*axioms*

MDPI

*Article*

# A Penalized Profile Quasi-Likelihood Method for a Semiparametric Varying Coefficient Spatial Autoregressive Panel Model with Fixed Effects

**Ruiqin Tian [1,*], Miaojie Xia [2] and Dengke Xu [3]**

[1] School of Mathematics, Hangzhou Normal University, Hangzhou 311121, China
[2] School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China; xiamiaojie@emails.bjut.edu.cn
[3] School of Economics, Hangzhou Dianzi University, Hangzhou 310018, China; xudengke@hdu.edu.cn
* Correspondence: ruiqintian@163.com

**Abstract:** This paper proposes a variable selection method for a semiparametric varying coefficient spatial autoregressive panel model with fixed effects based on a penalized profile quasi-likelihood method, which can simultaneously select significant variables in parametric components and nonparametric components without estimating fixed effects. With an appropriate selection of the tuning parameters and some mild assumptions, the consistency of this procedure and the oracle property of the obtained estimators are established. Then, we conduct some Monte Carlo simulations to assess the finite sample performance of the proposed variable selection method, and finally, we analyze a real dataset for further illustration.

**Keywords:** semiparametric varying coefficient; spatial autoregressive panel model; fixed effects; variable selection; profile quasi-likelihood

**MSC:** 62G05; 62F12

check for updates

Academic Editor: J. Alberto Conejero

Received: 11 January 2025
Revised: 3 February 2025
Accepted: 5 February 2025
Published: 7 February 2025

**Citation:** Tian, R.; Xia, M.; Xu, D. A Penalized Profile Quasi-Likelihood Method for a Semiparametric Varying Coefficient Spatial Autoregressive Panel Model with Fixed Effects. *Axioms* **2025**, *14*, 121. https://doi.org/10.3390/axioms14020121

**Copyright:** © 2025 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

Recently, there has been a surge in focus on spatial panel data model research. These models not only account for the spatial interdependencies of economic phenomena but also allow investigators to manage the unobservable heterogeneity among geographical units [1–5]. The basic spatial panel model is specified as follows:

$$y_{it} = \rho_0 \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{X}_{it}^{\tau} \boldsymbol{\beta}_0 + \mu_{0i} + \epsilon_{it}, i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \quad (1)$$

where $\{y_{it}, \boldsymbol{X}_{it}\}$ represents the observations of an individual $i$ in period $t$, $w_{ij}$ denotes the spatial weight between an individual $i$ and $j$, $\mu_{0i}$ denotes the unobserved and time-invariant individual effect, $\epsilon_{it}$ denotes random disturbance, and $\{\rho_0, \boldsymbol{\beta}_0\}$ denotes the unknown true parameter value. However, model (1) adopts a linear specification and conducting parametric statistical inference inherently requires a set of model assumptions, with linearity serving as one of the most practical options. Despite their robust theoretical foundations, linear models frequently fall short in practical applications. Furthermore, when a linear model is misspecified with respect to the data analysis process, it may result in significant modeling biases and misleading conclusions. Therefore, more flexible spatial panel models are required. Ai and Zhang [6] extended model (1) to a partially linear spatial

panel model with fixed effects by adding an unknown function, and they proposed a sieve two-stage least squares regression to consistently estimate their model. Zhang and Sun [7] considered a partially specified dynamic spatial panel model, which takes into account past information of the dependent variable. Furthermore, a model known as the semiparametric varying coefficient spatial autoregressive (hereafter, SVCSAR) panel model, which strikes a balance between flexibility and interpretability, is being popularly studied; see [8,9] for more details. The model is specified as follows:

$$y_{it} = \rho_0 \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{X}_{it}^{\tau} \boldsymbol{\beta}_0 + \boldsymbol{Z}_{it}^{\tau} \boldsymbol{\alpha}_0(u_{it}) + \mu_{0i} + \epsilon_{it}, i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \quad (2)$$

where $\{y_{it}, u_{it}, \boldsymbol{X}_{it}, \boldsymbol{Z}_{it}\}$ represents the observations of an individual $i$ in period $t$. $\boldsymbol{\alpha}_0(u)$ is a vector of unknown functions. Model (2) is a general model that can be reduced to some existing panel models. For instance, if $\rho_0 = 0$, this model is reduced to the varying coefficient panel model studied by [10–13]. If $\rho_0 = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}$, this model is reduced to a varying coefficient panel model [14]. If $\boldsymbol{\alpha}_0(\cdot) \equiv \mathbf{0}$, this model generalizes the spatial autoregressive panel model [5]. Moreover, if $\rho_0 = 0$, $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\boldsymbol{\alpha}_0(\cdot) \equiv \mathbf{0}$, this model becomes the classical panel model. However, while model (2) can be reduced to various panel models, it also increases the risk of model misspecification in practical applications. Specifying the model form becomes an inevitable issue, which is equivalent to detecting the zero components of $\{\rho_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0(\cdot)\}$. In other words, a variable selection method for model (2) is required. Additionally, when the number of covariates in model (2) is large, selecting important variables is also a purpose of the variable selection method.

Variable selection constitutes a crucial aspect of contemporary statistical inference. Over the years, numerous variable selection techniques have emerged for parametric models. LASSO [15], SCAD [16], and ALASSO [17] are the most popular methods among them. Based on those methods, variable selection methods for nonparametric or semiparametric models have been established in recent years. Wang et al. [18] considered variable selection for varying coefficient models using the SCAD penalty. Li and Liang [19] utilized the SCAD penalty to identify significant variables within the parametric components of a semiparametric varying coefficient partially linear model. Wang et al. [20] and Zhao and Xue [21] presented a variable selection procedure by combining basis function approximations with SCAD penalty for semiparametric varying coefficient partially linear models. The proposed procedure simultaneously selects significant variables in parametric components and nonparametric components. Tian et al. [22] introduced a novel method for variable selection, which integrates basis function approximations with quadratic inference functions. This approach enables the simultaneous identification of significant variables in both parametric and nonparametric components. For further development of the variable selection methods for nonparametric or semiparametric models, see [23–26], among others. However, there are few studies on the variable selection of panel model (2), in which spatial components and nonparametric components are simultaneously included.

In this paper, we propose a variable selection procedure for model (2). In order to avoid incidental parameter problems [27] brought by unknown fixed effects, this variable selection procedure combines a profile quasi-likelihood method with the basis function approximation and SCAD penalty to achieve estimation and variable selection simultaneously. The proposed procedure can shrink the spatial, linear, and functional coefficients of irrelevant covariates automatically to achieve variable selection. Moreover, by selecting appropriate tuning parameters, we demonstrate the consistency of our variable selection method. The regression coefficient estimators exhibit the oracle property, which implies that the nonparametric component estimators converge optimally, while the parametric component estimators share the same asymptotic behavior as those derived from the true

submodel. This suggests that our penalized estimators perform as effectively as if the true zero coefficients were known. Compared with Liu et al. [28] and Xie et al. [29], we consider panel data and varying coefficient components. In comparison, although Luo and Wu [30] took into account variable selection for the SVCSAR model, the model they analyzed was confined to cross-sectional data. Furthermore, they only selected parametric coefficients. However, our method enables the simultaneous selection of significant variables in both parametric and nonparametric components under panel data.

The rest of this paper is structured as follows. In Section 2, we introduce a variable selection process designed for the SVCSAR panel model with fixed effects. In Section 3, we establish the asymptotic properties of the resulting estimators. In Section 4, we detail the computational steps for obtaining these estimators and discuss the selection of tuning parameters. In Section 5, we conduct simulations to assess the performance of our method with finite samples. Additionally, in Section 6, we provide a real-data analysis to further demonstrate the application of our proposed methodology. Finally, in Section 7, we conclude the article with a concise discussion. All technical proofs supporting the asymptotic results are included in the Appendix A.

## 2. Penalized Profile Quasi-Likelihood Method

We first consider the SVCSAR panel model described in (2). Specifically, $y_{it}$ and $u_{it}$ are scalars, while $\boldsymbol{X}_{it}$ is a $p \times 1$ vector and $\boldsymbol{Z}_{it}$ is a $q \times 1$ vector, respectively; $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \ldots, \beta_{0p})^\tau$ is an unknown vector that reflects the linear effect of $\boldsymbol{X}_{it}$ on $y_{it}$, and it is assumed that only a subset of its elements are non-zero. Without loss of generality, we assume that the first $s$ elements specifically are non-zero. $\boldsymbol{\alpha}_0(u) \equiv (\alpha_{01}(u), \ldots, \alpha_{0k}(u), \ldots, \alpha_{0q}(u))^\tau$ is a vector of unknown functions. Similarly, we assume that the first $d$ function is non-zero. $\epsilon_{it}$ is an i.i.d. disturbance with zero mean and finite unknown variance $\sigma_0^2$; thus, $\{\sigma_0^2, \rho_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \mu_{01}, \ldots, \mu_{0N}\}$ are unknown components. Next, we show a penalized quasi-likelihood method that can skip the estimation of $\mu_{0i} (i = 1, \ldots, N)$ and shrink the remaining estimators.

Let $n = N \times T, \boldsymbol{Y}_n \equiv (\boldsymbol{Y}_1^\tau, \ldots, \boldsymbol{Y}_t^\tau, \ldots, \boldsymbol{Y}_T^\tau)^\tau$ with $\boldsymbol{Y}_t = (y_{1t}, \ldots, y_{Nt})^\tau; \boldsymbol{W}_n \equiv \boldsymbol{I}_T \otimes \boldsymbol{W}_N$, where "$\otimes$" demotes the Kronecker product symbol and $\boldsymbol{I}_T$ denotes a $T$-dimensional identity matrix; $\boldsymbol{X}_n \equiv (\boldsymbol{X}_1^\tau, \ldots, \boldsymbol{X}_t^\tau, \ldots, \boldsymbol{X}_T^\tau)^\tau$ with $\boldsymbol{X}_t = (\boldsymbol{X}_{1t}, \ldots, \boldsymbol{X}_{Nt})^\tau; \boldsymbol{A}_0 \equiv (\boldsymbol{A}_{01}^\tau, \ldots, \boldsymbol{A}_{0t}^\tau, \ldots, \boldsymbol{A}_{0T}^\tau)^\tau$ with $\boldsymbol{A}_{0t} = (\boldsymbol{Z}_{1t}^\tau \boldsymbol{\alpha}_0(u_{1t}), \ldots, \boldsymbol{Z}_{Nt}^\tau \boldsymbol{\alpha}_0(u_{Nt}))^\tau; \boldsymbol{V}_n \equiv (\boldsymbol{V}_1^\tau, \ldots, \boldsymbol{V}_t^\tau, \ldots, \boldsymbol{V}_T^\tau)^\tau$ with $\boldsymbol{V}_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})^\tau$; $\boldsymbol{\mu}_0 = (\mu_{01}, .., \mu_{0N})^\tau$ denotes the fixed effects vector; $\boldsymbol{D}_n \equiv \boldsymbol{\iota}_T \otimes \boldsymbol{I}_N$, where $\boldsymbol{\iota}_T$ represents a $T$-dimensional vector with all 1. Then, the model (2) can be expressed in matrix form:

$$\boldsymbol{Y}_n = \rho_0 \boldsymbol{W}_n \boldsymbol{Y}_n + \boldsymbol{X}_n \boldsymbol{\beta}_0 + \boldsymbol{A}_0 + \boldsymbol{D}_n \boldsymbol{\mu}_0 + \boldsymbol{V}_n, \tag{3}$$

where $E(\boldsymbol{V}_n) = \boldsymbol{0}, \mathrm{var}(\boldsymbol{V}_n) = \sigma_0^2 \boldsymbol{I}_n$. Thus, the unknown parametric component is $(\sigma_0^2, \rho_0, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0)$, and the unknown nonparametric component is $\boldsymbol{\alpha}_0(u)$.

Let $\boldsymbol{\theta} = (\sigma^2, \rho, \boldsymbol{\beta}^\tau)^\tau$, $\boldsymbol{\theta}_0 = (\sigma_0^2, \rho_0, \boldsymbol{\beta}_0^\tau)^\tau$, and $\boldsymbol{M}_n(\rho) = \boldsymbol{I}_n - \rho \boldsymbol{W}_n$ for any $\rho$ and $\boldsymbol{M}_n \equiv \boldsymbol{M}_n(\rho_0)$. Subsequently, we suggest optimizing the log-Gaussian quasi-likelihood, following the approach taken by [31,32], and the log-Gaussian quasi-likelihood of (3) is

$$\begin{aligned} \ln \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\alpha}(u)) = &-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln |\boldsymbol{M}_n(\rho)| \\ &- \frac{1}{2\sigma^2} (\boldsymbol{M}_n(\rho) \boldsymbol{Y}_n - \boldsymbol{X}_n \boldsymbol{\beta} - \boldsymbol{A} - \boldsymbol{D}_n \boldsymbol{\mu})^\tau (\boldsymbol{M}_n(\rho) \boldsymbol{Y}_n - \boldsymbol{X}_n \boldsymbol{\beta} - \boldsymbol{A} - \boldsymbol{D}_n \boldsymbol{\mu}), \end{aligned} \tag{4}$$

where $\boldsymbol{A} = \left(\boldsymbol{A}_1^\tau, \ldots, \boldsymbol{A}_t^\tau, \ldots, \boldsymbol{A}_T^\tau\right)^\tau$, $\boldsymbol{A}_t = \left[\boldsymbol{Z}_{1t}^\tau \boldsymbol{\alpha}(u_{1t}), \ldots, \boldsymbol{Z}_{Nt}^\tau \boldsymbol{\alpha}(u_{Nt})\right]^\tau$, and $\boldsymbol{\alpha}(u) = \left[\alpha_1(u), \ldots, \alpha_q(u)\right]^\tau$. When maximizing Equation (4), we encounter two main challenges: (i) Directly estimating $\boldsymbol{\mu}_0$ can lead to the incidental parameter problem [27], especially when

$\boldsymbol{\mu}_0$ becomes high-dimensional as $N$ increases. (ii) Estimating $\boldsymbol{\alpha}_0$ is difficult due to its infinite dimensionality. To address these issues, we follow the approach in Tian et al. [32] and let

$$\boldsymbol{B}(u) = (B_1(u), B_2(u), \ldots, B_{K_n+l+1}(u))^\tau$$

denote a vector comprising normalized B-spline basis functions of order $l$ with $K_n$ internal knots. Subsequently, we approximate $\alpha_{0k}(u)$ as a linear combination of $B_1(u), B_2(u), \ldots, B_{K_n+l+1}(u)$, i.e., $\alpha_{0k}(u) \approx \boldsymbol{B}^\tau(u)\gamma_{0k}$ for $k = 1, 2, \ldots, q$. Consequently, mode (3) can be written as follows:

$$\boldsymbol{Y}_n \approx \rho_0 \boldsymbol{W}_n \boldsymbol{Y}_n + \boldsymbol{X}_n \boldsymbol{\beta}_0 + \boldsymbol{S}_n \gamma_0 + \boldsymbol{D}_n \boldsymbol{\mu}_0 + \boldsymbol{V}_n, \tag{5}$$

where $\boldsymbol{S}_n = (\boldsymbol{S}_1^\tau, \ldots, \boldsymbol{S}_t^\tau, \ldots, \boldsymbol{S}_T^\tau)^\tau$, $\boldsymbol{S}_t = \left[\left(\boldsymbol{I}_q \otimes \boldsymbol{B}(u_{1t})\right)\boldsymbol{Z}_{1t}, \ldots, \left(\boldsymbol{I}_q \otimes \boldsymbol{B}(u_{Nt})\right)\boldsymbol{Z}_{Nt}\right]$, and $\gamma_0 = (\gamma_{01}^\tau, \ldots, \gamma_{0q}^\tau)^\tau$. The log-Gaussian quasi-likelihood of (5) is

$$\begin{aligned} \ln \hat{L}(\boldsymbol{\theta}, \boldsymbol{\mu}, \gamma) = & -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 + \ln |\boldsymbol{M}_n(\rho)| \\ & - \frac{1}{2\sigma^2}(\boldsymbol{M}_n(\rho)\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{S}_n\gamma - \boldsymbol{D}_n\boldsymbol{\mu})^\tau(\boldsymbol{M}_n(\rho)\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{S}_n\gamma - \boldsymbol{D}_n\boldsymbol{\mu}). \end{aligned} \tag{6}$$

To avoid the incidental parameter problem [27] caused by $\boldsymbol{\mu}$, we first concentrate $\boldsymbol{\mu}$ out and obtain the profile quasi-likelihood. Let $\boldsymbol{\eta} = (\boldsymbol{\theta}^\tau, \gamma^\tau)^\tau$. For given $\boldsymbol{\eta}$, from (6), we derive

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\eta}) = (\boldsymbol{D}_n^\tau \boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^\tau(\boldsymbol{M}_n(\rho)\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{S}_n\gamma)$$

and substitute this into (6). Then,

$$\begin{aligned} \ln L_n(\boldsymbol{\eta}) = & -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 + \ln |\boldsymbol{M}_n(\rho)| \\ & - \frac{1}{2\sigma^2}(\boldsymbol{M}_n(\rho)\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{S}_n\gamma)^\tau \boldsymbol{J}_n(\boldsymbol{M}_n(\rho)\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta} - \boldsymbol{S}_n\gamma), \end{aligned} \tag{7}$$

where $\boldsymbol{J}_n = (\boldsymbol{I}_n - \boldsymbol{Q}_n)^\tau(\boldsymbol{I}_n - \boldsymbol{Q}_n)$ and $\boldsymbol{Q}_n = \boldsymbol{D}_n(\boldsymbol{D}_n^\tau \boldsymbol{D}_n)^{-1}\boldsymbol{D}_n^\tau$.

Inspired by the concept of variable selection in semiparametric varying coefficient partially linear models [21], we introduce a penalized profile quasi-likelihood function defined as follows:

$$Q_n(\boldsymbol{\eta}) = \ln L_n(\boldsymbol{\eta}) - n\sum_{j=2}^{p+2} p_{\lambda_{1,n}}(|\theta_j|) - n\sum_{k=1}^{q} p_{\lambda_{2,n}}(\|\boldsymbol{B}^\tau(\cdot)\gamma_k\|), \tag{8}$$

where $\|\boldsymbol{B}^\tau(\cdot)\gamma_k\| = (\int(\boldsymbol{B}^\tau(u)\gamma_k)^2 du)^{1/2}$, and $p_\lambda(\cdot)$ is a SCAD penalty function [16] defined by

$$p'_\lambda(\omega) = \lambda\left\{I(\omega \le \lambda) + \frac{(a\lambda - \omega)_+}{(a-1)\lambda} I(\omega \ge \lambda)\right\},$$

with $a > 2, \omega > 0$ and $p_\lambda(0) = 0$. Throughout this paper, we adopt the suggestion by Fan and Li [16] that the choice of $a = 3.7$ performs well in a variety of situations. The tuning parameter $\lambda$ can be different for all $\theta_j$ and $\boldsymbol{B}^\tau(\cdot)\gamma_k$. Note that $\|\boldsymbol{B}^\tau(\cdot)\gamma_k\| = (\int(\boldsymbol{B}^\tau(u)\gamma_k)^2 du)^{1/2} = (\gamma_k^\tau \boldsymbol{H}\gamma_k)^{1/2} \equiv \|\gamma_k\|_{\boldsymbol{H}}$, where $\boldsymbol{H} = \int \boldsymbol{B}(u)\boldsymbol{B}^\tau(u)du$. Then, the penalized profile quasi-likelihood function can be written as follows:

$$Q_n(\boldsymbol{\eta}) = \ln L_n(\boldsymbol{\eta}) - n\sum_{j=2}^{p+2} p_{\lambda_{1,n}}(|\theta_j|) - n\sum_{k=1}^{q} p_{\lambda_{2,n}}(\|\gamma_k\|_{\boldsymbol{H}}). \tag{9}$$

Let $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\theta}}^\tau, \hat{\boldsymbol{\gamma}}^\tau)^\tau$ be the solution by maximizing (9). Then, $\hat{\boldsymbol{\theta}}$ is the penalized profile quasi-likelihood estimator (penalized profile QMLE) of $\boldsymbol{\theta}$, and the estimator of $\alpha_k(u)$ can be obtained by $\alpha_k(u) = \boldsymbol{B}^\tau(u)\hat{\gamma}_k$. Next, we study the asymptotic properties of the resulting penalized profile likelihood estimators. Without loss of generality, we assume that $\theta_{0j} = 0$ for $j = s+1, \ldots, p+2$, with the remaining $\theta_{0j} (j = 1, \ldots, s)$ being the non-zero components of $\boldsymbol{\theta}_0$. Similarly, we assume that $\alpha_{0k}(\cdot) \equiv 0$ for $k = d+1, \ldots, q$ and that the $\alpha_{0k}(\cdot)$, $(k = 1, \ldots, d)$ constitute all the non-zero components of $\boldsymbol{\alpha}_0(\cdot)$.

## 3. Asymptotic Results

Denote $\boldsymbol{G}_n = \boldsymbol{W}_n\boldsymbol{M}_n^{-1}$, $\boldsymbol{R}_n = \boldsymbol{G}_n(\boldsymbol{X}_n\boldsymbol{\beta}_0 + \boldsymbol{A}_0 + \boldsymbol{D}_n\boldsymbol{\mu}_0)$. The following assumptions are necessary prior to establishing the asymptotic properties.

C1 $T$ is finite and greater than 2, and $N$ is large.

C2 The disturbances $\{\epsilon_{it}\}$ for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ are independently and identically distributed with zero mean, and the finite variance is $\sigma_0^2$. Additionally, for some $v > 0$, $E|\epsilon_{it}|^{4+v}$ exists.

C3 The entries $\{w_{ij}\}$ of $\boldsymbol{W}_n$ satisfy $w_{ii} = 0$ and $w_{ij} = O(1/h_n)$, where $h_n/n \to 0$ as $n \to \infty$.

C4 The matrix $\boldsymbol{M}_n(\rho)$ is nonsingular for all $\rho$ in a compact parameter space $\Lambda$. The sequences $\{\boldsymbol{M}_n^{-1}(\rho)\}$ are uniformly bounded in either row or column sums for all $\rho \in \Lambda$. The true $\rho_0$ is in the interior of $\Lambda$.

C5 The sequences of matrices $\{\boldsymbol{W}_n\}$ and $\{\boldsymbol{M}_n^{-1}\}$ are uniformly bounded in both row and column sums.

C6 The elements of $\boldsymbol{X}_n$ and $\boldsymbol{S}_n$ are uniformly bounded for all $n$, and the limit $\lim_{n\to\infty} n^{-1}(\boldsymbol{X}_n, \boldsymbol{S}_n, \boldsymbol{R}_n)^\tau \boldsymbol{J}_n(\boldsymbol{X}_n, \boldsymbol{S}_n, \boldsymbol{R}_n)$ exists and is nonsingular.

C7 There exists a constant $\lambda_c$, such that $\lambda_c \boldsymbol{I}_n - \boldsymbol{\Gamma}_n\boldsymbol{\Gamma}_n^\tau$ is positive and semidefinite for all $n$, where $\boldsymbol{\Gamma}_n = \boldsymbol{R}_n, \boldsymbol{X}_n, \boldsymbol{G}_n$.

C8 The limits $\lim_{n\to\infty} E[n^{-1}\partial^2 \ln L_n(\boldsymbol{\eta}_T)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\tau]$ exist.

C9 Third derivatives $\partial^3 \ln L_n(\boldsymbol{\eta})/(\partial\eta_i\partial\eta_j\partial\eta_k)$ exist for all $\boldsymbol{\eta}$ in an open set $\mathbb{H}$ that contains the parameter point $\boldsymbol{\eta}_T$. Furthermore, there exist functions $M_{ijk}$, such that $|\partial^3 \ln L_n(\boldsymbol{\eta})/(\partial\eta_i\partial\eta_j\partial\eta_k)| \le M_{ijk}$ for all $\eta \in \mathbb{H}$, where $E(M_{ijk}) \le \infty$ for $i, j, k$.

C10 For $k = 1 \ldots q$, $\alpha_k(u) \in C^r(0,1)$, where $r \ge 2$. The distribution of $u_{it}$ is absolutely continuous, and its density is bound away from zero and infinity on $[0,1]$.

C11 Let $\pi_1, \ldots, \pi_{K_n}$ denote the interior knots within the interval $[0,1]$ and $\pi_0 = 0$, $\pi_{K_n+1} = 1$. Define $\varrho_i = \pi_i - \pi_{i-1}$. There exists a constant $C$, such that $\max \varrho_i / \min \varrho_i \le C$ and $\max\{\varrho_i\} = o(K_n^{-1})$.

C12 The knot number $K_n$ is assumed to satisfy $K_n = n^{\frac{1}{2r+1}}$.

C13 Let $b_n = \max_{j,k}\{p''_{\lambda_{1j}}(|\theta_{0j}), p''_{\lambda_{2,n}}(\|\gamma_{0k}\|_{\boldsymbol{H}}) : |\theta_{0j}| \ne 0, \|\gamma_{0k}\|_{\boldsymbol{H}} \ne 0\}$. Then, $b_n \to 0$, as $n \to \infty$.

C14 $j = s+1, \ldots, p, k = d+1, \ldots, q$, $\liminf_{n\to\infty} \liminf_{\theta_j\to 0^+} \lambda_{2j}^{-1} p'_{\lambda_{2j}}(|\theta_j|) > 0$ and $\liminf_{n\to\infty} \liminf_{\|\gamma_{0k}\|_{\boldsymbol{H}}\to 0^+} \lambda_{2k}^{-1} p'_{\lambda_{2k}}(\|\gamma_{0k}\|_{\boldsymbol{H}}) > 0$ hold.

**Remark 1.** *C1 excludes the scene where T approaches infinity. Essentially, this assumption implies a setting (large N, small T) aligning with many spatial data studies. Conversely, the scenarios in which only T increases indefinitely and where both N and T go to infinity closely resemble the scenario where only N goes to infinity. C2 is needed to apply the central limit theorem in Kelejian and Prucha [33]. C3–C5 are analogous to Assumption 2 in Lee [31], which focuses on the properties of the spatial weight matrix $\boldsymbol{W}_n$ and is essential for the identifiability of $\rho$. Specifically, $\Lambda = (-1, 1)$ when $\boldsymbol{W}_n$ satisfies $\sum_{j=1}^n w_{ij} = 1$ for all $i$. C6–C9 are applied for asymptotic normality of the profile QMLE. C10–C12 facilitate achieving the optimal convergence rate for $\hat{\alpha}_k(u)$. He et al. [34] suggested that cubic B-splines are adequate for accurately approximating nonparametric functions,*

*with the number of interior knots set to the integer part of* $n^{1/5}$. *Meanwhile, C13 and C14 present assumptions about the penalty function, which are comparable to those utilized in [16,18,19].*

Due to the projection matrix $\boldsymbol{J}_n$, a portion of the degrees of freedom is lost; therefore, the estimator $\hat{\sigma}^2$ derived from (9) is not a consistent estimator of $\sigma_0^2$. A correction is needed. Let $\sigma_T^2 = (T-1)\sigma_0^2/T$, $\boldsymbol{\theta}_T = (\sigma_T^2, \rho_0, \boldsymbol{\beta}_0^\tau)^\tau$ and $\boldsymbol{\eta}_T = (\boldsymbol{\theta}_T^\tau, \boldsymbol{\gamma}_0^\tau)^\tau$. Under the assumptions, the subsequent theorem establishes the consistency property of the penalized profile QMLE.

**Theorem 1.** *Suppose that Assumptions C1–C12 hold; then, we can have*

$$\|\hat{\alpha}_k(\cdot) - \alpha_{0k}(\cdot)\| = O_p(n^{-r/(2r+1)} + a_n), k = 1, \ldots, q,$$

*where* $a_n = \max_{j,k}\left\{p'_{\lambda_{1j}}(|\theta_{0j}|), p'_{\lambda_{2k}}(\|\boldsymbol{\gamma}_{0k}\|_{\boldsymbol{H}}) : |\theta_{0j}| \neq 0, \|\boldsymbol{\gamma}_{0k}\|_{\boldsymbol{H}} \neq 0\right\}$.

Furthermore, under some conditions, we show that such consistent estimators must possess the sparsity property, which is stated as follows.

**Theorem 2.** *Suppose that Assumptions C1–C14 hold, and let* $\lambda_{\max} = \max_{j,k}\{\lambda_{1j}, \lambda_{2k}\}$ *and* $\lambda_{\min} = \min_{j,k}\{\lambda_{1j}, \lambda_{2k}\}$. *If* $\lambda_{\max} \to 0$ *and* $n^{r/(2r+1)}\lambda_{\min} \to \infty$ *as* $n \to \infty$, *then, with probability tending to 1,* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\alpha}(\cdot)$ *must satisfy*

*(i)* $\hat{\beta}_j = 0, j = s+1, \ldots, p.$
*(ii)* $\hat{\alpha}_k(\cdot) \equiv 0, k = d+1, \ldots, q.$

According to Remark 1 in Fan and Li [16], if $\lambda_{\max} \to 0$ as $n \to \infty$, then $a_n = 0$. Consequently, based on Theorems 1 and 2, it becomes evident that by selecting appropriate tuning parameters, our variable selection approach is consistent. Furthermore, the estimators of the nonparametric components attain the optimal convergence rate as if the subset of true zero coefficients were already known [35]. Subsequently, we will demonstrate that the estimators for the non-zero coefficients in the parametric components share the same asymptotic distribution as those derived from the correct submodel. Let $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}_1, \ldots, \hat{\theta}_s)^\tau$, $\boldsymbol{\theta}_T^* = (\theta_{T1}, \ldots, \theta_{Ts})^\tau$, where the following result states the asymptotic normality of $\hat{\boldsymbol{\theta}}$.

**Theorem 3.** *Suppose that Assumptions C1–C14 and the conditions in Theorem 2 hold, then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_T^*) \to N\left(\mathbf{0}, \frac{T}{T-1}\left\{\boldsymbol{\Omega}^{-1} + \boldsymbol{\Omega}^{-1}[\boldsymbol{\Psi}_{\boldsymbol{\theta}}^* - 2\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\gamma}}^{*\tau}(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^*)^{-1}\boldsymbol{\Psi}_{\boldsymbol{\theta}}^*]\boldsymbol{\Omega}^{-1}\right\}\right),$$

*where* $\boldsymbol{\Omega}, \boldsymbol{\Psi}_{\boldsymbol{\theta}}^*, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^*, \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\gamma}}^*$ *are defined in the Appendix A.*

## 4. Some Issues in Practice

In the practical situation, we have to choose the proper tuning parameters $(\lambda_1, \lambda_2)$ and provide an effective calculation algorithm. In this section, we discuss these practical issues.

### *4.1. Selection of Tuning Parameters*

The tuning parameters $\lambda_{1k}$'s and $\lambda_{2k}$'s should be chosen. In practice, we suggest taking $\lambda_{1j} = \lambda_1/|\beta_j^{(0)}|$ and $\lambda_{2k} = \lambda_2/\|\boldsymbol{\gamma}_k^{(0)}\|_{\boldsymbol{H}}$, and the pair $(\lambda_1, \lambda_2)$ is derived by minimizing the following BIC-type criterion:

$$BIC(\lambda_1, \lambda_2) = -2\ln L_n(\hat{\boldsymbol{\eta}}_{\lambda_1,\lambda_2}) + df_{\lambda_1,\lambda_2} \times \ln n,$$

where $\hat{\boldsymbol{\eta}}_{\lambda_1,\lambda_2}$ is the $\hat{\boldsymbol{\eta}}$ derived for given $\lambda_1, \lambda_2$, and $df_{\lambda_1,\lambda_2}$ is the number of non-zero elements of both $\hat{\boldsymbol{\theta}}$ and $(\|\hat{\gamma}_1\|_{\boldsymbol{H}}, \ldots, \|\hat{\gamma}_q\|_{\boldsymbol{H}})$.

*4.2. Computational Algorithm*

Given that $Q_n(\boldsymbol{\eta})$ is not differentiable at the origin, the standard gradient method cannot be employed. Therefore, we have devised an iterative algorithm that relies on a local quadratic approximation of the penalty function $p_\lambda(\cdot)$, similar to the approach taken by Fan and Li [16]. Let

$$\begin{cases} \boldsymbol{f}(\boldsymbol{\eta}) = \dfrac{\partial \ln L_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}, \\ \boldsymbol{\Gamma}(\boldsymbol{\eta}) = \operatorname{diag}\left\{0, \dfrac{p'_{\lambda_{12}}(|\theta_1|)}{|\theta_1|}, \ldots, \dfrac{p'_{\lambda_{1p}}(|\theta_p|)}{|\theta_p|}, \dfrac{p'_{\lambda_{21}}(\|\gamma_1\|_{\boldsymbol{H}})}{\|\gamma_1\|_{\boldsymbol{H}}}\boldsymbol{H}, \ldots, \dfrac{p'_{\lambda_{2q}}(\|\gamma_q\|_{\boldsymbol{H}})}{\|\gamma_q\|_{\boldsymbol{H}}}\boldsymbol{H}\right\}, \\ \boldsymbol{U}(\boldsymbol{\eta}) = \boldsymbol{\Gamma}(\boldsymbol{\eta})\boldsymbol{\eta}, \\ \boldsymbol{\Sigma}(\boldsymbol{\eta}) = \dfrac{\partial^2 \ln L_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\tau}. \end{cases}$$

Then, a feasible algorithm is as follows:

Step 1 Initialize $\boldsymbol{\eta}^{(0)}$.

Step 2 Update $\boldsymbol{\eta}^{(m+1)} = \boldsymbol{\eta}^{(m)} - [\boldsymbol{\Sigma}(\boldsymbol{\eta}^{(m)}) + \boldsymbol{\Gamma}(\boldsymbol{\eta}^{(m)})]^{-1}[\boldsymbol{f}(\boldsymbol{\eta}^{(m)}) - \boldsymbol{U}(\boldsymbol{\eta}^{(m)})]$.

Step 3 Iterate Step 2 until convergence, and denote the final estimators as the penalized profile quasi-likelihood estimators.

The initial value $\eta^{(0)}$ in Step 1 is obtained from the profile QMLE, i.e., $\partial \ln L_n(\boldsymbol{\eta}^{(0)})/\partial \boldsymbol{\eta} = 0$.

## 5. Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to assess the finite sample performance of our proposed method. Following the methodology employed by Li and Liang in [19], we evaluate the estimator $\hat{\boldsymbol{\theta}}$ using the generalized mean square error (GMSE), which is defined as follows:

$$\text{GMSE} = \frac{1}{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_T)^\tau [(\mathbf{1}, \boldsymbol{W}_n\boldsymbol{Y}, \boldsymbol{X})^\tau (\mathbf{1}, \boldsymbol{W}_n\boldsymbol{Y}, \boldsymbol{X})](\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_T).$$

The performance of estimator $\hat{\boldsymbol{\alpha}}(\cdot)$ is assessed using average square errors (ASE):

$$\text{ASE} = \left\{ \frac{1}{S} \sum_{s=1}^{S} \sum_{k=1}^{q} [\hat{\alpha}_k(u_s) - \alpha_{0k}(u_s)]^2 \right\},$$

where $u_s, s = 1, \ldots, S$ are the grid points at which the function $\hat{\boldsymbol{\alpha}}(\cdot)$ is evaluated. In our simulation, $S = 200$ is used.

The spatial matrix is generated from the following procedure. (i) Calculate $G_N = \text{round}(N^{0.8})$ as the number of "groups" and $m = N^{0.2}$ as the average number of individuals in each group. (ii) Generate the "group" size $n_i \sim Uniform(0.8m, 1.2m)(i = 1, \ldots, G_N)$ and adjust $n_i$ so that it satisfies $\sum_{i=1}^{G_N} n_i = N$. (iii) Normalize the matrices $\boldsymbol{W}_i(i = 1, \ldots, G_N)$ with zero for diagonal elements and $1/(n_i - 1)$ for other elements. (iv) Generate the final spatial matrix $\boldsymbol{W}_N = \operatorname{diag}\{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{G_N}\}$.

Data are generated from model (2), and we construct two examples to simulate different scenarios. Example 1 is a regular example that is primarily used to verify the asymptotic properties of penalized profile likelihood estimators. Various sample sizes, degrees of spatial dependence, distributions of disturbances, and variances in disturbances are considered. Example 2 focuses on examining the performance of the proposed method in cases where the sample size is large, and the dimensions of parametric and nonparametric components are high. Unlike Example 1, Example 2 includes covariates with AR structure and different functions. All simulation results are obtained based on 500 repetitions.

### 5.1. Example 1: Regular Scenario

Let $\boldsymbol{\beta}_0 = (2, -1, \mathbf{0}_5)^\tau$, $\boldsymbol{\alpha}_0(u) = (2\sin(2\pi u), 2\cos(2\pi u), \mathbf{0}_5)^\tau$. To simulate different spatial degrees, we choose different $\rho_0 \in \{0, 0.3, 0.7\}$. We take two distribution settings for $\varepsilon_{it}$: (i) $\varepsilon_{it} \sim N(0, \sigma_0^2)$ and (ii) $\varepsilon_{it} \sim \sqrt{\sigma_0^2/1.5} \cdot t(6)$, with $\sigma_0^2 \in \{1, 2\}$. To perform this simulation, we take the covariates $\boldsymbol{X}_{it} \sim N(\mathbf{0}, \boldsymbol{I}_7)$, $\boldsymbol{Z}_{it} \sim N(\mathbf{0}, \boldsymbol{I}_7)$ and $u_{it} \sim U(0, 1)$. The fixed effect $\mu_{0i}, i = 1, \ldots, N$ is generated from $U(0, 1)$. We use the cubic B-splines in all simulations and generate $N = 30, 60; T = 10, 15$, respectively.

We compare the performance of the variable selection procedure based on the SCAD penalty (SCAD) proposed by this paper with that based on the adaptive LASSO (ALASSO) penalty [17]. Tables 1 and 2 report the effects of variable selection under normal disturbance. The column labeled "C" indicates the average number of true zeros correctly set to zero, while the column labeled "I" shows the average number of true non-zeros incorrectly set to zero. The row labeled "Oracle" refers to the oracle estimators computed using the true model when the zero coefficients are known.

**Table 1.** Variable selection under $\epsilon_{it} \sim N(0, \sigma_0^2)$ in Example 1 ($T = 10$).

| | | | $\sigma_0^2 = 1$ | | | | | | $\sigma_0^2 = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $\rho_0$ | Method | $\hat{\boldsymbol{\theta}}$ | | | $\hat{\boldsymbol{\alpha}}(\cdot)$ | | | $\hat{\boldsymbol{\theta}}$ | | | $\hat{\boldsymbol{\alpha}}(\cdot)$ | | |
| | | | GMSE | C | I | ASE | C | I | GMSE | C | I | ASE | C | I |
| 30 | 0 | SCAD | 0.020 | 5.970 | 0 | 0.158 | 4.950 | 0 | 0.049 | 5.966 | 0 | 0.234 | 4.722 | 0 |
| | | ALASSO | 0.025 | 5.772 | 0 | 0.158 | 4.958 | 0 | 0.062 | 5.654 | 0 | 0.210 | 4.924 | 0 |
| | | Oracle | 0.020 | 6.000 | 0 | 0.155 | 5.000 | 0 | 0.044 | 6.000 | 0 | 0.207 | 5.000 | 0 |
| | 0.3 | SCAD | 0.030 | 4.822 | 0 | 0.160 | 4.940 | 0 | 0.064 | 4.802 | 0 | 0.235 | 4.726 | 0 |
| | | ALASSO | 0.028 | 4.880 | 0 | 0.157 | 4.950 | 0 | 0.064 | 4.812 | 0 | 0.214 | 4.920 | 0 |
| | | Oracle | 0.025 | 5.000 | 0 | 0.156 | 5.000 | 0 | 0.052 | 5.000 | 0 | 0.206 | 5.000 | 0 |
| | 0.7 | SCAD | 0.026 | 4.984 | 0 | 0.159 | 4.962 | 0 | 0.069 | 4.966 | 0 | 0.233 | 4.758 | 0 |
| | | ALASSO | 0.030 | 4.916 | 0 | 0.158 | 4.972 | 0 | 0.081 | 4.828 | 0 | 0.218 | 4.934 | 0 |
| | | Oracle | 0.025 | 5.000 | 0 | 0.156 | 5.000 | 0 | 0.063 | 5.000 | 0 | 0.209 | 5.000 | 0 |
| 60 | 0 | SCAD | 0.011 | 5.990 | 0 | 0.128 | 4.992 | 0 | 0.023 | 5.986 | 0 | 0.160 | 4.826 | 0 |
| | | ALASSO | 0.014 | 5.834 | 0 | 0.128 | 4.998 | 0 | 0.032 | 5.742 | 0 | 0.154 | 4.990 | 0 |
| | | Oracle | 0.012 | 6.000 | 0 | 0.128 | 5.000 | 0 | 0.024 | 6.000 | 0 | 0.152 | 5.000 | 0 |
| | 0.3 | SCAD | 0.015 | 4.944 | 0 | 0.129 | 4.980 | 0 | 0.027 | 4.896 | 0 | 0.162 | 4.778 | 0 |
| | | ALASSO | 0.016 | 4.958 | 0 | 0.129 | 5.000 | 0 | 0.030 | 4.906 | 0 | 0.154 | 4.990 | 0 |
| | | Oracle | 0.014 | 5.000 | 0 | 0.128 | 5.000 | 0 | 0.026 | 5.000 | 0 | 0.152 | 5.000 | 0 |
| | 0.7 | SCAD | 0.015 | 4.988 | 0 | 0.129 | 4.978 | 0 | 0.029 | 4.990 | 0 | 0.159 | 4.848 | 0 |
| | | ALASSO | 0.018 | 4.956 | 0 | 0.129 | 4.996 | 0 | 0.036 | 4.890 | 0 | 0.155 | 4.992 | 0 |
| | | Oracle | 0.015 | 5.000 | 0 | 0.128 | 5.000 | 0 | 0.029 | 5.000 | 0 | 0.152 | 5.000 | 0 |

From Tables 1 and 2, we can infer the following consequences. (i) As $n = N \times T$ increases, the performance of all variable selection methods, both parametric and nonparametric, converges towards the oracle procedure in terms of model error and complexity. (ii) For the parametric component, SCAD-based variable selection outperforms ALASSO-based methods, while for the nonparametric component, the reverse is true. However, overall, the effects of variable selection based on these two penalty functions are satisfactory and comparable. (iii) When the true parameter $\rho_0 = 0$, reducing the model to a non-spatial varying coefficient panel model, the proposed variable selection methods can accurately identify the true model.

Table A1 in Appendix B reports the effect of variable selection under t disturbance. We can see that the conclusions drawn from Tables 1 and 2 also hold for Table A1. Comparing Tables 1 and 2 and Table A1, the GMSE and ASE under t disturbance are slightly larger than those under normal disturbance, but this difference is so small that it can be ignored, which implies the robustness of the proposed variable selection procedure.

**Table 2.** Variable selection under $\epsilon_{it} \sim N(0, \sigma_0^2)$ in Example 1 ($T = 15$).

| $N$ | $\rho_0$ | Method | $\sigma_0^2 = 1$ | | | | | | $\sigma_0^2 = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | |
| | | | GMSE | C | I | ASE | C | I | GMSE | C | I | ASE | C | I |
| 30 | 0 | SCAD | 0.014 | 5.974 | 0 | 0.138 | 4.930 | 0 | 0.031 | 5.960 | 0 | 0.179 | 4.796 | 0 |
| | | ALASSO | 0.018 | 5.810 | 0 | 0.136 | 4.990 | 0 | 0.045 | 5.730 | 0 | 0.168 | 4.984 | 0 |
| | | Oracle | 0.015 | 6.000 | 0 | 0.135 | 5.000 | 0 | 0.030 | 6.000 | 0 | 0.166 | 5.000 | 0 |
| | 0.3 | SCAD | 0.017 | 4.918 | 0 | 0.137 | 4.944 | 0 | 0.038 | 4.874 | 0 | 0.187 | 4.702 | 0 |
| | | ALASSO | 0.019 | 4.952 | 0 | 0.135 | 4.998 | 0 | 0.046 | 4.902 | 0 | 0.172 | 4.962 | 0 |
| | | Oracle | 0.016 | 5.000 | 0 | 0.135 | 5.000 | 0 | 0.035 | 5.000 | 0 | 0.168 | 5.000 | 0 |
| | 0.7 | SCAD | 0.019 | 4.980 | 0 | 0.139 | 4.934 | 0 | 0.045 | 4.978 | 0 | 0.185 | 4.740 | 0 |
| | | ALASSO | 0.023 | 4.930 | 0 | 0.137 | 4.988 | 0 | 0.058 | 4.854 | 0 | 0.172 | 4.964 | 0 |
| | | Oracle | 0.019 | 5.000 | 0 | 0.136 | 5.000 | 0 | 0.042 | 5.000 | 0 | 0.168 | 5.000 | 0 |
| 60 | 0 | SCAD | 0.011 | 5.994 | 0 | 0.118 | 4.988 | 0 | 0.018 | 5.984 | 0 | 0.137 | 4.910 | 0 |
| | | ALASSO | 0.013 | 5.864 | 0 | 0.119 | 4.988 | 0 | 0.025 | 5.836 | 0 | 0.135 | 5.000 | 0 |
| | | Oracle | 0.012 | 6.000 | 0 | 0.118 | 5.000 | 0 | 0.019 | 6.000 | 0 | 0.134 | 5.000 | 0 |
| | 0.3 | SCAD | 0.012 | 4.974 | 0 | 0.119 | 4.994 | 0 | 0.020 | 4.942 | 0 | 0.137 | 4.892 | 0 |
| | | ALASSO | 0.013 | 4.978 | 0 | 0.119 | 5.000 | 0 | 0.024 | 4.950 | 0 | 0.136 | 4.996 | 0 |
| | | Oracle | 0.012 | 5.000 | 0 | 0.119 | 5.000 | 0 | 0.020 | 5.000 | 0 | 0.134 | 5.000 | 0 |
| | 0.7 | SCAD | 0.013 | 4.998 | 0 | 0.120 | 4.994 | 0 | 0.022 | 4.996 | 0 | 0.138 | 4.904 | 0 |
| | | ALASSO | 0.016 | 4.976 | 0 | 0.120 | 5.000 | 0 | 0.031 | 4.954 | 0 | 0.138 | 4.994 | 0 |
| | | Oracle | 0.013 | 5.000 | 0 | 0.120 | 5.000 | 0 | 0.023 | 5.000 | 0 | 0.135 | 5.000 | 0 |

### *5.2. Example 2: High-Dimension Scenario*

Let $N \times T = 100 \times 20$, $\boldsymbol{\beta}_0 = (3, -1.5, 2, 1, \mathbf{0}_{46})^\tau$, and $\boldsymbol{\alpha}_0(u) = (2\sin(2\pi u), 4u(1-u)(u-3), \ln(16u^2-1), \mathbf{0}_{12})^\tau$. The covariates $(\boldsymbol{X}_{it}^\tau, \boldsymbol{Z}_{it}^\tau)^\tau$ are generated from $N(\mathbf{0}, \mathbf{AR}_{0.5})$, where

$$\mathbf{AR}_{0.5} = \begin{bmatrix} 1 & 0.5 & \cdots & 0.5^{65} \\ 0.5 & 1 & \cdots & 0.5^{64} \\ \vdots & \vdots & \ddots & \vdots \\ 0.5^{65} & 0.5^{64} & \cdots & 1 \end{bmatrix}.$$

To save space, we only considered the case where the disturbance term follows a normal distribution in this example. The remaining settings are identical to those in Example 1.

Table A2 in Appendix B presents the outcomes of Example 2. As depicted in Table A2, the proposed procedures demonstrate robust performance despite the substantial dimensions of both the parametric and nonparametric components.

## 6. A Real Example

In this section, we use China's provincial carbon emission panel dataset to illustrate our proposed methods. The dataset contains 14 annual variables of 30 provinces in China from 2007 to 2019. The model used is an extension of the STRIPAT model of Dietz and Rosa [36], which is specified as

$$\ln y_{it} = \rho_0 \sum_{j=1}^{30} w_{ij} \ln y_{jt} + \sum_{j=1}^{8} \beta_{0j} \ln x_{itj} + \sum_{k=1}^{6} \alpha_{0k}(u_{it}) \ln z_{itk} + \mu_{0i} + \epsilon_{it} \tag{10}$$

for $i = 1, \ldots, 30; t = 1, \ldots, 13$. The corresponding variables are described in Table 3. Such a model aims to analyze the social factors that have an impact on carbon dioxide emissions, especially the affluence factor. The spatial weight matrix $\boldsymbol{W}_N = w_{ij}$ is specified by contiguity rules; that is, let

$$w_{ij} = \begin{cases} 1, & \text{province } i \text{ and } j \text{ share the common graphic boundary} \\ 0, & \text{else} \end{cases}.$$

Then, adjust $w_{ij}$ so that the diagonal elements of $\boldsymbol{W}_N$ are 0 and the row sums are 1.

**Table 3.** Description of related variables.

| First-Tier Indicators | Second-Tier Indicators | Abbreviation | Symbol |
|---|---|---|---|
| Environmental impact | Carbon emissions per capita | CE | $y_{it}$ |
| Population | Urbanization rate | UR | $x_{it1}$ |
| Science and technology | Energy intensity | EI | $x_{it2}$ |
| | R&D funding intensity | RDF | $x_{it3}$ |
| Finance | Financial Interrelation Ratio | FIR | $x_{it4}$ |
| | Financial Efficiency | FE | $x_{it5}$ |
| Transport | Freight volume per capita | FV | $x_{it6}$ |
| | Public transport passenger volume per capita | PTPV | $x_{it7}$ |
| Ecology | Percentage of forest cover | PF | $x_{it8}$ |
| Energy Structure | Coal consumption proportion | CCP | $u_{it}$ |
| Affluence | GDP per capita | GDP | $z_{it1}$ |
| | Fiscal revenue per capita | FR | $z_{it2}$ |
| | Residents consumption level | RCL | $z_{it3}$ |
| | Disposable income per capita | DI | $z_{it4}$ |
| | Import and export amount per capita | IEA | $z_{it5}$ |
| | Fixed investment per capita | FINV | $z_{it6}$ |

In previous studies, the affluence factor always had significantly positive effects on carbon dioxide emissions, and the effects were always set as a constant, see [37,38]. This phenomenon can be described by the following mechanism: among the three major industries, the secondary industry exhibits the highest dependence on energy consumption, followed by the tertiary industry. However, these two industries are also the pillar industries that drive economic development, implying a significant reliance on economic growth on energy consumption. Currently, fossil fuels constitute the primary energy source in the world, and their consumption leads to a substantial release of carbon dioxide. As indicated by existing research analyses, economic development contributes significantly to carbon emissions. It is well known that coal consumption among fossil fuels generates the most carbon dioxide, suggesting that the energy structure may directly influence the impact of economic development on carbon emissions. However, considering the significant variations in energy structures across and within different provinces, cities, and over different years, it may be unreasonable to measure this impact using a constant. Therefore, we wish to capture the potential heterogeneity of this impact using the coal consumption proportion (CCP), which is why we set the CCP as $u_{it}$ and six variables describing affluence as $z_{it1}, \ldots, z_{it6}$.

Table 4 reports the estimated coefficient parameters, and Figure 1 depicts the estimated coefficient functions. There is not much difference between the results for SCAD and ALASSO, so we focus on the SCAD results in the following analysis. The spatial coefficient $\hat{\rho}$ is not shrunk to zero and is positive, which means that neighboring provinces exhibit a substantial spatial positive correlation. The constant coefficients $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$, and $\hat{\beta}_6$ are positive, and the mean values are estimated when EI, RDF, FIR, and FV change by 1% and CE changes by 0.35%, 0.087%, 0.245%, and 0.1%, respectively. $\hat{\beta}_7$ is the only negative constant coefficient, which implies that as the public transportation passenger volume

increases, carbon emissions tend to decrease. Public transportation is a low-carbon mode of transportation. The more people choose to travel by public transportation, the fewer people use private cars, leading to a reduction in carbon emissions. This provides the evidence for the negative coefficient $\hat{\beta}_7$. The remaining coefficients are estimated to be zero, which means that they are excluded through variable selection. In addition, most of the coefficient functions in Figure 1 show an upward trend, which means that as coal consumption proportion increases, the contribution of affluence factors to carbon emission increases. Meanwhile, two function coefficients are removed through variable selection. This is generally consistent with the previous analysis in this section.

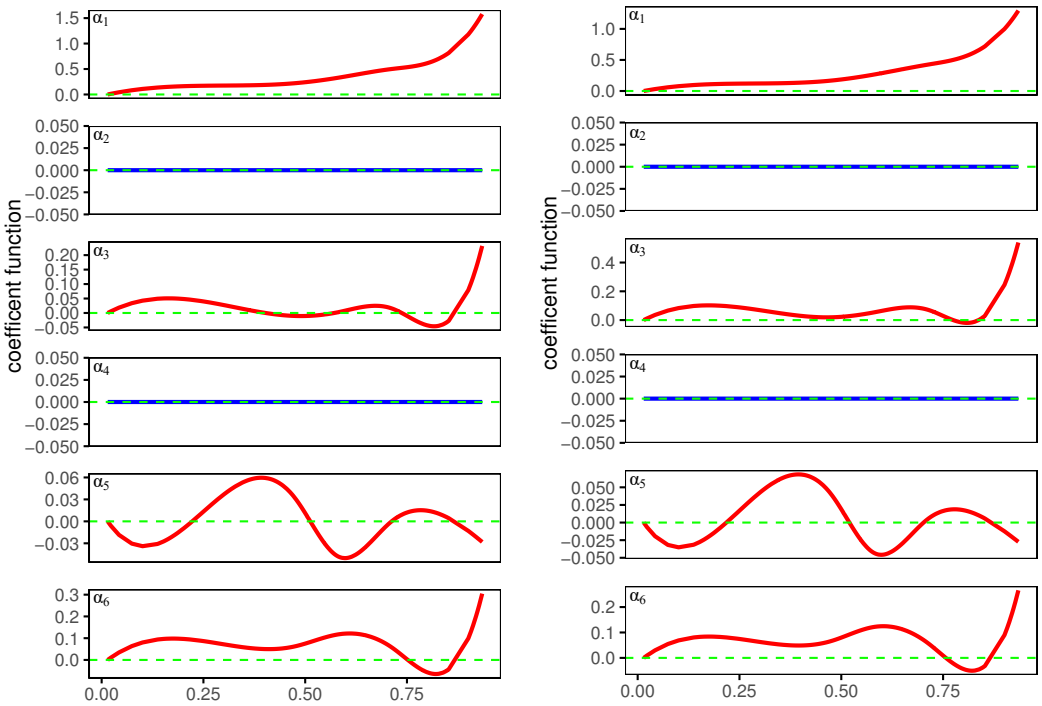



**Figure 1.** The estimated varying coefficient functions based on SCAD and ALASSO. (**a**) Varying coefficient derived from SCAD; (**b**) varying coefficient derived from ALASSO.

**Table 4.** Penalized estimators for the parametric components.

| **Method** | $\hat{\rho}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\beta}_6$ | $\hat{\beta}_7$ | $\hat{\beta}_8$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SCAD | 0.154 | 0 | 0.350 | 0.087 | 0.245 | 0 | 0.100 | −0.014 | 0 | 0.005 |
| ALASSO | 0.167 | 0 | 0.355 | 0.120 | 0.193 | 0 | 0.113 | −0.017 | 0 | 0.005 |

## 7. Discussion and Conclusions

Within the context of SVCSAR panel models with fixed effects, we developed a variable selection process that utilizes basis function approximations alongside the profile quasi-likelihood method. This approach allows for the simultaneous selection of significant variables in both parametric and nonparametric components, as well as the estimation of unknown coefficients. By selecting appropriate tuning parameters, we demonstrated that this selection process is consistent, and the estimators of constant coefficients exhibit the oracle property. Simulation results highlight the effectiveness of our proposed method in selecting variables and estimating both constant and varying coefficients. In this study, we assume that the dimensions of covariates $X$ and $Z$ remain fixed. Additionally, we also

presented simulations in Section 5 and obtained desired results when the dimensions $p$ and $q$ were large. However, it is worth noting that our variable selection process may not be applicable to scenarios involving ultra-high-dimensional covariates. As a potential area for future research, exploring variable selection techniques for SVCSAR panel models with ultra-high-dimensional covariates would be of great interest.

**Author Contributions:** Conceptualization, R.T. and M.X.; Methodology, R.T., M.X. and D.X.; Software, M.X.; Formal analysis, R.T. and M.X.; Writing—original draft, R.T., M.X. and D.X.; Supervision, R.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** R.T.'s work is supported by the Zhejiang Provincial Philosophy and Social Sciences Planning Project (No. 24NDJC014YB). D.X.'s work is supported by the Zhejiang Provincial Natural Science Foundation of China (No. LY23A010013).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Proof of Theorems

The following list summarizes some frequently used facts in the appendix.

**Fact A1.** *Let $\boldsymbol{M}_1, \boldsymbol{M}_2$ be $n \times n$ symmetric matrices, with $\boldsymbol{M}_2$ being positive semidefinite. Then, $\lambda_i(\boldsymbol{M}_1) \leq \lambda_i(\boldsymbol{M}_1 + \boldsymbol{M}_2)$ for $i = 1, \ldots, n$, where $\lambda_i$ denotes the i-th eigenvalue.*

**Fact A2.** *Let $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ be $n \times n$ matrices and uniformly bound in either row or column sums; then, the entries of their product $\boldsymbol{M}_1\boldsymbol{M}_2$ are also uniformly bound.*

Before proving the main theorems, we supplement several lemmas.

**Lemma A1.** *Let $\delta_n = \boldsymbol{A}_0 - \boldsymbol{S}_n\gamma_0$, supposing that Assumptions C1–C12 hold. Then,*

$$\frac{1}{\sqrt{n}}\boldsymbol{H}_n^\tau \boldsymbol{J}_n \delta_n = o_p(1),$$

*for $\boldsymbol{H}_n = \boldsymbol{V}_n, \boldsymbol{A}_0, \boldsymbol{X}_n$ and $\boldsymbol{W}_n\boldsymbol{Y}_n$.*

**Proof.** The proof is similar to that of Lemma 1 in Tian et al. [32]. □

**Lemma A2.** *Suppose that Assumptions C1–C4 hold, then*

$$\text{(I) } \boldsymbol{G}_n \, is uniformly bounded in either row or column sums; \text{(II) } \frac{\text{tr}\boldsymbol{G}_n}{n} = O(1); \text{(III) } \frac{\text{tr}(\boldsymbol{G}_n^\tau \boldsymbol{G}_n)}{n} = O(1); \text{(IV) } \frac{\text{tr}\boldsymbol{J}_n}{n} = \frac{T-1}{T}; \text{(V) } \frac{\text{tr}(\boldsymbol{G}_n^\tau \boldsymbol{J}_n)}{n} = \frac{T-1}{T}\frac{\text{tr}\boldsymbol{G}_n}{n}; \text{(VI) } \frac{\text{tr}(\boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{G}_n)}{n} = \frac{T-1}{T}\frac{\text{tr}\boldsymbol{G}_n^\tau \boldsymbol{G}_n}{n}.$$

**Proof.** The proof method is similar to that of Lemma 3 in Tian et al. [32]; therefore, will not be elaborated here. □

**Lemma A3.** *Suppose that Assumptions C1–C12 hold, then*

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}} = O_p(1).$$

**Proof.** The first order partial derivatives of the profile likelihood function is

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \sigma^2} &= -\frac{\sqrt{n}}{2\sigma_T^2} + \frac{1}{2\sigma_T^4\sqrt{n}}\boldsymbol{V}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1)\\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \rho} &= -\frac{1}{\sqrt{n}}\mathrm{tr}\boldsymbol{G}_n + \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{V}_n^\tau \boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1)\\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1).\\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \gamma} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1).
\end{aligned}$$

The variance of $n^{-1/2}\partial \ln L_n(\boldsymbol{\eta}_T)/\partial\sigma^2$ is

$$\begin{aligned}
\mathrm{var}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \sigma^2}\right) &= \frac{1}{4\sigma_T^8 n}\mathrm{var}(\boldsymbol{V}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n) + o(1)\\
&= \frac{1}{4\sigma_T^8}\left\{\frac{(T-1)^2(\mu_4^* - 3\sigma_0^4)}{T^2} + \frac{2(T-1)\sigma_0^4}{T}\right\} + o(1)\cdot\\
&= O(1)
\end{aligned}$$

Therefore, $n^{-1/2}\partial \ln L_n(\boldsymbol{\eta}_T)/\partial\sigma^2 = O_p(1)$. The variance of $n^{-1/2}\partial \ln L_n(\boldsymbol{\eta}_T)/\partial\rho$ is

$$\begin{aligned}
\mathrm{var}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \rho}\right) &\le \frac{2}{\sigma_T^4 n}[\mathrm{var}(\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n) + \mathrm{var}(\boldsymbol{V}_n^\tau \boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n)] + o(1)\\
&\le \frac{2}{\sigma_T^4}\left[\sigma_0^2\lambda_c + \frac{(T-1)^2(\mu_4^* - 3\sigma_0^4)}{T^2} + \frac{(T-1)\sigma_0^4\mathrm{tr}[(\boldsymbol{G}_n + \boldsymbol{G}_n^\tau)\boldsymbol{G}_n)]}{T}\right].\\
&= O(1).
\end{aligned}$$

Thus, $n^{-1/2}\partial \ln L_n(\boldsymbol{\eta}_T)/\partial\rho = O_p(1)$. Given the uniform boundedness of the elements of $\boldsymbol{X}_n$ and $\boldsymbol{S}_n$ are uniformly bound for all $n$, as stated in C6, it is evident that

$$\frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n = O_p(1), \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n = O_p(1).$$

□

**Lemma A4.** *Suppose that Assumptions C1–C12 hold, then*

$$\frac{1}{n}\left\{\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\tau} - E\left[\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\tau}\right]\right\} = o_p(1).$$

**Proof.** Utilizing a similar argument as in Theorem 3.2 of Lee [31], Lemma A4 is proven. □

**Proof of Theorem 1.** Let $\Delta_n = a_n + n^{-r/(2r+1)}, \boldsymbol{\theta} = \boldsymbol{\theta}_T + \Delta_n\boldsymbol{\zeta}_1, \gamma = \gamma_0 + \Delta_n\boldsymbol{\zeta}_2$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^\tau, \boldsymbol{\zeta}_2^\tau)^\tau$. It is sufficient to show that, for any given $\varepsilon > 0$, there exists a large constant $C$ such that

$$P\left\{\sup_{\|\boldsymbol{\zeta}\|=C} Q(\boldsymbol{\theta}, \gamma) < Q(\boldsymbol{\theta}_T, \gamma_0)\right\} > 1 - \varepsilon. \tag{A1}$$

Let $D(\boldsymbol{\zeta}) = Q(\boldsymbol{\theta}, \boldsymbol{\gamma}) - Q(\boldsymbol{\theta}_T, \boldsymbol{\gamma}_0)$; then, with a simple calculation, we see that

$$\begin{aligned} D(\boldsymbol{\zeta}) =& \Delta_n \left( \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}} \right)^\tau \boldsymbol{\zeta} + \frac{1}{2} \Delta_n^2 \boldsymbol{\zeta}^\tau \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\tau} \boldsymbol{\zeta} (1 + o_p(1)) \\ &+ n \sum_{j=2}^{p+2} [p_{\lambda_{1j}}(|\theta_j|) - p_{\lambda_{1j}}(|\theta_{Tj}|)] + n \sum_{k=1}^{q} [p_{\lambda_{2k}}(\|\boldsymbol{\gamma}_k\|_{\boldsymbol{H}}) - p_{\lambda_{2k}}(\|\boldsymbol{\gamma}_{0j}\|_{\boldsymbol{H}})] \\ \equiv& D_1 + D_2 + D_3 + D_4. \end{aligned}$$

It follows by Lemma A3 and the Cauchy inequality that

$$\begin{aligned} |D_1| = \left| \Delta_n \left( \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}} \right)^\tau \boldsymbol{\zeta} \right| &\le \Delta_n \sqrt{n} \left\| \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}} \right\| \cdot \|\boldsymbol{\zeta}\| \\ &= \|\boldsymbol{\zeta}\| O_p(\sqrt{n K_n} \Delta_n) = \|\boldsymbol{\zeta}\| O_p(n \Delta_n^2). \end{aligned}$$

According to Lemma A4, a simple calculation yields

$$D_2 = \frac{1}{2} \Delta_n^2 \boldsymbol{\zeta}^\tau \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\tau} \boldsymbol{\zeta} (1 + o_p(1)) = -n \Delta_n^2 \boldsymbol{\zeta}^\tau \boldsymbol{\Sigma}_{n,\boldsymbol{\eta}} \boldsymbol{\zeta} (1 + o_p(1)).$$

Hence, by choosing a sufficiently large $C$, $D_2$ dominates $D_1$ uniformly in $\|\boldsymbol{\zeta}\| = C$. Furthermore, invoking C13 and $p_{\lambda_{1j}}(0) = 0$, and by the standard argument of the Taylor expansion, we obtain that

$$\begin{aligned} D_3 &\le n \sum_{j=2}^{s} [p_{\lambda_{1j}}(|\theta_j|) - p_{\lambda_{1j}}(|\theta_{Tj}|)] \\ &\le n \sum_{j=2}^{s} [n \Delta_n p'_{\lambda_{1j}}(|\theta_{Tj}|) \mathrm{sgn}(\theta_{Tj}) |\zeta_{1j}| + n \Delta_n^2 p''_{\lambda_{1j}}(|\theta_{Tj}|) |\zeta_{1j}|^2 (1 + o_p(1))] \\ &\le \sqrt{s} n \Delta_n a_n \|\boldsymbol{\zeta}\| + n \Delta_n^2 b_n \|\boldsymbol{\zeta}\|^2 \\ &= \|\boldsymbol{\zeta}\| O_p(n \Delta_n^2) + \|\boldsymbol{\zeta}\| o_p(n \Delta_n^2). \end{aligned}$$

Then, it is easy to show that $D_3$ is dominated by $D_2$ uniformly in $\|\boldsymbol{\zeta}\| = C$. With the same argument, we can prove that $D_4$ is dominated by $D_2$. Hence, by choosing a sufficiently large $C$, (A1) holds. This implies that a local maximizer $\hat{\boldsymbol{\gamma}}$ exists such that $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\Delta_n) = O_p(a_n + n^{-r/(2r+1)})$. Then, $\|\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}\| = O_p(a_n + n^{-r/(2r+1)})$. Let $\delta_k = \alpha_k(u) - \boldsymbol{B}(u)^\tau \boldsymbol{\gamma}_{0k}, k = 1, \ldots, q$. Note that

$$\begin{aligned} \|\hat{\alpha}_k(u) - \alpha_{0k}(u)\|^2 &= \int_0^1 [\hat{\alpha}_k(u) - \alpha_{0k}(u)]^2 \mathrm{d}u \\ &= \int_0^1 [\boldsymbol{B}^{\boldsymbol{\tau}}(u) \hat{\boldsymbol{\gamma}}_k - \boldsymbol{B}^{\boldsymbol{\tau}}(u) \boldsymbol{\gamma}_{0k} - \delta_k]^2 \mathrm{d}u \\ &\le 2 \int_0^1 [\boldsymbol{B}^{\boldsymbol{\tau}}(u)(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})]^2 \mathrm{d}u + 2 \int_0^1 \delta_k^2 \mathrm{d}u \\ &= (\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})^\tau \boldsymbol{H} (\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}) + 2 \int_0^1 \delta_k^2 \mathrm{d}u, \end{aligned}$$

and by invoking $\|\boldsymbol{H}\| = O(1)$, a simple calculation yields

$$(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})^\tau \boldsymbol{H} (\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}) = O_p(a_n^2 + n^{-2r/(2r+1)}).$$

Together with

$$\int_0^1 \delta_k^2 \mathrm{d}u = O_p(n^{-2r/(2r+1)}),$$

the proof is completed. □

**Proof of Theorem 2.** We begin by proving part (i). Given $\lambda_{\max} \to 0$, it follows easily that $a_n = 0$ for large $n$. Then, according to Theorem 1, it suffices to show that for any $\theta_j$ satisfying $|\theta_j - \theta_{0j}| = O_p(n^{-r/(2r+1)}), j = 1, \ldots, s$ and any $\boldsymbol{\gamma}$ satisfying $\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| = O_p(n^{-r/(2r+1)})$, with some given small $\varepsilon = Cn^{-r/(2r+1)}$, as $n \to \infty$, the probability tends to 1 that

$$\frac{\partial Q_n(\boldsymbol{\eta})}{\partial \theta_j} < 0, \quad \text{for } 0 < \theta_j < \varepsilon, j = s+1, \ldots, p+2, \tag{A2}$$

and

$$\frac{\partial Q_n(\boldsymbol{\eta})}{\partial \theta_j} > 0, \quad \text{for } -\varepsilon < \theta_j < 0, j = s+1, \ldots, p+2. \tag{A3}$$

Thus, (A2) and (A3) imply that the maximizer of $\boldsymbol{Q}_n(\boldsymbol{\eta})$ attains $\theta_j = 0, j = s+1, \ldots, p+2$. Using a similar argument to the proof of Theorem 1, we obtain that

$$\begin{aligned}
\frac{\partial Q_n(\boldsymbol{\eta})}{\partial \theta_j} =& \frac{\partial \ln L_n(\boldsymbol{\eta})}{\partial \theta_j} - n p'_{\lambda_{1j}}(|\theta_{0j}|)\mathrm{sgn}(\theta_{0j}) \\
=& \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j} + \sum_{k=1} \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j \partial \theta_k}(\eta_k - \eta_{Tk}) \\
&+ \sum_{k=1}\sum_{l=1} \frac{\partial^3 \ln L_n(\boldsymbol{\eta}^*)}{\partial \theta_j \partial \theta_k \partial \theta_l}(\eta_k - \eta_{Tk})(\eta_l - \eta_{Tl}) - n p'_{\lambda_{1j}}(|\theta_j|)\mathrm{sgn}(\theta_j),
\end{aligned}$$

where $\boldsymbol{\eta}^*$ lies between $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_T$. From Lemmas A1 and A2 and assumption C9, we have

$$\begin{aligned}
&\frac{1}{n}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j} = O_p(n^{-\frac{1}{2}}), \\
&\frac{1}{n}\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\tau} = E\left[\frac{1}{n}\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\tau}\right] + o_p(1), \\
&\frac{1}{n}\frac{\partial^3 \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j \partial \eta_k \partial \eta_l} = O_p(1).
\end{aligned}$$

Then,

$$\begin{aligned}
\frac{\partial Q_n(\boldsymbol{\eta})}{\partial \theta_j} =& n\lambda_{1j}\Bigg\{\lambda_{1j}^{-1}\frac{1}{n}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j} + \lambda_{1j}^{-1}\frac{1}{n}\sum_{k=1}\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \theta_j \partial \theta_k}(\eta_k - \eta_{Tk}) \\
&+ \lambda_{1j}^{-1}\frac{1}{n}\sum_{k=1}\sum_{l=1}\frac{\partial^3 \ln L_n(\boldsymbol{\eta}^*)}{\partial \theta_j \partial \theta_k \partial \theta_l}(\eta_k - \eta_{Tk})(\eta_l - \eta_{Tl}) - \lambda_{1j}^{-1} p'_{\lambda_{1j}}(|\theta_j|)\mathrm{sgn}(\theta_j)\Bigg\} \\
=& n\lambda_{1j}\Big\{\lambda_{1j}^{-1}O_p\left(n^{\frac{-r}{2r+1}}\right) - \lambda_{1j}^{-1} p'_{\lambda_{1j}}(|\theta_j|)\mathrm{sgn}(\theta_j)\Big\}.
\end{aligned}$$

Since $\lim_{n\to\infty} \liminf_{\theta_j \to 0} \lambda_{1j}^{-1} p'_{\lambda_{1j}}(|\theta_j|) > 0$ and $n^{r/(2r+1)}\lambda_{1j} > n^{r/(2r+1)}\lambda_{\min} \to \infty$, the sign of the derivation is completely determined by that of $\theta_j$; then, (A2) and (A3) hold.

By applying similar techniques to our analysis of part (i) in this theorem, we find that with probability tending to 1, $\hat{\gamma}_k = 0, k = d+1, \ldots, q$. Then, using the fact that $\sup_u \|\boldsymbol{B}(u)\| = O(1)$, the result of this theorem follows immediately from $\hat{\alpha}_k(u) = \boldsymbol{B}^\tau(u)\hat{\boldsymbol{\gamma}}_k$. □

**Proof of Theorem 3.** Let $\gamma^* = (\gamma_1^\tau, \ldots, \gamma_d^\tau)^\tau$, $\gamma_0^* = (\gamma_{01}^\tau, \ldots, \gamma_{0d}^\tau)^\tau$, $\boldsymbol{\eta}^* = (\boldsymbol{\theta}^*, \gamma^*)$, and $\boldsymbol{\eta}_T^* = (\boldsymbol{\theta}_T^*, \gamma_0^*)$. In order to obtain the asymptotic distribution, we first write the components of $n^{-1/2}\partial \ln L_n(\boldsymbol{\eta}_T)/\partial \boldsymbol{\eta}$ as follows:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \sigma^2} &= -\frac{\sqrt{n}}{2\sigma_T^2} + \frac{1}{2\sigma_T^4\sqrt{n}}(\boldsymbol{\delta}_n + \boldsymbol{V}_n)^\tau \boldsymbol{J}_n(\boldsymbol{\delta}_n + \boldsymbol{V}_n) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \rho} &= -\frac{1}{\sqrt{n}}\mathrm{tr}\boldsymbol{G}_n + \frac{1}{\sigma_T^2\sqrt{n}}(\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + \boldsymbol{V}_n^\tau \boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + (\boldsymbol{W}_n \boldsymbol{Y}_n)^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{X}_n^\tau \boldsymbol{J}_n(\boldsymbol{\delta}_n + \boldsymbol{V}_n) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \gamma} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{S}_n^\tau \boldsymbol{J}_n(\boldsymbol{\delta}_n + \boldsymbol{V}_n),
\end{aligned}$$

where $\boldsymbol{\delta}_n = \boldsymbol{A}_0 - \boldsymbol{S}_n\gamma_0$. By Lemma A1, $\frac{1}{\sqrt{n}}\boldsymbol{H}_n^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n = o_p(1)$ for $\boldsymbol{H}_n = \boldsymbol{V}_n, \boldsymbol{\delta}_n, \boldsymbol{X}_n$ and $\boldsymbol{W}_n\boldsymbol{Y}_n$, and the formula above is rewritten as follows:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \sigma^2} &= -\frac{\sqrt{n}}{2\sigma_T^2} + \frac{1}{2\sigma_T^4\sqrt{n}}\boldsymbol{V}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \rho} &= -\frac{1}{\sqrt{n}}\mathrm{tr}\boldsymbol{G}_n + \frac{1}{\sigma_T^2\sqrt{n}}(\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + \boldsymbol{V}_n^\tau \boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n) + o_p(1) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1) \\
\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \gamma} &= \frac{1}{\sigma_T^2\sqrt{n}}\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{V}_n + o_p(1).
\end{aligned} \tag{A4}$$

Let $\mu_3^* = E|\epsilon_{it}|^3$ and $\mu_4^* = E|\epsilon_{it}|^4$, $\boldsymbol{T}_n^\tau \boldsymbol{J}_n \boldsymbol{T}_n/n \equiv \boldsymbol{\Phi}_{\boldsymbol{TT}}$ for $\boldsymbol{T}_n = \boldsymbol{X}_n, \boldsymbol{R}_n$ or $\boldsymbol{S}_n$. Then, using Lemma A2 we write the following:

$$\begin{aligned}
E\left[-\frac{1}{n}\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}^\tau}\right] &= \begin{bmatrix}
\frac{\sigma_0^2 \mathrm{tr}\boldsymbol{J}_n}{n\sigma_T^6} - \frac{1}{2\sigma_T^4} + \frac{\boldsymbol{\delta}_n^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n}{n\sigma_T^6} & * & * & * \\
\frac{\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n}{n\sigma_T^4} + \frac{\sigma_0^2 \mathrm{tr}(\boldsymbol{G}_n^\tau \boldsymbol{J}_n)}{n\sigma_T^4} & \frac{\boldsymbol{R}_n^\tau \boldsymbol{J}_n \boldsymbol{R}_n}{n\sigma_T^2} + \frac{\mathrm{tr}\boldsymbol{G}_n^2}{n} + \frac{\sigma_0^2 \mathrm{tr}(\boldsymbol{G}_n^\tau \boldsymbol{J}_n \boldsymbol{G}_n)}{n\sigma_T^2} & * & * \\
\frac{\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n}{n\sigma_T^4} & \frac{\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{R}_n}{n\sigma_T^2} & \frac{\boldsymbol{X}_n^\tau \boldsymbol{J}_n \boldsymbol{X}_n}{n\sigma_T^2} & * \\
\frac{\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{\delta}_n}{n\sigma_T^4} & \frac{\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{R}_n}{n\sigma_T^2} & \frac{\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{X}_n}{n\sigma_T^2} & \frac{\boldsymbol{S}_n^\tau \boldsymbol{J}_n \boldsymbol{S}_n}{n\sigma_T^2}
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{1}{2\sigma_T^4} & * & * & * \\
\frac{\mathrm{tr}\boldsymbol{G}_n}{n\sigma_T^2} & \frac{\boldsymbol{\Phi}_{\boldsymbol{RR}}}{\sigma_T^2} + \frac{\mathrm{tr}(\boldsymbol{G}_n^2 + \boldsymbol{G}_n^\tau \boldsymbol{G}_n)}{n} & * & * \\
0 & \frac{\boldsymbol{\Phi}_{\boldsymbol{XR}}}{\sigma_T^2} & \frac{\boldsymbol{\Phi}_{\boldsymbol{XX}}}{\sigma_T^2} & * \\
0 & \frac{\boldsymbol{\Phi}_{\boldsymbol{SR}}}{\sigma_T^2} & \frac{\boldsymbol{\Phi}_{\boldsymbol{SX}}}{\sigma_T^2} & \frac{\boldsymbol{\Phi}_{\boldsymbol{SS}}}{\sigma_T^2}
\end{bmatrix} + o(1) \equiv \boldsymbol{\Sigma}_{n,\boldsymbol{\eta}} + o(1),
\end{aligned} \tag{A5}$$

$$
\begin{aligned}
&E\left[\frac{1}{n}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\eta}^\tau}\right] \\
&= \begin{bmatrix}
\frac{T}{T-1}\frac{1}{2\sigma_T^4}+\frac{\mu_4^*-3\sigma_0^4}{4\sigma_T^4\sigma_0^4} & * & * & * \\
\frac{T}{T-1}\frac{\mathrm{tr}\boldsymbol{G}_n}{n\sigma_T^2}+\frac{(\mu_4^*-3\sigma_0^4)\mathrm{tr}\boldsymbol{G}_n}{2n\sigma_T^2\sigma_0^4} & \frac{T}{T-1}\left(\frac{\boldsymbol{\Phi}_{RR}}{\sigma_T^2}+\frac{\mathrm{tr}(\boldsymbol{G}_n\boldsymbol{G}_n+\boldsymbol{G}_n^\tau\boldsymbol{G}_n)}{n}\right) & & \\
\quad+\frac{\mu_3^*\boldsymbol{R}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \quad+\frac{(\mu_4^*-3\sigma_0^4)\sum g_{n,ii}^2}{4\sigma_T^4\sigma_0^4}+\frac{2\mu_3^*\boldsymbol{R}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & * & * \\
\frac{\mu_3^*\boldsymbol{X}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \frac{T}{T-1}\frac{\boldsymbol{\Phi}_{XR}}{\sigma_T^2}+\frac{\mu_3^*\boldsymbol{X}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & \frac{T}{T-1}\frac{\boldsymbol{\Phi}_{XX}}{\sigma_T^2} & * \\
\frac{\mu_3^*\boldsymbol{S}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \frac{T}{T-1}\frac{\boldsymbol{\Phi}_{SR}}{\sigma_T^2}+\frac{\mu_3^*\boldsymbol{X}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & \frac{T}{T-1}\frac{\boldsymbol{\Phi}_{SX}}{\sigma_T^2} & \frac{T}{T-1}\frac{\boldsymbol{\Phi}_{SS}}{\sigma_T^2}
\end{bmatrix}+o(1) \\
&= \frac{T}{T-1}(\boldsymbol{\Sigma}_{n,\boldsymbol{\eta}}+\boldsymbol{\Psi}_{n,\boldsymbol{\eta}})+o(1),
\end{aligned}
\tag{A6}
$$

where $g_{n,ii}$ is the $i$-th diagonal element of $\boldsymbol{G}_n$ and

$$
\boldsymbol{\Psi}_{n,\boldsymbol{\theta}} = \frac{T-1}{T}\begin{bmatrix}
\frac{\mu_4^*-3\sigma_0^4}{4\sigma_T^4\sigma_0^4} & * & * & * \\
\frac{(\mu_4^*-3\sigma_0^4)\mathrm{tr}\boldsymbol{G}_n}{2n\sigma_T^2\sigma_0^4}+\frac{\mu_3^*\boldsymbol{R}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \frac{(\mu_4^*-3\sigma_0^4)\sum g_{n,ii}^2}{4\sigma_T^4\sigma_0^4}+\frac{2\mu_3^*\boldsymbol{R}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & * & * \\
\frac{\mu_3^*\boldsymbol{X}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \frac{\mu_3^*\boldsymbol{X}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & 0 & * \\
\frac{\mu_3^*\boldsymbol{S}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{J}_n)}{2n\sigma_T^6} & \frac{\mu_3^*\boldsymbol{S}_n^\tau\boldsymbol{J}_n\mathrm{diag}(\boldsymbol{G}_n^\tau\boldsymbol{J}_n)}{n\sigma_T^4} & 0 & 0
\end{bmatrix}.
\tag{A7}
$$

To derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, we divide $\boldsymbol{\Sigma}_{n,\boldsymbol{\eta}}$ into four block matrices, which correspond to the second-order derivatives of the likelihood function with respect to $\boldsymbol{\theta}$, the cross-partial derivatives of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, and the second-order derivatives of $\boldsymbol{\gamma}$. The matrices are as follows:

$$
\boldsymbol{\Sigma}_{n,\boldsymbol{\eta}} \equiv \begin{bmatrix}
\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}^\tau \\
\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}} & \boldsymbol{\Sigma}_{n,\boldsymbol{\gamma}},
\end{bmatrix}
$$

where

$$
\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}} \equiv \begin{bmatrix}
\frac{1}{2\sigma_T^4} & * & * \\
\frac{\mathrm{tr}\boldsymbol{G}_n}{n\sigma_T^2} & \frac{\boldsymbol{\Phi}_{RR}}{\sigma_T^2}+\frac{\mathrm{tr}(\boldsymbol{G}_n^2+\boldsymbol{G}_n^\tau\boldsymbol{G}_n)}{n} & * \\
0 & \frac{\boldsymbol{\Phi}_{XR}}{\sigma_T^2} & \frac{\boldsymbol{\Phi}_{XX}}{\sigma_T^2}
\end{bmatrix}, \boldsymbol{\Sigma}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}} \equiv \begin{bmatrix}
0, \frac{\boldsymbol{\Phi}_{SR}}{\sigma_T^2} & \frac{\boldsymbol{\Phi}_{SX}}{\sigma_T^2}
\end{bmatrix}, \boldsymbol{\Sigma}_{n,\boldsymbol{\gamma}} \equiv \begin{bmatrix}
\frac{\boldsymbol{\Phi}_{XX}}{\sigma_T^2}
\end{bmatrix}.
$$

Let $\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^*$ be the first $s \times s$ upper-left submatrix of $\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}^*$ be the first $(K_n+l+1)d \times s$ upper-left submatrix of $\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}$, and $\boldsymbol{\Sigma}_{n,\boldsymbol{\gamma}}^*$ be the first $(K_n+l+1)d \times (K_n+l+1)d$ upper-left submatrix of $\boldsymbol{\Sigma}_{n,\boldsymbol{\gamma}}$. Using the same argument, we partition $\boldsymbol{\Psi}_{n,\boldsymbol{\eta}}$ into four block matrices, that is

$$
\boldsymbol{\Psi}_{n,\boldsymbol{\eta}} \equiv \begin{bmatrix}
\boldsymbol{\Psi}_{n,\boldsymbol{\theta}} & \boldsymbol{\Psi}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}^\tau \\
\boldsymbol{\Psi}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}} & \boldsymbol{\Psi}_{n,\boldsymbol{\gamma}}
\end{bmatrix},
\tag{A8}
$$

and obtain submatrices $\boldsymbol{\Psi}_{n,\boldsymbol{\theta}}^*$, $\boldsymbol{\Psi}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}^*$ and $\boldsymbol{\Psi}_{n,\boldsymbol{\gamma}}^*$. Let the notation of matrices without subscripts "n" represent the limiting version of the matrices. For example, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^* = \lim_{n\to\infty}\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^*$. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^* - \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\gamma}}^{*\tau}(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^*)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\gamma}}^*$, and we assume that $\boldsymbol{\Omega}$ is non-singular.

According to Theorems 1 and 2, as $n \to \infty$, with probability tending to 1, $Q_n(\boldsymbol{\eta})$ achieves its maximal value at $(\hat{\boldsymbol{\theta}}^{*\tau}, \mathbf{0})^\tau$ and $(\hat{\boldsymbol{\gamma}}^{*\tau}, \mathbf{0})^\tau$; then, $(\hat{\boldsymbol{\theta}}^{*\tau}, \mathbf{0})^\tau$ and $(\hat{\boldsymbol{\gamma}}^{*\tau}, \mathbf{0})^\tau$ must satisfy the following:

$$\frac{1}{\sqrt{n}}\frac{\partial Q_n((\hat{\boldsymbol{\theta}}^{*\tau},\mathbf{0})^\tau,(\hat{\boldsymbol{\gamma}}^{*\tau},\mathbf{0})^\tau)}{\partial \boldsymbol{\eta}}=0.$$

Applying the Taylor expansion, we have

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\theta}^*}\\ \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*}\end{pmatrix}-\left\{\frac{1}{n}\begin{pmatrix}\frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\theta}^*\partial \boldsymbol{\theta}^{*\tau}} & \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^{*\tau}\partial \boldsymbol{\theta}^*}\\ \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*\partial \boldsymbol{\theta}^{*\tau}} & \frac{\partial^2 \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*\partial \boldsymbol{\gamma}^{*\tau}}\end{pmatrix}+o_p(1)\right\}\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_T\\ \hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\end{pmatrix}+\begin{pmatrix}\boldsymbol{P}_{\boldsymbol{\theta}}\\ \boldsymbol{P}_{\boldsymbol{\gamma}}\end{pmatrix}=\mathbf{0}, \tag{A9}$$

where

$$\boldsymbol{P}_\theta=\begin{pmatrix}\sqrt{n}p'_{\lambda_{11}}(|\hat{\theta}_1|)\mathrm{sgn}(\hat{\theta}_1)\\ \vdots\\ \sqrt{n}p'_{\lambda_{1s}}(|\hat{\theta}_s|)\mathrm{sgn}(\hat{\theta}_s)\end{pmatrix},\boldsymbol{P}_\gamma=\begin{pmatrix}\sqrt{n}p'_{\lambda_{21}}(\|\hat{\gamma}_1\|_{\boldsymbol{H}})\frac{\boldsymbol{H}\hat{\gamma}_1}{\|\hat{\gamma}_1\|_{\boldsymbol{H}}}\\ \vdots\\ \sqrt{n}p'_{\lambda_{2d}}(\|\hat{\gamma}_d\|_{\boldsymbol{H}})\frac{\boldsymbol{H}\hat{\gamma}_d}{\|\hat{\gamma}_d\|_{\boldsymbol{H}}}\end{pmatrix}.$$

Applying the Taylor expansion to $p'_{\lambda_{1j}}(|\hat{\theta}_j|)$, we obtain that

$$p'_{\lambda_{1j}}(|\hat{\theta}_j|)=p'_{\lambda_{1j}}(|\theta_{0j}|)+\{p''_{\lambda_{1j}}(|\theta_{0j}|)+o_p(1)\}(\hat{\theta}_j-\theta_{0j}).$$

Furthermore, Assumption C13 implies that $p''_{\lambda_{1j}}(|\theta_{0j}|)=o_p(1)$, and note that $p'_{\lambda_{1j}}(|\theta_{0j}|)=0$ as $\lambda_{\max}\to 0$, then $\boldsymbol{P}_{\boldsymbol{\theta}}=o_p(\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}_T^*))$. Using similar arguments, we can prove that $\boldsymbol{P}_\gamma=o_p(\sqrt{n}(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}_0^*))$. Together with Lemma A4 and Equation (A9), a simple calculation yields the following:

$$\begin{aligned}&\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\theta}^*}-[\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}}+o_p(1)]\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}_T^*)-[\boldsymbol{\Sigma}^{*\tau}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}+o_p(1)]\sqrt{n}(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}_0^*)=0\\ &\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*}-[\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}+o_p(1)]\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}_T^*)-[\boldsymbol{\Sigma}^*_{n,\boldsymbol{\gamma}}+o_p(1)]\sqrt{n}(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}_0^*)=0.\end{aligned}$$

By substitution, the term $\sqrt{n}(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}_0^*)$ is eliminated, yielding

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}_T^*)=[\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}}-\boldsymbol{\Sigma}^{*\tau}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}(\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}\boldsymbol{\gamma}})^{-1}\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}+o_p(1)]^{-1}[\boldsymbol{I}_s,-\boldsymbol{\Sigma}^{*\tau}_{n,\boldsymbol{\theta}\boldsymbol{\gamma}}(\boldsymbol{\Sigma}^*_{n,\boldsymbol{\theta}\boldsymbol{\gamma}})^{-1}]\begin{pmatrix}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\theta}^*}\\ \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*}\end{pmatrix}+o_p(1). \tag{A10}$$

Furthermore, the central limit theorem for linear–quadratic forms [33] shows that

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\theta}^*}\\ \frac{\partial \ln L_n(\boldsymbol{\eta}_T)}{\partial \boldsymbol{\gamma}^*}\end{pmatrix}\xrightarrow{D}N\left(\mathbf{0},\frac{T}{T-1}\begin{bmatrix}\boldsymbol{\Sigma}^*_{\boldsymbol{\theta}}+\boldsymbol{\Psi}^*_{\boldsymbol{\theta}} & \boldsymbol{\Sigma}^{*\tau}_{\boldsymbol{\theta}\boldsymbol{\gamma}}+\boldsymbol{\Psi}^{*\tau}_{\boldsymbol{\theta}\boldsymbol{\gamma}}\\ \boldsymbol{\Sigma}^*_{\boldsymbol{\theta}\boldsymbol{\gamma}}+\boldsymbol{\Psi}^*_{\boldsymbol{\theta}\boldsymbol{\gamma}} & \boldsymbol{\Sigma}^*_{\boldsymbol{\gamma}}+\boldsymbol{\Psi}^*_{\boldsymbol{\gamma}},\end{bmatrix}\right) \tag{A11}$$

where the definition of the symbols in the covariance matrix is located in the context of (A8). Then, invoking (A10) and (A11), and using the Slutsky theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}_T^*)\to N\left(\mathbf{0},\frac{T}{T-1}\left\{\boldsymbol{\Omega}^{-1}+\boldsymbol{\Omega}^{-1}[\boldsymbol{\Psi}^*_{\boldsymbol{\theta}}-2\boldsymbol{\Sigma}^{*\tau}_{\boldsymbol{\theta}\boldsymbol{\gamma}}(\boldsymbol{\Sigma}^*_{\boldsymbol{\gamma}})^{-1}\boldsymbol{\Psi}^*_{\boldsymbol{\theta}}]\boldsymbol{\Omega}^{-1}\right\}\right).$$

□

# Appendix B. Supplementary Simulation Results

**Table A1.** Variable selection under $\epsilon_{it} \sim \sqrt{\sigma_0^2/1.5} \cdot t(6)$ in Example 1.

| $N \times T$ | $\rho_0$ | Method | $\sigma_0^2 = 1$ | | | | | | $\sigma_0^2 = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | |
| | | | GMSE | C | I | ASE | C | I | GMSE | C | I | ASE | C | I |
| $30 \times 10$ | 0 | SCAD | 0.027 | 5.944 | 0 | 0.161 | 4.944 | 0 | 0.076 | 5.938 | 0 | 0.234 | 4.716 | 0 |
| | | ALASSO | 0.032 | 5.748 | 0 | 0.160 | 4.960 | 0 | 0.086 | 5.630 | 0 | 0.211 | 4.914 | 0 |
| | | Oracle | 0.027 | 6.000 | 0 | 0.157 | 5.000 | 0 | 0.071 | 6.000 | 0 | 0.206 | 5.000 | 0 |
| | 0.3 | SCAD | 0.032 | 4.846 | 0 | 0.160 | 4.962 | 0 | 0.084 | 4.798 | 0 | 0.237 | 4.734 | 0 |
| | | ALASSO | 0.032 | 4.902 | 0 | 0.158 | 4.968 | 0 | 0.086 | 4.830 | 0 | 0.213 | 4.944 | 0 |
| | | Oracle | 0.029 | 5.000 | 0 | 0.157 | 5.000 | 0 | 0.074 | 5.000 | 0 | 0.207 | 5.000 | 0 |
| | 0.7 | SCAD | 0.032 | 4.970 | 0 | 0.158 | 4.954 | 0 | 0.092 | 4.966 | 0 | 0.236 | 4.750 | 0 |
| | | ALASSO | 0.036 | 4.904 | 0 | 0.158 | 4.960 | 0 | 0.102 | 4.798 | 0 | 0.218 | 4.918 | 0 |
| | | Oracle | 0.031 | 5.000 | 0 | 0.155 | 5.000 | 0 | 0.086 | 5.000 | 0 | 0.211 | 5.000 | 0 |
| $30 \times 15$ | 0 | SCAD | 0.021 | 5.972 | 0 | 0.140 | 4.918 | 0 | 0.052 | 5.966 | 0 | 0.188 | 4.712 | 0 |
| | | ALASSO | 0.025 | 5.836 | 0 | 0.138 | 4.980 | 0 | 0.066 | 5.752 | 0 | 0.172 | 4.946 | 0 |
| | | Oracle | 0.022 | 6.000 | 0 | 0.137 | 5.000 | 0 | 0.051 | 6.000 | 0 | 0.170 | 5.000 | 0 |
| | 0.3 | SCAD | 0.021 | 4.890 | 0 | 0.138 | 4.932 | 0 | 0.060 | 4.866 | 0 | 0.185 | 4.736 | 0 |
| | | ALASSO | 0.023 | 4.944 | 0 | 0.135 | 4.998 | 0 | 0.069 | 4.896 | 0 | 0.170 | 4.982 | 0 |
| | | Oracle | 0.020 | 5.000 | 0 | 0.135 | 5.000 | 0 | 0.059 | 5.000 | 0 | 0.167 | 5.000 | 0 |
| | 0.7 | SCAD | 0.025 | 4.994 | 0 | 0.139 | 4.922 | 0 | 0.063 | 4.966 | 0 | 0.183 | 4.762 | 0 |
| | | ALASSO | 0.030 | 4.920 | 0 | 0.138 | 4.994 | 0 | 0.076 | 4.858 | 0 | 0.174 | 4.956 | 0 |
| | | Oracle | 0.025 | 5.000 | 0 | 0.136 | 5.000 | 0 | 0.060 | 5.000 | 0 | 0.167 | 5.000 | 0 |
| $60 \times 10$ | 0 | SCAD | 0.015 | 5.992 | 0 | 0.128 | 5.000 | 0 | 0.039 | 5.976 | 0 | 0.152 | 4.988 | 0 |
| | | ALASSO | 0.018 | 5.832 | 0 | 0.129 | 4.996 | 0 | 0.053 | 5.812 | 0 | 0.153 | 4.992 | 0 |
| | | Oracle | 0.016 | 6.000 | 0 | 0.128 | 5.000 | 0 | 0.040 | 6.000 | 0 | 0.152 | 5.000 | 0 |
| | 0.3 | SCAD | 0.020 | 4.932 | 0 | 0.129 | 5.000 | 0 | 0.038 | 4.894 | 0 | 0.151 | 4.998 | 0 |
| | | ALASSO | 0.022 | 4.966 | 0 | 0.129 | 4.998 | 0 | 0.043 | 4.928 | 0 | 0.152 | 4.992 | 0 |
| | | Oracle | 0.020 | 5.000 | 0 | 0.129 | 5.000 | 0 | 0.036 | 5.000 | 0 | 0.151 | 5.000 | 0 |
| | 0.7 | SCAD | 0.020 | 4.988 | 0 | 0.128 | 5.000 | 0 | 0.049 | 4.990 | 0 | 0.151 | 4.998 | 0 |
| | | ALASSO | 0.024 | 4.954 | 0 | 0.129 | 4.998 | 0 | 0.060 | 4.926 | 0 | 0.154 | 4.996 | 0 |
| | | Oracle | 0.020 | 5.000 | 0 | 0.128 | 5.000 | 0 | 0.048 | 5.000 | 0 | 0.151 | 5.000 | 0 |
| $60 \times 15$ | 0 | SCAD | 0.014 | 6.000 | 0 | 0.118 | 4.994 | 0 | 0.028 | 5.998 | 0 | 0.139 | 4.888 | 0 |
| | | ALASSO | 0.016 | 5.906 | 0 | 0.119 | 5.000 | 0 | 0.036 | 5.878 | 0 | 0.136 | 4.998 | 0 |
| | | Oracle | 0.014 | 6.000 | 0 | 0.118 | 5.000 | 0 | 0.030 | 6.000 | 0 | 0.135 | 5.000 | 0 |
| | 0.3 | SCAD | 0.015 | 4.930 | 0 | 0.119 | 4.978 | 0 | 0.037 | 4.958 | 0 | 0.139 | 4.854 | 0 |
| | | ALASSO | 0.017 | 4.934 | 0 | 0.120 | 4.984 | 0 | 0.043 | 4.938 | 0 | 0.135 | 5.000 | 0 |
| | | Oracle | 0.015 | 5.000 | 0 | 0.119 | 5.000 | 0 | 0.037 | 5.000 | 0 | 0.134 | 5.000 | 0 |
| | 0.7 | SCAD | 0.015 | 4.998 | 0 | 0.119 | 4.998 | 0 | 0.030 | 4.990 | 0 | 0.136 | 4.920 | 0 |
| | | ALASSO | 0.018 | 4.970 | 0 | 0.119 | 4.998 | 0 | 0.040 | 4.940 | 0 | 0.136 | 4.994 | 0 |
| | | Oracle | 0.016 | 5.000 | 0 | 0.119 | 5.000 | 0 | 0.031 | 5.000 | 0 | 0.134 | 5.000 | 0 |

**Table A2.** Variable selection for $n = 100 \times 20$ in Example 2.

| $\rho_0$ | Method | $\sigma_0^2 = 1$ | | | | | | $\sigma_0^2 = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | | $\hat{\theta}$ | | | $\hat{\alpha}(\cdot)$ | | |
| | | GMSE | C | I | ASE | C | I | GMSE | C | I | ASE | C | I |
| 0 | SCAD | 0.004 | 46.986 | 0 | 0.037 | 11.670 | 0 | 0.010 | 46.982 | 0 | 0.049 | 11.508 | 0 |
| | ALASSO | 0.005 | 46.388 | 0 | 0.036 | 11.996 | 0 | 0.018 | 45.818 | 0 | 0.048 | 11.982 | 0 |
| | Oracle | 0.003 | 47.000 | 0 | 0.036 | 12.000 | 0 | 0.009 | 47.000 | 0 | 0.046 | 12.000 | 0 |
| 0.3 | SCAD | 0.004 | 45.814 | 0 | 0.039 | 10.972 | 0 | 0.011 | 45.602 | 0 | 0.053 | 11.110 | 0 |
| | ALASSO | 0.006 | 45.506 | 0 | 0.037 | 11.914 | 0 | 0.018 | 44.648 | 0 | 0.049 | 11.672 | 0 |
| | Oracle | 0.004 | 46.000 | 0 | 0.036 | 12.000 | 0 | 0.009 | 46.000 | 0 | 0.046 | 12.000 | 0 |
| 0.7 | SCAD | 0.004 | 45.986 | 0 | 0.038 | 11.450 | 0 | 0.011 | 45.964 | 0 | 0.049 | 11.420 | 0 |
| | ALASSO | 0.006 | 45.568 | 0 | 0.037 | 11.992 | 0 | 0.021 | 44.846 | 0 | 0.048 | 11.964 | 0 |
| | Oracle | 0.004 | 46.000 | 0 | 0.036 | 12.000 | 0 | 0.010 | 46.000 | 0 | 0.046 | 12.000 | 0 |

## References


1. Anselin, L.; Hudak, S. Spatial econometrics in practice: A review of software options. *Reg. Sci. Urban Econ.* **1992**, *22*, 509–536. [CrossRef]
2. Baltagi, B.H.; Heun Song, S.; Cheol Jung, B.; Koh, W. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *J. Econom.* **2007**, *140*, 5–51. [CrossRef]
3. Kapoor, M.; Kelejian, H.H.; Prucha, I.R. Panel data models with spatially correlated error components. *J. Econom.* **2007**, *140*, 97–130. [CrossRef]
4. Baltagi, B.H.; Kao, C.; Liu, L. Asymptotic properties of estimators for the linear panel regression model with random individual effects and serially correlated errors: The case of stationary and non-stationary regressors and residuals. *Econom. J.* **2008**, *11*, 554–572. [CrossRef]
5. Lee, L.F.; Yu, J. Estimation of spatial autoregressive panel data models with fixed effects. *J. Econom.* **2010**, *154*, 165–185. [CrossRef]
6. Ai, C.; Zhang, Y. Estimation of partially specified spatial panel data models with fixed-effects. *Econom. Rev.* **2017**, *36*, 6–22. [CrossRef]
7. Zhang, Y.; Sun, Y. Estimation of partially specified dynamic spatial panel data models with fixed-effects. *Reg. Sci. Urban Econ.* **2015**, *51*, 37–46. [CrossRef] [PubMed]
8. Zhang, Y.; Shen, D. Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects. *J. Stat. Plan. Inference* **2015**, *159*, 64–80. [CrossRef]
9. Feng, S.; Tong, T.; Chiu, S.N. Statistical Inference for Partially Linear Varying Coefficient Spatial Autoregressive Panel Data Model. *Mathematics* **2023**, *11*, 4606. [CrossRef]
10. Hu, X. Estimation in a semi-varying coefficient model for panel data with fixed effects. *J. Syst. Sci. Complex.* **2014**, *27*, 594–604. [CrossRef]
11. He, B.; Hong, X.; Fan, G. Empirical likelihood for semi-varying coefficient models for panel data with fixed effects. *J. Korean Stat. Soc.* **2016**, *45*, 395–408. [CrossRef]
12. Feng, S.; He, W.; Li, F. Model detection and estimation for varying coefficient panel data models with fixed effects. *Comput. Stat. Data Anal.* **2020**, *152*, 107054. [CrossRef]
13. Feng, S.; Li, G.; Peng, H.; Tong, T. Varying coefficient panel data model with interactive fixed effects. *Stat. Sin.* **2021**, *31*, 935–957. [CrossRef]
14. Sun, Y.; Carroll, R.; Li, D. Semiparametric estimation of fixed-effects panel data varying coefficient models. *Adv. Econom.* **2009**, *25*, 101–129.
15. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
16. Fan, J.; Li, R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
17. Zou, H. The Adaptive Lasso and Its Oracle Properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [CrossRef]
18. Wang, L.; Li, H.; Huang, J.Z. Variable Selection in Nonparametric Varying-Coefficient Models for Analysis of Repeated Measurements. *J. Am. Stat. Assoc.* **2008**, *103*, 1556–1569. [CrossRef]
19. Li, R.; Liang, H. Variable selection in semiparametric regression modeling. *Ann. Stat.* **2008**, *36*, 261–286. [CrossRef]
20. Wang, H.J.; Zhu, Z.; Zhou, J. Quantile regression in partially linear varying coefficient models. *Ann. Stat.* **2009**, *37*, 3841–3866. [CrossRef]
21. Zhao, P.; Xue, L. Variable selection for semiparametric varying coefficient partially linear models. *Stat. Probab. Lett.* **2009**, *79*, 2148–2157. [CrossRef]
22. Tian, R.; Xue, L.; Liu, C. Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *J. Multivar. Anal.* **2014**, *132*, 94–110. [CrossRef]
23. Tian, R.; Xue, L.; Hu, Y. Smooth-threshold GEE variable selection for varying coefficient partially linear models with longitudinal data. *J. Korean Stat. Soc.* **2015**, *44*, 419–431. [CrossRef]
24. Li, R.; Mu, S.; Hao, R. Estimation and variable selection for partially linear additive models with measurement errors. *Commun. Stat. Theory Methods* **2021**, *50*, 1416–1445. [CrossRef]
25. Ma, X.; Du, Y.; Wang, J. Model detection and variable selection for mode varying coefficient model. *Stat. Methods Appl.* **2022**, *31*, 321–341. [CrossRef]
26. Liu, Y.; Wang, Z.; Tian, M.; Yu, K. Estimation and variable selection for generalized functional partially varying coefficient hybrid models. *Stat. Pap.* **2024**, *65*, 93–119. [CrossRef]
27. Neyman, J.; Scott, E.L. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **1948**, *16*, 1–32. [CrossRef]
28. Liu, X.; Chen, J.; Cheng, S. A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. *Spat. Stat.* **2018**, *25*, 86–104. [CrossRef]
29. Xie, T.; Cao, R.; Du, J. Variable selection for spatial autoregressive models with a diverging number of parameters. *Stat. Pap.* **2020**, *61*, 1125–1145. [CrossRef]

30. Luo, G.; Wu, M. Variable selection for semiparametric varying-coefficient spatial autoregressive models with a diverging number of parameters. *Commun. Stat. Theory Methods* **2021**, *50*, 2062–2079. [CrossRef]
31. Lee, L.F. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* **2004**, *72*, 1899–1925. [CrossRef]
32. Tian, R.; Xia, M.; Xu, D. Profile quasi-maximum likelihood estimation for semiparametric varying-coefficient spatial autoregressive panel models with fixed effects. *Stat. Pap.* **2024**, *65*, 5109–5143. [CrossRef]
33. Kelejian, H.H.; Prucha, I.R. On the asymptotic distribution of the Moran I test statistic with applications. *J. Econom.* **2001**, *104*, 219–257. [CrossRef]
34. He, X.; Fung, W.K.; Zhu, Z. Robust Estimation in Generalized Partial Linear Models for Clustered Data. *J. Am. Stat. Assoc.* **2005**, *100*, 1176–1184. [CrossRef]
35. Stone, C.J. Optimal Global Rates of Convergence for Nonparametric Regression. *Ann. Stat.* **1982**, *10*, 1040–1053. [CrossRef]
36. Dietz, T.; Rosa, E. Effects of population and affluence on $CO_2$ emissions. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 175–179. [CrossRef]
37. Wen, L.; Shao, H. Analysis of influencing factors of the carbon dioxide emissions in China's commercial department based on the STIRPAT model and ridge regression. *Environ. Sci. Pollut. Res.* **2019**, *26*, 27138–27147. [CrossRef] [PubMed]
38. Li, W.; Wang, W.; Wang, Y.; Qin, Y. Industrial structure, technological progress and $CO_2$ emissions in China: Analysis based on the STIRPAT framework. *Nat. Hazards* **2017**, *88*, 1545–1564. [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.