Theoretical foundations for spatial econometric research

Xingbai Xu a,∗, Lung-fei Lee b

a Wang Yanan Institute for Studies in Economics (WISE), Department of Statistics, School of Economics, MOE Key Lab of Econometrics and Fujian Key Lab of Statistics, Xiamen University, Xiamen, 361005, China

b Department of Economics, The Ohio State University, Columbus, OH, USA

Abstract

This paper reviews the development of large sample theories for spatial econometric models. These theories form important parts of statistical foundations for spatial econometrics. Another important component is the theoretical economics foundation for spatial econometric model specifications. We discuss how spatial econometric models can be regarded as the Nash equilibrium of some complete information games. Moran’s I test for spatial dependence is based on a statistic with a linear-quadratic form. Scores of the ML and moments for 2SLS and GMM are also in linear-quadratic form. A statistic with a linear-quadratic form can be characterized as a sum of martingale differences, so the central limit theorem for martingale difference arrays is crucial for asymptotic distributions of such statistics. For linear spatial models, statistics on linear-quadratic forms are the basis of spatial econometrics. For nonlinear spatial models, near-epoch dependent random fields play a crucial role. We summarize some important properties of near-epoch dependent random fields and illustrate how they are used in studying nonlinear spatial models such as spatial Tobit and spatial binary choice models.

1. Introduction

Spatial econometrics study estimation, tests, and inference of econometric models where spatial correlation exists among economic data across spatial units, which can be individuals, cities, counties, etc. Several books, e.g., Cliff and Ord (1973), Anselin (1988) and LeSage and Pace (2009), summarize some important developments in economic and statistical theories, computation, estimation, and empirical applications. Various estimation approaches have been developed for spatial econometric models. Kelejian and Prucha (1998) study the two-stage least squares (2SLS) estimation of spatially autoregressive (SAR) models with possibly spatially autoregressive disturbances. Kelejian and Prucha (1999) investigate the generalized method of moments (GMM) for a spatial error (SE) model. Lee (2004) establishes the consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) of SAR models. Lee (2007) discovers a type of quadratic moment conditions for SAR models to obtain a best GMM estimator. And spatial panel data models are also a field with many empirical applications, see, e.g., Lee and Yu (2010) and Qu et al. (2017).

In recent years, large sample theories for nonlinear spatial econometrics attract interest. Linear spatial models are not suitable for some empirical applications, e.g., discrete choice and limited dependent variables in spatial econometrics are two obvious cases. Linear models usually have a closed form expression in terms of exogenous regressors and disturbances. However, for nonlinear models, that will not be the case. As a result, large sample properties of estimators are expected to be more challenging to be investigated. In this regard, recent developments on weak laws of large numbers (WLLN) and central limit theorems (CLT) for nonlinear spatial econometrics are important. Jenish and Prucha (2009, 2012) provide some limiting laws for mixing and near-epoch dependent (NED) random fields. Based on the spatial NED theory, Xu and Lee (2015b, 2018) examine asymptotic theories of estimators for spatial Tobit models, Qu and Lee (2015) investigate a SAR model with endogenous spatial
weight matrix, and Xu and Lee (2016) study a binary choice SAR model.\(^1\)

This paper reports foundations and techniques on analyzing large sample properties of estimators for linear and nonlinear spatial econometrics. In Section 2, we discuss relevant economic foundations on games to derive linear SAR models, and review large sample theories for linear spatial econometric models. Model stability would require uniform boundedness of spatial weight matrices in norms. Statistics on linear-quadratic form characterize various estimation methods for linear SAR models. Martingale CLT provides an essential tool for statistical inference. In Section 3, we extend the game foundations of linear SAR models to nonlinear ones. For nonlinear spatial econometric models, we discuss some spatial weak dependence concepts and their properties, especially NED random fields. We point out some other tools relevant for analyzing nonlinear spatial models. In Section 4, we illustrate the use of NED random fields to investigate spatial Tobit and binary choice models and their estimation problems. Conclusions are drawn in Section 5.

2. Theoretical foundations for linear spatial econometric models

2.1. Some popular linear spatial econometric models

Among spatial econometrics models, linear ones are the most useful ones, where linearity means that dependent variables are affine functions of disturbances in a model. There are several popular linear spatial econometric models. A SAR model describes the interaction of endogeneous variables of spatial units:

\[ y_{in} = \lambda_0 \sum_{j=1}^{n} w_{ij} y_{jn} + x_{in}' \beta_0 + \epsilon_{in}, \tag{1} \]

where \( y_{in} \) is the dependent variable, \( x_{in} \) is an exogenous variable (column) vector, \( W_n \equiv (w_{ij}) \) is a specified spatial weight matrix, \( \epsilon_{in}'s \) are disturbance terms that are usually assumed to be independently or even independently and identically distributed (iid), and \( \lambda_0 \) and \( \beta_0 \) are true coefficients. We can write Eq. (1) in a matrix form:

\[ Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \epsilon_n, \tag{2} \]

where \( Y_n \equiv (y_{1n}, \ldots, y_{nn})' \) and \( X_n \equiv (x_{1n}, \ldots, x_{nn})' \).

A SE model assumes that spatial interactions take place among disturbance terms:

\[ Y_n = X_n \beta_0 + \epsilon_n, \quad \epsilon_n = \lambda_0 W_n a_n + u_n, \tag{3} \]

where \( u_n \)'s are set to be independent or iid with zero mean and finite variances. A SAR model with SAR disturbances (SARAR) model combines a SAR model and a SE model:

\[ Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \epsilon_n, \quad \epsilon_n = \lambda_0 M_n \epsilon_n + u_n, \tag{4} \]

where the two spatial weight matrices \( W_n \) and \( M_n \) can be the same or different.

2.2. Game foundations for SAR models

A SAR model can be regarded as a model on the Nash equilibrium of a static complete information game with linear-quadratic utilities. Suppose that there are \( n \) individuals, and they choose their actions (e.g., efforts) to maximize their utilities. Let the action for individual \( i \) be \( y_{in} \), and its cost equal \( \frac{1}{2} \epsilon_{in}' \epsilon_{in} \). Suppose individual \( i \)'s benefit from his action is proportional to his action, and it depends on his characteristics and others’ actions: \( y_{in} \left( \lambda_0 \sum_{j=1}^{n} w_{ij} y_{jn} + x_{in}' \beta_0 + \epsilon_{in} \right) \), which can be substitute or complementary depending on the sign of \( \lambda_0 \). Then his utility is

\[ u_i(y_{in}) = y_{in} \left( \lambda_0 \sum_{j=1}^{n} w_{ij} y_{jn} + x_{in}' \beta_0 + \epsilon_{in} \right) - \frac{1}{2} \epsilon_{in}' \epsilon_{in}. \tag{5} \]

where \( \epsilon_{i,n} \) and \( \epsilon_{in}'s \) are known to all individuals. The optimal action for \( i \) will be characterized by Eq. (1). When Eq. (1) has a solution, the solution is a Nash equilibrium of this game. Depending on applications of this model, there are possibly other theoretical justifications. For social interactions, one may have a private and social utility:

\[ u_i(y_{in}) = y_{in} \left( x_{in}' \beta_0 + \epsilon_{in} \right) - \frac{1}{2} (y_{in} - \lambda_0 \sum_{j=1}^{n} w_{ij} y_{jn})^2, \]

where the first component represents private utility associated with an action \( y_{in} \) and the second component captures a conformity effect with friends (see, Brock and Durlauf, 2001).

2.3. Linear spatial models and stability conditions

We will use a SAR model as an example to discuss some widely adopted structures in spatial econometrics. As one would like to exclude self-influence, the diagonal elements of \( W_n \) are specified to be zero. For the model to be stable in the sense that \( \lambda_0 \)'s do not have unbounded variances as \( n \) becomes large, we need to add some restrictions on \( W_n \).

As in Kelejian and Prucha (1998, 1999), \( W_n \) is often set to be uniformly bounded in both the row sum norm (\( \infty \)-norm) and column sum norm (1-norm), i.e.,

\[ \sup_n \| W_n \|_{\infty} \equiv \sup_n \left\| \sum_{i=1}^{n} |w_{ij}| \right\| < \infty \quad \text{and} \quad \sup_n \| W_n \|_1 \equiv \sup_n \left\| \sum_{j=1}^{n} |w_{ij}| \right\| < \infty. \]

Sometimes, \( W_n \) is assumed to be row-normalized: \( \sum_{i=1}^{n} |w_{ij}| = 1 \) for every nonzero row of \( W_n \). In this case, \( \| W_n \|_{\infty} \equiv 1 \). One may interpret the influence on \( y_{in} \) under interactions is the average of neighboring units’ activities. In this case, the uniform boundedness is solely imposed on the column sums of \( W_n \). In general, uniform boundedness is imposed on both rows and columns of \( W_n \). The uniform boundedness in a norm for the sequence of \( W_n \), however, is not sufficient for a SAR process to be stable. An example is a unit root process \( y_{it} = y_{i,t-1} + \epsilon_i, \quad t = 2, \ldots, n \), with \( y_1 = \epsilon_1 \). The implied \( W_n \) is uniformly bounded in both row and column sum norms, but \( y_n = \sum_{t=1}^{n} \epsilon_t \) is not a stable process. To rule out this situation, Kelejian and Prucha (1998, 1999) assume the additional condition that \( S_n(\lambda_0) \equiv (I_n - \lambda_0 W_n)^{-1} \) is uniform bounded in both row and column sum norms. A sufficient condition for model stability is that \( \| \lambda_0 W_n \|_1 < 1 \) for some norm \( \| \| \). With the additional condition, the Neumann’s expansion,

\[ S_n(\lambda_0)^{-1} \equiv (I_n - \lambda_0 W_n)^{-1} = \sum_{l=0}^{\infty} \lambda_0^l W_n^l \]

Then, Eq. (2) implies

\[ Y_n = (X_n \beta_0 + \epsilon_n) + \lambda_0 W_n (X_n \beta_0 + \epsilon_n) + \lambda_0^2 W_n^2 (X_n \beta_0 + \epsilon_n) + \cdots \tag{6} \]

The second term on the right hand side (RHS) of Eq. (6) can be regarded as the impact from the first-order contiguous neighbors, and the third term is the impact from the second-order contiguous neighbors (neighbors’ neighbors), etc. \( \| \lambda_0 W_n \|_1 < 1 \) implies that the impact of the \( m \)-th order contiguous neighbors decreases exponentially as \( m \) increases. This model is thus stable. With uniform boundedness \( \| W_n \|_{\infty} = 1 \), the additional condition \( \| \lambda_0 W_n \|_1 < 1 \) gives rise to \( |\lambda_0| < 1 \). This setting is similar to that for a stationary AR(1) process for which we restrict the value of \( \lambda_0 \). As an extension, for a more general model \( Y_n = W_n(\lambda_0) Y_n + X_n \beta_0 + \epsilon_n \), where unknown parameters are involved in spatial weights, a stable condition for the system is to assume \( \| W_n(\lambda_0) \|_{\infty} < 1 \) for some norm \( \| \| \) for all \( n \).

\(^1\) Beyond classical statistical approaches, LeSage and Pace (2009) and Greene (2011, 2013) discuss the Bayesian MCMC estimation approach for nonlinear spatial econometrics.
دریافت فوری
متن کامل مقاله

| امکان دانلود نسخه تمام متن مقالات انگلیسی | ✓ |
| امکان دانلود نسخه ترجمه شده مقالات | ✓ |
| پذیرش سفارش ترجمه تخصصی | ✓ |
| امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله | ✓ |
| امکان دانلود رایگان ۲ صفحه اول هر مقاله | ✓ |
| امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب | ✓ |
| دانلود فوری مقاله پس از پرداخت آنلاین | ✓ |
| پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات | ✓ |