sum_{t=1}^T R'_{ns} R_{ns} + \frac{1}{n} [\text{tr}(F_n^2) + \text{tr}(F'_{ns} F_{ns})] & \frac{1}{n\sigma_0^2} \text{tr}[F_{ns}] \\ * & * & \frac{1}{2\sigma_0^4} \end{bmatrix}$$

where $F_{ns} = [I_n - S(\hat{\beta})]F_n$ and $R_{ns} = F_{ns}(Z_{nt}\delta_0 + \hat{G}_{IN})$.

Assumption 10. $\lim_{n \rightarrow \infty} \frac{1}{n} [\text{tr}(F_n^2) + \text{tr}(F'_{ns} F_{ns}) - \frac{2}{n} (\text{tr} F_{ns})^2] \neq 0$.

Assumption 10 is an important condition for the non-singularity of the limiting information matrix $\Sigma_{\theta_0^\beta, nT}$ in addition to the global identification in Lemma 4 and Theorem 1.

Lemma 4. The information matrix $\Sigma_{\theta_0^\beta, nT}$ is non-singular.

Theorem 1. Under Assumptions 1–10, θ_0 is globally identifiable and $\hat{\theta}$ is a consistent estimator of θ_0 (similar to Yu (2008)).

Theorem 2. Under Assumptions 1–10, θ_0 is globally identifiable and $\hat{\theta}_{nT} \xrightarrow{P} \theta_0$ if $\lim_{n \rightarrow \infty} \left(\frac{1}{n} \ln |\sigma_0^2 A_n^{-1} A_n^{-1}| - \frac{1}{n} \ln |\sigma_n^2(\rho) A_n^{-1}(\rho) A_n^{-1}(\rho)| \right) \neq 0$ for $\rho \neq \rho_0$ (similar to Yu (2008)).

Lemma 5. Under Assumptions 1–10, $\frac{1}{n} M(u, \beta)' \kappa(u, \beta) M(u, \beta) \xrightarrow{P} f(x) \begin{pmatrix} 1 & 0 \\ 0 & \omega_2 I_n \end{pmatrix}$.

Lemma 6. Under Assumptions 1–10, $S(\beta) V_{nt} = o_p(1)$.

Lemma 7. Under Assumptions 1–9, $(I_n - S) G_{nt} = o_p(1)$.

Theorem 3. Under Assumptions 1–9, $\hat{G}_{nt} \xrightarrow{P} G_{nt}$.

4. Monte Carlo Results

In this section, the Monte Carlo experiment is carried out on the estimation method pertaining to the previously constructed Si-SDPD model, and the simulation results are evaluated to test the performance of the PML method under limited samples. Then, the practical application value of the PMLEs is evaluated. All experiments are compiled using R language and plotted using the 'ggplot2' package.

For the parametric part, we generate samples from (1) and use $\theta_1^\beta = (-0.3, 0.4, 1, -0.2)'$ and $\theta_2^\beta = (0.25, -0.6, 0.5, 0.1)'$, where $\theta_0^\beta = (\gamma_0, \rho_0, \sigma_0^2, \tau_0)'$. The component of the two-dimensional random variable $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{pmatrix}' = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}'$ and the random error term $V_{nt} = (v_{1t}, v_{2t}, \cdots, v_{nt})'$ are generated from the uniform distribution $U[-3, 3]$ and the independent normal distribution $N(0, \sigma_0^2)$, respectively. The spatial weight matrix W_n that we use is the Rook matrix, which is one of the main types of spatial weight matrix in spatial econometrics. For the non-parametric part, we use the commonly used Gaussian kernel function, $g(X) = 2e^{-(\beta_1 X_1 + \beta_2 X_2)^2}$, where $(\beta_1, \beta_2) = (\sqrt{1/3}, \sqrt{2/3}) \approx (0.5774, 0.8165)$. As it is difficult to select the optimal window width, we simply select the window width using the rule-of-thumb method. Finally, we use $n = 10, 49, 100$ as the sample size and $T = 4, 10$ as the number of periods. For each set of n and T , the sampling observations are generated with the Metropolis–Hastings sampling algorithm.

The evaluation of simulation results should also be divided into parametric and non-parametric parts. In the parametric part, for each estimator, we calculate the standard deviation (Std) and the root-mean-squared error (RMSE), where $RMSE = \left[\frac{1}{n} \sum_{i=1}^{reps} (\hat{\theta}_i - \theta_0)^2 \right]^{1/2}$, $reps$ is the number of simulations and $\hat{\theta}_i$, $i = 1, \cdots, reps$ are the parameter estimates obtained from each simulation. In order to accurately estimate the parameter values, according to Su (2012), we take the window width $h = std(x) \cdot n^{1/5}$ here, where $std(x)$ represents the standard deviation of sequence x . In the non-parametric part, we refer to Chen (2012) when choosing the mean absolute deviation error (MADE) as the evaluation standard, which is $MADE = \frac{1}{M} \sum_{j=1}^M |\hat{g}_j(x_m) - g_j(x_m)|$, $j = 1, \cdots, reps$, where $\{x_m\}_{m=1}^M$ is the M fixed grid points selected within the support set of x . We select 20 fixed lattice points in $(-2, 2)$, namely $M = 20$. When estimating the non-parametric part, we use the method of leave-one-out cross-validation to select the window width, i.e., the window width minimizes $\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\hat{y}_i - \hat{g}_{-i})^2$, where $\hat{y}_i$ is the i th element of the $\hat{Y}_t = A_n(\hat{\rho})Y_{nt} - Z_{nt}\hat{\delta}$ after the estimated value $\hat{\theta}$ and $\hat{g}_{-i}$ is the $g(x_i)$ estimate obtained with the observation value other than the i th observation.

For different cases of n and T , 100 simulations are carried out with R Language. In each simulation, the Metropolis–Hastings sampling algorithm is used to conduct 1000 samples in the PML function. In order to obtain the distribution of samples close to reality and in order to achieve stability in the state, the first 200 sampling results are discarded. With two different values of θ_0 for each n and T , finite sample properties of both estimators are summarized in Tables 1 and 2, in which we report the means, variances (Vars), root mean square error (RMSE), and coverage probability (CP).

For each case, the estimated value of the parameter, i.e., the mean, is relatively close to the real value, and we can see that for each given n , when T is larger, the variance of estimators will be smaller; for each given T , when n is larger, the biases between the real value and the estimators will be nearly the same, but the variance will be smaller. When both n and T are maximized, i.e., $n = 100$, $T = 10$, the variance and RMSEs of the parameter estimators are the smallest in all cases, which indicates that the parameter estimators will converge with the increase in the n value, which is consistent with the large sample property, as demonstrated. For different values of θ_0 , the variances are almost all less than 0.01, which indicates that the fitting error is small and the fitting results are good. In addition, the variance and root mean square error have little changes and will not change with an increase in the simulation time and the sample size, which further indicates that the variance of parameters is stable. Due to the small estimation variance, the mean hardly fluctuates around the true value, and the range of the confidence interval is also relatively stable. In only a few cases, the confidence intervals do not cover the true value, and with an increase in n and T , the coverage degree becomes higher and higher, i.e., the CP gradually approaches 1.

Table 1. The performance of spatial coefficients estimators with θ_1^β .

n		$T=4$				$T=10$			
		RMSE	Means	Vars	CP	RMSE	Means	Vars	CP
10	β_1	0.0337	0.562	0.0074	0.82	0.0323	0.584	0.0066	0.84
	β_2	0.0341	0.832	0.0072	0.81	0.0319	0.794	0.0069	0.85
	γ	0.0282	−0.319	0.0058	0.84	0.0254	−0.313	0.0052	0.86
	ρ	0.0317	0.419	0.0064	0.82	0.0301	0.416	0.0054	0.84
	σ^2	0.1088	1.021	0.0069	0.85	0.0867	1.019	0.0062	0.88
	τ	0.0246	−0.224	0.0057	0.84	0.0238	−0.218	0.0051	0.87
49	β_1	0.0312	0.586	0.0068	0.85	0.0298	0.571	0.0052	0.88
	β_2	0.0325	0.828	0.0067	0.83	0.0296	0.809	0.0053	0.89
	γ	0.0234	−0.315	0.0054	0.86	0.0222	−0.288	0.0049	0.89
	ρ	0.0241	0.413	0.0057	0.84	0.0225	0.389	0.0048	0.88
	σ^2	0.0834	0.988	0.0059	0.87	0.0676	1.008	0.0054	0.90
	τ	0.0235	−0.179	0.0053	0.87	0.0226	−0.211	0.0047	0.91
100	β_1	0.0298	0.569	0.0053	0.87	0.0282	0.572	0.0045	0.90
	β_2	0.0302	0.807	0.0055	0.87	0.0283	0.811	0.0049	0.91
	γ	0.0217	−0.309	0.0049	0.88	0.0205	−0.297	0.0043	0.92
	ρ	0.0225	0.391	0.0047	0.89	0.0208	0.405	0.0039	0.91
	σ^2	0.0649	1.009	0.0048	0.90	0.0521	0.994	0.0041	0.93
	τ	0.0221	−0.185	0.0047	0.89	0.0213	−0.192	0.0042	0.92

Table 2. The performance of spatial coefficients estimators with θ_2^β .

n		$T = 4$				$T = 10$			
		RMSE	Means	Vars	CP	RMSE	Means	Vars	CP
10	β_1	0.0406	0.564	0.0088	0.79	0.0395	0.569	0.0077	0.82
	β_2	0.0512	0.829	0.0091	0.81	0.0494	0.806	0.0086	0.83
	γ	0.0332	0.226	0.0076	0.83	0.0318	0.234	0.0069	0.85
	ρ	0.0343	−0.620	0.0081	0.82	0.0335	−0.617	0.0077	0.83
	σ^2	0.1039	0.531	0.0084	0.82	0.1021	0.526	0.0078	0.84
	τ	0.0342	0.122	0.0078	0.84	0.0329	0.115	0.0071	0.86
49	β_1	0.0372	0.586	0.0071	0.82	0.0368	0.571	0.0056	0.85
	β_2	0.0433	0.802	0.0074	0.84	0.0419	0.812	0.0063	0.85
	γ	0.0325	0.239	0.0068	0.87	0.0299	0.258	0.0057	0.90
	ρ	0.0301	−0.612	0.0068	0.84	0.0298	−0.592	0.0055	0.88
	σ^2	0.0914	0.518	0.0069	0.85	0.0867	0.489	0.0062	0.90
	τ	0.0329	0.114	0.0064	0.88	0.0314	0.089	0.0054	0.92
100	β_1	0.0355	0.581	0.0061	0.84	0.0316	0.573	0.0056	0.89
	β_2	0.0367	0.813	0.0057	0.87	0.0345	0.810	0.0051	0.91
	γ	0.0307	0.243	0.0052	0.90	0.0281	0.253	0.0049	0.92
	ρ	0.0281	−0.592	0.0053	0.87	0.0282	−0.604	0.0046	0.90
	σ^2	0.0823	0.512	0.0061	0.89	0.0606	0.494	0.0057	0.92
	τ	0.0298	0.092	0.0058	0.91	0.0289	0.105	0.0051	0.94

Table 3 shows the average absolute error and variance of unknown function estimators $g(X) = 2e^{-(\beta_1 X_1 + \beta_2 X_2)^2}$ under different samples. As shown in Table 3, with an increase in the sample size and total number of periods under the same parameter setting, both the value and estimation error of the unknown function at 20 fixed lattice points decrease, indicating that the estimation of the non-parametric part is convergent. This is consistent with the theoretical results of Theorem 3.

Table 3. The performance of unknown function estimators $\hat{G}_{nt}$.

n	$\theta_0^a = (-0.3, 0.4, 1, -0.2)'$				$\theta_0^b = (0.25, -0.6, 0.5, 0.1)'$			
	$T = 4$		$T = 10$		$T = 4$		$T = 10$	
	MADE	Std.dev	MADE	Std.dev	MADE	Std.dev	MADE	Std.dev
10	0.0406	0.0216	0.0344	0.0187	0.0397	0.0187	0.0321	0.0145
49	0.0314	0.0125	0.0229	0.0112	0.0252	0.0138	0.0176	0.0098
100	0.0198	0.0092	0.0106	0.0076	0.0142	0.0096	0.0094	0.0071

5. Conclusions

In this paper, we propose a Si-SDPD model that can overcome the “Curse of Dimensionality” problem and deal with spatial and temporal correlation. With this model, we construct a PML method when the traditional maximum likelihood estimation method is not applicable. The theoretical results show that, under certain regularity conditions, both parametric and nonparametric estimators are consistent. Numerical simulation results show that PMLEs have good small sample characteristics and estimation accuracy increases with an increase in the sample size and time periods. Our research results will enrich and improve the estimation methods of single-index panel data models in spatial econometrics, and provide a new research tool and perspective for the applied research of related disciplines.

It is a complicated problem to prove the asymptotic normality of PML estimation because the number of parameters and unknown variables in the Si-SDPD model is too large, and the representation of the covariance matrix will be very complicated. The cross-sectional heteroskedasticity (space-varying error variances) in the Si-SDPD model is another interesting extension to consider. Most estimators are generally inconsistent in the presence of an unknown form of heteroskedasticity in the disturbance term in the SDPD model, much less in the Si-SDPD model with more unknown parameters. These works would be much more challenging than the already quite challenging works presented in this paper, and will be the topics of our future research.

Author Contributions: Conceptualization, M.Z. and B.T.; methodology, M.Z. and B.T.; software, M.Z.; validation, M.Z. and B.T.; formal analysis, M.Z. and B.T.; investigation, M.Z. and B.T.; resources, M.Z.; data curation, M.Z.; writing—original draft preparation, M.Z.; writing—review and editing, M.Z. and B.T.; visualization, M.Z. and B.T.; supervision, B.T.; project administration, B.T.; funding acquisition, B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 91646106).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the requirements of related projects supported by the National Natural Science Foundation of China.

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

Appendix A. Some Basic Lemmas

Proof of Lemma 2. Based on $[I_n - S(\hat{\beta})]Y_t = V_{nt}(\zeta) = A_n(\rho)Y_{nt} - Z_{nt}\delta - \hat{G}_{IN}$, we have $V_{nt}(\zeta) = V_{nt} - Z_{nt}(\delta - \delta_0) - (\rho - \rho_0)W_n Y_{nt}$. Hence,

$$\begin{aligned}
 V_{nt}(\zeta)'V_{nt}(\zeta) &= V_{nt}'V_{nt} - 2(\delta - \delta_0)'Z_{nt}'V_{nt} - 2(\rho - \rho_0)(W_n Y_{nt})'V_{nt} + (\rho - \rho_0)^2(W_n Y_{nt})'(W_n Y_{nt}) \\
 &\quad + (\delta - \delta_0)'Z_{nt}'Z_{nt}(\delta - \delta_0) \\
 &\quad + 2(\rho - \rho_0)(W_n Y_{nt})'Z_{nt}(\delta - \delta_0)
 \end{aligned} \tag{A1}$$

where, using $W_n Y_{nt} = W_n A_n^{-1} \{Z_{nt} \delta + [I_n - S(\hat{\beta})]^{-1} V_{nt}\}$ and $F_n = W_n A_n^{-1}$,

$$\begin{aligned} [W_n Y_{nt}]' [W_n Y_{nt}] = & \delta_0' Z_{nt}' F_n' F_n Z_{nt} \delta_0 + \left\{ F_n [I_n - S(\hat{\beta})]^{-1} V_{nt} \right\}' \left\{ F_n [I_n - S(\hat{\beta})]^{-1} V_{nt} \right\} \\ & + 2\delta_0' Z_{nt}' F_n' \left\{ F_n [I_n - S(\hat{\beta})]^{-1} V_{nt} \right\} \end{aligned} \quad (A2)$$

Using Lemma 1,

$$\begin{aligned} \frac{1}{nT} \sum_{t=1}^T V_{nt}' V_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T V_{nt}' V_{nt} \right) & \xrightarrow{P} 0 \\ \frac{1}{nT} \sum_{t=1}^T Z_{nt}' Z_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' Z_{nt} \right) & \xrightarrow{P} 0 \\ \frac{1}{nT} \sum_{t=1}^T Z_{nt}' V_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T Z_{nt}' V_{nt} \right) & \xrightarrow{P} 0 \\ \frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' V_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' V_{nt} \right) & \xrightarrow{P} 0 \\ \frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' Z_{nt} - E \left(\frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' Z_{nt} \right) & \xrightarrow{P} 0 \\ \frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' [W_n Y_{nt}] - E \left(\frac{1}{nT} \sum_{t=1}^T [W_n Y_{nt}]' [W_n Y_{nt}] \right) & \xrightarrow{P} 0 \end{aligned} \quad (A3)$$

As δ and ρ are bounded in Θ , we have $\frac{1}{nT} \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) - \frac{1}{nT} E \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) \xrightarrow{P} 0$ that is uniform in θ and Θ . Using the fact that σ^2 is bounded away from zero in Θ and $Q_{n,T}(\theta) = \frac{1}{nT} E[\ln L_{n,T}(\theta)]$,

$$\begin{aligned} \frac{1}{nT} \ln L_{n,T}(\theta) - Q_{n,T}(\theta) &= \frac{1}{2\sigma^2} \left(\sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) - \frac{1}{nT} E \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) \right) \xrightarrow{P} 0 \end{aligned} \quad (A4)$$

uniformly in θ . $\square$

Proof of Lemma 3 (similar to Ref. [3]). Given $\hat{G}_{IN} = S(\hat{\beta}) Y_t$ and $V_{nt}(\zeta) = [I_n - S(\hat{\beta})] Y_t$, we have

$$\begin{aligned} V_{nt}(\zeta) &= A_n(\rho) A_n^{-1} (Z_{nt} \delta_0 + \hat{G}_{IN} + V_{nt}) - Z_{nt} \delta - \hat{G}_{IN} \\ &= A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta + A_n(\rho) A_n^{-1} V_{nt} + A_n(\rho) A_n^{-1} \hat{G}_{IN} - \hat{G}_{IN} \\ &= A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta + [A_n(\rho) A_n^{-1} - S(\hat{\beta})] [I_n - S(\hat{\beta})]^{-1} V_{nt} \end{aligned} \quad (A5)$$

And $Q_{n,T}(\theta) = \frac{1}{nT} E[\ln L_{n,T}(\theta)] = -\frac{1}{2} \ln(2\pi + 1) + \frac{1}{n} \ln |A_n(\rho)| - \frac{1}{2\sigma^2} E \frac{1}{nT} \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta)$, where

$$\begin{aligned} E \frac{1}{nT} \sum_{t=1}^T V_{nt}'(\zeta) V_{nt}(\zeta) &= \frac{1}{nT} E \sum_{t=1}^T [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta]' [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta] \\ &+ \frac{1}{n} \sigma_0^2 \text{tr} \left\{ \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right]' \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right] \right\} \\ &+ \frac{2}{nT} \sum_{t=1}^T [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta]' \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right] V_{nt} \end{aligned} \quad (A6)$$

According to Lemma 1, the third term is $\frac{2}{nT} \sum_{t=1}^T [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta]' \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right] V_{nt} = O\left(\frac{1}{T}\right)$, and $O\left(\frac{1}{T}\right)$ is uniform in θ and Θ because it is a polynomial function in θ and Θ is a bounded set. The second term is equal to $\sigma_n^2(\rho)$, where

$\sigma_n^2(\rho) = \frac{1}{n} \sigma_0^2 \operatorname{tr} \left\{ \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right]' \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right] \right\}$, which are polynomial functions of θ and are uniform in Θ . Using $A_n(\rho) A_n^{-1} = I_n - (\rho - \rho_0) F_n$ in the first term, we have

$$\begin{aligned} & \frac{1}{nT} E \sum_{t=1}^T [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta]' [A_n(\rho) A_n^{-1} Z_{nt} \delta_0 - Z_{nt} \delta] \\ &= \frac{1}{nT} E \sum_{t=1}^T [Z_{nt}(\delta - \delta_0) + (\rho - \rho_0) F_n Z_{nt} \delta_0]' [Z_{nt}(\delta - \delta_0) + (\rho - \rho_0) F_n Z_{nt} \delta_0] \\ &= (\delta' - \delta_0', \rho - \rho_0) E \Lambda_{nT} (\delta' - \delta_0', \rho - \rho_0) \end{aligned}$$

where

$$\Lambda_{nT} = \begin{bmatrix} \frac{1}{nT} \sum_{t=1}^T Z_{nt}' Z_{nt} & \frac{1}{nT} \sum_{t=1}^T (F_n Z_{nt} \delta_0)' Z_{nt} \\ \frac{1}{nT} \sum_{t=1}^T (F_n Z_{nt} \delta_0)' Z_{nt} & \frac{1}{nT} \sum_{t=1}^T (F_n Z_{nt} \delta_0)' (F_n Z_{nt} \delta_0) \end{bmatrix} \quad (A7)$$

To prove that $Q_{n,T}(\zeta)$ is uniformly equicontinuous for $\theta \in \Theta$, the following four conditions must be true. (1) $\ln \sigma^2$ is uniformly continuous; σ^2 is bounded away from zero in Θ , so (1) is obvious. (2) $\frac{1}{n} \ln |A_n(\rho)|$ is uniformly equicontinuous; we know $\frac{1}{n} \ln |A_n(\rho_2)| - \frac{1}{n} \ln |A_n(\rho_1)| = \frac{1}{n} \operatorname{tr} (W_n A_n^{-1}(\bar{\rho})) (\rho_2 - \rho_1)$, where $\bar{\rho}$ falls between ρ_1 and ρ_2 . Because $A_n^{-1}(\rho)$ is UB, and θ is uniform in Θ , $\frac{1}{n} \operatorname{tr} (W_n A_n^{-1}(\bar{\rho}))$ is bounded. Hence, (2) is true. (3) The first term, i.e., $(\delta' - \delta_0', \rho - \rho_0) E \Lambda_{nT} (\delta' - \delta_0', \rho - \rho_0)$, is uniformly equicontinuous; both ρ and δ are bounded, with $E \Lambda_{nT} = O(1)$, meaning (3) is true. (4) $\sigma_n^2(\rho)$ is uniformly equicontinuous given that

$$\sigma_n^2(\rho) = \frac{1}{n} \sigma_0^2 \operatorname{tr} \left\{ \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right]' \left[(A_n(\rho) A_n^{-1} - S(\hat{\beta})) (I_n - S(\hat{\beta}))^{-1} \right] \right\}$$

and $A_n(\rho) A_n^{-1} = I_n - (\rho - \rho_0) F_n$, we have

$$\sigma_n^2(\rho_2) - \sigma_n^2(\rho_1) = \sigma_0^2(\rho_2 - \rho_1) \frac{1}{n} [(\rho_2 + \rho_1 - 2\rho_0) \operatorname{tr} F_n' F_n - \operatorname{tr} (F_n' + F_n)]$$

And because $F_n' F_n$ and F_n are UB, (4) is true. $\square$

Proof of Lemma 4. In line with Lee's study (2004), we can use a contradiction to prove the result. Firstly, we assume $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$, where α_1, α_2 and α_3 are scalars. Next, for $\Sigma_{\theta_0^\beta} \equiv \lim_{T \rightarrow \infty} \Sigma_{\theta_0^\beta, nT}$, we need to prove that $\Sigma_{\theta_0^\beta} \alpha = 0$ implies $\alpha = 0$. If this is true, then, columns of $\Sigma_{\theta_0^\beta}$ would be linear-independent and $\Sigma_{\theta_0^\beta}$ would be nonsingular.

From (21),

$$\Sigma_{\theta_0^\beta} = \frac{1}{\sigma_0^2} \times \begin{bmatrix} E\Phi_{ZZ} & E\Phi_{ZR} & 0 \\ E\Phi_{RZ} & E\Phi_{RR} + \lim_{n \rightarrow \infty} \frac{\sigma_0^2}{n} [\operatorname{tr}(F_n^2) + \operatorname{tr}(F_{ns}' F_{ns})] & \lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr}[F_{ns}] \\ 0 & \lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr}[F_{ns}] & \frac{1}{2\sigma_0^2} \end{bmatrix}$$

where $\Phi_{ZZ} = \lim_{T \rightarrow \infty} \frac{1}{nT} \sum_{t=1}^T \{ [I_n - S(\hat{\beta})] Z_{nt} \}' [I_n - S(\hat{\beta})] Z_{nt}$, $\Phi_{RZ} = \Phi_{ZR}'$, $\Phi_{ZR} = \lim_{T \rightarrow \infty} \frac{1}{nT} \sum_{t=1}^T \{ [I_n - S(\hat{\beta})] Z_{nt} \}' R_n$, and $\Phi_{RR} = \lim_{T \rightarrow \infty} \frac{1}{nT} \sum_{t=1}^T R_n' R_n$. Hence, $\Sigma_{\theta_0^\beta} \alpha = 0$ implies

$$\begin{aligned} & \frac{1}{\sigma_0^2} E\Phi_{ZZ} \times \alpha_1 + \frac{1}{\sigma_0^2} E\Phi_{ZR} \times \alpha_2 = 0 \\ & \frac{1}{\sigma_0^2} E\Phi_{RZ} \times \alpha_1 + \left\{ \frac{1}{\sigma_0^2} E\Phi_{RR} + \lim_{n \rightarrow \infty} \frac{1}{n} [\operatorname{tr}(F_n^2) + \operatorname{tr}(F_{ns}' F_{ns})] \right\} \times \alpha_2 \\ & + \lim_{n \rightarrow \infty} \frac{1}{n\sigma_0^2} \operatorname{tr}[F_{ns}] \times \alpha_3 = 0 \\ & \lim_{n \rightarrow \infty} \frac{1}{n\sigma_0^2} \operatorname{tr}[F_{ns}] \times \alpha_2 + \frac{1}{2\sigma_0^2} \times \alpha_3 = 0 \end{aligned}$$

The first and third equations imply, respectively, $\alpha_1 = -(E\Phi_{ZZ})^{-1}E\Phi_{ZR} \times \alpha_2$ and $\alpha_3 = -2 \lim_{n \rightarrow \infty} \frac{\sigma_0^2}{n} \text{tr}[F_{ns}] \times \alpha_2$. By eliminating α_1 and α_3 , the second equation becomes $\left\{ \frac{1}{\sigma_0^2} [E\Phi_{RR} - E\Phi_{RZ}(E\Phi_{ZZ})^{-1}E\Phi_{ZR}] + \lim_{n \rightarrow \infty} \frac{1}{n} [\text{tr}(F_n^2) + \text{tr}(F'_{ns}F_{ns}) + \frac{2}{n}(\text{tr}(F_{ns}))^2] \right\} \times \alpha_2 = 0$.

Under Assumption 10, we assume that $\Phi_{nT} = \frac{1}{\sigma_0^2} \begin{bmatrix} \Phi_{ZZ} & \Phi_{ZR} \\ \Phi_{RZ} & \Phi_{RR} \end{bmatrix}$ and that $E\Phi_{nT}$ is nonsingular, meaning that the above formula is only true if $\alpha_2 = 0$, i.e., $\alpha = 0$. The information matrix $\Sigma_{\theta_0^{\beta}, nT}$ is nonsingular. $\square$

Proof of Lemma 5. As known from the previous definition,

$$\begin{aligned} \frac{1}{n} M(u, \beta)' \kappa(u, \beta) M(u, \beta) &= \\ &\begin{pmatrix} \frac{1}{n} \sum_{i=1}^n k_h(u_i - u) & \frac{1}{n} \sum_{i=1}^n k_h(u_i - u) \left(\frac{u_i - u}{h} \right)' \\ \frac{1}{n} \sum_{i=1}^n k_h(u_i - u) \left(\frac{u_i - u}{h} \right) & \frac{1}{n} \sum_{i=1}^n k_h(u_i - u) \left(\frac{u_i - u}{h} \right)^2 \end{pmatrix} \\ &\stackrel{\text{def}}{=} \begin{pmatrix} \Psi_{11}^1 & \Psi_{12}^1 \\ \Psi_{21}^1 & \Psi_{22}^1 \end{pmatrix} = \Psi^1 \end{aligned}$$

From Assumption 4, $\{x_i\}_{i=1}^n$ is an independent, identically distributed random sequence, so $\{u_i\}_{i=1}^n$ is an independent, identically distributed sequence.

$$\begin{aligned} E\Psi_{11}^1 &= \int h^{-1} k\left(\frac{u_1 - u}{h}\right) f(u_1 - u) du_1 = \int k(m) f(u + hm) dm = \int k(m) [f(u) + hf'(u)m + h^2 f''(\xi)m^2] dm \\ &= f(u) + O(h^2) \end{aligned}$$

where $m = \frac{u_1 - u}{h}$ is a variable substitution, and ξ falls between u_1 and u .

Similarly, $E\Psi_{21}^1 = Ek_h(u_1 - u) \left(\frac{u_1 - u}{h} \right) = \int h^{-1} k\left(\frac{u_1 - u}{h}\right) f(u_1) \left(\frac{u_1 - u}{h} \right) du_1 = \int k(m) f(u + hm) m dm = \int k(m) [f(u) + hf'(u)m + h^2 f''(\xi)m^2] m dm = O(h)$ and $E\Psi_{12}^1 = (E\Psi_{21}^1)' = O(h)$.

Moreover, $E\Psi_{22}^1 = Ek_h(u_1 - u) \left(\frac{u_1 - u}{h} \right)^2 = \int h^{-1} k\left(\frac{u_1 - u}{h}\right) f(u_1) \left(\frac{u_1 - u}{h} \right)^2 du_1 = \int k(m) f(u + hm) m^2 dm = \int k(m) [f(u) + hf'(u)m + h^2 f''(\xi)m^2] m^2 dm = f(u) \int k(m) m^2 dm + O(h^2)$.

According to the Khintchine law of large numbers, $\Psi_{ij}^1 \xrightarrow{P} E\Psi_{ij}^1$, $i = 1, 2$ and

$$\Psi^1 \xrightarrow{P} \begin{pmatrix} E\Psi_{11}^1 & E\Psi_{12}^1 \\ E\Psi_{21}^1 & E\Psi_{22}^1 \end{pmatrix} = \begin{pmatrix} f(u) + O(h^2) & O(h) \\ O(h) & f(u) \int k(m) m^2 dm \cdot I_n + O(h^2) \end{pmatrix}$$

Based on Assumption 6, when $n \rightarrow \infty$, $h \rightarrow 0$, meaning that

$$\Psi^1 \xrightarrow{P} f(u) \begin{pmatrix} 1 & 0 \\ 0 & f(u) \int k(m) m^2 dm \cdot I_n \end{pmatrix}.$$

$\square$

Proof of Lemma 6. As $s(u, \beta) V_{nt} = e_1 \left(M(u, \beta)' \kappa(u, \beta) M(u, \beta) \right)^{-1} M(u, \beta)' \kappa(u, \beta) V_{nt} = e_1 \left(\frac{1}{n} M(u, \beta)' \kappa(u, \beta) M(u, \beta) \right)^{-1} \frac{1}{n} M(u, \beta)' \kappa(u, \beta) V_{nt}$, we describe $\frac{1}{n} M(u, \beta)' \kappa(u, \beta) V_{nt} = \left(\frac{1}{n} \sum_{i=1}^n k_h(u_i - u) v_{it} \right) \stackrel{\text{def}}{=} \begin{pmatrix} \Psi_1^2 \\ \Psi_2^2 \end{pmatrix} = \Psi^2$. Based on Assumptions 2 and 4, since $\{x_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ are independent, $\{u_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ are independent, and, $E\Psi_1^2 = 0$, $E\Psi_2^2 = 0$. Furthermore, $\{u_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ denote an independent, identically distributed sequence, so

$$\begin{aligned} \text{Var}\Psi_1^2 &= \frac{1}{n}Ek_h^2(u_i - u)v_i^2 = \frac{1}{n}E[k_h^2(u_i - u)]Ev_i^2 = \frac{\sigma_0^2}{n}Ek_h^2(u_i - u) = \frac{\sigma_0^2}{n}h^{-1} \int k^2(m)f(u + hm)dm = O((nh)^{-1}); \\ \text{Var}\Psi_2^2 &= \frac{1}{n}Ek_h^2(u_i - u)\left(\frac{u_i - u}{h}\right)^2v_i^2 = \frac{1}{n}E\left[k_h^2(u_i - u)\left(\frac{u_i - u}{h}\right)^2E(v_i^2)\right] = \frac{\sigma_0^2}{n}E\left[k_h^2(u_i - u)\left(\frac{u_i - u}{h}\right)^2\right] \\ &= \frac{\sigma_0^2}{n}h^{-1} \int k^2(m)f(u + hm)m^2dm = O((nh)^{-1}). \end{aligned}$$

Based on Assumption 5, $\Psi_1^2 = O_p((nh)^{-1/2}) = o_p(1)$ and $\Psi_2^2 = O_p((nh)^{-1/2}) = o_p(1)$, so $\frac{1}{n}M(u, \beta)' \kappa(u, \beta)V_{nt} = O_p((nh)^{-1/2}) = o_p(1)$. Then, from Lemma 5, $S(\beta)V_{nt} = o_p(1)$. $\square$

Proof of Lemma 7. As (6),

$$\begin{aligned} s(u, \beta)G_{nt} &= e_1 \left(\frac{1}{n}M(u, \beta)' \kappa(u, \beta)M(u, \beta) \right)^{-1} \frac{1}{n}M(u, \beta)' \kappa(u, \beta)G_{nt}, \\ \frac{1}{n}M(u, \beta)' \kappa(u, \beta)G_{nt} &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n k_h(u_i - u)g(u_i) \\ \frac{1}{n} \sum_{i=1}^n k_h(u_i - u)\left(\frac{u_i - u}{h}\right)g(u_i) \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \Psi_1^3 \\ \Psi_2^3 \end{pmatrix} = \Psi^3. \end{aligned}$$

Based on Assumptions 2 and 4,

$$\begin{aligned} E\Psi_1^3 &= Ek_h(u_i - u)g(u_i) = \int h^{-p}k\left(\frac{u_i - u}{h}\right)g(u_i)f(u_i)du_i \\ &= \int k(m)f(u + hm)g(u + hm)dm \\ &= \int k(m)[g(u) + hg'(u)m + h^2g''(\xi_1)m][f(u) + hf'(x)m + h^2f''(\xi_2)m^2]dm \\ &= g(u)f(u) + O(h^2), \end{aligned}$$

$$\text{Var}\Psi_1^3 = \frac{1}{n}\text{Var}k_h(u_i - u)g(u_i) = \frac{1}{n}\left\{Ek_h^2(u_i - u)g^2(u_i) - [Ek_h(u_i - u)g(u_i)]^2\right\},$$

where ξ_i $i = 1, 2$ falls between u_i and u . And then,

$$\begin{aligned} Ek_h^2(u_i - u)g^2(u_i) &= h^{-1} \int k^2(m)g(u + hm)f(u + hm)dm = O(h^{-1}), \\ \text{Var}\Psi_1^3 &= \frac{1}{n}\left\{O(h^{-1}) + [g(u)f(u) + O(h^2)]^2\right\} = O((nh)^{-1}), \\ \Psi_1^3 &= g(u)f(u) + O(h^2) + O_p((nh)^{-1/2}) = g(u)f(u) + O(h^2 + (nh)^{-1/2}). \end{aligned}$$

Similarly,

$$\begin{aligned} E\Psi_2^3 &= Ek_h(u_i - u)\left(\frac{u_i - u}{h}\right)g(u_i) = \int k(m)f(u + hm)g(u + hm)m dm \\ &= \int k(m)[g(u) + hg'(\xi_1)m][f(u) + hf'(\xi_2)m]m dm = O(h), \\ (E\Psi_2^3)^2 &= \frac{1}{n}Ek_h^2(u_i - u)\left(\frac{u_i - u}{h}\right)^2g^2(u_i) \\ &= (nh)^{-1} \int k^2(m)f(u + hm)g^2(u + hm)m^2dm = O((nh)^{-1}). \end{aligned}$$

In that way, $\text{Var}\Psi_2^3 = E(\Psi_2^3)^2 - (E\Psi_2^3)^2 = O(h^2 + (nh)^{-1})$ and $\Psi_2^3 = O_p(h + (nh)^{-1/2}) = o_p(1)$. This suggests that

$$\frac{1}{n}M(u, \beta)' \kappa(u, \beta)G_{nt} = \begin{pmatrix} g(u)f(u) + O(h^2 + (nh)^{-1/2}) \\ O_p(h + (nh)^{-1/2}) \end{pmatrix} \xrightarrow{P} \begin{pmatrix} g(u)f(u) \\ 0 \end{pmatrix}.$$

And then, according to Assumption 4,

$$s(u, \beta)G_{nt} = e_1 \left(\frac{1}{n} M(u, \beta)' \kappa(u, \beta) M(u, \beta) \right)^{-1} \frac{1}{n} M(u, \beta)' \kappa(u, \beta) G_{nt} = g(u) + o_p(1), \text{ i.e., } s(u_i, \beta)G_{nt} = g(u_i) + o_p(1), i = 1, 2, \dots, n$$

$$S(\beta)G_{nt} = \left(s(x'_1 \beta)' G_{nt}, s(x'_2 \beta)' G_{nt}, \dots, s(x'_n \beta)' G_{nt} \right)' = G_{nt} + o_p(1). \text{ Hence, } (I_n - S(\beta))G_{nt} = o_p(1).$$

□

Appendix B. Proof of Theoretical Results

Proof of Theorem 1. As $E \sum_{t=1}^T V'_{nt} V_{nt} = nT\sigma_0^2$, at θ_0 , (10) implies $E[\ln L_{n,T}(\theta_0)] = -\frac{nT}{2} \ln 2\pi - \frac{nT}{2} \ln \sigma_0^2 + T \ln |A_n| - \frac{nT}{2}$. Then, we have

$$\begin{aligned} & \frac{1}{nT} E[\ln L_{n,T}(\theta)] - \frac{1}{nT} E[\ln L_{n,T}(\theta_0)] \\ &= -\frac{1}{2} (\ln \sigma_0^2 - \ln \sigma^2) + \frac{1}{n} \ln |A_n(\rho)| - \frac{1}{n} \ln |A_n| - \left[\frac{1}{nT} \frac{1}{2\sigma^2} E \sum_{t=1}^T V_{nt}(\zeta)' V_{nt}(\zeta) - \frac{1}{2} \right] \\ &= T_{1,n}(\rho, \sigma^2) - \frac{1}{2\sigma^2} T_{2,n}(\rho, \delta) + o(1), \text{ where } T_{1,n}(\rho, \sigma^2) = -\frac{1}{2} (\ln \sigma_0^2 - \ln \sigma^2) + \frac{1}{n} \ln |A_n(\rho)| - \frac{1}{n} \ln |A_n| - \frac{1}{2\sigma^2} \\ & \quad [\sigma_n^2(\rho) - \sigma^2] \text{ and } T_{2,n}(\rho, \delta) = \frac{1}{nT} E \sum_{t=1}^T [Z_{nt}(\delta - \delta_0) + (\rho - \rho_0) F_n Z_{nt} \delta_0]' [Z_{nt}(\delta - \delta_0) + (\rho - \rho_0) F_n Z_{nt} \delta_0] \end{aligned}$$

We consider the process $Y_{nt} = \rho_0 W_n Y_{nt} + V_{nt}$, for a period t , the log-likelihood function of which is

$$\ln L_{p,n}(\rho, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) + \ln |A_n(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^T [A_n(\rho) Y_{nt}]' [A_n(\rho) Y_{nt}]$$

By letting $E_p(\cdot)$ be the expectation operator for Y_{nt} , we have

$$\begin{aligned} & E_p \left[\frac{1}{n} \ln L_{p,n}(\rho, \sigma^2) \right] - E_p \left[\frac{1}{n} \ln L_{p,n}(\rho_0, \sigma_0^2) \right] \\ &= -\frac{1}{2} [\ln \sigma^2 - \ln \sigma_0^2] + \frac{1}{n} \ln |A_n(\rho)| - \frac{1}{n} \ln |A_n| - \frac{1}{2\sigma^2} [\sigma_n^2(\rho) - \sigma^2] \\ &= T_{1,n}(\rho, \sigma^2) \end{aligned}$$

Based on information inequality, $\ln L_{p,n}(\rho, \sigma^2) - \ln L_{p,n}(\rho_0, \sigma_0^2) \leq 0$. Thus, $T_{1,n}(\rho, \sigma^2) \leq 0$ for any (ρ, σ^2) . Also, $T_{2,n}(\rho, \delta)$ is a quadratic function of ρ and δ . Under the condition that $E\Phi_{nT}$ is nonsingular, $T_{2,n}(\rho, \delta) > 0$ whenever $(\rho, \delta) \neq (\rho_0, \delta_0)$, so (ρ, δ) is globally identified, given that ρ_0, σ_0^2 is a unique maximizer of $T_{1,n}(\rho, \sigma^2)$. Hence, (ρ, δ, σ^2) is globally identified. Combined with uniform convergence and equicontinuity in Lemmas 3 and 4, the consistency follows. □

Proof of Theorem 2. From the proof of Theorem 1, $\frac{1}{nT} E[\ln L_{n,T}(\theta)] - \frac{1}{nT} E[\ln L_{n,T}(\theta_0)] = T_{1,n}(\rho, \sigma^2) - \frac{1}{2\sigma^2} T_{2,n}(\rho, \delta) + o(1)$. When $E\Phi_{nT}$ is singular, δ_0 and ρ_0 cannot be identified from $T_{2,n}(\rho, \delta)$. Global identification requires that the limit of $T_{1,n}(\rho, \sigma^2)$ is strictly less than zero. As $T_{1,n}(\rho, \sigma^2) \leq 0$ based on information inequality, $T_{1,n}(\rho, \sigma^2) \neq 0$ is equivalent to

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \ln |\sigma_0^2 A_n^{-1} A_n'^{-1}| - \frac{1}{n} \ln |\sigma_n^2(\rho) A_n^{-1}(\rho) A_n'^{-1}(\rho)| \right) \neq 0$$

(see Lee (2004)). After ρ_0 and σ_0^2 are identified, given ρ_0, δ_0 can be identified from $T_{2,n}(\rho, \delta)$. Combined with uniform convergence and equicontinuity in Lemmas 3 and 4, the consistency follows. □

Proof of Theorem 3. As (6), we have $\hat{G}_{nt} = S(\hat{\beta})\hat{Y}_t = S(\hat{\beta})(Y_{nt} - \hat{\rho}W_nY_{nt} - Z_{nt}\hat{\delta}) = S(\hat{\beta})(Y_{nt} - \hat{\rho}W_nY_{nt} + \rho_0W_nY_{nt} - \rho_0W_nY_{nt} - Z_{nt}\hat{\delta} + Z_{nt}\delta_0 - Z_{nt}\delta_0) = S(\hat{\beta})[(Y_{nt} - \rho_0W_nY_{nt} - Z_{nt}\delta_0) - (\hat{\rho} - \rho_0)W_nY_{nt} - Z_{nt}(\hat{\delta} - \delta_0)] = S(\hat{\beta})[G_{nt} + V_{nt} - (\hat{\rho} - \rho_0)W_nY_{nt} - Z_{nt}(\hat{\delta} - \delta_0)]$, so that, $\hat{G}_{nt} - G_{nt} = S(\hat{\beta})V_{nt} - [I_n - S(\hat{\beta})]G_{nt} - [(\hat{\rho} - \rho_0)S(\hat{\beta})W_nY_{nt} + S(\hat{\beta})Z_{nt}(\hat{\delta} - \delta_0)]$.

From Theorems 1 and 2, $\hat{\rho} - \rho_0 = o_p(1)$, $\hat{\delta} - \delta_0 = o_p(1)$. And from Lemmas 6 and 7, $[I_n - S(\hat{\beta})]G_{nt} = o_p(1)$, $S(\hat{\beta})V_{nt} = o_p(1)$. Otherwise determined by the boundedness in the assumptions, $\hat{G}_{nt} - G_{nt} = o_p(1)$. $\square$

References

1. Ai, C.; Chen, X. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **2003**, *71*, 1795–1843. [\[CrossRef\]](#)
2. Baltagi, B.H.; Song, S.H.; Koh, W. Testing panel data regression models with spatial error correlation. *Econometrics* **2003**, *117*, 123–150. [\[CrossRef\]](#)
3. Chen, J.; Gao, J.T.; Li, D.G. Estimation in partially linear single-index panel data models with fixed effects. *J. Bus. Econ. Stat.* **2013**, *31*, 315–330. [\[CrossRef\]](#)
4. Elhorst, J.P. Specification and estimation of spatial panel data models. *Int. Reg. Sci. Rev.* **2003**, *26*, 244–268. [\[CrossRef\]](#)
5. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
6. Jin, B.; Wu, Y.; Rao, C.R.; Hou, L. Estimation and model selection in general spatial dynamic panel data models. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5235–5241. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Lee, L.F.; Yu, J. Estimation of spatial panel model with fixed effects. *Econometrics* **2010**, *154*, 165–185. [\[CrossRef\]](#)
8. Lee, L.F.; Yu, J. A spatial dynamic panel data model with both time and individual fixed effects. *Econom. Theory* **2010**, *26*, 564–597. [\[CrossRef\]](#)
9. Lee, L.F.; Yu, J. Some recent developments in spatial panel data models. *Reg. Sci. Urban Econ.* **2010**, *40*, 255–271. [\[CrossRef\]](#)
10. Pang, Z.; Xue, L.G. Estimation for the single-index models with random effects. *Comput. Stat. Data Anal.* **2012**, *56*, 1837–1853. [\[CrossRef\]](#)
11. Parent, O.; LeSage, J.P. A space-time filter for panel data models containing random effects. *Comput. Stat. Data Anal.* **2011**, *55*, 475–490. [\[CrossRef\]](#)
12. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
13. Su, L.; Yang, Z. *QML Estimation of Dynamic Panel Data Models with Spatial Errors*; Singapore Management University: Singapore, 2007.
14. Su, L.; Jin, S.N. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *J. Econom.* **2010**, *157*, 18–33. [\[CrossRef\]](#)
15. Wang, J.L.; Xue, L.G.; Zhu, L.X.; Chong, Y.S. Estimation for a partial-linear single-index model. *Ann. Stat.* **2010**, *38*, 246–274. [\[CrossRef\]](#)
16. Yu, J.; de Jong, R.; Lee, L.F. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *J. Econom.* **2008**, *146*, 118–134. [\[CrossRef\]](#)
17. Yu, J.; Lee, J.F. Efficient GMM estimation of spatial dynamic panel data models. *J. Econom.* **2010**, *180*, 174–197.
18. Yu, J.; Lee, J.F. Estimation of unit root spatial dynamic panel data models. *Econom. Theory* **2010**, *26*, 1332–1362. [\[CrossRef\]](#)
19. Yu, J.; de Jong, R.; Lee, L.F. Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *J. Econom.* **2012**, *167*, 16–37. [\[CrossRef\]](#)
20. Zhang, Y.Q.; Shen, D.M. Estimation of semiparametric varying-coefficient spatial panel data models with random effects. *J. Stat. Plan. Inference* **2015**, *159*, 64–80. [\[CrossRef\]](#)

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.