mathematics

Article

Variable Selection for Spatial Logistic Autoregressive Models

Jiaxuan Liang, Yi Cheng, Yuqi Su, Shuyue Xiao and Yunquan Song *

School of Science, China University of Petroleum, Qingdao 266580, China
* Correspondence: syqfly1980@upc.edu.cn

Abstract: When the spatial response variables are discrete, the spatial logistic autoregressive model adds an additional network structure to the ordinary logistic regression model to improve the classification accuracy. With the emergence of high-dimensional data in various fields, sparse spatial logistic regression models have attracted a great deal of interest from researchers. For the high-dimensional spatial logistic autoregressive model, in this paper, we propose a variable selection method with for the spatial logistic model. To identify important variables and make predictions, one efficient algorithm is employed to solve the penalized likelihood function. Simulations and a real example show that our methods perform well in a limited sample.

Keywords: spatial logistic autoregressive model; variable selection; maximum likelihood

MSC: 62F12; 62G08; 62G20; 62J07T07

check for updates

Citation: Liang, J.; Cheng, Y.; Su, Y.; Xiao, S.; Song, Y. Variable Selection for Spatial Logistic Autoregressive Models. *Mathematics* **2022**, *10*, 3095. https://doi.org/10.3390/math10173095

Academic Editor: Dan Vilenchik

Received: 21 June 2022
Accepted: 22 August 2022
Published: 29 August 2022

1. Introduction

As a branch of modern econometrics, spatial econometrics has been widely used in many traditional economic fields such as regional economy, real estate economy, demand analysis, and labor economics as well as ecology, epidemiology, and other disciplines. At present, many modeling methods have been used to deal with spatial econometrics. In many models, our commonly used models are Spatial Autoregressive Model (SAR), Spatial-Lag of X Model (SLX), Spatial Error Model (SEM), Spatial Autoregressive Combined Model (SAC), and Spatial Durbin Model (SDM), etc. Among them, the spatial autoregressive (SAR) model proposed by Ord (1975) [1] has been popularly used. In the framework of spatial autoregression, Anselin (1980) [2] discussed the estimation method of parameters. Cliff and Ord (1981) [3] investigated the maximum likelihood estimation based on this work. Lee (2004) [4] applied the maximum likelihood estimation and quasimaximum likelihood estimation to the spatial econometric model and strictly deduced the asymptotic distribution of the estimated parameters. SAR models can be applied to many fields including social Sciences (Ma (2020) [5], Darmofal (2015) [6]), real estate (Osland (2010) [7]), crime incidents (Ahmar et al. (2018) [8]), analyzing poverty (Islamy et al. (2021) [9]) and ecological analysis (Jay et al. (2018) [10]). According to the SAR model, we can view the factors affecting dependent variables as a natural combination of the independent variables and the spatial spillover effects of the dependent variables. Thence, the model can conveniently deal with traditional covariates and network dependence problems.

The main focus of the existing spatial econometrics literature is the statistical inference of the spatial mean regression model whose dependent variable is continuous, which only reflects the location information of the conditional distribution of the explained variable. As many scholars have found that many dependent variables in practical problems are discrete variables in empirical research, spatial logistic autoregressive model has attracted the attention of theoretical econometrists and applied researchers. Spatial logistic regression model studies the influence of covariates on the correlation response of spatial discrete values. Spatial logistic regression model based on classification technology to model spatial data is a new field of spatial econometrics, and the related research is still limited.

In recent years, many studies on spatial autoregressive model have proposed several methods to analyze the regression under high-dimensional data. Penalized techniques and their variants have attracted people's attention to high-dimensional data analysis by shrinking inactive coefficients to 0. Such as LASSO (Tibshirani (1996) [11]), SCAD (Fan and Li (2001) [12]) and MCP (Zhang (2010) [13]) for mean regression. For high dimensional spatial data, Han et al. (2017) [14] proposed the estimation and model selection of higher-order spatial autoregressive models through an efficient Bayesian approach. They developed a more efficient algorithm based on the exchange algorithm, in order to solve the problem of computing the Jacobian determinant in the likelihood function of the parametric posterior distribution when the number of cross-sectional spatial units is large. Liu et al. (2018) [15] developed a penalized quasimaximum likelihood method for simultaneous model selection and parameter estimation in the spatial autoregressive model with independent and identical distributed errors. Michael (2019) [16] proposed two global-local shrinkage priors under the background of high-dimensional matrix exponential spatial specifications. Especially when the number of parameters to be estimated surpasses the number of observations, both simulations and real data results reveal that they perform particularly well in high-dimensional settings. Song et al. (2021) [17] propose a class of penalized robust regression estimators on the basis of exponential squared loss with independent and identical distributed errors for general spatial autoregressive models.Numerical studies demonstrate that the proposed method is especially robust and applicable when there are outliers or intensive noise in the observations or when the estimated spatial weight matrix is imprecise. To alleviate the problems of computational time for Bayesian model-averaging for spatial autoregressive models, Justin M. Leach et al. (2022) [18] proposed a novel approach to using the spike-and-slab prior with the elastic net when predictors display spatial structure. The elastic net may outperform LASSO when the number of predictors far exceeds the sample size and the predictors behave strong correlations. Romina et al. (2022) [19] studied the variable selection for spatial regression models with locations on irregular lattices and errors according to Conditional or Simultaneous Autoregressive (CAR or SAR) models. The strategy is to whiten the residuals by estimating their spatial covariance matrix and then proceed by performing the standard L1-penalized regression LASSO for independent data on the transformed model. The above studies are only for spatial autoregressive models with continuous response variables. As far as we know, there are still no studies on variable selection for spatial logistic autoregression in high dimensional space data.

In this paper, we put forward a class of penalized regression estimators on the basis of quasimaximum likelihood with independent and identical distributed errors for general spatial logistic autoregressive models. Consider estimating $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ by solving the following optimization problem,

$$\mathbf{min}_{(\boldsymbol{\beta},\rho)} - \ln[L(\boldsymbol{\beta},\rho)] + 2n\sum_j^p p_\lambda(|\beta_j|).$$

In this work, we presented a variable selection method for spatial logistic autoregressive based on the quasimaximum likelihood loss function and the SCAD penalty. This method was capable of selecting significant predictors while estimating regression coefficients. The following are the main contributions of this work.

1. We construct a variable selection method for high-dimensional spatial logistic regression model.
2. We propose a new optimization algorithm to solve the penalized spatial logistic regression model and then construct the model selection criteria to select the optimal tuning parameter.
3. We conducted specific numerical studies and verified the effectiveness of the proposed method in selecting significant variables. Numerical studies indicate that the proposed method far outperforms the comparative methods in terms of the number of correctly identified zero coefficients, the number of incorrectly identified nonzero coefficients and ME.

The outline of the remainder of this paper is as follows. Section 2 discusses the general specification of the spatial autoregression model and the spatial logistic regression model. Section 3 proposes a penalized spatial logistic autoregressive model and a optimization algorithm to solve this model. Section 4 performs a simulation study to evaluate the effect of variable selection in spatial logistic regression. Section 5 applies the model to a real example and Section 6 concludes.

2. Materials and Methods

2.1. Spatial Autoregressive Model (SAR)

Consider a network with n nodes. We can describe the structure of the network by the matrix $A \in R^{n\times n}$. Define $a_{ij} = 1$ when node i follows node j, and $a_{ij} = 0$ otherwise. If we have a $n \times 1$ vector of observations on the dependent variable Y and a $n \times d$ matrix of regressors X, we can establish the following SAR model:

$$Y = \rho WY + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\rho \in \mathcal{R}$ is network autocorrelation coefficient and $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)^T \in \mathcal{R}^d$ is the regression coefficient vector. W is the row-normalized version of A such that $w_{ij} = a_{ij} / \sum_{j=1}^{n} a_{ij}$. Let $\theta = (\rho, \boldsymbol{\beta}^T)^T \in \mathcal{R}^{d+1}$ be the estimator and denote $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T$ be the error vector, we assume it is *i.i.d.* with zero mean and finite variance σ^2.

Denote $G = I - \rho W, S = Y - \rho WY - X\boldsymbol{\beta}$, then we have the log likelihood function of SAR model:

$$\ln L(\theta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|G| - \frac{1}{2\sigma^2}S^T S. \tag{2}$$

2.2. Spatial Logistic Regression Model

Spatial logistic regression model is a combination of spatial autoregressive model and logistic regression model, and the response variables of logistic regression model can be binary classification or multi-classification, nevertheless, we only consider that the response variables are binary.

The model (1) can be written as:

$$\begin{aligned} \boldsymbol{y}^* &= (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon} \\ &= \boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \boldsymbol{e} \sim MVN(0, \Omega) \end{aligned} \tag{3}$$

To make it easily distinguish, we use $\boldsymbol{y}^*$ instead of $\boldsymbol{Y}$, where $\boldsymbol{H}$ is an $(n \times n)$ matrix, define $\boldsymbol{H} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$, define the ith component of $\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta}$ as $[\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta}]_i$, $\boldsymbol{e}$ is an $(n \times 1)$ vector, define $\boldsymbol{e} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}$. Latent variable $\boldsymbol{y}^*$ has binary category which is defined as variable y:

$$y_i = \begin{cases} 1, \text{ for } y_i^* > 0 \\ 0, \text{ for } y_i^* \leq 0 \end{cases} \tag{4}$$

Therefore, the probability of $P(y_i = 1)$ and $P(y_i = 0)$ is:

$$\begin{aligned} P(y_i = 1 \mid \mathrm{X}_i) &= P(y_i^* > 0) \\ &= P([\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta}]_i + \boldsymbol{e} > 0) \\ &= P(-\boldsymbol{e} \leq [\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta}]_i) \\ &= \frac{1}{1 + \exp(-[\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta}]_i)} \end{aligned} \tag{5}$$

$$
\begin{aligned}
P(y_i = 0 \mid X_i) &= P(y_i^* \leq 0) \\
&= P([\boldsymbol{HX\beta}]_i + \boldsymbol{e} \leq 0) \\
&= P(-\boldsymbol{e} > [\boldsymbol{HX\beta}]_i) \\
&= 1 - P(-\boldsymbol{e} \leq [\boldsymbol{HX\beta}]_i) \\
&= 1 - \frac{1}{1 + \exp(-[\boldsymbol{HX\beta}]_i)}
\end{aligned} \tag{6}
$$

When we assume the mean value of $\boldsymbol{e}$ is 0 and the variance is $\boldsymbol{\Omega}$, then we get

$$
P(y_i = 1) = \frac{1}{1 + \exp\left(-\frac{[\boldsymbol{HX\beta}]_i}{\Omega_{ii}}\right)} \tag{7}
$$

where $\boldsymbol{\Omega}_{ii}$ is diagonal element of $\boldsymbol{\Omega}$, which is formed as $\boldsymbol{\Omega} = \left[(\boldsymbol{I} - \rho\boldsymbol{W})^T(\boldsymbol{I} - \rho\boldsymbol{W})\right]^{-1}$, so the same can be obtained $P(y_i = 0)$.

The estimation of spatial logistic regression parameters can be acquired by maximum likelihood estimation (MLE). The parameter is estimated by maximizing likelihood function of random variable y_i, which follow Bernoulli distribution:

$$
L(\boldsymbol{\beta}, \rho) = \prod_{i=1}^{n} \left(\left[\frac{1}{1 + \exp\left(-\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Omega}_{ii}}\right)} \right]^{y_i} \left[1 - \frac{1}{1 + \exp\left(-\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Omega}_{ii}}\right)} \right]^{1-y_i} \right) \tag{8}
$$

Then, the likelihood function is transformed by natural log(ln) as follows:

$$
\ln[L(\boldsymbol{\beta}, \rho)] = \sum_{i=1}^{n} y_i \ln\left[\frac{1}{1 + \exp\left(-\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Omega}_{ii}}\right)} \right] + \sum_{i=1}^{n} (1 - y_i) \ln\left[1 - \frac{1}{1 + \exp\left(-\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Omega}_{ii}}\right)} \right] \tag{9}
$$

To estimate $\boldsymbol{\beta}$, we use the maximization Formula (9), then define $\hat{\boldsymbol{\beta}} = argmax \ln[L(\boldsymbol{\beta}, \rho)]$.

3. Results

3.1. Variable Selection with Linear Constraints

In this section, we consider the variable selection of high-dimensional spatial logistic regression model. The objective function of the model consists of the likelihood function of the Spatial Logistic regression Model($\ln[L(\boldsymbol{\beta}, \rho)]$) and the penalty function.By minimizing the objective function, we can get the estimated parameters.

We will study the variable selection of high-dimensional spatial logistic regression model:

$$
(\hat{\boldsymbol{\beta}}, \hat{\rho}) = \mathbf{arg}\min_{(\boldsymbol{\beta}, \rho)} - \ln[L(\boldsymbol{\beta}, \rho)] + 2n \sum_{j}^{p} p_\lambda(|\beta_j|). \tag{10}
$$

where $p_\lambda(\bullet)$ is the penalty function, the shrinkage degree of penalty is determined by the tune parameter λ in the penalty term. Some possible choices include:

(1) the LASSO penalty with $p_\lambda(t) = \lambda|t|$;
(2) the SCAD penalty with $p_\lambda(t) = \lambda \int_0^{|t|} \min\{1, (a - t/\lambda)_+/(a-1)\}dt, a > 2$ where v_+ denotes its positive part, that is, $vI(v \geq 0)$;
(3) the MCP with $p_\lambda(t) = \lambda \int_0^{|t|} (1 - t/(\lambda a))_+ \mathrm{d}t, a > 1$.

Fan and Li (2001) [12] use unbiasedness, sparsity and continuity to evaluate penalty functions. LASSO is not unbiased, and MCP calculation is relatively complex. Fan and Li (2001) [12] pointed out that LASSO does not have the properties of Oracle, but SCAD has them. Compared with ridge regression, SCAD method reduces the prediction variance of the model. At the same time, compared with LASSO, SCAD method reduces the deviation of parameter estimation, so it has received extensive attention. so we choose to use SCAD penalty here.

3.2. A Feasible Algorithm

This is a nonconcave optimization problem that maximizes the penalty likelihood function $Q(\theta)$. In the study of classical linear regression models, some kinds of algorithms have been developed to find and compute the local solutions of nonconcave penalized function, such as local quadratic approximation (LQA) algorithm (Fan, 2001 [12]), local linear approximation (LLA) algorithm and coordinate descent algorithm. Unfortunately, owing to the spatial correlation of the model, we find that the algorithm mentioned above unable to be straight forward calculate the minimum value of the nonconcave penalized likelihood function $Q(\theta)$. Therefore, we propose a new iterative algorithm:

Step 1. Initialize $\boldsymbol{\theta}^{(0)} = \left(\sigma^{(0)}, \rho^{(0)}, \boldsymbol{\beta}^{(0)}\right)$.

Step 2. Update $\sigma^{(m+1)} = \arg\min_{\sigma\in(0,\infty)} \left\{ l_1(\sigma) = -\ln L_n\left(\sigma, \rho^{(m)}, \boldsymbol{\beta}^{(m)}\right)\right\}$.

Step 3. Update $\rho^{(m+1)} = \arg\min_{\rho\in(-1,1)} \left\{ l_2(\rho) = -Q\left(\sigma^{(m+1)}, \rho, \boldsymbol{\beta}^{(m)}\right)\right\}$.

Step 4. Update $\boldsymbol{\beta}^{(m+1)} = \arg\min_{\beta\in R^k} \left\{ l_3(\boldsymbol{\beta}) = -Q\left(\sigma^{(m+1)}, \rho^{(m+1)}, \boldsymbol{\beta}\right)\right\}$.

Step 5. Iterate Step 2 to Step 4 until convergence and denote the final estimator of $(\sigma^2, \rho, \boldsymbol{\beta})$ as $\left(\hat{\sigma}^2, \hat{\rho}, \hat{\boldsymbol{\beta}}\right)$, then $\hat{\boldsymbol{\theta}} = \left(\hat{\sigma}^2, \hat{\rho}, \hat{\boldsymbol{\beta}}^{\mathrm{T}}\right)^{\mathrm{T}}$.

In Steps 2 and 3, since they are both one-dimensional nonlinear optimization problems, they can be solved by the Brent method (Press et al. known). Therefore, the model (1) can be written as the following linear model:

$$\boldsymbol{Y}_n^* = \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{E}_n,$$

where $\boldsymbol{Y}_n^* = \boldsymbol{Y}_n - \rho \boldsymbol{W}_n \boldsymbol{Y}_n$. Thence, we can apply the LQA algorithm in the classic linear regression model to accomplish this step. We also need to determine the tuning parameters a and λ in the SCAD function. Here, we accept the suggestion of Fan and Li (2001) and set $a = 3.7$.

3.3. The Selection of Tuning Parameter

Based on the above, we chose to use SCAD penalty. The penalty function is defined as:

$$p_\lambda(|\boldsymbol{\beta}|) = \begin{cases} \lambda|\beta_j|, & 0 \le |\beta_j| < \lambda, \\ -\left(|\beta_j|^2 - 2a\lambda|\beta_j| + \lambda^2\right)/(2a-2), & \lambda \le |\beta_j| < a\lambda, \\ (a+1)\lambda^2/2, & |\beta_j| \ge a\lambda, \end{cases} \tag{11}$$

where $\lambda \ge 0$ and $a > 2$ are tune parameters. In the article of Fan and Li (2001), it is suggested that a should be 3.7, and λ determines the shrinkage strength of parameter estimation. In this paper, λ is determined by Bayesian information criterion (BIC) .

The selection of tuning parameter λ is an important application of degrees of freedom. We will use Bayesian information criterion (BIC) (Schwarz (1978)) as the criteria for model selection in this paper: For determining the value of λ, we use the Bayesian information criterion

$$\mathrm{BIC}(\lambda) = -2\ln L_n(\hat{\boldsymbol{\theta}}) + \alpha(\lambda)\log n$$

where $\alpha(\lambda) = \sum_{j=1}^{k+2} I\left(\hat{\theta}_j \ne 0\right)$. Then a choice of λ is $\hat{\lambda} = \arg\min_{\lambda}\{\mathrm{BIC}(\lambda)\}$.

4. Simulation Studies

In former sections, we put forward an advanced spatial logistic regression model, and here use R code to fulfil Monte Carlo simulations to assess and test the performance of variable selection of this model. The data for the simulated experiments are originated from Formula (1), in which the covariates are identified as following a $(q+3)$-dim normal distribution with zero mean and a covariance matrix σ_{ij}. Its concrete expression is $\sigma_{ij} = 0.5^{|i-j|}$. We make up our mind to set the sample size $n \in \{60, 90, 120\}$ and the number of inessential covariates $q \in \{5, 10, 35, 85\}$ in the subsequent simulation studies. Consequently, X can be thought of as an $n \times (q+3)$ matrix.

In regard to SAR model, the network autocorrelation coefficient ρ arises from the uniform distribution ranging from $(\rho_1 - 1)$ to $(\rho_1 + 1)$. The feasible value of ρ_1 can be 0.2, 0.5 and 0.8, representing the spatial coefficient of different intensities. To compare model performance, we also think over setting $\rho = 0$, that suggests no spatial dependence in this model, thence model (2.1) will be a classic linear model.

Additionally, let the spatial weight matrix is $W = \boldsymbol{I}_R \otimes \boldsymbol{B}_m$. In which, B_m is $(1/(m-1))\left(\mathbf{1}_m \cdot \mathbf{1}_m^T - \boldsymbol{I}_m\right)$, and "$\otimes$" expresses Kronecker product. $\mathbf{1}_m$ is an m-dim column vector of ones. In this formula, we premeditate $m = 3$ and several not the same values of R, for instance $R \in \{10, 20, 30, 40\}$.

The regression coefficients are set to $\beta = \left(3, 2, 1.6, \mathbf{0}_q\right)^T$, where $(\beta_1, \beta_2, \beta_3)$ is generated from a 3-dimensional normal distribution with mean vector (3, 2, 1.6) and covariance matrix $0.001\boldsymbol{I}_3$, $\mathbf{0}_q$ is a zero vector of q dimension. The response variable y^* is given by the following formula:

$$y^* = (I_n - \rho W)^{-1}(X\beta + \varepsilon_n) \tag{12}$$

Then we turn the response variable into category variable by the following formula:

$$Y_i = \begin{cases} 1, \text{ for } y_i^* > 0 \\ 0, \text{ for } y_i^* \leq 0 \end{cases} \tag{13}$$

Thus, the response variable Y of the binary classification is obtained. In order to verify the robustness of the model, we consider two error distributions: $\varepsilon_n \sim N(0, \sigma^2 I_n)$, which is denoted as ε_0, and Mixed Gaussian distribution: $\varepsilon_n \sim 0.5N(-1, 2.5^2 I_n) + 0.5N(1, 0.5^2 I_n)$, denoted as ε_1. σ^2 is generated by the Uniform distribution on the interval $[\sigma_1 - 0.1, \sigma_1 + 0.1]$, where $\sigma_1 \in \{1, 2\}$. In the second case, $E(\varepsilon) = 0, Mode(\varepsilon) = 1$.

4.1. Simulation Indicators

For each case, we repeat 100 times. In order to evaluate the variable selection ability of the model, we define three indicators:

- **Correct**: the average number of coefficients, of the true zeros correctly set to zero;
- **Incorrect**: the average number of coefficients, of the true nonzeros incorrectly set to zero;
- **ME**: the mean error between the true and estimator, which is defined by;

$$\frac{1}{100}\sum_{i=1}^{100}\|\theta_i - \hat{\theta}_i\| \tag{14}$$

- **MAD**: the median absolute deviations of parameter estimation;
- **MEAN**: the means of parameter estimation;
- **SD**: the standard deviations of parameter estimation.

4.2. Simulation Results

For each case, the following reference is based on 100 simulations. To facilitate comparison of model results, the variable selection results by our algorithm in the SAR model are written as SLR. Meanwhile, LLA represents the variable selection results presented by the LLA algorithm, where simulated samples are arised from the classical regression model. Furthermore, to contrast the effects of different penalty functions, LASSO penalty $p_\lambda(\delta) = \lambda|\delta|$ is introduced for variable selection. In Tables 1 and 2, it is clear that as the number of samples n increases, the accuracy of variable selection in both models gradually improves. When the sample size n is 120, the performance of the SLR model reaches the ideal state. In this case, the model has a higher "Correct" and a lower "Incorrect" and ME, which are consistent with our speculations. When the spatial effect is weak or moderate (such as $\rho_1 = 0.2$ and 0.5) or does not exist $(\rho = 0)$, the ME of the SLR model is significantly lower than that of the LLA model, and the number of variables which are correctly selected is evidently higher than that of the LLA. These show that under low to medium intensity spatial effects, the SLR model has high accuracy, and the influence of spatial effects on the model is weakened. When we set error is ε_1, these two models exhibit good robustness. When the spatial effect is strong $(\rho_1 = 0.8)$, the estimations are more inexact, suggesting that neglecting spatial effect will seriously bias the estimate.

Table 1. Simulation results of variable selection via SCAD penalty function (q = 5).

Method		$n = 60, q = 5$		$n = 90, q = 5$		$n = 120, q = 5$	
		SLR	**LLA**	**SLR**	**LLA**	**SLR**	**LLA**
ρ = 0.0	Correct	4.7500	3.0600	4.7500	2.8700	4.8000	3.0900
ε_0	Incorrect	0.0300	0.0300	0.0900	0.0300	0.0300	0.0000
	ME	0.5443	39.710	0.5667	27.523	0.4189	7.7846
ρ = 0.0	Correct	4.7700	2.6600	4.8300	2.7800	4.8400	3.6100
ε_1	Incorrect	0.0300	0.0300	0.0500	0.0200	0.0000	0.0000
	ME	0.5267	26.015	0.5089	16.697	0.4398	5.2516
ρ_1 = 0.2	Correct	4.7700	2.5400	4.7400	3.0500	4.8300	3.5100
ε_0	Incorrect	0.0200	0.0200	0.0500	0.0200	0.0100	0.0000
	ME	0.6629	26.553	0.5714	13.238	0.5226	4.7464
ρ_1 = 0.2	Correct	4.7300	2.6800	4.7500	3.2400	4.8500	3.7800
ε_1	Incorrect	0.0500	0.0300	0.0700	0.0000	0.0200	0.0000
	ME	0.6204	20.937	0.5862	5.3993	0.5180	2.5285
ρ_1 = 0.5	Correct	4.4700	2.6500	4.7500	3.3900	4.7600	4.1400
ε_0	Incorrect	0.0300	0.0300	0.0200	0.0300	0.0100	0.0000
	ME	1.4812	10.299	1.1671	1.8735	1.2150	1.2959
ρ_1 = 0.5	Correct	4.5400	2.8800	4.7700	3.6800	4.7400	4.3900
ε_1	Incorrect	0.0700	0.0300	0.0400	0.0300	0.0100	0.0100
	ME	1.3525	6.7171	1.1989	1.6030	1.2130	1.2991
ρ_1 = 0.8	Correct	2.0400	3.4000	3.6900	4.2100	3.9000	4.7600
ε_0	Incorrect	0.0400	0.3200	0.0100	0.2700	0.0100	0.2300
	ME	6.4407	2.2473	5.6574	2.1468	5.7883	2.1607
ρ_1 = 0.8	Correct	3.2800	3.4200	3.3600	4.3500	3.8000	4.7100
ε_1	Incorrect	0.0300	0.3300	0.0100	0.3600	0.0200	0.2600
	ME	6.4877	2.2987	5.9324	2.2557	5.7459	2.1878

Moreover, by comparing the ME between the two penalty functions, we find that the SCAD penalty is better than the LASSO penalty by the SLR algorithm, although they get closer as the sample size increases. From the perspective of Correct and Incorrect, although LASSO has a lower error rate, the accuracy rate of SCAD is always at a high level, and its estimation error is significantly lower. In Table 1, the number of coefficients with zero values chosen correctly is close to the true value, and as the number of samples increases, the average number of coefficients with zero values incorrectly chosen is close to zero. All simulation results are consistent with theoretical analysis. However, under the same conditions, the number of correctly chosen zero coefficients for models with LASSO penalties is less than half of the true value. Under the same number of observations, the average number of correct selecting zero-value coefficients by LASSO penalty is remarkably lower than by SCAD penalty, which may imply that SCAD penalty tends to give smaller models than LASSO penalty. These results are in accordance with the studies obtained by Fan and Li (2001) [12].

Table 2. Simulation results of variable selection via LASSO penalty function (q = 5).

Method		$n = 60, q = 5$		$n = 90, q = 5$		$n = 120, q = 5$	
		SLR	**LLA**	**SLR**	**LLA**	**SLR**	**LLA**
ρ = 0.0	Correct	2.2800	0.8000	2.4000	1.0700	2.5800	1.4000
ε_0	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	0.9315	20.399	0.6583	13.610	0.4613	7.1323
ρ = 0.0	Correct	2.2700	0.7800	2.0900	1.1600	2.1600	1.8800
ε_1	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	1.0551	15.274	0.6850	7.6718	0.6105	3.4521
ρ_1 = 0.2	Correct	2.3000	0.8000	2.5200	1.1900	2.5000	1.7200
ε_0	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	0.9735	16.272	0.6776	8.5473	0.4623	4.6528
ρ_1 = 0.2	Correct	2.3400	0.8700	2.2600	1.6100	2.1500	1.9100
ε_1	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	1.0599	13.332	0.6879	4.5913	0.5618	2.4616
ρ_1 = 0.5	Correct	2.5200	1.0800	2.6300	1.7300	2.5600	2.2600
ε_0	Incorrect	0.0100	0.0100	0.0000	0.0000	0.0000	0.0000
	ME	1.0038	5.6760	0.7466	1.8178	0.7139	1.2749
ρ_1 = 0.5	Correct	2.3500	1.2000	2.5100	1.8100	2.4800	2.3700
ε_1	Incorrect	0.0000	0.0100	0.0000	0.0000	0.0000	0.0100
	ME	1.1404	5.1790	0.7768	1.5665	0.7014	1.2904
ρ_1 = 0.8	Correct	2.2700	1.4700	2.2600	2.2000	2.3100	2.8300
ε_0	Incorrect	0.1800	0.0900	0.0900	0.0600	0.0300	0.0200
	ME	2.4780	2.2362	2.4000	2.1520	2.4363	2.1664
ρ_1 = 0.8	Correct	2.2400	1.5900	2.3200	2.3900	2.4000	2.9700
ε_1	Incorrect	0.1900	0.0500	0.0900	0.0600	0.0400	0.0500
	ME	2.4328	2.2905	2.2099	2.2588	2.3256	2.1934

In terms of the Correct and Incorrect, Tables 1 and 2 present the results with fixed $q = 5$. In this case, the sample dimension is low. To explore the performance of the models in high-dimensional situations, we set $q = 10$ and $q = 35$. In Tables 3 and 4, in high-dimensional cases, we found that SLR model is significantly better than LLA model. As a result of the inferior performance of the LLA model, we only present the SLR model in Table 5 in order to save space. On the whole, the variable selection effect of the model is reduced, and the number of correctly selected zero coefficients is slightly different from the true value. However, it is not difficult to see that as the sample size increases, the maximum likelihood function of the penalty with SCAD and LASSO penalties effectively reduces the complicacy of the model, and the performance of the SCAD penalty is better than that of the LASSO penalty.

To measure the robustness of the model, we not only set up two error distributions, but also calculated the median absolute deviations of different parameters. MAD is a robust statistic that is more resilient than standard deviations for the treatment of outliers in a dataset. It is a good measure of model robustness.Based on the above conclusions, we have compared the performance of the SLR model under two different penalty functions in Tables 6 and 7. We use means, standard deviations, median absolute deviations to measure the effect of the models. In the simulation study, we find that the means and variances of the errors for the different study combinations do not differ significantly and are not shown in the tables.

According to Tables 6 and 7, we observe that the estimate of nonzero parameter $(\beta_1, \beta_2, \beta_3)$ gradually tends to true value as the sample size n increases, indicating that our model has higher accuracy. In Table 6, we are surprised to find that the model performed better when we set ε_1. It is reflected in lower MAD values, lower variances and parameter estimates that are closer to the true value. In contrast to Table 6, the estimates of the nonzero coefficients are generally lower than the true values in the models with the LASSO penalty, indicating that the penalty is less effective in shrinkage degree than the SCAD penalty. By analyzing the estimation results of nonzero parameters, it is found that as the spatial effect increases, the estimation of the standard deviation becomes less precise.

Table 3. Simulation results of variable selection via SCAD penalty function ($q = 35$).

Method		$n = 60, q = 35$		$n = 90, q = 35$		$n = 120, q = 35$	
		SLR	**LLA**	**SLR**	**LLA**	**SLR**	**LLA**
$\rho = 0.0$	Correct	29.3500	22.0400	32.1000	28.6700	32.8500	32.0000
ε_0	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	0.9802	1.3659	0.6939	0.9054	0.6388	0.7051
$\rho = 0.0$	Correct	27.2200	18.2900	28.1300	24.9300	30.3100	29.1100
ε_1	Incorrect	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	1.1501	1.7382	0.9268	1.1563	0.7181	0.8876
$\rho_1 = 0.2$	Correct	28.0500	19.7400	30.7000	26.3700	31.6500	30.1600
ε_0	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	1.3621	1.6003	0.8765	1.0687	0.7747	0.8413
$\rho_1 = 0.2$	Correct	27.5200	17.0400	28.2200	23.0000	30.4600	27.3400
ε_1	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000
	ME	1.5050	1.9408	1.0588	1.2988	0.9202	1.0093
$\rho_1 = 0.5$	Correct	22.9000	10.7400	27.0000	15.2200	33.8000	18.4600
ε_0	Incorrect	0.0000	0.0100	0.0000	0.0000	0.0100	0.0000
	ME	3.2759	3.4503	2.8857	2.3748	1.8003	1.9243
$\rho_1 = 0.5$	Correct	21.1000	9.9700	27.1500	14.1000	30.1700	17.3200
ε_1	Incorrect	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ME	3.2655	3.7152	2.3088	2.5436	2.0182	2.0512
$\rho_1 = 0.8$	Correct	12.8300	3.3600	15.6700	4.9000	22.5000	6.2600
ε_0	Incorrect	0.1600	0.0100	0.0000	0.0100	0.0100	0.0000
	ME	11.8324	12.7776	11.1446	9.2003	7.3722	7.8126
$\rho_1 = 0.8$	Correct	12.6500	3.1700	17.4500	4.8400	21.1100	6.2300
ε_1	Incorrect	0.0500	0.0300	0.0100	0.0200	0.1100	0.0000
	ME	11.5900	13.1540	9.4916	9.4230	7.0805	7.9880

Table 4. Simulation results of variable selection via LASSO penalty function ($q = 35$).

Method		$n = 60, q = 35$		$n = 90, q = 35$		$n = 120, q = 35$	
		SLR	**LLA**	**SLR**	**LLA**	**SLR**	**LLA**
ρ=0.0	Correct	26.5900	21.2400	30.2800	16.7200	31.4000	14.4300
ε_0	Incorrect	0.1100	0.0000	0.0700	0.0000	0.0100	0.0000
	ME	3.4254	6.4100	3.4241	10.1052	3.4143	13.2199
ρ = 0.0	Correct	26.2600	19.7400	30.1700	14.3700	31.2600	11.3600
ε_1	Incorrect	0.1400	0.0000	0.0500	0.0000	0.0600	0.0000
	ME	3.4459	6.5985	3.4320	11.5742	3.4218	16.0966
ρ_1 = 0.2	Correct	26.1800	20.4600	30.1800	15.0000	31.4000	12.0700
ε_0	Incorrect	0.0800	0.0000	0.1200	0.0000	0.0300	0.0000
	ME	3.4386	6.5198	3.4394	10.8455	3.4181	14.9343
ρ_1 = 0.2	Correct	25.8600	19.2600	29.9200	13.0900	31.6100	9.9100
ε_1	Incorrect	0.1200	0.0100	0.0800	0.0000	0.0200	0.0000
	ME	3.4500	6.6266	3.4472	11.8648	3.4280	17.3861
ρ_1 = 0.5	Correct	25.1900	16.1500	29.3600	9.6100	31.0000	5.7400
ε_0	Incorrect	0.2700	0.0500	0.1700	0.0000	0.0700	0.0000
	ME	3.5079	8.2365	3.4832	17.2251	3.4817	24.5215
ρ_1 = 0.5	Correct	25.3100	15.4600	29.1600	8.8100	30.5300	5.4500
ε_1	Incorrect	0.2200	0.0800	0.1600	0.0200	0.0800	0.0000
	ME	3.5032	8.3032	3.4932	18.5977	3.4925	24.0086
ρ_1 = 0.8	Correct	24.8300	11.5900	28.6400	4.2000	30.3300	7.5600
ε_0	Incorrect	0.5400	0.2300	0.3800	0.0200	0.3000	0.0100
	ME	3.6149	12.5654	3.5891	27.8645	3.5918	12.1578
ρ_1 = 0.8	Correct	24.8000	11.7400	28.5800	3.2500	30.2600	7.9300
ε_1	Incorrect	0.5100	0.2400	0.4300	0.0200	0.2800	0.0200
	ME	3.6146	13.1417	3.6000	28.7948	3.5946	11.8813

Table 5. Simulation results of variable selection when the number of components of $\boldsymbol{\beta}$ is 10.

Method		$n = 60, q = 10$		$n = 90, q = 10$		$n = 120, q = 10$	
		SCAD	**LASSO**	**SCAD**	**LASSO**	**SCAD**	**LASSO**
ρ = 0.0	Correct	7.6600	5.1800	7.6000	5.3900	7.4900	5.2300
ε_0	Incorrect	0.0100	0.0000	0.0300	0.0000	0.0200	0.0000
	ME	0.5453	0.8422	0.5249	0.5510	0.4791	0.4343
ρ = 0.0	Correct	7.1200	4.5200	7.4100	4.7400	7.4800	4.6100
ε_1	Incorrect	0.0500	0.0000	0.0500	0.0000	0.0500	0.0000
	ME	0.6408	1.0402	0.5247	0.7471	0.5418	0.5791
ρ_1 = 0.2	Correct	7.3100	5.1000	7.3700	5.2400	7.4800	5.2300
ε_0	Incorrect	0.0300	0.0000	0.0600	0.0000	0.0100	0.0000
	ME	0.5960	0.8643	0.5856	0.5680	0.5250	0.4817
ρ_1 = 0.2	Correct	6.9100	4.1600	7.1900	4.6200	7.4500	4.6500
ε_1	Incorrect	0.1000	0.0000	0.0700	0.0000	0.0400	0.0000
	ME	0.8297	1.0863	0.5988	0.7330	0.5598	0.5998
ρ_1 = 0.5	Correct	6.7600	4.4800	7.0800	4.9400	7.0000	5.1300
ε_0	Incorrect	0.0300	0.0100	0.0000	0.0000	0.0100	0.0000
	ME	1.3827	1.0727	1.1509	0.8280	1.2536	0.8056

Table 5. *Cont.*

Method		$n = 60, q = 10$		$n = 90, q = 10$		$n = 120, q = 10$	
		SCAD	**LASSO**	**SCAD**	**LASSO**	**SCAD**	**LASSO**
$\rho_1 = 0.5$	Correct	6.5000	4.7100	7.0400	4.7000	7.2100	4.6100
ε_1	Incorrect	0.0500	0.0100	0.0100	0.0000	0.0200	0.0000
	ME	1.5347	1.1668	1.1242	0.8354	1.2782	0.8670
$\rho_1 = 0.8$	Correct	5.1700	3.3200	5.3400	4.0700	5.3400	4.0700
ε_0	Incorrect	0.0000	0.1300	0.0000	0.0500	0.0400	0.0300
	ME	6.9527	3.2176	6.0424	2.9194	5.7429	2.9715
$\rho_1 = 0.8$	Correct	4.8400	3.4100	5.2900	3.9900	5.4200	3.8900
ε_1	Incorrect	0.0400	0.1700	0.0100	0.0500	0.0100	0.0300
	ME	7.0023	3.3527	6.0356	2.8764	5.9172	3.1053

Table 6. Standard deviations and means of estimators of the nonzero regression coefficients.

Method		$n = 60, q = 5$			$n = 90, q = 5$			$n = 120, q = 5$		
		MAD	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**
SCAD	β_1	0.2455	3.1868	0.3518	0.2636	3.2012	0.3604	0.2059	3.1560	0.2301
$\rho_1 = 0.0$	β_2	0.2596	2.1391	0.3367	0.1700	2.1222	0.3150	0.2003	2.1257	0.2507
ε_0	β_3	0.2545	1.6710	0.4275	0.3215	1.5604	0.5628	0.1856	1.6082	0.3618
	σ	0.0697			0.0739			0.0771		
SCAD	β_1	0.3278	3.1061	0.3140	0.2193	3.1143	0.2818	0.2097	3.1453	0.3095
$\rho_1 = 0.0$	β_2	0.2980	2.0998	0.3197	0.2506	2.1097	0.3547	0.2600	2.1116	0.2416
ε_1	β_3	0.3092	1.6416	0.3999	0.2528	1.5930	0.4137	0.2382	1.6805	0.2548
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		
SCAD	β_1	0.3405	3.3175	0.4782	0.2345	3.2538	0.2998	0.2424	3.2824	0.4111
$\rho_1 = 0.2$	β_2	0.3215	2.2035	0.3670	0.2928	2.1281	0.3042	0.1988	2.2157	0.3676
ε_0	β_3	0.2729	1.7262	0.4074	0.2850	1.6297	0.4549	0.2329	1.7127	0.3269
	σ	0.0697			0.0739			0.0771		
SCAD	β_1	0.3066	3.1977	0.3551	0.3634	3.2155	0.3451	0.2703	3.2236	0.3035
$\rho_1 = 0.2$	β_2	0.3218	2.1285	0.3340	0.2726	2.1541	0.3093	0.2447	2.1187	0.2909
ε_1	β_3	0.2900	1.6381	0.4931	0.2387	1.5575	0.4864	0.2433	1.6999	0.3470
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		
SCAD	β_1	0.5887	3.8562	0.9985	0.4864	3.7980	0.5011	0.4398	3.7895	0.5038
$\rho_1 = 0.5$	β_2	0.5312	2.5957	0.6553	0.4291	2.4581	0.4526	0.4217	2.5749	0.4995
ε_0	β_3	0.4764	2.0705	0.7608	0.4020	1.9487	0.4694	0.3718	1.9845	0.4893
	σ	0.0697			0.0739			0.0771		
SCAD	β_1	0.6080	3.7618	0.6917	0.5305	3.7769	0.5967	0.4413	3.8122	0.4579
$\rho_1 = 0.5$	β_2	0.4864	2.5355	0.5607	0.4128	2.3989	0.5558	0.4526	2.5171	0.4376
ε_1	β_3	0.4859	1.8453	0.6894	0.4985	1.8754	0.5335	0.3339	2.0543	0.4598
	σ_1	0.0697			0.0697			0.0771		
	σ_2	0.0602			0.0602			0.0848		
SCAD	β_1	2.0179	7.0442	2.3790	1.5072	7.0772	2.3910	1.7351	7.1679	2.1987
$\rho_1 = 0.8$	β_2	1.9750	4.9501	2.2412	1.3111	4.2876	1.5123	1.1471	4.6002	1.8087
ε_0	β_3	1.6558	3.7185	2.2376	1.1559	3.6419	1.5232	1.1865	3.8710	1.5290
	σ	0.0697			0.0739			0.0771		

Table 6. *Cont.*

Method		$n = 60, q = 5$			$n = 90, q = 5$			$n = 120, q = 5$		
		MAD	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**
SCAD	β_1	2.1170	7.2101	2.5888	1.9399	7.1463	2.4596	1.8336	7.1832	2.1051
$\rho_1 = 0.8$	β_2	1.8740	4.9594	2.1258	1.4001	4.5172	1.6663	1.2885	4.6123	1.7462
ε_1	β_3	1.7771	3.6942	2.0929	1.3543	3.6027	1.5921	1.2216	3.7060	1.4169
	σ_1	0.0697			0.0697			0.0771		
	σ_2	0.0602			0.0602			0.0848		

Table 7. Standard deviations and means of estimators of the nonzero regression coefficients.

Method		$n = 60, q = 5$			$n = 90, q = 5$			$n = 120, q = 5$		
		MAD	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**	**MAD**	**MEAN**	**SD**
LASSO	β_1	0.4058	2.4638	0.3234	0.3058	2.8603	0.3533	0.2666	2.9453	0.2548
$\rho_1 = 0.0$	β_2	0.4386	1.6110	0.3471	0.2665	1.7970	0.2771	0.1943	1.9427	0.2309
ε_0	β_3	0.3268	1.2491	0.3228	0.2853	1.3814	0.3217	0.2121	1.5010	0.2249
	σ	0.0697			0.0739			0.0771		
LASSO	β_1	0.3199	2.4047	0.3250	0.3159	2.7176	0.2929	0.3114	2.7647	0.3266
$\rho_1 = 0.0$	β_2	0.3726	1.5360	0.3728	0.3192	1.7868	0.2897	0.2994	1.8135	0.2906
ε_1	β_3	0.3604	1.1914	0.3440	0.2847	1.3243	0.3018	0.2531	1.4356	0.2365
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		
LASSO	β_1	0.3247	2.4977	0.3587	0.3301	2.8398	0.2992	0.2347	3.0282	0.2689
$\rho_1 = 0.2$	β_2	0.3406	1.5942	0.3675	0.3332	1.7930	0.3526	0.2108	1.9934	0.2039
ε_0	β_3	0.2805	1.2211	0.3387	0.3017	1.3784	0.3036	0.2691	1.5115	0.2332
	σ	0.0697			0.0739			0.0771		
LASSO	β_1	0.3436	2.4367	0.3680	0.3421	2.7136	0.3099	0.3185	2.8873	0.3071
$\rho_1 = 0.2$	β_2	0.3795	1.5498	0.3694	0.3200	1.7739	0.2954	0.2543	1.8448	0.2794
ε_1	β_3	0.3847	1.1731	0.3862	0.2699	1.3189	0.2642	0.2747	1.4865	0.2548
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		
LASSO	β_1	0.4018	2.7972	0.5157	0.3895	3.1733	0.3737	0.3060	3.3872	0.3614
$\rho_1 = 0.5$	β_2	0.5035	1.7161	0.4279	0.4342	1.9538	0.4156	0.2779	2.1410	0.3373
ε_0	β_3	0.5346	1.2690	0.4965	0.3798	1.4383	0.3857	0.2761	1.6648	0.2872
	σ	0.0697			0.0739			0.0771		
LASSO	β_1	0.4514	2.6579	0.5536	0.4769	3.0253	0.4331	0.3343	3.2222	0.3860
$\rho_1 = 0.5$	β_2	0.4932	1.6128	0.5226	0.3960	1.8848	0.3754	0.3472	2.0364	0.3742
ε_1	β_3	0.4378	1.1378	0.4317	0.3315	1.3910	0.3836	0.3460	1.6178	0.3657
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		
LASSO	β_1	1.1758	3.2846	1.3361	1.1406	4.3800	1.8818	0.7971	4.6060	1.8911
$\rho_1 = 0.8$	β_2	1.0994	1.9483	1.3684	1.0162	2.1504	1.1529	0.8062	2.4770	1.2061
ε_0	β_3	1.0369	1.2004	1.3797	1.0446	1.3117	0.9479	0.7605	1.7002	0.8070
	σ	0.0697			0.0739			0.0771		
LASSO	β_1	0.9996	3.4737	1.5004	1.0690	4.2106	1.4199	1.1367	4.5226	1.7562
$\rho_1 = 0.8$	β_2	1.1422	1.7939	1.4225	1.1369	2.0650	1.0878	0.8946	2.3502	1.0537
ε_1	β_3	0.9546	0.9301	1.0104	0.9934	1.2820	0.9865	0.6433	1.8759	0.9692
	σ_1	0.0697			0.0739			0.0771		
	σ_2	0.0602			0.0605			0.0848		

5. Data Example

In this section, we provide a real-world example to demonstrate the performance of the variable selection procedure proposed in this paper for spatial logistic regression models.

5.1. The Sample Data

The dataset is a collection of different types of land area data (recorded every five years) from 1954 to 2012 for the 48 states of the United States, using the Spatial Logistic regression models to analyse land utilization data. The dependent variables are binary ("1" indicates that the land utilization rate is low, that is, most of the land has not been effectively developed; "0" indicates that the land utilization rate is high, that is, most of the land has been developed and utilized). The independent variables have eight attributes, including: Cropland used for crops, Cropland used for pasture, Cropland idled, Grassland pasture and range, Forest-use land grazed, Land in rural transportation facilities, Land in urban areas and Other idle land (showed in Table 8).

The spatial weight matrix is generally set by many basic principles, including Rook and Queen Contiguity, binary distance bands, inverse distance, k-nearest neighbors and kernel weights (Yrigoyen 2013 [20]). We choose to generate the spatial weight matrix by defining the common boundary (Queen criterion), which is the contiguity-based spatial weights matrix. For sensitivity and robustness analysis, we decided to use a different spatial weight matrix (distance-based spatial weight matrix) to estimate our model (showed in Appendix A). More ordinarily, the method to establish a space matrix is as follows: if they share common boundaries, the weight is 1, otherwise the weight is zero. Furthermore, spatial weight matrices are usually row-normalised in practice.

Table 8. Summary of variables.

Variable Name	Description
CLand_C	Cropland used for crops
CLand_P	Cropland used for pasture
CLand_I	Cropland idled
Grass_P	Grassland pasture and range
Land_G	Forest-use land grazed
Land_T	Land in rural transportation facilities
Land_U	Land in urban areas
Land_I	Other idle land

5.2. Model Selection and Estimation

Tables 9 and 10 presents the results of the maximum likelihood estimate via the SCAD and LASSO penalties under the fitting of a Spatial Logistic regression model. For the Spatial Logistic regression model, the penalized maximum likelihood with the SCAD penalty shows Forest-use land grazed, Cropland used for crop and Land in rural transportation facilitie are unimportant.These three variables have a small effect on land utilisation rate (the absolute value of the coefficient is less than 0.0001) and can be ignored.In Table 9, it is significant to note that Cropland idled, Grassland pasture and range, Other idle land, these three variables are significant, with Cropland idled and Other idle land being the most significant. The results of this experiment are in line with our prediction.

Comparing the effects of these two penalty functions in practice, the SCAD penalty has a more significant shrinkage degree on the model. This is mainly due to the fact that the model with the SCAD penalty is able to select most of the important variables and discard the less important ones as much as possible. However, the LASSO penalty does not perform well in this respect. We can visually see in Table 10 that there are many coefficients with small estimates, but the limitation of the shrinkage degree of LASSO penalty causes them not to be estimated as 0 (Here, we consider coefficient estimates less than 0.0001 to be judged as 0).

Table 9. Parameter estimates using penalty estimates via SCAD under a spatial logistic model.

Year	CLand_C	CLand_P	CLand_I	Grass_P	Land_G	Land_T	Land_U	Land_I
1954	–	–	0.9820	0.4230	–	–	–	0.6020
1959	–	1.7000	–	−1.7400	–	–	0.3050	1.2700
1964	–	1.2800	–	−0.6150	–	–	–	0.8920
1969	–	1.5500	0.7150	−1.4900	–	−0.3850	–	1.0600
1974	0.8540	1.5200	1.3900	−2.2600	–	−0.5960	−0.6510	1.5200
1978	0.6630	–	1.9000	−1.7200	−0.2490	−0.2870	0.6140	1.2600
1982	0.3020	1.4500	1.1200	−1.1200	–	–	−0.9840	1.1400
1987	–	1.1500	0.6640	–	–	–	−1.1100	1.2500
1992	0.9810	–	0.1740	2.8200	–	−0.9490	−2.4100	1.8000
1997	0.5680	–	0.2620	2.7500	–	−0.8770	−1.9300	1.8700
2002	2.4200	–	1.6100	−1.2100	–	−1.0000	–	2.3900
2007	1.3000	0.8520	1.7600	−1.2000	–	–	−0.3830	2.4700
2012	–	–	2.5800	−0.4680	–	−0.7990	–	1.0200

Table 10. Parameter estimates using penalty estimates via LASSO under a spatial logistic model.

Year	CLand_C	CLand_P	CLand_I	Grass_P	Land_G	Land_T	Land_U	Land_I
1954	–	–	0.7070	0.6400	−0.4510	–	0.2930	0.1510
1959	0.0607	–	0.8950	0.0101	−0.3660	0.1550	0.0085	0.4150
1964	−0.0019	0.0007	1.3279	−0.2782	−0.0156	–	−0.0010	0.3816
1969	0.0010	0.1430	0.3649	0.0333	0.0797	0.0035	–	0.8547
1974	−0.2485	–	0.6534	0.2852	−0.0021	−0.0021	−0.0047	0.7925
1978	–	0.0797	0.2419	–	–	–	0.2834	1.3200
1982	−0.1152	−0.0899	0.2129	0.0859	−0.0549	0.0592	−0.1792	1.9704
1987	–	0.0403	–	1.0158	0.0306	0.0148	0.0374	1.4981
1992	–	0.3098	0.5501	–	0.0806	0.4132	−0.1816	0.8606
1997	−0.0018	0.1444	1.0449	–	−0.0427	−0.1067	–	0.6831
2002	–	−0.0816	1.2938	−0.0104	–	0.0248	0.0180	0.6495
2007	0.2985	−0.0101	−0.0109	0.4834	0.0063	0.3495	0.0231	0.6077
2012	0.0132	0.0014	0.5038	0.0145	0.0420	0.0051	–	1.0461

From the empirical results, we are able to find that the model is less stable in terms of parameter estimation. For example, in Table 9, the parameter of the variable Cropland used for pasture is obtained as 0.423 through the land utilization data of 1954. However, when we use the land utilization data of 1992, the parameter estimate for this variable is 2.82. The reason for this analysis may be that the sample size in the dataset is too small and that there are some anomalies or errors in the data.

6. Conclusions

In this paper, a spatial autoregressive model is used as the basis and a logistic regression model is combined with it to generate a spatial logistic model. Because of the potential endogeneity issue of the SAR model, we choose to apply the penalized maximum likelihood method to select significant covariates and simultaneously estimate the unknown parameters. Owing to the complexness of the penalized likelihood function, we have put forward a sort of more appropriate iterative algorithm to optimise the objective function. Both simulated experiments and the real case illustrate that our method performs well in limited samples. Additionally, we have contrasted not the same procedures of variable selection between SCAD and LASSO penalty. Comparison with LASSO, we find that SCAD is more valid and outstanding in nearly all instances in which our algorithm is used. It is unclear whether the submited method has similar results for other more nimble spatial models, incorporating parametric, nonparametric and semiparametric spatial regression models. What is more, it is possible to consider whether constraints could be introduced into this model. Nonetheless, prime framework and substantial foundation have been established and we will proceed to work on these subjects henceforth.

Author Contributions: Conceptualization, Y.S. (Yunquan Song) and J.L.; software, J.L.; resources, Y.S. (Yunquan Song) and J.L.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L., Y.C., Y.S. (Yuqi Su) and S.X.; project administration, Y.S. (Yunquan Song); funding acquisition, Y.S. (Yunquan Song). All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Fundamental Research Funds for the Central Universities (No. 20CX05012A), NSF project (ZR2019MA016) of Shandong Province of China, Statistical research project(KT028) of Shandong Province of China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the source code could be freely downloaded from https://github.com/Lmargery/SAR-Project (accessed on 20 June 2022).

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

Appendix A

In the Data Example, we use the Queen criterion to generate the spatial weight matrix. For the sake of sensitivity analysis and to avoid possible biases in the model results due to the choice of a single spatial weight matrix, we decided to estimate our model with a different spatial weight matrix, the minimum distance matrix (a distance-based spatial weight matrix). The results of the model are shown in Tables A1 and A2. It is obvious that the model can still select significant variables and verify the robustness of the model.

Table A1. Parameter estimates using penalty estimates via SCAD under a spatial logistic model.

Year	CLand_C	CLand_P	CLand_I	Grass_P	Land_G	Land_T	Land_U	Land_I
1954	0.0041	–	1.1816	−1.6001	–	0.1438	–	1.5291
1959	–	–	1.6730	−2.2342	–	–	–	1.6182
1964	–	–	1.4018	−1.7498	–	–	−0.7698	1.4987
1969	–	–	1.4567	−2.0923	–	–	0.0028	0.8858
1974	−0.1548	–	1.4706	−1.9466	−0.0278	–	–	1.4508
1978	−0.0434	–	1.3910	−1.4376	–	–	–	0.8428
1982	0.0248	−0.0056	0.7797	−0.0050	−0.0055	−0.0761	−0.0314	0.0037
1987	−0.0119	−0.0013	1.9232	−1.8952	0.0100	−0.0021	−0.7012	0.8860
1992	–	–	1.5005	−1.6278	–	–	−0.8751	1.3734
1997	–	–	1.0053	−1.7434	–	–	–	1.0357
2002	–	–	0.8515	−0.8234	–	–	−0.6124	1.4318
2007	–	–	1.6466	−1.3704	–	–	−1.2806	1.0864
2012	–	0.0010	0.8936	−0.8096	0.0047	–	−0.0026	0.8552

Table A2. Parameter estimates using penalty estimates via LASSO under a spatial logistic model.

Year	CLand_C	CLand_P	CLand_I	Grass_P	Land_G	Land_T	Land_U	Land_I
1954	−0.2383	–	0.5704	−0.9573	–	–	−0.0104	1.4060
1959	−0.3976	–	1.0474	−0.6412	−0.0475	–	–	0.8030
1964	–	0.0086	0.5267	−0.0309	−0.0976	–	−1.2473	0.9454
1969	–	−0.6377	1.2268	−0.7067	0.0019	–	–	0.9814
1974	−0.1366	−0.0079	0.3802	−0.0814	–	–	−0.7113	0.9834
1978	−0.0578	−0.1920	0.4416	−0.0011	−0.1774	−0.0066	−0.9952	1.1208
1982	−0.3635	−0.2077	0.6977	–	–	0.0027	−0.7201	0.9518
1987	−0.1601	−1.0231	0.5397	–	–	–	–	0.6990
1992	−0.0132	−0.5343	0.2966	−0.0582	−0.0428	–	−0.1322	1.5956
1997	−0.2571	−0.3083	0.5104	–	−0.0752	–	−0.6651	1.0377
2002	−0.0231	−0.8781	0.2262	–	–	−0.3388	−0.1083	1.0491
2007	−0.1737	–	0.4287	−0.6469	−0.4591	–	–	1.2079
2012	0.2314	–	0.4552	−0.3429	–	–	−0.0710	0.7424

References

1. Ord, J. Estimation methods for models of spatial interaction. *J. Am. Stat. Assoc.* **1975**, *70*, 120–126. [CrossRef]
2. Anselin, L. *Estimation Methods for Spatial Autoregressive Structures: A Study in Spat1al Econometrics*; Cornell University: New York, NY, USA, 1980.
3. Cliff, A.D.; Ord, J.K. *Spatial Processes: Models & Applications*; Taylor & Francis: Abingdon, UK, 1981.
4. Lee, L.F. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for SpatialAutoregressive Models. *Econometrica* **2004**, *72*, 1899–1925. [CrossRef]
5. Ma, Y.; Pan, R.; Zou, T. A Naive Least Squares Method for Spatial Autoregression with Covariates. *Stat. Sin.* **2020**, *30*, 653–672. [CrossRef]
6. Darmofal, D. *Spatial Analysis for the Social Sciences*; Cambridge University Press: New York, NY, USA, 2015.
7. Osland, L. An application of spatial econometrics in relation to hedonic house price modelling. *J. Real Estate Res.* **2010**, *32*, 289–320. [CrossRef]
8. Ahmar, A.S.; Aidid, M.K. Crime Modeling using Spatial Regression Approach. *J. Phys. Conf. Ser.* **2018**, *954*, 012013. [CrossRef]
9. Islamy, U.; Novianti, A.; Hidayat, F.P.; Kurniawan, M.H.S. Application of the Spatial Autoregressive (SAR) Method in Analyzing Poverty in Indonesia and the Self Organizing Map (SOM) Method in Grouping Provinces Based on Factors Affecting Poverty. *Enthusiastic Int. J. Appl. Stat. Data Sci.* **2021**, *1*. 76–83. [CrossRef]
10. Ver Hoef, J.M.; Peterson, E.E.; Hooten, M.B.; Hanks, E.M.; Fortin, M.J. Spatial autoregressive models for statistical inference from ecological data. *Ecol. Monogr.* **2018**, *88*, 36–59. [CrossRef]
11. Tibshirani, R.J. Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]
12. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]
13. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [CrossRef]
14. Han, X.; Hsieh, C.S.; Lee, L.F. Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach. *Reg. Sci. Urban Econ. March* **2017**, *63*, 97–120. [CrossRef]
15. 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.
16. Pfarrhofer, M.; Piribauer, P. Flexible shrinkage in high-dimensional Bayesian spatial autoregressive models. *Spat. Stat.* **2019**, *29*, 109–128. [CrossRef]
17. Song, Y.; Liang, X.; Zhu, Y.; Lin, L. Robust variable selection with exponential squared loss for the spatial autoregressive model. *Comput. Stat. Data Anal.* **2021**, *155*, 107094. [CrossRef]
18. Leach, J.M.; Aban, I.;Yi, N. Incorporating spatial structure into inclusion probabilities for Bayesian variable selection in generalized linear models with the spike-and-slab elastic net. *J. Stat. Plan. Inference* **2022**, *217*, 141–152. [CrossRef]
19. Gonella, R.; Bourel, M.; Bel, L. Facing spatial massive data in science and society: Variable selection for spatial models. *Spat. Stat.* 2022, *in press*. [CrossRef]
20. Yrigoyen, C.C. GeoDaSpace: A resource for teaching spatial regression models. *Rect@* **2013**, *4*, 119–144.