Hindawi
Journal of Mathematics
Volume 2023, Article ID 9837117, 16 pages
https://doi.org/10.1155/2023/9837117

Research Article

Variable Selection of High-Dimensional Spatial Autoregressive Panel Models with Fixed Effects

Miaojie Xia, Yuqi Zhang, and Ruiqin Tian

School of Mathematics, Hangzhou Normal University, Hangzhou 311121, China

Correspondence should be addressed to Ruiqin Tian; ruiqintian@163.com

Received 22 March 2023; Revised 29 April 2023; Accepted 5 May 2023; Published 31 May 2023

Academic Editor: Antonio Di Crescenzo

This paper studies the variable selection of high-dimensional spatial autoregressive panel models with fixed effects in which a matrix transformation method is applied to eliminate the fixed effects. Then, a penalized quasi-maximum likelihood is developed for variable selection and parameter estimation in the transformed panel model. Under some regular conditions, the consistency and oracle properties of the proposed estimator are established. Some Monte-Carlo experiments and a real data analysis are conducted to examine the finite sample performance of the proposed variable selection procedure, showing that the proposed variable selection method works satisfactorily.

1. Introduction

Panel data, which contains dimensions of time and space, are becoming more and more common under the circumstances of big data. Compared with traditional cross-sectional data and time series data, a great advantage of panel data is that it can effectively expand the sample size. A large number of studies have given a variety of panel models, of which the panel model with fixed effects is widely studied and applied because it can capture non-time-varying and unobservable exogenous variables. In fact, these models are still based on the independence assumption which is improper in the real situation, so in this paper, we consider spatial autoregressive panel (SARP) models with fixed effects that permit interdependence between spatial units in panel data. Most data in the context of big data are high-dimensional, which also means the number of covariates can diverge with the growth of sample size and lead to a rapid increase in model complexity. In order to reduce the amount of calculation and model complexity, penalized methods are indispensable to remove irrelevant variables. However, the variable selection of high-dimensional spatial panel data is more complex than that of high-dimensional cross-sectional data due to spatial terms and fixed effects. The ordinary penalized methods for variable selection, such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani [1]), the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li [2]), which are proposed for classical linear regression models, cannot be used in the high-dimensional SARP models directly. Therefore, we propose a penalized estimation method for high-dimensional SARP models with fixed effects and establish corresponding asymptotic properties.

Up to now, there are a large number of related studies which have given statistical inferences about spatial models. Since Cliff and Ord [3] proposed a structure with spatial correlation, spatial models have been receiving increasing attention. Anselin [4] proposed the maximum likelihood estimation of the spatial autoregressive (SAR) model and constructed an LM test for the spatial term. Kelejian and Prucha [5] proposed a generalized spatial two-stage least squares procedure for instrument matrices and studied its properties. Lee [6] established the asymptotic distributions of quasi-maximum likelihood for SAR models. Wei et al. [7] constructed the partially linear varying coefficient SAR models and approximated the nonlinear part locally by a linear function. Du et al. [8] proposed the estimator for the asymptotic covariance matrix of the parameter estimator of partially linear additive SAR models and established the asymptotic properties for the resulting estimators. Other research results on SAR models can also be referred to

Cheng et al. [9], Dai et al. [10], Gupta and Robinson [11], Lin and Lee [12], Tian et al. [13], Tian et al. [14], and so on. These studies based on cross-sectional data are not applicable to panel data, and variable selection is rarely involved.

Recently, panel data have attracted tremendous attention, especially since a number of works have studied the relevant statistical inference of spatial panel data models. For example, Baltagi et al. [15] considered panel regression models with SAR disturbances and LM tests under five hypotheses. Lee and Yu [16] proposed an orthonormal transformation for spatial autoregressive panel models with fixed effects and provided a method which allows the estimation of coefficients without estimating fixed effects. Ju et al. [17] estimated parameters of spatial dynamic panel data models by the Bayesian method, and their method can adapt to a skew-normal distribution. These research studies investigated the estimators and corresponding large sample properties of spatial panel models; however, little work has been performed on the variable selection of models.

To the best of our knowledge, Liu et al. [18] investigated variable selection in the SAR model with independent and identically distributed errors, but their model was not under the situation of a diverging number of parameters and the asymptotic properties they established were not available for high-dimensional data. Xie et al. [19] considered the penalized estimation for SAR models with a diverging number of parameters and established the oracle properties; however, their method was available for high-dimensional cross-sectional data but not for panel data. Therefore, we consider variable selection for the high-dimensional SARP model with fixed effect, present the penalized estimators, and establish related asymptotic properties thoroughly.

This paper is organized as follows: in Section 2, we introduce a high-dimensional SARP model with fixed effect and eliminate the fixed effects term by transformation matrix. In Section 3, we consider the penalized quasi-maximum likelihood estimators (QMLE) which are based on the SCAD penalty function for SARP models with fixed effects and establish its consistency and oracle property. Besides, we introduce a feasible iterative algorithm for the penalized QMLE in this section. In Section 4, some Monte-Carlo simulations are carried out to examine the finite sample performance of QMLE. In Section 5, a real data application of China's carbon emission is provided for illustrative purposes. In Section 6, we give a brief conclusion of this paper. The detailed proofs of theoretical results are provided in the Appendix.

2. Matrix Transformation for SARP Models with Fixed Effects

We consider the following SARP models with fixed effects:

$$\mathbf{Y}_t = \rho \mathbf{W}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u} + \mathbf{V}_t, t = 1, 2, \ldots, T, \tag{1}$$

where $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})^\tau$ is an $N \times 1$ vector of observations on the dependent variables, ρ is an unknown spatial autoregressive coefficient, the spatial weight matrix $\mathbf{W}$ is an $N \times N$ matrix of known constants with zero diagonal elements and satisfies that the sum of rows is 1. $\mathbf{X}_t = [\mathbf{X}_{t1}, \ldots, \mathbf{X}_{tk}, \ldots, \mathbf{X}_{td_N}]$ is an $N \times d_N$ matrix of observations on d_N linear regressors, where d_N is divergent as $N \longrightarrow \infty, \beta = (\beta_1, \ldots, \beta_{d_N})^\tau$ is an unknown $d_N \times 1$ vector of regression coefficients, $\mathbf{u} = (u_1, \ldots, u_N)^\tau$ is an unknown $N \times 1$ vector of fixed effects, and $\mathbf{V_t} = (V_{t1}, \ldots, V_{tN})^\tau$ is a vector i.i.d across t with zero mean and finite covariance matrix $\sigma^2\mathbf{I}_N$, where σ^2 is an unknown parameter and $\mathbf{I}_N$ is the identity matrix. Therefore, the unknown parameters to be estimated can be expressed as $(\sigma^2, \rho, \beta^\tau, \mathbf{u}^\tau)^\tau$.

The fixed effects $\mathbf{u}$ is an unknown $N \times 1$ vector which means that there is inconsistency regarding N; however, if we just focus on $\sigma^2, \rho,$ and β, a transformation method is available to eliminate $\mathbf{u}$. According to the method used by Lee and Yu [16], we let $\mathbf{J}_T = \mathbf{I}_T - L_T L_T^T/T$, where ι_T represents a T-dimensional vector with all 1 and let $\mathbf{F}_{T,T-1}$ be the orthonormal eigenvector matrix of $\mathbf{J}_T$ which corresponds to the eigenvalues of one. Thus, we define the transformed matrices $[\mathbf{Y}_1^*, \ldots, \mathbf{Y}_{T-1}^*] = [\mathbf{Y}_1, \ldots, \mathbf{Y}_T]\mathbf{F}_{T,T-1}$, $[\mathbf{V}_1^*, \ldots, \mathbf{V}_T^*] = [\mathbf{V}_1, \ldots, \mathbf{V}_T]\mathbf{F}_{T,T-1}$, and apparently $[\mathbf{u}, \ldots, \mathbf{u}]\mathbf{F}_{T,T-1} = \mathbf{0}$. For $k = 1, 2, \ldots, d_N$, we similarly define $[\mathbf{X}_{1k}^*, \ldots, \mathbf{X}_{(T-1)k}^*] = [\mathbf{X}_{1k}, \ldots, \mathbf{X}_{Tk}]\mathbf{F}_{T,T-1}$. Then, (1) can be rewritten as follows:

$$\mathbf{Y}^* = \rho \mathbb{W}\mathbf{Y}^* + \mathbf{X}^*\boldsymbol{\beta} + \mathbf{V}^*, \tag{2}$$

where $\mathbf{Y}^* = [\mathbf{Y}_1^{*\tau}, \ldots, \mathbf{Y}_{T-1}^{*\tau}]^\tau$, $\mathbb{W} = \mathbf{I}_{T-1} \otimes \mathbf{W}$, $\otimes$ is Kronecker's product symbol, $\mathbf{X}^* = [\mathbf{X}_1^*, \ldots, \mathbf{X}_k^*, \ldots, \mathbf{X}_{d_N}^*]$, $\mathbf{X}_k^* = [\mathbf{X}_{1k}^{*\tau}, \mathbf{X}_{2k}^{*\tau}, \ldots, \mathbf{X}_{(T-1)k}^{*\tau}]^\tau$, and $\mathbf{V}^* = [\mathbf{V}_1^{*\tau}, \ldots, \mathbf{V}_{T-1}^{*\tau}]^\tau$. Then, it is easy to know $[\mathbf{V}_1^{*\tau}, \ldots, \mathbf{V}_{T-1}^{*\tau}]^\tau = (\mathbf{F}_{T,T-1} \otimes I)[\mathbf{V}_1^\tau, \ldots, \mathbf{V}_T^\tau]^\tau$, so we can obtain $E(\mathbf{V}^*\mathbf{V}^{*\tau}) = \sigma^2(\mathbf{F}_{T,T-1} \otimes \mathbf{I}_N)^\tau(\mathbf{F}_{T,T-1} \otimes \mathbf{I}_N) = \sigma^2\mathbf{I}_{N(T-1)}$. Besides, $E(\mathbf{V}^*) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{V}^*) = \sigma^2\mathbf{I}_{N(T-1)}$, and the parameter vector to be estimated is marked as $\boldsymbol{\theta} = (\sigma^2, \rho, \beta^\tau)^\tau$.

3. Methods and Large Sample Properties

3.1. Penalized Quasimaximum Likelihood Estimator. Let $\mathbf{M}_N(\rho) = \mathbf{I}_{N(T-1)} - \rho\mathbb{W}$, and if $\mathbf{V}_t$ is normally distributed, the log-likelihood function $\ln L(\theta)$ of model (2) is obtained in the following equation:

$$\ln L(\boldsymbol{\theta}) = -\frac{N(T-1)}{2}\ln(2\pi) - \frac{N(T-1)}{2}\ln\left(\sigma^2\right) + \ln\left|\mathbf{M}_N(\rho)\right| - \frac{1}{2\sigma^2}[\mathbf{M}_N(\rho)\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}]^\tau[\mathbf{M}_N(\rho)\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}]. \tag{3}$$

In addition, we let $p_\lambda(\cdot)$ be the SCAD penalty function, and of course, other penalty functions can also be considered here. Then, we renumber parameter vector as $\boldsymbol{\theta} = (\sigma^2, \rho, \beta^\tau)^\tau = (\theta_1, \theta_2, \theta_3, \ldots, \theta_{d_N+2})^\tau$, thus the penalized quasimaximum likelihood function can be obtained as follows:

$$Q(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) - N(T-1)\sum_{j=2}^{d_N+2} p_\lambda\left(\left|\theta_j\right|\right), \tag{4}$$

where the SCAD penalty function $p_\lambda(\cdot)$ is defined by its first derivative:

$$p_\lambda'(\vartheta) = \lambda\left\{I(\vartheta \le \lambda) + \frac{(a\lambda - \vartheta)_+}{(a-1)\lambda} I(\vartheta > \lambda)\right\} \text{for } \vartheta > 0. \quad (5)$$

We set $a = 3.7$ as Fan and Li [2] recommended in their paper. To determine the tuning parameter λ, a Bayesian information criterion can be applied:

$$\mathrm{BIC}_\lambda = -2\ln L(\widehat{\boldsymbol\theta}) + \ln[N(T-1)]df_\lambda, \quad (6)$$

where df_λ is the number of nonzero parameters, and $\widehat{\lambda} = \operatorname{argmin}\mathrm{BIC}_\lambda$. Then, we get the penalized QMLE $\widehat{\boldsymbol\theta} = (\widehat{\sigma}^2, \widehat{\rho}, \widehat{\beta}^\tau)^\tau = \operatorname*{argmax}_{\boldsymbol\theta}[Q(\boldsymbol\theta)]$.

3.2. Asymptotic Properties. Let $\boldsymbol\theta_0 = (\sigma_0^2, \rho_0, \beta_0^\tau)^\tau = (\theta_{10}, \theta_{20}, \theta_{30}, \ldots, \theta_{(d_N+2)0})^\tau$ be the true parameter vector, without loss of generality, we assume that the first $s\,(s > 1)$ of $\boldsymbol\theta_0$ is nonzero parameters and zero for the remainder. Remark all the nonzero parameters as $\boldsymbol\theta_{10}$, then we can rewrite $\boldsymbol\theta_0$ as $(\boldsymbol\theta_{10}^\tau, \boldsymbol\theta_{20}^\tau)^\tau = (\boldsymbol\theta_{10}^\tau, \mathbf{0}^\tau)^\tau$. Let $\mathbf{M}_0 = \mathbf{I}_{N(T-1)} - \rho_0\mathbb{W}$, $\mathbf{V}_0^* = \mathbf{M}_0\mathbf{Y}^* - \mathbf{X}^*\boldsymbol\beta_0$, in order to obtain the asymptotic properties of the penalized QMLE, there are some regular assumptions as follows:

A1. The T is finite and $T > 2$.

A2. The moment $E(|\mathbf{V}_0^*|_i^{4+v})$ exists for some $v > 0$.

A3. The elements $\{w_{ij}\}$ in $\mathbb{W}$ satisfy $w_{ij} = O(1/h_N)$, where h_N can be divergent or bounded, $h_N/N \longrightarrow 0$ when $N \longrightarrow \infty$.

A4. The matrix $M_N(\rho)$ is a nonsingular matrix for all ρ in $(-1, 1)$.

A5. The sequences of matrices $\{\mathbb{W}\}$ and $\{M_0^{-1}\}$ are uniformly bounded in both row and column sums for all N.

A6. The $\lim_{N\longrightarrow\infty} 1/(N(T-1))\mathbf{X}^{*\tau}\mathbf{X}^*$ exists and is nonsingular. The $\lim_{N\longrightarrow\infty} 1/(Nd_N) tr\mathbf{X}^{*\tau}\mathbf{X}^*$ exists. The elements of $\mathbf{X}^*$ are uniformly bounded constants for all N.

A7. $\mathbf{M}_N^{-1}(\rho)$ is bounded in both row and column sums, uniformly in ρ in a closed subset ϱ of $(-1, 1)$ and ρ_0 which is an interior point of ϱ.

A8. The $\lim_{N\longrightarrow\infty} 1/(N(T-1))(\mathbf{X}^*, \mathbb{W}\mathbf{M}_0^{-1}\mathbf{X}^*\beta_0)^\tau (\mathbf{X}^*, \mathbb{W}\,\mathbf{M}_0^{-1}\mathbf{X}^*\beta_0)$ exists and is nonsingular.

A9. The $\lim_{N\longrightarrow\infty} 1/(N(T-1))E(\partial^2 \ln L(\boldsymbol\theta_0)/\partial\boldsymbol\theta\partial\boldsymbol\theta^\tau)$ exists and is nonsingular.

A10. The third derivatives $\partial^3 \ln L(\boldsymbol\theta_0)/\partial\theta_i\partial\theta_j\partial\theta_k$ exist for all θ in an open set Θ that contains true parameter point $\boldsymbol\theta_0$. There exist functions M_{ijk} that satisfy $|1/(N(T-1))\partial^3 \ln L(\boldsymbol\theta_0)/\partial\theta_i\partial\theta_j\partial\theta_k| \le M_{ijk} \le \infty$, where $E(M_{ijk}^2) < \infty$ for n, p and i, j, k.

A11. The eigenvalues of the Hessian matrix $E(\partial \ln L(\boldsymbol\theta_0)/\partial\boldsymbol\theta\partial \ln L(\boldsymbol\theta_0)/\partial\boldsymbol\theta^\tau)$ satisfy the following:

$$\begin{aligned}
&0 < C_1 < \lambda_{\min}\left\{\frac{1}{N(T-1)}E\left(\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\boldsymbol\theta}\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\boldsymbol\theta^\tau}\right)\right\}\\
&\le \lambda_{\max}\left\{\frac{1}{N(T-1)}E\left(\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\boldsymbol\theta}\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\boldsymbol\theta^\tau}\right)\right\} < C_2 < \infty,\\
&\frac{1}{N(T-1)}E\left[\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\theta_i}\frac{\partial \ln L(\boldsymbol\theta_0)}{\partial\theta_j}\right]^2 < C_3 < \infty,\\
&\frac{1}{N(T-1)}E\left[\frac{\partial^2 \ln L(\boldsymbol\theta_0)}{\partial\theta_i\partial\theta_j}\right]^2 < C_4 < \infty \text{ for all } N, i, j.
\end{aligned} \quad (7)$$

A12. $\liminf_{n\longrightarrow\infty} \liminf_{\beta_j\longrightarrow 0^+} \lambda^{-1}p_\lambda'(|\beta_j|) > 0$ for $j = s+1, \ldots, d_N$.

A13. $\max\left\{|p_\lambda''(\beta_{j0})\colon \beta_{j0} \ne 0|\right\} \triangleq b_N$, $b_N \longrightarrow 0$ as $N \longrightarrow \infty$.

Remark 1. Assumptions A1–A8 are set for spatial term and regressors, A9–A11 are imposed on the likelihood functions, and A12 and A13 are for penalty functions. In addition, A1 shows that $N \longrightarrow \infty$ is the only large sample scenario; A2 ensures the moment exists; A3–A8 ensure the QMLE of SARP exists; and A9–A13, which are similar to the assumptions provided by Fan and Peng [20], are necessary for obtaining the consistency and the oracle property of PQLME.

Let $\mathbf{G} = \mathbb{W}\mathbf{M}_0^{-1}$, then the first derivative $1/\sqrt{N(T-1)}\partial Q(\theta_0)/\partial\theta$ is as follows:

$$\begin{cases}
\dfrac{\partial Q(\boldsymbol\theta_0)}{\partial\sigma^2} = -\dfrac{N(T-1)}{2\sigma_0^2} + \dfrac{1}{2\sigma_0^4}\mathbf{V}_0^\tau\mathbf{V}_0^*,\\[2ex]
\dfrac{\partial Q(\boldsymbol\theta_0)}{\partial\rho} = -tr\mathbf{G} + \dfrac{1}{\sigma_0^2}\mathbf{V}_0^{*\tau}\mathbb{W}\mathbf{Y}^* - N(T-1)p_\lambda'(|\rho_0|)\operatorname{sgn}(\rho_0),\\[2ex]
\dfrac{\partial Q(\boldsymbol\theta_0)}{\partial\boldsymbol\beta} = \dfrac{1}{\sigma_0^2}\mathbf{X}^{*\tau}\mathbf{V}_0^* - N(T-1)p_\lambda'(|\boldsymbol\beta_0|),
\end{cases} \quad (8)$$

where $p_\lambda'(|\beta_0|) = [p_\lambda'(|\beta_{10}|)\operatorname{sgn}(\beta_{10}), \ldots, p_\lambda'(|\beta_{d_N0}|)\operatorname{sgn}(\beta_{d_N0})]^\tau$. Then, we have theorems as follows:

Theorem 2 (Consistency). *Under assumptions A1–A12, we suppose that $d_N^4/N \longrightarrow 0$ as $N \longrightarrow \infty$, then there is a local maximizer $\widehat{\theta}$ of $Q(\theta)$ that satisfies the following:*

$$\|\widehat{\boldsymbol\theta} - \boldsymbol\theta_0\| = O_p\left(\sqrt{d_N}\left(N^{-1/2} + a_N\right)\right), \quad (9)$$

where $a_N = \max\left\{p_\lambda'(|\theta_{j0}|)\colon \theta_{j0} \ne 0\right\}$.

According to Theorem 2, we can choose a proper λ to achieve $\sqrt{N/d_N}$ consistent penalized QMLE under A1–A12.

We know that $E(\partial \ln L(\boldsymbol\theta_0)/\partial\theta\partial \ln L(\boldsymbol\theta_0)/\partial\boldsymbol\theta^\tau)$ is the Hessian matrix under θ_0. Note that $E(\mathbf{V}_0^*\mathbf{V}_0^{*\tau}) = \sigma_0^2\mathbf{I}_{N(T-1)}$, then the covariance matrix of $1/\sqrt{(N(T-1))}\partial \ln L(\boldsymbol\theta_0)/\partial\boldsymbol\theta$ is obtained as follows:

$$\operatorname{var}\left(\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) = E\left(\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\cdot\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\tau}}\right),$$
$$\triangleq \frac{1}{N(T-1)}I(\boldsymbol{\theta}_0) + \frac{1}{N(T-1)}\Delta(\boldsymbol{\theta}_0), \tag{10}$$

where

$$I(\boldsymbol{\theta}_0) = -E\left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}\right)$$
$$= \begin{bmatrix} tr\mathbf{G}^2 + tr[\mathbf{G}^{\tau}\mathbf{G}] + \frac{1}{\sigma_0^2}(\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0)^{\tau}\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0 & \frac{1}{\sigma_0^2}tr\mathbf{G} & \frac{1}{\sigma_0^2}(\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0)^{\tau}\mathbf{X}^* \\ \frac{1}{\sigma_0^2}tr\mathbf{G} & \frac{N(T-1)}{2\sigma_0^4} & \mathbf{0}_{1\times d_N} \\ \frac{1}{\sigma_0^2}\mathbf{X}^{*\tau}\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0 & \mathbf{0}_{d_N\times 1} & \frac{1}{\sigma_0^2}\mathbf{X}^{*\tau}\mathbf{X}^* \end{bmatrix}, \tag{11}$$

and $\Delta(\boldsymbol{\theta}_0)$ is a symmetric matrix:

$$\Delta(\boldsymbol{\theta}_0) = \begin{bmatrix} \frac{2\mu_3}{\sigma_0^4}\sum_{i=1}^{N(T-1)}\mathbf{G}_{ii}\mathbf{G}_{i\cdot}\mathbf{X}^*\boldsymbol{\beta}_0 + \frac{\mu_4-3\sigma_0^4}{\sigma_0^4}\sum_{i=1}^{N(T-1)}\mathbf{G}_{ii} & * & * \\ \frac{1}{2\sigma_0^6}\left[\mu_3\boldsymbol{\iota}_{N(T-1)}^{\tau}\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0 + \left(\mu_4-3\sigma_0^4\right)tr\mathbf{G}\right] & \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} & * \\ \frac{\mu_3}{\sigma_0^4}\sum_{i=1}^{N(T-1)}\mathbf{G}_{ii}\mathbf{X}_{i\cdot}^* & \frac{\mu_3}{2\sigma_0^6}\boldsymbol{\iota}_{N(T-1)}^{\tau}\mathbf{X}^* & 0 \end{bmatrix}. \tag{12}$$

In addition, $\mu_3 = E[(\mathbf{V}_0^*)_i^3], \mu_4 = E[(\mathbf{V}_0^*)_i^4]$, $\mathbf{G}_{i\cdot}$ represents the ith row of $\mathbf{G}$, $\mathbf{G}_{ij}$ is the (i, j) element of $\mathbf{G}$ and $\mathbf{X}_{i\cdot}^*$ is the ith row of $\mathbf{X}^*$. A8 ensures that $1/(N(T-1))I(\boldsymbol{\theta}_0)$ is nonsingular as N goes infinite. Apparently, $\Delta(\boldsymbol{\theta}_0) = 0$ that provided $\mathbf{V}_0^*$ is normally distributed, so the covariance matrix of $1/\sqrt{N(T-1)}\partial \ln L(\boldsymbol{\theta}_0)/\partial\theta$ is $1/(N(T-1))[I(\theta_0) + \Delta(\theta_0)]$.

Theorem 3. *Under A1–A13, we suppose that $\lambda \longrightarrow 0$, $\sqrt{N/d_N}\lambda \longrightarrow \infty$, $d_N^5/N \longrightarrow 0$ as $N \longrightarrow \infty$, then with probability tending to 1, for any θ_1 that satisfy $\|\theta_1 - \theta_{10}\| = O_p(\sqrt{d_N/N})$ and any constants C, the following equation holds:*

$$Q[(\boldsymbol{\theta}_1^{\tau}, \mathbf{0}^{\tau})^{\tau}] = \max_{\|\boldsymbol{\theta}_2\|\le C\sqrt{d_N/N}} Q[(\boldsymbol{\theta}_1^{\tau}, \boldsymbol{\theta}_2^{\tau})^{\tau}]. \tag{13}$$

Let $I(\theta_{10}, \mathbf{0})$ and $\Delta(\theta_{10}, \mathbf{0})$ be $I(\theta)$ and $\Delta(\theta)$ knowing $\theta_{20} = \mathbf{0},$ respectively, and

$$\begin{aligned} \sum(\boldsymbol{\theta}_{10}) &= \operatorname{diag}\left\{0, p''_\lambda(|\theta_{20}|), p''_\lambda(|\theta_{30}|), \ldots, p''_\lambda(|\theta_{s0}|)\right\}, \\ \mathbf{b} &= \left[0, p'_\lambda(|\theta_{20}|)\operatorname{sgn}(|\theta_{20}|), p'_\lambda(|\theta_{30}|)\operatorname{sgn}(|\theta_{30}|), \ldots, p'_\lambda(|\theta_{s0}|)\operatorname{sgn}(|\theta_{s0}|)\right]^\tau. \end{aligned} \tag{14}$$

Then, we have the following oracle property of PQMLE.

Theorem 4 (Oracle property). *Under A1–A13, we suppose that* $\lambda \longrightarrow 0$, $\sqrt{N/d_N}\lambda \longrightarrow \infty$, $d_N^5/N \longrightarrow 0$ *as* $N \longrightarrow \infty$, *then with probability tending to 1, the root-*(N/d_N)*-consistent local maximizer* $\widehat{\theta} = (\widehat{\theta}_1^\tau, \widehat{\theta}_2^\tau)^\tau$ *must satisfy the following:*

(i) (Sparsity) $\widehat{\theta}_2 = \mathbf{0}$.

(ii) (Asymptotic normality)

$$\sqrt{N(T-1)}\left\{\left[\frac{1}{N(T-1)}I_1(\boldsymbol{\theta}_{10}, \mathbf{0}) + \sum(\boldsymbol{\theta}_0)\right]\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}\right) + \mathbf{b}\right\} \longrightarrow N(0, \mathbf{H} + \mathbf{K}), \tag{15}$$

where $I_1(\theta_{10}, \mathbf{0})$ *and* $\Delta_1(\theta_{10}, \mathbf{0})$ *is the first s upper-left submatrix of* $I(\theta_{10}, \mathbf{0})$ *and* $\Delta(\theta_{10}, \mathbf{0})$*, respectively;* $\mathbf{H} = \lim_{N\longrightarrow\infty} 1/(N(T-1)) I_1(\theta_{10}, \mathbf{0})$ *and* $\mathbf{K} = \lim_{N\longrightarrow\infty} 1/(N(T-1))\Delta_1(\theta_{10}, \mathbf{0})$.

3.3. Numerical Algorithm. The analytical solution for maximizing $Q(\boldsymbol{\theta})$ cannot be obtained due to nonconcave penalty function and spatial term. Although the local quadratic approximation (LQA) algorithm (Fan and Li [2]) cannot be applied to SARP models directly, there are still some ideas that can be used for reference. Let $\Gamma(\boldsymbol{\theta}) = \operatorname{diag}(0, p'_\lambda(|\rho|)/|\rho|, p'_\lambda(|\beta_1|)/|\beta_1|, \ldots, p'_\lambda(|\beta_{d_N}|)/|\beta_{d_N}|)$ which can be regarded as the approximation matrix of $p''_\lambda(|\theta_j|)$, $U(\boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta})\boldsymbol{\theta}$ is the vector form of $p'_\lambda(|\theta_j|)$, and $f(\boldsymbol{\theta}) = \partial \ln L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. A feasible fisher's scoring algorithm is as follows:

Step 1: We initialize $\boldsymbol{\theta}^{(0)} = (\sigma^{2\,(0)}, \rho^{(0)}, \boldsymbol{\beta}^{(0)})$.

Step 2: We update $\theta^{(m+1)} = \boldsymbol{\theta}^{(m)} + [I(\boldsymbol{\theta}^{(m)}) + \Gamma(\boldsymbol{\theta}^{(m)})]^{-} 1[f(\boldsymbol{\theta}^{(m)}) - U(\boldsymbol{\theta}^{(m)})]$.

Step 3: If $\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}\| < \epsilon$, where ϵ is the error limit, we take $\boldsymbol{\theta}^{(m+1)}$ as the final estimator, i.e., $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(m+1)}$, otherwise iterate **Step2.**

The initial value $\boldsymbol{\theta}^{(0)}$ in **Step1** is obtained by using the ordinary least squares method based on (2), and the inverse matrix in **Step2** can be approximated by the generalized inverse matrix.

4. Monte-Carlo Experiments

In this section, we conduct some Monte-Carlo experiments to examine the performance of estimation and variable selection. The simulate data are generated from model (1) as follows:

$$\mathbf{Y}_t = \rho\mathbf{W}\mathbf{Y}_t + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u} + \mathbf{V}_t, t = 1, 2, \ldots T. \tag{16}$$

4.1. Parameters and Regressors Setting. We consider $T = 3$ and $N = 50, 100, 150$ as different sample sizes at three levels. We set the spatial autoregressive coefficient $\rho = 0.3, 0.5, 0.7$ as three degrees of spatial dependence. Moreover, we set the case of $\rho = 0$ which implies that the proposed model reduces to the panel model with fixed effects to examine the penalized estimator performance. In order to make d_N diverge, we set $d_N = 5, 10, 15$, respectively, when $N = 50, 100, 150$, i.e., $\boldsymbol{\beta}_0 = (5, -2, 1, \mathbf{0}_2^\tau)^\tau$, $(5, -2, 1, \mathbf{0}_7^\tau)^\tau$, and $(5, -2, 1, \mathbf{0}_{12}^\tau)^\tau$, where $\mathbf{0}_m$ is a $m \times 1$ zero vector. For the disturbance term, we set three variance levels $\sigma^2 = 0.5, 1, 2$ and two types of disturbance distributions as follows: (i) $\varepsilon_{it} \sim N(0, \sigma^2)$ and (ii) $\varepsilon_{it} \sim \sigma/\sqrt{3}t(3)$ to explore the influence of the disturbance on the proposed estimator. We set $u_i \sim U(0, 1)$ and $\mathbf{X}_t \sim N(0, \Sigma_{0.5})$, where $\Sigma_{0.5}$ is the AR (1) matrix and is shown as follows:

$$\begin{bmatrix} 1 & \cdots & 0.5^{d_N - 1} \\ \vdots & \ddots & \vdots \\ 0.5^{d_N - 1} & \cdots & 1 \end{bmatrix}. \tag{17}$$

In addition, we consider another setting, that is, $N = 50$ and $T = 6, 9, 12$, and $\beta_0 = (5, -2, 1, \mathbf{0}_5^\tau)^\tau$, $(5, -2, 1, \mathbf{0}_6^\tau)^\tau$, and $(5, -2, 1, \mathbf{0}_7^\tau)^\tau$ for $T = 6, 9, 12$, respectively, which is just to simulate large T scenarios, and it only considers the normal disturbance. Except for N, T, and disturbance settings, the others are same as the setting mentioned earlier.

Referring to Baltagi and Yang [21] for the generation of spatial weight matrix, the main idea is that all individuals in a "group" are regarded as "neighbors" to each other and each individual has equal influence on their "neighbor." The steps of the procedure are as follows: (a) we set a constant $c \in (0, 1)$ and let $G_N = \text{round}(N^c)$ be the number of "groups" and $m = N^{1-c}$ be the average number of individuals in each area. (b) We generate "group" size as $n_i \sim U(0.8m, 1.2m)\,(i = 1, \ldots, G_N)$ and adjust n_i so that it satisfies $\sum_{i=1}^{G_N} n_i = N$. (c) We set matrices $\mathbf{W}_i\,(i = 1, 2, \ldots, G_N)$ with zero for diagonal elements and $1/(n_i - 1)$ for others. (d) We set matrix $\mathbf{W} = \mathrm{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_{G_N})$ as the final spatial weight matrix in model (1) which also satisfies A3–A5. In this paper, we set $c = 0.8$ and generate $\mathbf{W}$ by the method mentioned previously.

Considering that choosing different penalty function will not produce additional computation, we use the Adaptive Lasso (AdLasso) penalty function (Zou [22]) as a comparison, and the form of AdLasso penalty function is as follows:

$$p_\lambda(\vartheta) = \lambda \frac{|\vartheta|}{|\vartheta_{OLS}|^2}. \tag{18}$$

Referring to Zhao and Xue [23], we construct a generalized mean square error (GMSE) to compare the estimation accuracy, which is defined as follows:

$$\mathrm{GMSE} = \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^\tau E\left(\mathbf{D}_{i\cdot}^\tau \mathbf{D}_{i\cdot}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right), \tag{19}$$

where $\mathbf{D} = [\mathbf{0}_{N(T-1)}, \mathbb{W}\mathbf{Y}^*, \mathbf{X}^*]$ and $\mathbf{D}_{i\cdot}$ is the ith row of $\mathbf{D}$.

4.2. Monte-Carlo Results. This subsection shows the results of Monte-Carlo simulations which are reported in Table 1–3. Lable "C" in table means the average number of zero regression coefficients that are correctly estimated as zero, and "I" depicts the average number of nonzero regression coefficients that are erroneously set to zero. These two indicators indicate the effects of variable selection.

Table 1 presents the results of the penalized QMLE of a SARP model with fixed effects under normal disturbance when T is fixed. There are some conclusions we can derive from Table 1: (a) The GMSEs reduce, and the effects of variable selection improve as N increases, which proves the consistency and sparsity of estimation. (b) The performance of the penalized QMLE under disturbances of small variance is better than the large one. Small variance means less uncertainty which results in higher estimator accuracy. (c) According to the scenario of $\rho = 0$, the variable selection imposed on the spatial autoregressive coefficient is effective. (d) The GMSEs and the variable selection effect of the SCAD method are almost the same as the AdLasso method in all simulations. Table 2 presents the results of the penalized QMLE of a SARP model with fixed effects under t disturbance and shows that all conclusions derived from Table 1 are still available. Comparing Tables 1 and 2, we know that although the performance of the penalized QMLE under t disturbance is not as good as the normal one, it is still well below the misspecified distribution. Table 3 presents the results of the penalized QMLE of an SARP model with fixed effects under normal disturbance when N is fixed with different T. From Table 3, we can derive the conclusion that the proposed penalized QMLE also performs well under large T.

5. Application

Now, we apply the proposed procedure to analyze the dataset of China's carbon emissions. The dataset contains the carbon emission data and 10 other relevant indicators for 30 provinces and cities in China from 2008 to 2019 (except for Hong Kong, Macao, Taiwan, and Tibet, due to the difficulty of data collection) which means $N = 30, T = 12$, and $d_N = 10$, and is partially shown in Figure 1. The raw data are obtained from the China Energy Statistics Yearbook, National Bureau of Statistics of China (https://data.stats.gov.cn/). The carbon emission data are calculated by using the formula based on energy consumption and energy carbon content which is proposed by the IPCC.

We consider a spatial autoregressive panel model for the analysis of factors affecting carbon emissions as follows:

$$\ln y_{it} = \rho \sum_{j=1}^{N} w_{ij} \ln y_{it} + \sum_{k=1}^{10} \beta_k \ln x_{itk} + u_i + V_{it}. \tag{20}$$

We set up the standardized spatial weight matrix $\mathbf{W}$ based on seven geographical divisions (East China, Northeast China, North China, Central China, South China, Northwest China, and Southwest China), and the variables represented by $X_1 \sim X_{10}, Y$ (original data before logarithm transformation) are listed in Table 4. First, we need to test whether the real data conform to our proposed model. In fact, the QMLE proposed in this paper have penalized the spatial autoregressive coefficient ρ, which sufficiently substitutes for the test of spatial autocorrelations. To test the existence of individual fixed effects, we use the chow test for the poolability of the data and the Hausman test [24]. The p value of the chow test is much less than 0.01, which means there indeed exist some individual effects in the data; the p value of the Hausman test is also much less than 0.01, which means the individual effects should be fixed rather than random. Then, we fit the dataset for (20), ordinary fixed effects panel models (i.e., $\rho = 0$) and classical linear models (i.e., $\rho = 0, u_i = u$) are fitted for comparison. The tuning parameters λ are determined by BIC_λ. All results are given in Table 5. From the analysis results, we can draw the following conclusions: (i) The SARP model with fixed effects based on the SCAD penalty function shows similar results to the model based on the ALASSO penalty function, and they all reduce $\beta_4, \beta_9, \beta_{10}$ to zero. (ii) The penalized estimators $\beta_1, \beta_2, \beta_3, \beta_5, \beta_6$, and β_7 of the SARP model with fixed effects are positive, which is basically reasonable; β_8 is negative, which may be because the investment in the energy industry of the state economy promotes emission reduction technology. (iii) The spatial autoregressive coefficient ρ is not 0 under the SCAD and ALASSO penalty functions based on the SARP model with fixed effects, thus showing the existence of spatial dependence. (iv) Panel models with fixed effects under the penalty function reduce β_1 to zero, which is obviously unreasonable; the R square also indicates that the SARP models with fixed effects are better for China's carbon emission dataset.

Table 1: Variable selections for ρ and β under different N and different σ_0^2 when $V_{it} \sim N(0, \sigma_0^2)$.

(ρ, d_N, N)	Method	$\sigma_0^2 = 2$			$\sigma_0^2 = 1$			$\sigma_0^2 = 0.5$		
		GMSE	C	I	GMSE	C	I	GMSE	C	I
	AdLasso	0.0662	2.934	0	0.0307	2.982	0	0.0156	2.996	0
(0, 5, 50)	SCAD	0.0678	2.730	0	0.0356	2.752	0	0.0198	2.830	0
	Oracle	0.0621	3.000	0	0.0299	3.000	0	0.0151	3.000	0
	AdLasso	0.0367	7.986	0	0.0170	7.996	0	0.0082	8.000	0
(0, 10, 100)	SCAD	0.0341	7.988	0	0.0176	7.990	0	0.0082	7.990	0
	Oracle	0.0292	8.000	0	0.0153	8.000	0	0.0075	8.000	0
	AdLasso	0.0220	12.946	0	0.0101	12.976	0	0.0053	12.976	0
(0, 15, 150)	SCAD	0.0224	12.972	0	0.0119	12.982	0	0.0069	12.984	0
	Oracle	0.0192	13.000	0	0.0100	13.000	0	0.0050	13.000	0
	AdLasso	0.1102	1.910	0	0.0507	1.990	0	0.0233	1.994	0
(0.3, 5, 50)	SCAD	0.0795	1.938	0	0.0485	1.964	0	0.0236	1.964	0
	Oracle	0.0765	2.000	0	0.0413	2.000	0	0.0194	2.000	0
	AdLasso	0.0584	6.942	0	0.0244	6.978	0	0.0114	7.000	0
(0.3, 10, 100)	SCAD	0.0435	6.970	0	0.0204	6.976	0	0.0125	6.986	0
	Oracle	0.0383	7.000	0	0.0201	7.000	0	0.0097	7.000	0
	AdLasso	0.0327	11.934	0	0.0156	11.992	0	0.0075	12.000	0
(0.3, 15, 150)	SCAD	0.0291	11.998	0	0.0174	11.998	0	0.0075	11.998	0
	Oracle	0.0259	12.000	0	0.0128	12.000	0	0.0069	12.000	0
	AdLasso	0.0908	1.978	0	0.0486	1.998	0	0.0221	2.000	0
(0.5, 5, 50)	SCAD	0.0779	1.950	0	0.0447	1.974	0	0.0233	1.988	0
	Oracle	0.0764	2.000	0	0.0388	2.000	0	0.0206	2.000	0
	AdLasso	0.0436	6.970	0	0.0210	7.000	0	0.0108	7.000	0
(0.5, 10, 100)	SCAD	0.0417	6.980	0	0.0201	6.980	0	0.0106	6.996	0
	Oracle	0.0381	7.000	0	0.0199	7.000	0	0.0091	7.000	0
	AdLasso	0.0289	11.990	0	0.0146	11.998	0	0.0070	11.998	0
(0.5, 15, 150)	SCAD	0.0268	11.998	0	0.0145	11.998	0	0.0074	12.000	0
	Oracle	0.0262	12.000	0	0.0129	12.000	0	0.0066	12.000	0
	AdLasso	0.0967	1.984	0	0.0442	1.998	0	0.0206	2.000	0
(0.7, 5, 50)	SCAD	0.0844	1.945	0	0.0453	1.992	0	0.0226	1.994	0
	Oracle	0.0827	2.000	0	0.0402	2.000	0	0.0192	2.000	0
	AdLasso	0.0467	6.996	0	0.0199	7.000	0	0.0096	7.000	0
(0.7, 10, 100)	SCAD	0.0446	6.998	0	0.0210	7.000	0	0.0098	7.000	0
	Oracle	0.0395	7.000	0	0.0191	7.000	0	0.0094	7.000	0
	AdLasso	0.0265	11.980	0	0.0138	11.998	0	0.0070	12.000	0
(0.7, 15, 150)	SCAD	0.0259	11.996	0	0.0136	12.000	0	0.0074	12.000	0
	Oracle	0.0255	12.000	0	0.0126	12.000	0	0.0065	12.000	0

Table 2: Variable selections for ρ and β under different N and different σ_0^2 when $V_{it} \sim \sigma_0/\sqrt{3}t(3)$.

(ρ, d_N, N)	Method	$\sigma_0^2 = 2$			$\sigma_0^2 = 1$			$\sigma_0^2 = 0.5$		
		GMSE	C	I	GMSE	C	I	GMSE	C	I
	AdLasso	0.0708	2.932	0	0.0318	2.976	0	0.0162	2.996	0
(0, 5, 50)	SCAD	0.0742	2.690	0.006	0.0404	2.740	0.002	0.0233	2.810	0
	Oracle	0.0641	3.000	0	0.0299	3.000	0	0.0153	3.000	0
	AdLasso	0.0369	7.974	0.002	0.0172	7.984	0	0.0090	7.986	0
(0, 10, 100)	SCAD	0.0405	7.980	0.002	0.0168	7.980	0.002	0.0089	7.982	0
	Oracle	0.0297	8.000	0	0.0159	8.000	0	0.0080	8.000	0
	AdLasso	0.0246	12.936	0.002	0.0121	12.974	0	0.0055	12.976	0
(0, 15, 150)	SCAD	0.0272	12.968	0	0.0146	12.976	0	0.0073	12.980	0
	Oracle	0.0220	13.000	0	0.0102	13.000	0	0.0055	13.000	0
	AdLasso	0.1230	1.910	0.002	0.0538	1.972	0.002	0.0234	1.992	0
(0.3, 5, 50)	SCAD	0.1033	1.914	0	0.0561	1.952	0	0.0237	1.952	0
	Oracle	0.0768	2.000	0	0.0472	2.000	0	0.0205	2.000	0
	AdLasso	0.0584	6.942	0	0.0288	6.978	0	0.0120	6.998	0
(0.3, 10, 100)	SCAD	0.0460	6.956	0.002	0.0263	6.964	0	0.0136	6.982	0
	Oracle	0.0440	7.000	0	0.0205	7.000	0	0.0110	7.000	0

Table 2: Continued.

(ρ, d_N, N)	Method	$\sigma_0^2 = 2$ GMSE	$\sigma_0^2 = 2$ C	$\sigma_0^2 = 2$ I	$\sigma_0^2 = 1$ GMSE	$\sigma_0^2 = 1$ C	$\sigma_0^2 = 1$ I	$\sigma_0^2 = 0.5$ GMSE	$\sigma_0^2 = 0.5$ C	$\sigma_0^2 = 0.5$ I
	AdLasso	0.0403	11.934	0	0.0167	11.984	0	0.0075	11.998	0
(0.3, 15, 150)	SCAD	0.0384	11.992	0	0.0201	11.992	0	0.0101	11.994	0
	Oracle	0.0272	12.000	0	0.0132	12.000	0	0.0069	12.000	0
	AdLasso	0.1016	1.968	0.006	0.0514	1.998	0.002	0.0224	2.000	0
(0.5, 5, 50)	SCAD	0.0961	1.938	0.002	0.0470	1.972	0.002	0.0254	1.980	0
	Oracle	0.0790	2.000	0	0.0394	2.000	0	0.0216	2.000	0
	AdLasso	0.0466	6.968	0	0.0226	6.998	0	0.0116	7.000	0
(0.5, 10, 100)	SCAD	0.0456	6.964	0	0.0236	6.972	0	0.0123	6.994	0
	Oracle	0.0389	7.000	0	0.0129	7.000	0	0.0098	7.000	0
	AdLasso	0.0285	11.986	0	0.0158	11.992	0	0.0078	11.998	0
(0.5, 15, 150)	SCAD	0.0278	11.996	0	0.0161	11.998	0	0.0089	12.000	0
	Oracle	0.0277	12.000	0	0.0132	12.000	0	0.0069	12.000	0
	AdLasso	0.0984	1.970	0.004	0.0447	1.998	0	0.0214	2.000	0
(0.7, 5, 50)	SCAD	0.0971	1.940	0.002	0.0463	1.982	0	0.0236	1.988	0
	Oracle	0.0827	2.000	0	0.0405	2.000	0	0.0210	2.000	0
	AdLasso	0.0515	6.974	0	0.0346	6.988	0	0.0096	7.000	0
(0.7, 10, 100)	SCAD	0.0530	6.992	0	0.0275	6.994	0	0.0100	6.998	0
	Oracle	0.0398	7.000	0	0.0195	7.000	0	0.0095	7.000	0
	AdLasso	0.0291	11.980	0	0.0142	11.996	0	0.0082	12.000	0
(0.7, 15, 150)	SCAD	0.0318	11.994	0	0.0157	12.000	0	0.0091	12.000	0
	Oracle	0.0273	12.000	0	0.0127	12.000	0	0.0068	12.000	0

Table 3: Variable selections for ρ and β under different T and different σ_0^2 when $V_{it} \sim N(0, \sigma_0^2)$.

(ρ, d_N, T)	Method	$\sigma_0^2 = 2$ GMSE	$\sigma_0^2 = 2$ C	$\sigma_0^2 = 2$ I	$\sigma_0^2 = 1$ GMSE	$\sigma_0^2 = 1$ C	$\sigma_0^2 = 1$ I	$\sigma_0^2 = 0.5$ GMSE	$\sigma_0^2 = 0.5$ C	$\sigma_0^2 = 0.5$ I
	AdLasso	0.0273	4.988	0	0.0125	4.990	0	0.0061	4.994	0
(0, 7, 6)	SCAD	0.0265	4.992	0	0.0123	4.992	0	0.0061	5.000	0
	Oracle	0.0239	5.000	0	0.0116	5.000	0	0.0061	5.000	0
	AdLasso	0.0143	5.970	0	0.0072	5.976	0	0.0037	6.000	0
(0, 8, 9)	SCAD	0.0153	5.970	0	0.0072	5.970	0	0.0037	6.000	0
	Oracle	0.0140	6.000	0	0.0070	6.000	0	0.0037	6.000	0
	AdLasso	0.0111	6.978	0	0.0055	7.000	0	0.0027	7.000	0
(0, 9, 12)	SCAD	0.0116	6.976	0	0.0055	7.000	0	0.0028	7.000	0
	Oracle	0.0108	7.000	0	0.0051	7.000	0	0.0027	7.000	0
	AdLasso	0.0350	3.938	0	0.0173	3.994	0	0.0082	3.996	0
(0.3, 7, 6)	SCAD	0.0350	3.950	0	0.0166	3.994	0	0.0081	4.000	0
	Oracle	0.0308	4.000	0	0.0162	4.000	0	0.0079	4.000	0
	AdLasso	0.0213	4.970	0	0.0099	4.988	0	0.0052	5.000	0
(0.3, 8, 9)	SCAD	0.0206	4.974	0	0.0096	4.990	0	0.0055	4.992	0
	Oracle	0.0204	5.000	0	0.0096	5.000	0	0.0051	5.000	0
	AdLasso	0.0158	5.986	0	0.0079	6.000	0	0.0046	6.000	0
(0.3, 9, 12)	SCAD	0.0152	5.984	0	0.0079	6.000	0	0.0046	6.000	0
	Oracle	0.0143	6.000	0	0.0077	6.000	0	0.0038	6.000	0
	AdLasso	0.0329	3.958	0	0.0163	3.994	0	0.0081	3.996	0
(0.5, 7, 6)	SCAD	0.0345	3.970	0	0.0165	3.996	0	0.0081	3.998	0
	Oracle	0.0326	4.000	0	0.0159	4.000	0	0.0080	4.000	0
	AdLasso	0.0192	4.970	0	0.0095	4.998	0	0.0048	5.000	0
(0.5, 8, 9)	SCAD	0.0201	4.990	0	0.0098	4.998	0	0.0057	5.000	0
	Oracle	0.0185	5.000	0	0.0090	5.000	0	0.0046	5.000	0
	AdLasso	0.0147	5.988	0	0.0072	6.000	0	0.0039	6.000	0
(0.5, 9, 12)	SCAD	0.0152	5.992	0	0.0074	6.000	0	0.0040	6.000	0
	Oracle	0.0139	6.000	0	0.0069	6.000	0	0.0037	6.000	0
	AdLasso	0.0331	3.974	0	0.0164	3.996	0	0.0082	4.000	0
(0.7, 7, 6)	SCAD	0.0353	3.984	0	0.0167	3.998	0	0.0081	4.000	0
	Oracle	0.0326	4.000	0	0.0156	4.000	0	0.0079	4.000	0

Table 3: Continued.

(ρ, d_N, T)	Method	$\sigma_0^2 = 2$			$\sigma_0^2 = 1$			$\sigma_0^2 = 0.5$		
		GMSE	C	I	GMSE	C	I	GMSE	C	I
(0.7, 8, 9)	AdLasso	0.0193	4.964	0	0.0095	4.998	0	0.0048	5.000	0
	SCAD	0.0212	4.984	0	0.0099	4.998	0	0.0052	5.000	0
	Oracle	0.0191	5.000	0	0.0093	5.000	0	0.0046	5.000	0
(0.7, 9, 12)	AdLasso	0.0148	5.994	0	0.0072	6.000	0	0.0038	6.000	0
	SCAD	0.0162	6.000	0	0.0077	6.000	0	0.0040	6.000	0
	Oracle	0.0144	6.000	0	0.0071	6.000	0	0.0038	6.000	0

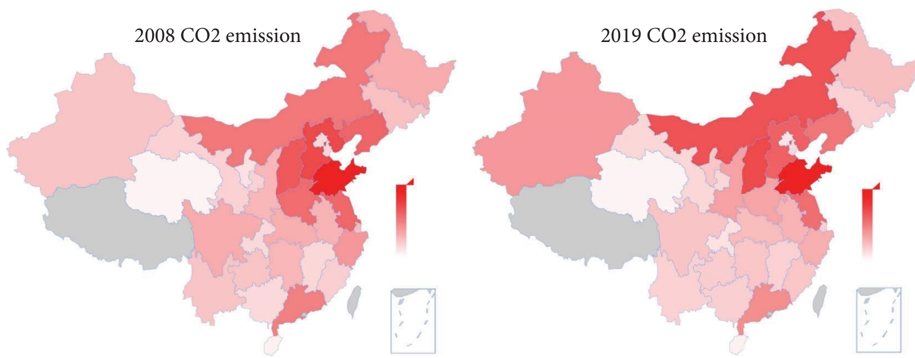

Figure 1: Carbon emissions in 30 provinces and cities in China.

Table 4: Variable descriptions of China's carbon emission dataset.

Variables	Meaning	Symbol	Unit
Y	CO_2 emission	CE	Mt
X_1	Population urbanization	PU	%
X_2	Electric consumption	EC	TWh
X_3	Energy intensity	EI	tce/10 Kyuan
X_4	Space urbanization	SU	%
X_5	GDP per capita	GDP	10 Kyuan
X_6	Population	P	10 K
X_7	Invention patent per 10 K capita	IPC	/10 K
X_8	Fixed investment in the energy industry of state economy	FIEISE	100Myuan
X_9	Volume of freight transport	VFT	10 Kt
X_{10}	Proportion of increase in the tertiary industry	PITI	%

Table 5: Different models fitting China's carbon emission dataset.

Coefficient	SARPF			PF		
	SCAD	ALASSO	Nonpenalty	SCAD	ALASSO	Nonpenalty
ρ	0.2080	0.2000	0.2065	0.0000	0.0000	0.0000
β_1-PU	0.2385	0.1978	0.2231	0.0000	0.0000	0.0672
β_2-EC	0.3476	0.3457	0.3434	0.3648	0.3674	0.3530
β_3-EI	0.5908	0.5942	0.6006	0.6416	0.6397	0.6847
β_4-SU	0.0000	0.0000	0.0243	0.0000	0.0000	0.0191
β_5-GDP	0.1327	0.1514	0.1413	0.2428	0.2448	0.2275
β_6-P	0.2222	0.2016	0.2354	0.1893	0.1798	0.2736
β_7-IPC	0.0668	0.0659	0.0680	0.0810	0.0789	0.0835
β_8-FIEISE	−0.0288	−0.0249	−0.0311	0.0000	0.0000	−0.0292
β_9-VFT	0.0000	0.0000	0.0063	0.0000	0.0000	0.0235
β_{10}-PITI	0.0000	0.0000	−0.0282	0.0000	0.0000	0.0269
R square	0.8410	0.8413	0.8413	0.8245	0.8244	0.8299

[1]SARPF: SARP model with fixed effects and PF: panel model with fixed effects.

6. Conclusion

Within the framework of high-dimensional SARP models with fixed effects, we propose a penalized quasi-maximum likelihood approach based on matrix transformation. This approach can achieve parameter estimation and variable selection simultaneously, and we have proven that the proposed estimators are asymptotically consistent and normally distributed under some conditions. The Monte-Carlo simulations and a real data analysis of China's carbon emissions are conducted to prove the proposed properties, and their results show the effectiveness of the proposed method.

This paper focuses only on the variable selection problem of high-dimensional SARP models with fixed effects which are still linear. There may not be similar results for nonlinear panel models and other more flexible spatial models. Furthermore, we use two penalty functions for variable selections, but the best method remains unknown. We will continue to study these issues in the future.

Appendix

Proof of theorems

In order to prove the theorems, we need the following lemmas:

Lemma A.1. *Under A1–A8, we can have*

$$\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = O_p(1). \tag{A.1}$$

Proof. $1/\sqrt{N(T-1)}\partial \ln L(\theta_0)/\partial\theta$ can be written as

$$\begin{aligned}
\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \rho} &= -\frac{1}{\sqrt{N(T-1)}}tr\mathbf{G} + \frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{V}_0^{*\tau}\mathbb{W}\mathbf{Y}^*,\\
\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \sigma^2} &= -\frac{\sqrt{N(T-1)}}{2\sigma_0^2} + \frac{1}{2\sigma_0^4\sqrt{N(T-1)}}\mathbf{V}_0^{*\tau}\mathbf{V}_0^*,\\
\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{X}^{*\tau}\mathbf{V}_0^*.
\end{aligned} \tag{A.2}$$

Apparently, $E(1/\sqrt{N(T-1)}\,\partial \ln L(\theta_0)/\partial\theta) = \mathbf{0}$. According to A2–A8, we can have

$$\begin{aligned}
&\operatorname{var}\left(\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \rho}\right)\\
&\leq \operatorname{var}\left(\frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{V}_0^{*\tau}\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0\right) + \operatorname{var}\left(\frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{V}_0^{*\tau}\mathbf{G}\mathbf{V}_0^*\right)\\
&= \frac{1}{\sigma_0^2 N(T-1)}(\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0)^\tau\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0 + \frac{1}{\sigma_0^4 N(T-1)}\operatorname{var}(\mathbf{V}_0^{*\tau}\mathbf{G}\mathbf{V}_0^*)\\
&= O(1) + O\left(\frac{1}{h_N}\right) = O(1),\\
&\operatorname{var}\left(\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \sigma^2}\right) = \frac{\mu_4-\sigma_0^4}{4\sigma_0^8} = O(1),\\
&\operatorname{var}\left(\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}}\right)\\
&= E\left(\frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{X}^{*\tau}\mathbf{V}_0^*\mathbf{V}_0^{*\tau}\mathbf{X}^*\right)\\
&= \frac{1}{\sigma_0^2\sqrt{N(T-1)}}\mathbf{X}^{*\tau}\mathbf{X}^* = O(1).
\end{aligned} \tag{A.3}$$

So that all the elements of $1/\sqrt{N(T-1)}\partial \ln L(\theta_0)/\partial\theta$ are $O_p(1)$, then Lemma A.1 holds. □

Lemma A.2. *Under A1–A8, we have*

$$\left\|\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\|^2 = O_p(Nd_N). \tag{A.4}$$

Proof. According to A2–A8 and Lemma A.1, we can have the following equation:

$$\begin{aligned}\left\|\frac{1}{\sqrt{N(T-1)}}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\|^2 &= \frac{1}{N(T-1)}\left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^{\tau}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\\ &= \frac{1}{N(T-1)}\left[\left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \rho}\right)^2 + \left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \sigma^2}\right)^2 + \left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}}\right)\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}^{\tau}}\right]\\ &= O_p(1) + O_p(1) + O_p(d_N) = O_p(d_N).\end{aligned} \tag{A.5}$$

According to A1, we have $\|\partial \ln L(\theta_0)/\partial\theta\|^2 = O_p(Nd_N)$. Thus, Lemma A.2 holds. □

Lemma A.3. *Under A1–A8, we have*

$$\frac{1}{N(T-1)}\left[\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\tau}} + I(\boldsymbol{\theta}_0)\right] = o_p(1). \tag{A.6}$$

Proof. The elements of matrix are linear or quadratic forms of $\mathbf{V}_0^*$, and the matrix is as follows:

$$\frac{1}{N(T-1)}\begin{bmatrix} tr\mathbf{G}^{\tau}\mathbf{G} - \frac{1}{\sigma_0^2}\left[2(\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0)^{\tau}\mathbf{V}_0^* + \mathbf{V}_0^{*\tau}\mathbf{G}^{\tau}\mathbf{G}\mathbf{V}_0^*\right] & * & * \\ \frac{1}{\sigma_0^4}\left(\sigma_0^2 tr\mathbf{G} - \mathbf{V}_0^{*\tau}\mathbf{G}\mathbf{X}^*\boldsymbol{\beta}_0 - \mathbf{V}_0^{*\tau}\mathbf{G}\mathbf{V}_0^*\right) & \frac{N(T-1)}{\sigma_0^4} - \frac{1}{\sigma_0^6}\mathbf{V}_0^{*\tau}\mathbf{V}_0^* & * \\ \frac{1}{\sigma_0^2}\mathbf{X}^{*\tau}\mathbf{G}\mathbf{V}_0^* & -\frac{1}{\sigma_0^4}\mathbf{X}^{*\tau}\mathbf{V}_0^* & \mathbf{0}\end{bmatrix}. \tag{A.7}$$

According to the law of large numbers, we know that all the elements are $o_p(1)$; thus, Lemma A.3 holds. □

Proof of Theorem 2. Let $\alpha_N = \sqrt{d_N}(N^{-1/2} + a_N)$ and $\|\mathbf{u}\| = C$, where C is a large enough constant. Similar to the proof of Theorem 1 in Fan and Peng [2], it is sufficient to prove it if we can prove that for any given $\epsilon > 0$, there is a large enough constant C such that

$$P\left\{\sup_{\|\mathbf{u}\|=C} Q(\boldsymbol{\theta}_0 + \alpha_N \mathbf{u}) < Q(\boldsymbol{\theta}_0)\right\} \geq 1 - \epsilon. \tag{A.8}$$

That means that with probability tending to 1, there is a local maximum $\widehat{\theta}$ in the ball $\{\theta_0 + \alpha_N \mathbf{u}\colon \|\mathbf{u}\| \leq C\}$ such that $\|\widehat{\theta} - \theta_0\| = O_p(\alpha_N)$. Note that $p_\lambda(0) = 0$, then we have

$$\begin{aligned} D(\mathbf{u}) &= Q(\boldsymbol{\theta}_0 + \alpha_N \mathbf{u}) - Q(\boldsymbol{\theta}_0) \\ &= [\ln L(\boldsymbol{\theta}_0 + \alpha_N \mathbf{u}) - \ln L(\boldsymbol{\theta}_0)] - \left\{ N(T-1) \sum_{j=2}^{d_N+2} \left[p_\lambda\left(\left|\theta_{j0} + \alpha_N u_j\right|\right) - p_\lambda\left(\left|\theta_{j0}\right|\right)\right]\right\} \\ &\leq [\ln L(\boldsymbol{\theta}_0 + \alpha_N \mathbf{u}) - \ln L(\boldsymbol{\theta}_0)] - \left\{ N(T-1) \sum_{j=2}^{s} \left[p_\lambda\left(\left|\theta_{j0} + \alpha_N u_j\right|\right) - p_\lambda\left(\left|\theta_{j0}\right|\right)\right]\right\} \\ &\triangleq A_1 + A_2. \end{aligned} \tag{A.9}$$

By the Taylor expansion of $D(\mathbf{u})$, we have

$$\begin{aligned} A_1 &= \alpha_N \left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^\tau \mathbf{u} + \frac{1}{2}\alpha_N^2 \mathbf{u}^\tau \left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau}\right)\mathbf{u} + \frac{1}{6}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\mathbf{u}^\tau \left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau}\right)\mathbf{u}\right]\mathbf{u}\alpha_N^3 \\ &\triangleq A_{11} + A_{12} + A_{13}, \\ A_2 &= -N(T-1)\sum_{j=2}^{s}\left\{\alpha_N p_\lambda'\left(\left|\theta_{j0}\right|\right)\operatorname{sgn}\left(\theta_{j0}\right)u_j + \alpha_N^2 p_\lambda''\left(\left|\theta_{j0}\right|\right)u_j^2[1+o(1)]\right\} \\ &\triangleq A_{21} + A_{22}, \end{aligned} \tag{A.10}$$

where θ^* lies between θ_0 and $\widehat{\theta}$. According to Lemma A.2, we have

$$\begin{aligned} \left|A_{11}\right| &= \left|\alpha_N \left(\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^\tau \mathbf{u}\right| \leq \alpha_N \left\|\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\| \|\mathbf{u}\| \\ &= O_p\left(\alpha_N \sqrt{N d_N}\right)\|\mathbf{u}\| \leq O_p\left(N\alpha_N^2\right)\|\mathbf{u}\|. \end{aligned} \tag{A.11}$$

By Lemma A.3 and A9, we have

$$\begin{aligned} A_{12} &= \frac{1}{2}N(T-1)\alpha_N^2 \mathbf{u}^\tau \left[\frac{1}{N(T-1)}\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau}\right]\mathbf{u} \\ &= \frac{1}{2}N(T-1)\alpha_N^2 \mathbf{u}^\tau \left[-\frac{1}{N(T-1)}I(\boldsymbol{\theta}_0) + o_p(1)\right]\mathbf{u} \\ &= -O_p\left(N\alpha_N^2\right)\|\mathbf{u}\|^2. \end{aligned} \tag{A.12}$$

By the Cauchy–Schwarz inequality, A10, $d_N^4/N \longrightarrow 0$ and $d_N^2 a_N \longrightarrow 0$ as $N \longrightarrow \infty$, we have

$$\begin{aligned}\left\|A_{13}\right\| &= \left|\frac{1}{6}\sum_{i,j,k=1}^{d_N+2}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_i\partial\theta_j\partial\theta_k}u_i u_j u_k \alpha_N^3\right| \\ &\leq \frac{1}{6}\left(N^2\sum_{i,j,k=1}^{d_N+2}M_{ijk}^2\right)^{1/2}\|\mathbf{u}\|^3\alpha_N^3 \\ &= O_p\left(d_N^{3/2}\alpha_N\right)N\alpha_N^2\|\mathbf{u}\|^3 = o_p\left(N\alpha_N^2\right)\|\mathbf{u}\|^2.\end{aligned} \tag{A.13}$$

Thus, $|A_{13}| = o_p(N\alpha_N^2)\|\mathbf{u}\|^2$. For A_{21} and A_{22}, we have

$$\begin{aligned}\left|A_{21}\right| &\leq N\sum_{j=2}^{s}\left|\alpha_N p_\lambda'\left(\left|\theta_{j0}\right|\right)\operatorname{sgn}\left(\theta_{j0}\right)u_j\right| \leq \sqrt{s}N\alpha_N a_N\|\mathbf{u}\| = O_p\left(N\alpha_N^2\right)\|\mathbf{u}\|, \\ \left|A_{22}\right| &= N\sum_{j=2}^{s}\alpha_N^2 p_\lambda''\left(\left|\theta_{j0}\right|\right)\operatorname{sgn}\left(\theta_{j0}\right)u_j^2[1+o(1)] \leq N b_N\alpha_N^2\|\mathbf{u}\|^2 = o_p\left(N\alpha_N^2\right)\|\mathbf{u}\|^2.\end{aligned} \tag{A.14}$$

Thus, A_{12} is negative and dominates all terms when C is large enough. Then, Theorem 2 holds. □

Proof of Theorem 3. Let $\eta = C\sqrt{d_N/N}$, we just need to prove that with probability tending to 1 as $N\longrightarrow\infty$ for any θ_1 satisfying $\|\theta_1-\theta_{10}\| = O_p(\sqrt{d_N/N})$ we have, for $j = s+1,\ldots,d_N$:

$$\begin{aligned}&\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} < 0, \quad 0<\theta_j<\eta, \\ &\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} > 0, -\eta<\theta_j<0.\end{aligned} \tag{A.15}$$

By Taylor expansion, we can have

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} &= \frac{\partial \ln L(\boldsymbol{\theta})}{\partial\theta_j} - N(T-1)p_\lambda'\left(\left|\theta_{j0}\right|\right)\operatorname{sgn}\left(\theta_{j0}\right) \\ &= \frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j} + \sum_{k=1}^{d_N+2}\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k}(\theta_k-\theta_{k0}) + \sum_{l,k=1}^{d_N+2}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_k\partial\theta_l}(\theta_k-\theta_{k0})(\theta_l-\theta_{l0}) \\ &\quad - N(T-1)p_\lambda'\left(\left|\theta_{j0}\right|\right)\operatorname{sgn}\left(\theta_j\right) \\ &\triangleq B_1+B_2+B_3+B_4,\end{aligned} \tag{A.16}$$

where θ^* lies between θ_0 and $\widehat{\theta}$. We consider B_1, B_2, B_3 first, and by Lemma A.1, we can have

$$B_1 = O_p(\sqrt{N}) = O_p\left(\sqrt{Nd_N}\right). \tag{A.17}$$

Term B_2 can be written as

$$\begin{aligned}B_2 &= \sum_{k=1}^{d_N+2}\left[\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k} - E\left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k}\right)\right](\theta_k-\theta_{k0}) + \sum_{k=1}^{d_N+2}E\left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k}\right)(\theta_k-\theta_{k0}) \\ &\triangleq B_{21}+B_{22}.\end{aligned} \tag{A.18}$$

According to Lemma A.3, A11 and $\|\theta-\theta_0\| = O_p(\sqrt{d_N/N})$, by the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
\left|B_{21}\right| &\leq \left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\| \sqrt{\sum_{k=1}^{d_N+2}\left[\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k}-E\left(\frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\theta_j\partial\theta_k}\right)\right]^2} \\
&= O_p\left(\sqrt{\frac{d_N}{N}}\right)O_p\left(\sqrt{Nd_N}\right) = O_p(d_N) = O_p\left(\sqrt{Nd_N}\right).
\end{aligned}
\tag{A.19}
$$

So, $|B_{21}| = O_p\sqrt{Nd_N}$. From Lemma A.3, the Cauchy–Schwarz inequality and A11, we have

$$
\begin{aligned}
\left|B_{22}\right| &= N(T-1)\left|\sum_{k=1}^{d_N+2}\left[\frac{1}{N(T-1)}I(\boldsymbol{\theta}_0)(j,k)\right](\theta_k-\theta_{k0})\right| \\
&\leq N(T-1)\sqrt{\sum_{k=1}^{d_N+2}\left[\frac{1}{N(T-1)}I(\boldsymbol{\theta}_0)(j,k)\right]^2}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\| \\
&= O(N)O(1)O_p\left(\sqrt{\frac{d_N}{N}}\right) = O_p\left(\sqrt{Nd_N}\right).
\end{aligned}
\tag{A.20}
$$

From the Cauchy–Schwarz inequality and A10, we can have

$$
\begin{aligned}
\left|B_3\right| &= \left|\sum_{l,k=1}^{d_N+2}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_k\partial\theta_l}(\theta_k-\theta_{k0})(\theta_l-\theta_{l0})\right| \\
&= N(T-1)\left|\sum_{l,k=1}^{d_N+2}\frac{1}{N(T-1)}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_k\partial\theta_l}(\theta_k-\theta_{k0})(\theta_l-\theta_{l0})\right| \\
&\leq N(T-1)\sqrt{\sum_{l,k=1}^{d_N+2}\left(\frac{1}{N(T-1)}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_k\partial\theta_l}\right)^2}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\|^2. \\
&= O_p\left(Nd_N\frac{d_N}{N}\right) = O_p\left(d_N^2\right) = o_p\left(\sqrt{Nd_N}\right).
\end{aligned}
\tag{A.21}
$$

Then, $B_1 + B_2 + B_3 = O_p(\sqrt{Nd_N})$, i.e.,

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} &= O_p\left(\sqrt{Nd_N}\right) - N(T-1)p_\lambda'\left(\left|\theta_j\right|\right)\mathrm{sgn}\left(\theta_j\right) \\
&= N\lambda\left[O_p\left(\frac{\sqrt{d_N/N}}{\lambda}\right) - (T-1)\frac{p_\lambda'\left(\left|\theta_j\right|\right)}{\lambda}\mathrm{sgn}\left(\theta_j\right)\right].
\end{aligned}
\tag{A.22}
$$

By the condition of the theorem and A12, we have $p_\lambda'(|\theta_j|)/\lambda > 0$, so it is easy to know that the sign of $\partial Q(\theta)/\partial\theta_j$ is completely determined by the sign of θ_j which implies Theorem 2. □

Proof of Theorem 4. Theorem 2 shows that there is a $\sqrt{d_N/N}$ consistent local maximum $\widehat{\theta}$ if we choose a proper tune parameter λ. According to Theorem 3, part (i) holds, so we only need to prove part (ii). As we know, there is an estimator $\widehat{\theta} = (\widehat{\theta}_1^\tau, 0^\tau)^\tau$ which satisfies the following equation:

$$\left.\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0 \quad (j = 1, \ldots, d_N + 2). \tag{A.23}$$

We note it as $\partial Q(\widehat{\theta})/\partial\theta$, and by the Taylor expansion, we have

$$\begin{aligned}&\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{\partial^2 \ln L(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + \frac{1}{2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^\tau \frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\partial^2 \ln L(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\\ &= N(T-1)\left[p_\lambda'(\boldsymbol{\theta}_0) + p_\lambda''(\boldsymbol{\theta}_0)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + o_p(1)\right],\end{aligned} \tag{A.24}$$

where θ^* is between θ_0 and $\widehat{\theta}$, $p_\lambda''(\theta_0) = \operatorname{diag}\{0, p_\lambda''(|\rho_0|), p_\lambda''(|\beta_{10}|), \ldots, p_\lambda''(|\beta_{(d_N)0}|)\}$, $p_\lambda'(\theta_0) = [0, p_\lambda'(|\rho_0|)\operatorname{sgn}(\rho_0), p_\lambda'(|\beta_{10}|)\operatorname{sgn}(\beta_{10}), \ldots, p_\lambda'(|\beta_{(d_N)0}|)\operatorname{sgn}(\beta_{(d_N)0})]^\tau$. Mark $1/2(\widehat{\theta} - \theta)^\tau\partial/\partial\theta(\partial^2 \ln L(\theta^*)/\partial\theta\partial\theta^\tau)(\widehat{\theta} - \theta_0)$ as $\nabla^3 \ln L(\theta^*)$. According to A10 and the Cauchy–Schwarz inequality, we can have

$$\left\|\nabla^3 \ln L(\boldsymbol{\theta}^*)\right\| \leq \frac{N(T-1)}{2}\sqrt{\sum_{j,k,l=1}^{d_N+2}\left(\frac{1}{N(T-1)}\frac{\partial^3 \ln L(\boldsymbol{\theta}^*)}{\partial\theta_j\partial\theta_k\partial\theta_l}\right)^2}\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\|^2 = O_p\left(d_N^{5/2}\right). \tag{A.25}$$

Let $\mathbf{B} = [I_{s\times s}, \mathbf{0}_{s\times(d_N+2-s)}]$ and $\mathbf{A} = (-1/(N(T-1))\partial^2 \ln L(\theta_{\mathbf{0}})/\partial\theta\partial\theta^\tau)_{s\times s}$ which is the first upper-left $s\times s$ submatrix of $-1/(N(T-1))\partial^2 \ln L(\theta_{\mathbf{0}})/\partial\theta\partial\theta^\tau$, then from part (i), (A.24) and (A.25), we know

$$\mathbf{B}\frac{\partial \ln L(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = N(T-1)\left[\mathbf{A} + \sum(\boldsymbol{\theta}_0) + o_p(1)\right]\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}\right) + N(T-1)\mathbf{b} - O_p\left(d_N^{5/2}\right). \tag{A.26}$$

We multiply both sides by $1/\sqrt{N(T-1)}$ and denote $1/\sqrt{N(T-1)}\mathbf{B}\partial \ln L(\theta_0)/\partial\theta$ as $\mathbf{Z}_N$

$$\mathbf{Z}_N = \sqrt{N(T-1)}\left\{\left[\mathbf{A} + \sum(\boldsymbol{\theta}_0) + o_p(1)\right]\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}\right) + \mathbf{b}\right\} - O_p\left(\sqrt{\frac{d_N^5}{N}}\right). \tag{A.27}$$

From Lemma A.3, we have $\mathbf{A} = 1/(N(T-1))I_1(\theta_{10}, \mathbf{0}) + o_p(1)$, then

$$\mathbf{Z}_N = \sqrt{N(T-1)}\left\{\left[\frac{1}{N(T-1)}I_1(\boldsymbol{\theta}_{10}, \mathbf{0}) + \sum(\boldsymbol{\theta}_0) + o_p(1)\right]\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}\right) + \mathbf{b}\right\} - O_p\left(\sqrt{\frac{d_N^5}{N}}\right). \tag{A.28}$$

By the condition that $d_N^5/N \longrightarrow 0$ and Slutsky's theorem, as $N \longrightarrow \infty$, we have

$$\sqrt{N(T-1)}\left\{\left[\frac{1}{N(T-1)}I_1(\boldsymbol{\theta}_{10},\mathbf{0})+\sum(\boldsymbol{\theta}_0)\right]\left(\widehat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{10}\right)+\mathbf{b}\right\}\xrightarrow{D}\mathbf{Z}_N, \tag{A.29}$$

where " $\longrightarrow^D$ " means convergence in distribution. Furthermore, the central limit theorem for linear-quadratic forms of Kelejian and Prucha [25] and assumption 9 shows

$$\mathbf{Z}_N=\frac{1}{\sqrt{N(T-1)}}\mathbf{B}\frac{\partial\ln L(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\xrightarrow{D}N(0,\mathbf{H}+\mathbf{K}). \tag{A.30}$$

Thus,

$$\sqrt{N(T-1)}\left\{\left[\frac{1}{N(T-1)}I_1(\boldsymbol{\theta}_{10},\mathbf{0})+\sum(\boldsymbol{\theta}_0)\right]\left(\widehat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_{10}\right)+\mathbf{b}\right\}\xrightarrow{D}N(0,\mathbf{H}+\mathbf{K}). \tag{A.31}$$

□

Data Availability

The research data are available from the corresponding author upon request on the website https://data.stats.gov.cn/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Statistical Science Research Project (2021LY061), the Startup Foundation for Talents at Hangzhou Normal University (2019QDL039), and the Postgraduate Research Innovation Promotion Project of the Hangzhou Normal University (2022HSDYJSKY181).

References

[1] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[2] J. Q. Fan and R. Z. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[3] A. Cliff and J. K. Ord, *Spatial Autocorrelation*, Pion, London, UK, 1973.

[4] L. Anselin, "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity," *Geographical Analysis*, vol. 20, pp. 1–17, 1998.

[5] H. H. Kelejian and I. R. Prucha, "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *The Journal of Real Estate Finance and Economics*, vol. 17, no. 1, pp. 99–121, 1998.

[6] L. F. Lee, "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, vol. 72, no. 6, pp. 1899–1925, 2004.

[7] C. H. Wei, S. Guo, and S. F. Zhai, "Statistical inference of partially linear varying coefficient spatial autoregressive models," *Economic Modelling*, vol. 64, pp. 553–559, 2017.

[8] J. Du, X. Q. Sun, R. Y. Cao, and Z. Z. Zhang, "Statistical inference for partially linear additive spatial autoregressive models," *Spatial Statistics*, vol. 25, pp. 52–67, 2018.

[9] S. L. Cheng, J. B. Chen, and X. Liu, "GMM estimation of partially linear single-index spatial autoregressive model," *Spatial Statistics*, vol. 31, Article ID 100354, 2019.

[10] X. W. Dai, L. B. Jin, M. Z. Tian, and L. Shi, "Bayesian local influence for spatial autoregressive models with heteroscedasticity," *Statistical Papers*, vol. 60, no. 5, pp. 1423–1446, 2019.

[11] A. Gupta and P. M. Robinson, "Inference on higher-order spatial autoregressive models with increasingly many parameters," *Journal of Econometrics*, vol. 186, no. 1, pp. 19–31, 2015.

[12] X. Lin and L. F. Lee, "GMM estimation of spatial autoregressive models with unknown het-eroskedasticity," *Journal of Econometrics*, vol. 157, no. 1, pp. 34–52, 2010.

[13] R. Q. Tian, M. J. Xia, and D. K. Xu, "Variable selection of heterogeneous spatial autoregressive models via double-penalized likelihood," *Symmetry*, vol. 14, no. 6, pp. 1–17, 2022.

[14] R. Q. Tian, D. K. Xu, J. Du, and J. F. Zhang, "Bayesian estimation for partially linear varying coefficient spatial autoregressive models," *Statistics and Its Interface*, vol. 15, no. 1, pp. 105–113, 2022.

[15] B. H. Baltagi, S. H. Song, and W. Koh, "Testing panel data regression models with spatial error correlation," *Journal of Econometrics*, vol. 117, no. 1, pp. 123–150, 2003.

[16] L. F. Lee and J. H. Yu, "Estimation of spatial autoregressive panel data models with fixed effects," *Journal of Econometrics*, vol. 154, no. 2, pp. 165–185, 2010.

[17] Y. Y. Ju, N. S. Tang, and X. X. Li, "Bayesian local influence analysis of skew-normal spatial dynamic panel data models," *Journal of Statistical Computation and Simulation*, vol. 88, no. 12, pp. 2342–2364, 2018.

[18] X. Liu, J. B. Chen, and S. L. Cheng, "A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model," *Spatial Statistics*, vol. 25, pp. 86–104, 2018.

[19] T. F. Xie, R. Y. Cao, and J. Du, "Variable selection for spatial autoregressive models with a diverging number of parameters," *Statistical Papers*, vol. 61, no. 3, pp. 1125–1145, 2020.

[20] J. Q. Fan and H. Peng, "Nonconcave penalized likelihood with a diverging number of parameters," *Annals of Statistics*, vol. 32, no. 3, pp. 928–961, 2004.

[21] B. H. Baltagi and Z. Yang, "Standardized LM tests for spatial error dependence in linear or panel regressions," *The Econometrics Journal*, vol. 16, no. 1, pp. 103–134, 2013.

[22] H. Zou, "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[23] P. X. Zhao and L. G. Xue, "Variable selection for semi-parametric varying coefficient partially linear errors-in-variables models," *Journal of Multivariate Analysis*, vol. 101, no. 8, pp. 1872–1883, 2010.

[24] J. A. Hausman, "Specification tests in econometrics," *Econometrica*, vol. 46, no. 6, pp. 1251–1271, 1978.

[25] H. H. Kelejian and I. R. Prucha, "On the asymptotic distribution of the moran I test statistic with application," *Econometrics*, vol. 104, pp. 219–257, 2012.