Survey on Spatial Econometric Model for Cross-Sectional Data

Xiaolong Wang [1, a]

[1]Rutgers Business School, Rutgers, The State University of New Jersey, Newark, NJ 07102, USA

[a]kevin20081013@hotmail.com

Keywords: Spatial Econometrics; Cross-sectional Data; Maximum Likelihood Estimation.

Abstract. Spatial econometrics is a branch discipline of econometrics, which mainly studies the traditional econometric models with spatial effect. Based on the development of spatial econometrics context, this paper gives a brief survey about the main spatial econometric model form, estimation and testing methods and the main applications of spatial econometric. And from the cross-sectional data point of view, this paper discusses a set of several classic spatial econometric models and maximum likelihood estimation method in detail.

1. Introduction

Generally speaking, the classical spatial econometrics is developed from the econometric model. If there is some kind of spatial interaction among the variables in the model, we can introduce the space effect to the econometric model in a proper form, changing the general econometric model to spatial econometrics model, so we can make econometric study of the spatial interaction between variables. From this point of view, by introducing the spatial effects, almost all present econometric models can be changed into spatial econometrics model. At present, the classical econometric model analysis framework has been introduced into the study of spatial econometrics, such as cross-sectional data spatial econometrics model, panel data spatial econometrics model, and other spatial model form such as constrained variable space model, spatial dynamic model, spatial nonlinear model, these corresponding estimation methods and testing theories are developing and improving now [1]. At first, most of the work in spatial econometrics was used to study regional economics and economic geography. However, Anselin indicated that in mainstream economics, there is more and more attention paid to the relationship of space and society, so from the perspective of theoretical research or application research, the spatial econometrics has received much attention [2].

This paper briefly introduces and discusses the cross-sectional data spatial econometrics model of which the theoretical framework is relatively mature, and the application is comparatively wide, and the estimation of the model and the related test are discussed emphatically. The spatial econometric model is developed from the general econometric model, as for linear regression model of general cross-sectional data, its expression is:

$$y = X\beta + e \tag{1}$$

In the above model, the spatial effect can be added by the spatial lag variables, spatial lag variables can be expressed in the form of dependent variable, explanatory variables and error terms. According to the different spatial mechanism, the models of spatial cross-sectional data mainly include three models, Spatial Autoregressive Model (SAR), Spatial Error Model (SEM) and Spatial Durbin Model (SDM). In this paper, these three models are discussed respectively.

2. Spatial Autoregressive Model

SAR is mainly used to study the situation in which the behavior of individuals in the adjacent space has influence on all the behavior of other individuals in the whole space, its expression is:

$$y = \rho W y + X\beta + \varepsilon \tag{2}$$

where W is $n \times n$ rank symmetric matrix, it reflects the interaction mechanism between individuals. In order to estimate conveniently, we generally standardize the weight matrix, making the sum of

each row of weight matrix 1. ρ is spatial autoregressive coefficient, the estimation of ρ is statistically significant indicates the existence of spatial correlation, positive symbol of ρ indicates the existence of positive spatial correlation, otherwise indicates the existence of negative spatial correlation. Wy is space lagged dependent variable, X is $n \times k$ explanatory variable vector, β is the corresponding coefficient vector, the error term is ε, and it assumes $\varepsilon \sim N(0, \sigma^2)$.

The traditional estimation of spatial autoregressive model will be biased, due to the space lagged dependent variable Wy in equation (2) is endogenous. Even if the error term ε is independent identically distributed, it is correlated with the space lagged dependent variable Wy. From equation (2), we can get

$$y = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} \varepsilon \tag{3}$$

Reversible matrix $(I - \rho W)^{-1}$ can be expressed as the sum of infinite terms, containing explanatory variable and error term for all position, also called Spatial Mutiplier, so space lagged term Wy must be treated as an endogenous variable, we need appropriate methods to estimate the model (2) and handle endogenous problem well.

From the view of the data generation process of the model, there is a Jacobi term in the transformation from error term ε to dependent variable y. Ord firstly gave the likelihood function of SAR model, in which ω is a n×1 vector containing the eigenvalues of space weight matrix W. If ω contains only real feature vectors, spatial correlation coefficient ρ is between the minimum eigenvalue and maximum eigenvalue of matrix W, so this ensures variance covariance matrix is positive definite. The likelihood function of SAR model is:

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e^T e}{2\sigma^2} \tag{4}$$

$$e = y - \rho Wy - X\beta \tag{5}$$

$$\rho \in [\min(\omega)^{-1}, \max(\omega)^{-1}] \tag{6}$$

3. Spatial Error Model

In Spatial Error Model, the relationship between institutions or regions is reflected by the error term. When the interactions between institutions or regions exist difference because of different relative position, we can use this model. The spatial form of the error term is generally autoregressive, so this model is also called Spatial Error Autoregressive Model, its specific setting is as follows:

$$y = X\beta + \varepsilon \tag{7}$$

$$\varepsilon = \rho W \varepsilon + v \tag{8}$$

$$y = X\beta + (I - \rho W)^{-1} v \tag{9}$$

Where ρ is spatial error autoregressive coefficient, $W\varepsilon$ is space delay error term, v is independent identically distributed error term. The model is also estimated using the maximum likelihood method, the likelihood function of the model is given by Anselin [3]:

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e^T e}{2\sigma^2} \tag{10}$$

$$e = (I_n - \rho W)(y - X\beta) \tag{11}$$

Careful observing the last term of equation (10) we can know that this likelihood function is equivalent to minimizing the sum of residual squares of the other model, that is dependent variable after spatial filtering $y^* = y - \rho Wy$ do regression on filtered explanatory variable $X^* = X - \rho WX$. Same as SAR model, by optimizing likelihood function we can get maximum likelihood estimator of parameter ρ, ρ^*. Then we can further get the maximum likelihood estimator of parameter β and σ^2:

$$\beta^* = [(X - \rho^* WX)^T (X - \rho^* WX)]^{-1} (X - \rho^* WX)^T (y - \rho^* Wy) \tag{12}$$

$$\hat{\sigma}^2 = (e - \rho^* We)^T (e - \rho^* We) n^{-1} \tag{13}$$

where $e = y - X\beta^*$

4. Spatial Durbin Model

4.1 The Setting of Spatial Durbin Model

SDM is first proposed by Anselin, the model's structure is similar to self-correlation of the time series model, so the equation (14) is called Spatial Durbin Model:

$$y = \alpha\tau_n + X\beta + \rho Wy + WX\gamma + \varepsilon \tag{14}$$

The corresponding data generating process of this model is

$$y = (I - \rho W)^{-1}(\alpha\tau_n + X\beta + \rho Wy + WX\gamma + \varepsilon) \tag{15}$$

We can calculate the endogenous variable's expectation:

$$E(y) = (I_n - \rho W)^{-1}(\alpha\tau_n + X\beta + WX\gamma) \tag{16}$$

where $\varepsilon \sim N(0, \sigma^2 I_n)$.

Spatial Durbin Model occupy a very interesting position in spatial regression analysis, because this model nests many other widely used spatial regression model, such as SAR，SEM and SLX et al.

4.2 The explanation of Spatial Durbin Model

An important reason for spatial Durbin model received more and more attention in application research is the explanatory power of this model is very strong. From the model (14) can be seen, the change of independent variable x in a certain area not only have influence on the dependent variable of this area (direct influence), but also have influence on other area by feedback effect (indirect influence), the model (14) can be reformulated as:

$$y = \sum_{q=1}^{Q} S_q(W)X_q + (I_n - \rho W)^{-1}\alpha\tau_n + (I_n - \rho W)^{-1}\varepsilon \tag{17}$$

$$S_q(W) = (I_n - \rho W)^{-1}(I_n\beta_q + W\gamma_q) \tag{18}$$

where q is [1,Q],Q is the number of explanatory variable, X_q represents the q th explanatory variable, vector β is the coefficient vector of non-constant explanatory variable, vector γ is the coefficient vector of space lag variable WX, the dimension of these two vectors is both （Q×1）. In order to observe the function of $S_q(W)$ clearly, consider a certain value of dependent variable, equation (17) can be shown as:

$$y_i = \sum_{q=1}^{Q}[S_q(W)_{i1}X_{1q} + S_q(W)_{i2}X_{2q} + ... + S_q(W)_{in}X_{nq}] + (I_n - \rho W)^{-1}\alpha\tau_n + (I_n - \rho W)^{-1}\varepsilon \tag{19}$$

Do partial derivation for the variable we can get:

$$\frac{\partial y_i}{\partial X_{jq}} = S_q(W)_{ij} \tag{20}$$

From equation (20), it can be seen that when the q th explanatory variable in area j changes, it not only has effect on the dependent variable y of region j, but also has effect on variable of other region. This is the indirect effects mentioned above. The direct effect can be also obtained by doing partial derivative:

$$\frac{\partial y_i}{\partial X_{iq}} = S_q(W)_{ii} \tag{21}$$

So matrix $S_q(W)$ not only reflects the direct effect (elements in diagonal line), but also reflects the indirect effect (other elements). Fischer et al. researched on the direct and indirect effects of human capital on regional labor productivity in the region of Europe by

using the model SDM [4], and thus it reveals the effect of human capital on labor productivity more profoundly.

5. Conclusions

Based on the development of spatial econometrics context, this paper gives a brief survey about the main spatial econometric model form, estimation and testing methods and the main applications of spatial econometric. And from the cross-sectional data point of view, this paper discusses a set of several classic space econometric models and maximum likelihood estimation method in detail, including SAR，SEM and SLX. This paper provides some foreshadowing and references to carry out follow-up studies and further research.

References

[1] LeSage J P. *An introduction to spatial econometrics*[J]. Revue d'économie industrielle, 2008 (123): 19-44.

[2] Anselin L. *Spatial econometrics*[J]. A companion to theoretical econometrics, 2001, 310330.

[3] Anselin L. *The future of spatial analysis in the social sciences*[J]. Geographic information sciences, 1999, 5(2): 67-76.

[4] Fischer M M, Bartkowska M, Riedl A, et al. *The impact of human capital on regional labor productivity in Europe*[J]. Letters in Spatial and Resource Sciences, 2009, 2(2-3): 97-108.