

Statistical Inference of Spatial Econometric Models Based on Likelihood Functions

Ziheng Zhang

School of Economics, Huazhong University of Science and Technology, Wuhan 050000, China
zhangziheng202207@163.com

Abstract. This paper is devoted to studying the probabilistic and statistical properties of maximum likelihood estimation of parameters in spatially mixed autoregressive models. When the response variable in the mixed autoregressive model obeys a continuous distribution, this paper verifies the monotonicity of the likelihood function of the mixed autoregressive model with respect to the autoregressive parameter P , which proves the existence and uniqueness of the maximum likelihood estimation of the autoregressive parameter. The results show that when the condition $n > \text{rank}(x+1)$ is satisfied, the quasi-likelihood function of the mixed autoregressive model has a unique maximum value with probability 1 in the parameter space; when $n > \text{rank}(x+1)$ and the regression coefficient matrix column is full rank, the maximum likelihood estimates of all parameters in the spatial mixed autoregressive model exist with probability 1 and are unique in the parameter space. In order to detect the strong influence points and abnormal points in the spatial mixed autoregressive model, this paper uses the first derivative method in the local influence analysis to obtain the statistics of the strong influence points and abnormal points in the spatial mixed autoregressive model in the form of variance perturbation. Simulation studies have shown that the first derivative of the maximum likelihood estimate of variance in the spatial mixed autoregressive model should be chosen to obtain the test statistic, which can effectively avoid the mearing and masking effects that often occur and are difficult to handle in local influence analysis. . As application and verification, this paper analyzes a real data to show that the conclusions obtained in this paper are reliable and practical.

Keywords: Maximum likelihood estimation; Existence; Econometrics.

1. Introduction

1.1 Background of topic selection

Spatial econometrics studies how cross-sectional data and spatial autocorrelation and spatial structure (spatial inhomogeneity) are dealt with in regression models for panel data. The basic idea is to add spatial effects of variables on the basis of traditional econometric models, and to perform a series of parameter estimation, model testing and model prediction for the models with spatial effects. With the continuous improvement of the theoretical basis of spatial econometrics, the application fields of spatial econometrics are becoming more and more extensive. On the one hand, it is reflected in the emergence of some specialized courses combined with spatial econometrics in the teaching field, such as regional economics, real estate economics, economic geography, spatial statistics, geographic information science, etc. Factors are introduced into traditional econometric analysis, for example, using spatial econometric models to analyze knowledge spillovers, economic growth convergence, urban-rural income gap, agricultural regional economic growth, industry innovation spatial correlation, provincial labor employment, and regional gaps in agricultural productivity.

However, at present, there are not many domestic reference books on spatial econometrics, the theoretical research on spatial models is insufficient, and domestic scholars have little understanding of the theoretical significance and application value of spatial econometric models, so the application of spatial econometric models is not widespread. This paper studies the existence and uniqueness of parameter maximum likelihood estimation in the spatial mixed autoregressive model (SARA) and the detection of strong influence points and abnormal points.

1.2 The meaning of the topic

The setting of spatial unit, the test of spatial correlation, the setting of spatial model structure and the estimation of spatial model are the main problems to be solved by spatial econometrics. Among them, parameter estimation can analyze and infer the essential laws of data response, and then infer the overall distribution or numerical characteristics, which is the most critical part of all problems.

With the wide application of econometrics, the statistical properties of the parameter estimators of econometric models have also attracted much attention. There are many parameter estimation methods for spatial econometric models. In the book "Spatial Econometrics: Methods and Models", it is first proposed to use the maximum likelihood estimation method to estimate the parameters of spatial econometric models. After that, some scholars put forward some new methods for parameter estimation of spatial econometric models, such as instrumental variable method, least squares method, two-stage least squares method, generalized distance method, etc. Anselin pointed out in 1988 that the least squares estimate of spatial econometric models are biased and inconsistent, but the statistical properties of maximum likelihood estimate of spatial econometric models have not been studied.

Statistical models are only approximate descriptions of economic phenomena, and the imprecision of statistical models themselves determines the importance of model sensitivity analysis. The sensitivity analysis of the model is to study the degree of influence of the small changes assumed by the model on the conclusions of statistical data analysis. At present, the sensitivity analysis method for the model is mainly the local influence analysis method, but there is no clear conclusion on the discussion of the Cook distance and the maximum curvature in the Cook method, and the Cook method often cannot effectively avoid the masking effect and the smearing effect.

In this paper, the existence and uniqueness of the maximum likelihood estimation of SAR model parameters are discussed first, and then the variance perturbation of the SAR model is carried out, and the local influence of the perturbed model is analyzed. The first derivative method is used instead of the Cook distance method. It is of great significance for the improvement of the theoretical system of spatial econometrics and the application of local influence analysis methods.

2. Econometric concepts

2.1 Spatial Econometrics

Spatial econometrics is a branch of econometrics, which is an econometric method that focuses on dealing with some special problems caused by the spatial characteristics of variables in econometric models. Theoretically, existing econometric models belong to the research category of spatial econometrics if they have spatial characteristics in the sense of economics or from the judgment of data characteristics. However, when there are spatial factors in the model structure setting, the traditional econometric estimation methods and testing methods either need to be improved or are no longer applicable. This requires the development of new model setting, estimation and testing methods; at the same time, how to introduce spatial characteristics into the model structure in an appropriate way is also a very important issue.

Traditional econometrics largely ignores spatial effects, whose existence violates the Gauss-Markov assumption of regression models. First, the Gauss-Markov assumption that explanatory variables are sampled with fixed repetition, and the existence of spatial correlation violates this assumption, which raises the need for estimable research methods. Secondly, in the case of a univariate linear relationship in the sampling data, due to the existence of spatial heterogeneity, if the spatial sampling data is changed, the relationship will also change, it is necessary to establish a model that satisfies this change and can draw appropriate conclusions. Model. Spatial econometrics is the study of how to incorporate spatial effects into traditional econometric models.

Since everything has a certain connection with the things around it, the closer the distance, the greater the correlation between things. If the spatial effect of the data is not considered, it will inevitably lead to the setting error of the model, and spatial econometrics will occur. for consideration

of this relationship. In 1974, Aelink put forward the concept of spatial measurement for the first time at the annual statistical conference in the Netherlands, and then scholars have gradually increased the discussion of spatial measurement.

2.2 Spatial effects

Spatial effects include spatial correlation and spatial heterogeneity. Spatial correlation is used to describe the spatial interdependence, interaction and mutual influence between economic phenomena. There are many factors that cause spatial correlation. It may be that the spatial location is similar, and the data tends to be consistent due to the influence of certain local policies or living habits; it is also possible that there are errors in the measurement of spatial data, and the errors make the data more consistent. tend to be consistent.

Spatial heterogeneity is a property used to describe the spatial differences that exist between economic phenomena. There are many factors that cause spatial heterogeneity. On the one hand, there are great differences in the shape and area of the spatial units; on the other hand, due to economic phenomena, the wooden body lacks a stable structure in space.

3. Maximum likelihood estimation of SAR model parameters

3.1 Overview of SAR Model

Spatial data has characteristics different from general data, namely spatial correlation of data, multi-scale in space and time, uncertainty of data expression, etc., which makes the measurement and knowledge discovery based on spatial data need spatial econometric theory model. Supported by , Anselin proposed a mixed autoregressive model, the model form is:

$$\begin{cases} Y_n = \rho W_n Y_n + X_n \beta + V_n \\ V_n \sim N_n(0, \sigma^2 I_n) \end{cases}$$

W_n is an $n \times n$ -dimensional spatial weight matrix that can be transformed into a row sum of 1 . Here W_n quantifies the associations between different regions. In this way, the i th element of Y_n can be written as:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + x_i^T \beta + v_i$$

3.2 Spatial Weight Matrix

The spatial weight matrix W_n is the embodiment of the spatial characteristics in the spatial econometric model. The construction of the spatial weight matrix will directly affect the parameter estimation of the spatial econometric model.

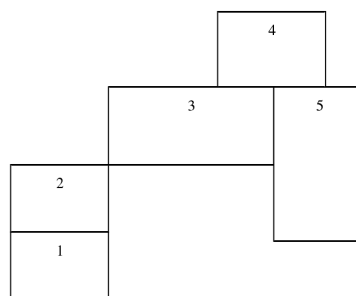


Figure 1. Spatial weight matrix diagram

As shown in Figure 1, area 1 and area 2, area 3 and area 4, area 3 and area 5, area 4 and area 5 have a common boundary, so $W_{12} = W_{34} = W_{35} = W_{45} = 1$, and other areas because There is no common boundary, so the elements of the corresponding weight matrix are 0.

The setting of the spatial weight matrix in this paper is obtained by using the spatial neighbor method, but in practical applications, the spatial neighbor matrix should be "standardized". The so-called standardization here is to make the sum of each row of the spatial weight matrix equal to 1. The action taken is to divide each row of the original spatial weight matrix by the sum of the corresponding row elements. This is why the original spatial weight matrix is symmetric, but the spatial weight matrix in practical applications is often asymmetric. The spatial neighbor matrix used in this paper is a symmetric matrix or a matrix that can be transformed into a symmetric form, and the related problems of a spatial matrix that is asymmetric but cannot be transformed into a symmetric form have not been solved.

3.3 Jordan decomposition of spatial weight matrix

It is assumed that the weight matrix W_n has m different eigenvalues, namely:

$$\begin{aligned} \tilde{\lambda}_1(W_n) &= \lambda_1(W_n) = \dots = \lambda_{n_1}(W_n) \\ &\dots \dots \\ \tilde{\lambda}_m(W_n) &= \lambda_{n-n_m+1}(W_n) = \dots = \lambda_n(W_n) \end{aligned}$$

where $n_1 + \dots + n_m = n$, noting that:

$$W_n = T_n J_n T_n^{-1}$$

Its form is as follows:

$$J_n = \begin{pmatrix} \tilde{J}_1 & & 0 \\ & \ddots & \\ 0 & & \tilde{J}_m \end{pmatrix}, \tilde{J}_j = \begin{pmatrix} \tilde{\lambda}_j(W_n) & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \tilde{\lambda}_j(W_n) \end{pmatrix}_{n_j \times n_j}, j = 1, \dots, m.$$

therefore

$$\begin{aligned} |I_n - \rho W_n| &= |I_n - \rho J_n| = \begin{vmatrix} I_{n_1} - \rho \tilde{J}_1 & & 0 \\ & \ddots & \\ & & I_{n_m} - \rho \tilde{J}_m \end{vmatrix} \\ &= \prod_{j=1}^m |I_{n_j} - \rho \tilde{J}_j| = \prod_{j=1}^m (1 - \rho \tilde{\lambda}_j(W_n))^{n_j} \end{aligned}$$

4. Local influence analysis based on the first derivative method

4.1 Overview of Local Impact Analysis

There are many forms of perturbation of the model. As far as the response variable n is concerned, a constant can be directly added to perturb the mean value, or it can be multiplied by a constant to perturb the variance. , semi-parametric perturbation. In terms of random error term $\{n$, mean, variance, parametric, nonparametric, semiparametric perturbations are also possible. The form of perturbation adopted in this paper is variance perturbation, that is, perturbation of the covariance matrix of the random error term changes from homoscedasticity to heteroscedasticity.

Masking and smearing effects often appear in local impact analysis. When there are many outliers, an outlier may not be detected due to the influence of other outliers, which is the so-called masking effect. Correspondingly, when there are many outliers, points that are not originally outliers may be mistaken for outliers and detected, which is the so-called smearing effect. Many scholars have studied methods to avoid these two effects. Marasinghe, Kianifard and Swallow proposed to use sequential test to detect K outliers, but the premise of this method is that the number of outliers K must be given in advance. Atkinson, Rousseuw and Leroy and Rousseuw and Zomeren proposed to use robust estimation methods to estimate autoregressive parameters to avoid masking effects. Hawkins tried to avoid the masking effect with a diagnostic method, however, that was based on repeated sampling. Gray and Ling proposed to use cluster analysis to identify outliers.

In this paper, the specific form of the test statistic is given on the basis of the variance estimate. The simulation study shows that the first derivative of the maximum likelihood estimate of variance in the spatial mixed autoregressive model should be selected to obtain the test statistic. It can effectively avoid the smearing and masking effects that often appear in local influence analysis and are difficult to handle.

4.2 Local influence analysis of SAR model

In this paper, the method of variance perturbation is used to analyze the local influence of the mixed autoregressive model. The model form is the same as the following formula:

$$\begin{cases} Y_n = \rho W_n Y_n + X_n \beta + V_n, \\ V_n \sim N_n(0, \sigma^2 I_n), \end{cases}$$

The perturbation method used in this paper is

$$V_n \sim N_n(0, \sigma^2 \Omega)$$

From the above equation, it is equivalent to

$$\begin{cases} \tilde{\beta}(\rho, \omega) = [X_n^T \Omega^{-1} X_n]^{-1} X_n^T \Omega^{-1} (I_n - \rho W_n) Y_n \\ \tilde{\sigma}^2(\rho, \omega) = \frac{1}{n} Y_n^T (I_n - \rho W_n)^T Z (Z^T \Omega Z)^{-1} Z^T (I_n - \rho W_n) Y_n \end{cases}$$

Therefore, solving the following equation gives a maximum likelihood estimate of the parameter ρ :

$$\left. \left\{ \frac{n}{2} \frac{\partial \tilde{\sigma}^2(\rho, \omega)}{\partial \rho} + \tilde{\sigma}^2(\rho, \omega) \text{tr}[W_n (I_n - \rho W_n)^{-1}] \right\} \right|_{\rho=\hat{\rho}(\omega)} = 0$$

Using the chain derivation rule to derive the vector t on both sides of the equation, we can get:

$$\left. \begin{aligned} & \left\{ \frac{n}{2} \left[\frac{\partial \hat{\rho}(1_n + td)}{\partial t} \frac{\partial \tilde{\sigma}^2(\rho, \omega)}{\partial \rho} + d^T \frac{\partial \tilde{\sigma}^2(\rho, \omega)}{\partial \omega} \right] \text{tr}[W_n (I_n - \rho W_n)^{-1}] \right\} \\ & + \left\{ \left[\frac{\partial \hat{\rho}(1_n + td)}{\partial t} \frac{\partial^2 \tilde{\sigma}^2(\rho, \omega)}{\partial \rho \partial \rho} + d^T \frac{\partial^2 \tilde{\sigma}^2(\rho, \omega)}{\partial \rho \partial \omega} \right] \right\} \\ & \left. + \left\{ \tilde{\sigma}^2(\rho, \omega) \text{tr}\{[W_n (I_n - \rho W_n)^{-1}]^2\} \frac{\partial \hat{\rho}(1_n + td)}{\partial t} \right\} \right\} \end{aligned}$$

4.3 Examples of local impact analysis

First, use Matlab to simulate a set of data, set the sample size $n=100$, $p=3$, the generation method of the spatial weight matrix splash adopts the generation method in Appendix B, and the elements of the design matrix X_n obey the exponential distribution.

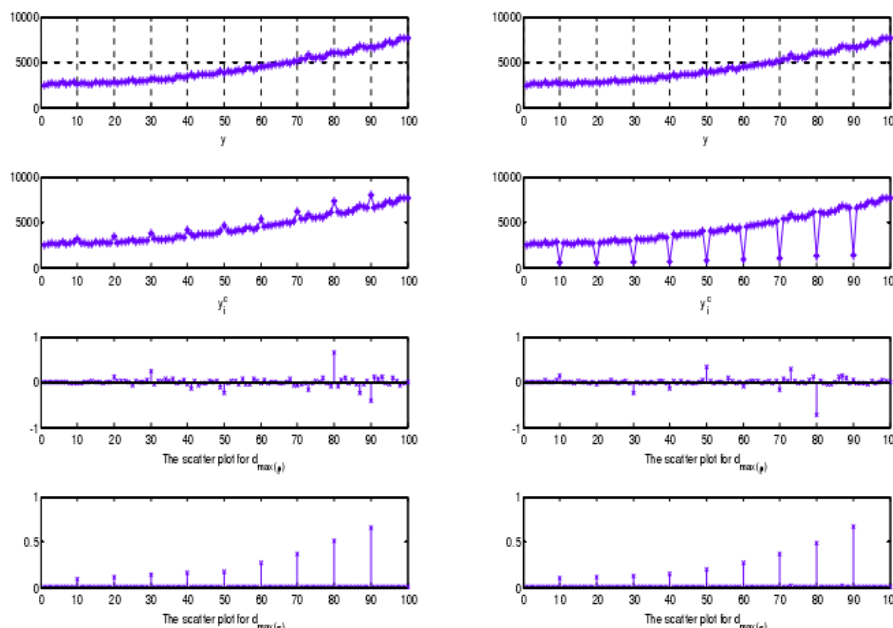


Figure 2. Scatter plot

And when $|a - 1| < 10^{-2}$, $d_{max(\sigma)}$ can no longer detect all strong influence points. The figure below demonstrates this.

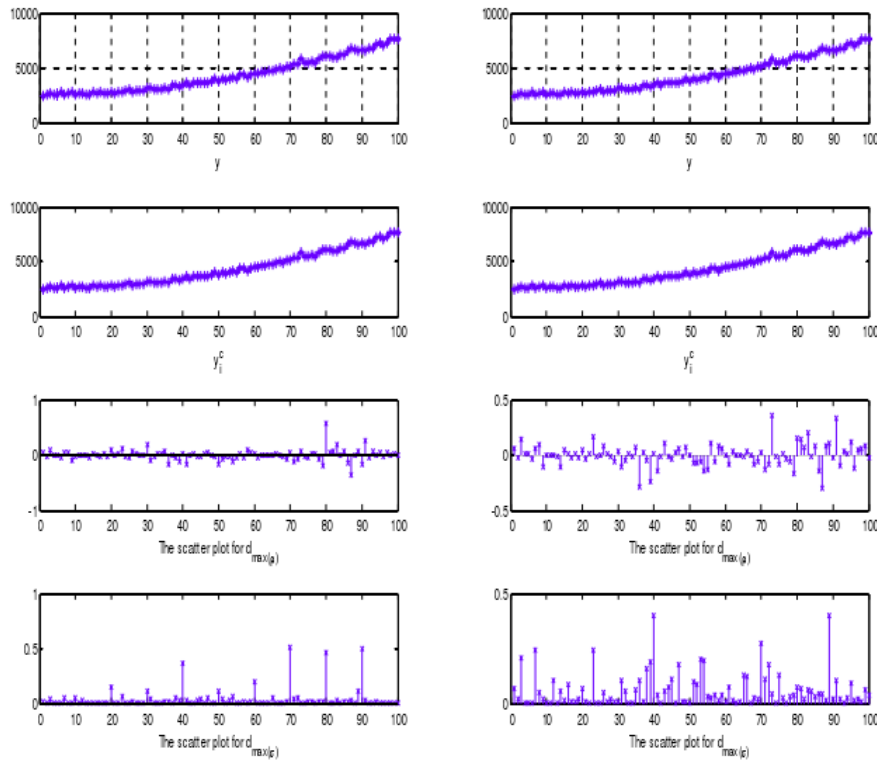


Figure 3. Adjusted scatterplot

5. Conclusion

This paper presents a mathematical proof of the existence and uniqueness of SAR model parameter estimation when the response variable is always distributed and the coefficient matrix Y_i is full rank. Mainly relying on the zero value theorem, the method is simple and easy to understand, and the calculation steps are simple. This paper presents a new detection statistic based on the variance estimate, which can effectively avoid masking and smearing effects that are often unavoidable in local influence analysis. Specific examples also demonstrate the effectiveness of the detection statistics proposed in this paper.

References

- [1] Statistical Diagnosis of Spatial Econometric Models [D]. Xiao Yunpeng. Yunnan University of Construction and Application of Shanghai Cooperation Organization International Link Econometric Model [D]. Chen Lei. Xinjiang University of Finance and Economics 2013.
- [2] Statistical Diagnosis of Spatial Econometric Models [D]. Xiao Yunpeng. Yunnan University of Finance and Economics 2013.
- [3] Statistical Inference of Spatial Econometric Models [D]. Li Mengyuan. Yunnan University of Finance and Economics 2013.
- [4] Research on Estimation of Ridge Regression Parameters in Econometric Models [D]. Zhang Bo. Tianjin University of Finance and Economics 2016.
- [5] Non-equilibrium analysis and control research of Wuhan real estate market [D]. Bao Junsheng. Central China Normal University 2012.
- [6] Research on regional economic differences based on spatial econometric model [D]. Li Jianbao. Northwest Normal University 2012.

- [7] Research on Customs Tax Prediction Based on Econometric Model [D]. Zou Wei. Tianjin University 2010.
- [8] Analysis of the impact of the international financial crisis on my country's economic growth and countermeasures [D]. Hou Na. Shandong University of Economics 2010.
- [9] The application of some economic models and their computer implementation [D]. Wang Jianbin. Lanzhou University 2009.
- [10] Spatial Statistical Analysis of Economy and Environment in Jiangsu Province [D]. Xu Ying. Nanjing University of Information Technology 2013.
- [11] Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins[J]. Cell Stem Cell. 2009 (5)
- [12] Robustness of Spatial Autocorrelation Specifications: Some Monte Carlo Evidence[J]. RobinDubin. Journal of Regional Science. 2003 (2).
- [13] Spurious Spatial Regression: Some Monte Carlo Results with a Spatial Unit Root and Spatial Cointegration[J]. BernardFingleton. Journal of Regional Science. 2002 (1).
- [14] Conformal normal curvature and assessment of local influence[J]. W.-Y.Poon,Y. S.Poon. Journal of the Royal Statistical Society: Series B (Statistical Methodology)2002 (1).
- [15] Rao's score test in spatial econometrics[J]. Luc Anselin. Journal of Statistical Planning and Inference. 2001 (1).
- [16] A Bayesian analysis of finite mixtures in the LISREL model[J]. Hong-Tu Zhu,Sik-Yum Lee. Psychometrika. 2001 (1).
- [17] A consistent nonparametric test for serial independence[J]. Joris Pinkse. Journal of Econometrics. 1998 (2).
- [18] Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs[J]. Danny T. Quah. Journal of Economic Growth. 1997 (1).
- [19] Spatial statistical analysis and geographic information systems[J]. Luc Anselin, Arthur Getis. The Annals of Regional Science. 1992 (1).
- [20] Assessment of Local Influence[J]. R. Dennis Cook. Journal of the Royal Statistical Society. Series. 1986 (2).