A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances[1]

Harry H. Kelejian and Ingmar R. Prucha[2]

April 1997
Revised August 1997

[1]We would like to thank two anonymous referees for helpful comments. We assume, however, full responsibility for any shortcomings.
[2]Department of Economics, University of Maryland, College Park, MD 20742

Abstract

Cross sectional spatial models frequently contain a spatial lag of the dependent variable as a regressor, or a disturbance term which is spatially autoregressive. In this paper we describe a computationally simple procedure for estimating cross sectional models which contain both of these characteristics. We also give formal large sample results.

1 Introduction

Cross sectional spatial regression models are often formulated such that they permit interdependence between spatial units. This interdependence complicates the estimation of such models. One form of interdependence arises when the value of the dependent variable corresponding to each cross sectional unit is assumed, in part, to depend upon a weighted average of that dependent variable corresponding to neighboring cross sectional units. This weighted average is often described in the literature as a spatial lag of the dependent variable, and the model is then referred to as a spatially autoregressive model – see, e.g., Blommestein (1983) and Anselin (1988, p. 35).[1] The spatially lagged dependent variable is typically correlated with the disturbance term – see, e.g., Ord(1975) and Anselin(1988, p. 58) – and hence the ordinary least squares estimator is typically not consistent in such situations. Another form of interdependence that arises in such models is that the disturbance term is often assumed to be spatially autoregressive. Consistent procedures, other than maximum likelihood, have been suggested in the literature for models which contain one of these interdependencies.[2] Unfortunately, such procedures are not available for models which have both of these characteristics. This shortcoming is of consequence because maximum likelihood procedures are often "computationally very challenging" when the sample size is large.[3]

[1] As an example, in a spatial model explaining property values, the property value at each location could relate to, among other things, the property values of neighboring locations. For empirical studies in which spatial lags of the dependent variable are considered see, e.g., Case (1991,1992), Case, Hines and Rosen (1993), and Kelejian and Robinson (1993).

[2] An early procedure which is partially based on maximum likelihood principles and which relates to models which have a spatially autoregressive disturbance term was suggested by Ord(1975); a more recent procedure for such models which is partially based on a generalized moments approach was suggested by Kelejian and Prucha (1995). An instrumental variable estimator for models which contain a spatially lagged dependent variable is described in Anselin(1982). See also Anselin (1990), and Anselin, Bera, Florax, and Yoon (1996) for a wide variety of tests relating to models which contain either a spatially autoregressive error term, a spatially lagged dependent variable, or both.

[3] These "computationally challenging issues" can be moderated by using Ord's (1975) eigenvalue approach to the evaluation of the likelihood function. Further simplifications can be realized by the use of sparse matrix routines if the weighting matrix involved is indeed sparse – see, e.g., Pace and Barry (1996). Our experience is that the computation of eigenvalues for general nonsymmetric matrices by standard subroutines in the IMSL program library may be inaccurate for matrices as small as 400×400. The accuracy

Furthermore, the maximum likelihood procedure requires distributional assumptions which the researcher may not wish to specify.[4]

The purpose of this paper is to suggest an estimation procedure for cross sectional spatial models which contain a spatially lagged dependent variable as well as a spatially autocorrelated error term. Our procedure is computationally simple, even in large samples. In addition, our procedure is conceptually simple in that its rational is obvious. We give formal large sample results with modest assumptions regarding the distribution of the disturbances.

The model is specified in Section 2. That section also contains a discussion of the assumptions involved. Our procedure is described in Section 3. Concluding remarks are given in Section 4. Technical details are relegated to the appendix.

2 The Model

In this section we first specify the regression model and all of its assumptions; we then provide a discussion and interpretation of these assumptions. It proves helpful to introduce the following notation: Let A_n with $n \in \mathbf{N}$ be some matrix; we then denote the (i, j)-th element of A_n as $a_{ij,n}$. Similarly, if v_n with $n \in \mathbf{N}$ is a vector, then $v_{i,n}$ denotes the i-th element of v_n. An analogous convention is adopted for matrices and vectors that do not depend on the index n, in which case the index n is suppressed on the elements. If A_n is a square matrix, then A_n^{-1} denotes the inverse of A_n. If A_n is singular, then A_n^{-1} should be interpreted as the generalized inverse of A_n. Further, let $(B_n)_{n\in\mathbf{N}}$ be some sequence of $n \times n$ matrices. Then we say the row and column sums of the (sequence of) matrices B_n are bounded uniformly in absolute value if there exists a constant $c_B < \infty$ (that does not dependent of

improves if the matrix involved is symmetric and that information is used. Bell and Bockstael (1997) report accuracy problems in determining eigenvalues for matrices of, roughly, order 2000×2000, even though sparse matrix routines in MATLAB were used. On the other hand, Pace and Barry (1996) were able to work with matrices of, approximately, order $20,000 \times 20,000$.

[4] Given appropriate conditions the maximum likelihood estimator should be consistent and asymptotically normally distributed. However, to the best of our knowledge, formal results establishing these properties for spatial models of the sort considered here under a specific set of low level assumptions do not seem to be available in the literature; cp. Kelejian and Prucha (1995) on this point.

n) such that

$$\max_{1\leq i\leq n}\sum_{j=1}^{n}|b_{ij,n}|\leq c_B \text{ and } \max_{1\leq j\leq n}\sum_{i=1}^{n}|b_{ij,n}|\leq c_B \text{ for all } n\in\mathbf{N}$$

holds. As a point of interest, we note that the above condition is identical to the condition that the sequences of the maximum column sum matrix norms and maximum row sum matrix norms of B_n are bounded; cp. Horn and Johnson (1985, pp.294-5).

2.1 Model Specification

Consider the following cross sectional (first order) autoregressive spatial model with (first order) autoregressive disturbances ($n\in\mathbf{N}$):

$$\begin{aligned} y_n &= X_n\beta+\lambda W_n y_n+u_n, && |\lambda|<1 \\ u_n &= \rho M_n u_n+\varepsilon_n, && |\rho|<1 \end{aligned} \tag{1}$$

where y_n is the $n\times 1$ vector of observations on the dependent variable, X_n is the $n\times k$ matrix of observations on k exogenous variables, W_n and M_n are $n\times n$ spatial weighting matrices of known constants, β is the $k\times 1$ vector of regression parameters, λ and ρ are scalar autoregressive parameters, u_n is the $n\times 1$ vector of regression disturbances, and ε_n is an $n\times 1$ vector of innovations. The variables $W_n y_n$ and $M_n u_n$ are typically referred to as spatial lags of y_n and u_n, respectively. For reasons of generality we permit the elements of X_n, W_n, M_n and ε_n to depend on n, i.e., to form triangular arrays. We condition our analysis on the realized values of the exogenous variables and so, henceforth, the matrices X_n will be viewed as a matrices of constants.

In scalar notation the spatial model (1) can be rewritten as

$$\begin{aligned} y_{i,n} &= \sum_{j=1}^{k}x_{ij,n}\beta_j+\lambda\sum_{j=1}^{n}w_{ij,n}y_{j,n}+u_{i,n}, && i=1,...,n, \\ u_{i,n} &= \rho\sum_{j=1}^{n}m_{ij,n}u_{j,n}+\varepsilon_{i,n}. \end{aligned} \tag{2}$$

The spatial weights $w_{ij,n}$ and $m_{ij,n}$ will typically be specified to be nonzero if cross sectional unit j relates to i in a meaningful way. In such cases, units

i and j are said to be neighbors. Usually neighboring units are taken to be those units which are close in some dimension – e.g., geographic, technological, etc. We allow for the possibility that $W_n = M_n$.

We maintain the following assumptions concerning the spatial model (1).

Assumption 1 *All diagonal elements of the spatial weighting matrices W_n and M_n are zero.*

Assumption 2 *The matrices $(I - \lambda W_n)$ and $(I - \rho M_n)$ are nonsingular for all $|\lambda| < 1$ and $|\rho| < 1$.*

Assumption 3 *The row and column sums of the matrices W_n, M_n, $(I - \lambda W_n)^{-1}$, and $(I - \rho M_n)^{-1}$ are bounded uniformly in absolute value.*

Assumption 4 *The regressor matrices X_n have full column rank (for n large enough). Furthermore, the elements of the matrices X_n are uniformly bounded in absolute value.*

Assumption 5 *The innovations $\{\varepsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ are distributed identically. Further, the innovations $\{\varepsilon_{i,n} : 1 \leq i \leq n\}$ are for each n distributed (jointly) independently with $E(\varepsilon_{i,n}) = 0$, $E(\varepsilon_{i,n}^2) = \sigma_\varepsilon^2$, where $0 < \sigma_\varepsilon^2 < b$ with $b < \infty$. Additionally the innovations are assumed to possess finite fourth moments.*

In estimating the spatial model (1) we will utilize a set of instruments. Let H_n denote the $n \times p$ matrix of those instruments, and let $Z_n = (X_n, W_n y_n)$ denote the matrix of regressors in the first equation of (1). We maintain the following assumptions concerning the instrument matrices H_n.

Assumption 6 [5] *The instrument matrices H_n have full column rank $p \geq k+1$ (for all n large enough). They are composed of a subset of the linearly independent columns of $(X_n, W_n X_n, W_n^2 X_n, \ldots, M_n X_n, M_n W_n X_n, M_n W_n^2 X_n, \ldots)$, where the subset contains at least the linearly independent columns of $(X_n, M_n X_n)$.*

[5] In principle we could have different instrument matrices for the first and third step of the estimation procedure discussed below, but this would further complicate our notation without expanding the results in an essential way.

Assumption 7 *The instruments H_n satisfy furthermore:*

(a)

$$Q_{HH} = \lim_{n\to\infty} n^{-1} H_n' H_n$$

where Q_{HH} is finite, and nonsingular.

(b)

$$Q_{HZ} = \operatorname*{plim}_{n\to\infty} n^{-1} H_n' Z_n$$

and

$$Q_{HMZ} = \operatorname*{plim}_{n\to\infty} n^{-1} H_n' M_n Z_n$$

where Q_{HZ} and Q_{HMZ} are finite, and have full column rank. Furthermore

$$Q_{HZ} - \rho Q_{HMZ} = \operatorname*{plim}_{n\to\infty} n^{-1} H_n'(I - \rho M_n) Z_n$$

has full column rank for all $|\rho| < 1$.

(c)

$$\Phi = \lim_{n\to\infty} n^{-1} H_n'(I - \rho M_n)^{-1}(I - \rho M_n')^{-1} H_n$$

is finite, and nonsingular for all $|\rho| < 1$.

The following assumption ensures that the autoregressive parameter ρ is "identifiably unique", cp. Kelejian and Prucha (1995).

Assumption 8 *The smallest eigenvalue of $\Gamma_n'\Gamma_n$ is bounded away from zero, i.e., $\lambda_{\min}(\Gamma_n'\Gamma_n) \geq \lambda_* > 0$, where*

$$\Gamma_n = \frac{1}{n}\begin{pmatrix} 2E(u_n'\overline{u}_n) & -E(\overline{u}_n'\overline{u}_n) & 1 \\ 2E(\overline{\overline{u}}_n'\overline{u}_n) & -E(\overline{\overline{u}}_n'\overline{\overline{u}}_n) & tr(M_n'M_n) \\ E(u_n'\overline{\overline{u}}_n + \overline{u}_n'\overline{u}_n) & -E(\overline{u}_n'\overline{\overline{u}}_n) & 0 \end{pmatrix} \tag{3}$$

and $\overline{u}_n = M_n u_n$ and $\overline{\overline{u}}_n = M_n \overline{u}_n = M_n^2 u_n$.

2.2 Some Implications of the Model Specification

The specifications in (1) and Assumption 2 imply that[6]

$$\begin{aligned} y_n &= (I - \lambda W_n)^{-1} X_n \beta + (I - \lambda W_n)^{-1} u_n \\ u_n &= (I - \rho M_n)^{-1} \varepsilon_n. \end{aligned} \tag{4}$$

[6]We note that, in general, the elements of $(I - \lambda W_n)^{-1}$ and $(I - \rho M_n)^{-1}$ will depend on the sample size n, even if the elements of W_n and M_n do not depend on n. Consequently, in general, the elements of y_n and u_n will also depend on n, and thus form a triangular array, even in the case where the innovations $\varepsilon_{i,n}$ do not depend on n.

Assumption 5 implies further that $E(u_n) = 0$, and that the variance-covariance matrix of u_n is

$$\Omega_{u_n} = E(u_n u_n') = \sigma_\varepsilon^2 (I - \rho M_n)^{-1} (I - \rho M_n')^{-1}. \tag{5}$$

Thus the disturbance terms are generally both spatially correlated and heteroskedastic. It follows from (4) and (5) that $E(y_n) = (I - \lambda W_n)^{-1} X_n \beta$, and that the variance-covariance matrix of y_n is

$$\Omega_{y_n} = \sigma_\varepsilon^2 (I - \lambda W_n)^{-1} (I - \rho M_n)^{-1} (I - \rho M_n')^{-1} (I - \lambda W_n')^{-1}. \tag{6}$$

Furthermore

$$\begin{aligned} E[(W_n y_n) u_n'] &= W_n (I - \lambda W_n)^{-1} \Omega_{u_n} \qquad (7) \\ &= \sigma_\varepsilon^2 W_n (I - \lambda W_n)^{-1} (I - \rho M_n)^{-1} (I - \rho M_n')^{-1} \\ &\neq 0. \end{aligned}$$

Thus, in general, the elements of the spatially lagged dependent vector, $W_n y_n$, are correlated with those of the disturbance vector. One implication of this is, of course, that the parameters of (1) can not be consistently estimated by ordinary least squares.

2.3 Further Interpretations of the Model Specification

Assumption 1 is a normalization of the model; it also implies that no unit is viewed as its own neighbor. Assumption 2 indicates that the model is complete in that it determines y_n and u_n. Next consider Assumption 3. In practice, weighting matrices are often specified to be row normalized in that $\sum_{j=1}^n w_{ij,n} = \sum_{j=1}^n m_{ij,n} = 1$ – see, e.g., Kelejian and Robinson (1993), and Anselin and Rey (1990). In many of these cases no unit is assumed to be a neighbor to more than a given number, say q, of other units – i.e., for every j the number of $m_{ij,n} \neq 0$ is less than or equal to q. Clearly in such cases Assumption 3 is satisfied for W_n and M_n. Also, often the weights are formulated such that they decline as a function of some measure of "distance" between neighbors. Again, in such cases Assumption 3 will typically be satisfied for W_n and M_n. Assumption 3 also maintains that the row and column sums of $(I - \rho M_n)^{-1}$ and $(I - \lambda W_n)^{-1}$ are uniformly bounded in absolute value. In light of (5) and (6) this assumption is reasonable in that it implies that the row and column sums of the covariance matrices Ω_{u_n} and

Ω_{y_n} are uniformly bounded in absolute value, thus limiting the degree of correlation between, respectively, the elements of u_n and y_n.[7] Our results relate to the large sample; the extent of correlation is limited in virtually all large sample analysis – see, e.g., Amemiya (1985, ch. 3,4) and Pötscher and Prucha (1997, ch. 5,6). Assumptions 4 and 5 regarding the regressor matrices X_n and the innovations ε_n seem in line with typical specifications – see, e.g., Schmidt (1976, p.2, 56).

The instrument matrices H_n will be used to instrument $Z_n = (X_n, W_n y_n)$ and $M_n Z_n = (M_n X_n, M_n W_n y_n)$ in terms of their predicted values from a least squares regression on H_n, i.e., $\widehat{Z}_n = P_{H_n} Z_n$ and $\widehat{M_n Z_n} = P_{H_n} M_n Z_n$ with $P_{H_n} = H_n (H_n' H_n)^{-1} H_n'$. The ideal instruments are $E(Z_n) = (X_n, W_n E(y_n))$ and $E(M_n Z_n) = (M_n X_n, M_n W_n E(y_n))$ where $E(y_n) = (I - \lambda W_n)^{-1} X_n \beta$. In principle we would like $\widehat{Z}_n$ and $\widehat{M_n Z_n}$ to approximate $E(Z_n)$ and $E(M_n Z_n)$ as closely as possible. Assumption 6 assumes that H_n contains, at least, the linearly independent columns of X_n and $M_n X_n$, which ensures that $\widehat{Z}_n = (X_n, \widehat{W_n y_n})$ and $\widehat{M_n Z_n} = (M_n X_n, M_n \widehat{W_n y_n})$with $\widehat{W_n y_n} = P_{H_n} W_n y_n$ and $M_n \widehat{W_n y_n} = P_{H_n} M_n W_n y_n$. Furthermore, suppose all eigenvalues of W_n are less than or equal to one in absolute value – which is, e.g., the case if W_n is row normalized. Then, observing that $|\lambda| < 1$, it is readily seen that[8]

$$\begin{aligned} E(y_n) &= (I - \lambda W_n)^{-1} X_n \beta \\ &= \left[\sum_{i=0}^{\infty} \lambda^i W_n^i \right] X_n \beta, \qquad W_n^0 = I. \end{aligned} \tag{8}$$

Consequently, in this case, $W_n E(y_n)$ and $M_n W_n E(y_n)$ are seen to be formed as a linear combination of the columns of the matrices X_n, $W_n X_n$, $W_n^2 X_n$, $\ldots$, $M_n X_n$, $M_n W_n X_n$, $M_n W_n^2 X_n$, $\ldots$ It is for this reason that we postulate in Assumption 6 that H_n is composed of a subset of the linearly independent columns of those matrices. In practice that subset might be the linearly

[7] This follows from the following fact: Let A_n and B_n be matrices which are conformable for multiplication and whose row and column sums are uniformly bounded in absolute value. Then the row and column sums of $A_n B_n$ are also uniformly bounded in absolute value - see, e.g., Kelejian and Prucha (1995).

[8] If all eigenvalues of W_n are less than or equal to one in absolute value, then $|\lambda| < 1$ implies that all eigenvalues of λW_n are less than one in absolute value. This in turn ensures that $(I - \lambda W_n)^{-1} = \sum_{i=0}^{\infty} \lambda^i W_n^i$ – see, e.g., Horn and Johnson (1985, pp. 296-301). The claim that all eigenvalues of W_n are less than or equal to one in absolute value, given W_n is row normalized, follows from Geršgorin's theorem – see, e.g., Horn and Johnson (1985, p. 344).

independent columns of $[X_n,\ W_nX_n,\ W_n^2X_n,\ M_nX_n,\ M_nW_nX_n,\ M_nW_n^2X_n]$, or if the number of regressors is large, just those of $[X_n,\ W_nX_n,\ M_nX_n,\ M_nW_nX_n]$.[9] We also note that the assumption that the matrices H_n have full column rank could be relaxed at the expense of working with generalized inverses, since the orthogonal projection of any vector onto the space spanned by the columns of H_n is unique even if H_n does not have full column rank. Finally, for future reference we note that the elements of H_n are in light of Assumptions 3 and 4 bounded in absolute value.

Consider now Assumption 7. This assumption will ensure that the estimators defined below remain well defined asymptotically. Assumption 7(a) is standard. Assumption 6 and Assumption 7(a) imply that $n^{-1}H_n'X_n$ converges to a full column rank matrix. Because of this and since $n^{-1}H_n'Z_n = (n^{-1}H_n'X_n, n^{-1}H_n'W_ny_n)$ the force of the first part of Assumption 7(b) relates to the probability limit of $n^{-1}H_n'W_ny_n$ and its linear independence from the limit of $n^{-1}H_n'X_n$. In the appendix we show that

$$\operatorname*{plim}_{n\to\infty} n^{-1}H_n'W_ny_n = \lim_{n\to\infty} n^{-1}H_n'W_n(I-\lambda W_n)^{-1}X_n\beta. \tag{9}$$

Two points should be noted. First, Assumption 7(b) clearly rules out models in which $\beta = 0$. That is, Assumption 7(b) rules out models in which all of the parameters corresponding to the exogenous regressors – including the intercept parameter, if an intercept is present – are zero. We note that in this case the mean of y_n is zero and hence this case may be of limited interest in practice. Second, as shown in more detail below, if W_n is row normalized the first part of Assumption 7(b) will also fail if the only nonzero element of β corresponds to the constant term. Thus, in this case, Assumption 7(b) requires that the generation of y_n involve at least one nonconstant regressor. One implication of this is that if the weighting matrix in the regression model is row normalized the hypothesis that all slopes are zero can not be tested in terms of the results provided in this paper.

We now give more detail concerning the case in which W_n is row normalized, and its relation to Assumption 7(b). Let $\mathbf{e}_n$ be the $n \times 1$ vector of unit elements. Also, suppose that the first column of X_n is $\mathbf{e}_n$ and the remaining columns are denoted by the $n \times (k-1)$ matrix $X_{1,n}$ so that $X_n = (\mathbf{e}_n, X_{1,n})$. Partition β correspondingly as $\beta = (\beta_0, \beta_1')'$. Then the first equation in (1)

[9] While we believe that our suggestion for selecting instruments is reasonable, permitting other instruments would not affect the subsequent analysis in any essential way.

can be expressed as

$$y_n = \mathbf{e}_n\beta_0 + X_{1,n}\beta_1 + \lambda W_n y_n + u_n. \tag{10}$$

If W_n is row normalized it follows that $W_n\mathbf{e}_n = \mathbf{e}_n$. Now, if $\beta_1 = 0$, then it follows from (8) that

$$E(W_n y_n) = W_n \sum_{i=0}^{\infty} \lambda^i W_n^i \mathbf{e}_n \beta_0 = \mathbf{e}_n\kappa, \qquad \kappa = \beta_0/(1-\lambda). \tag{11}$$

Thus, the mean of $W_n y_n$ is not linearly independent of $\mathbf{e}_n$. In the appendix, we demonstrate that

$$\operatorname*{plim}_{n\to\infty} n^{-1}H_n'(\mathbf{e}_n, W_n y_n) = \lim_{n\to\infty} n^{-1}H_n'(\mathbf{e}_n, \mathbf{e}_n\kappa). \tag{12}$$

Clearly this matrix does not have full column rank, and thus the first part of Assumption 7(b) is violated. In a similar fashion it is not difficult to show that analogous statements hold for the second and third part of Assumption 7(b).

In a sense, our Assumptions 7(b) are similar to the rank condition for identification in linear simultaneous equation systems. Among other things, that condition implies that a certain number of predetermined variables which are excluded from a given equation appear elsewhere in the system with nonzero coefficients. However, there is an important difference between our Assumption 7(b) and the rank condition for identification in linear simultaneous systems. Specifically, suppose our Assumption 7(b) does not hold because W_n is row weighted and $\beta_1 = 0$. Then, the estimation procedure we suggest in Section 3 is not consistent. However, the model's coefficients may still be identified and there may exist another procedure which, although perhaps more complex, is consistent – see, e.g., Kelejian and Prucha (1995) and note that the parameters of their autoregressive model can be consistently estimated but yet a condition corresponding to Assumption 7(b) would clearly not hold. We note that if W_n is not row normalized, then in general $W_n\mathbf{e}_n$ will be linearly independent of $\mathbf{e}_n$ and the development in (12) no longer holds. Thus in this case Assumption 7(b) does not require the existence of a nonconstant regressor in the generation of y_n.

Finally, consider Assumption 8. This assumption was made in Kelejian and Prucha (1995) in proving consistency of their estimator for ρ, which is used in the second step of the estimation procedure proposed below. Our development in the next section will indicate the role of Γ_n in that procedure.

3 A Generalized Spatial Two Stage Least Squares Procedure

Consider again the model in (1). Essentially, we propose a three step procedure. In the first step the regression model in (1) is estimated by two stage least squares (2SLS) using the instruments H_n. In the second step the autoregressive parameter, ρ, is estimated in terms of the residuals obtained via the first step and the generalized moments procedure suggested in Kelejian and Prucha (1995). We note that ρ can be consistently estimated in this manner whether or not W_n and M_n are equal. Finally, in the third step, the regression model in (1) is re-estimated by 2SLS after transforming the model via a Cochrane-Orcutt type transformation to account for the spatial correlation. In analogy to the generalized least squares estimator we refer to this estimation procedure as a generalized spatial two stage least squares (GS2SLS) procedure.[10]

For the following discussion it proves helpful to rewrite (1) more compactly as

$$\begin{aligned} y_n &= Z_n\delta + u_n, \\ u_n &= \rho M_n u_n + \varepsilon_n, \end{aligned} \tag{13}$$

where $Z_n = (X_n, W_n y_n)$ and $\delta = (\beta', \lambda)'$. Applying a Cochorane-Orcutt type transformation to this model yields furthermore

$$y_{n*} = Z_{n*}\delta + \varepsilon_n, \tag{14}$$

where $y_{n*} = y_n - \rho M_n y_n$ and $Z_{n*} = Z_n - \rho M_n Z_n$. In the following we may also express y_{n*} and Z_{n*} as $y_{n*}(\rho)$ and $Z_{n*}(\rho)$ to indicate the dependence of the transformed variables on ρ.

3.1 The First Step of the Procedure

We have previously indicated in (7) that $E[(W_n y_n)u_n'] \neq 0$ and so δ in (13) can not be consistently estimated by ordinary least squares. Therefore consider the following 2SLS estimator:

$$\widetilde{\delta}_n = (\widehat{Z}_n'\widehat{Z}_n)^{-1}\widehat{Z}_n' y_n, \tag{15}$$

[10] Of course, if no spatially lagged dependent variable is present in (1) we can estimate the model in the first and third steps by ordinary least squares; in this case the estimator computed in the third step would be the feasible generalized least squares estimator.

where $\widehat{Z}_n = P_{H_n} Z_n = (X_n, \widehat{W_n y_n})$, where $\widehat{W_n y_n} = P_{H_n} W_n y_n$ and $P_{H_n} = H_n(H_n' H_n)^{-1} H_n'$. The proof of the following theorem is given in the appendix.

Theorem 1 *Suppose the setup and the assumptions of Section 2 hold. Then* $\widetilde{\delta}_n = \delta + O_p(n^{-1/2})$ *and hence* $\widetilde{\delta}_n$ *is consistent for* δ, *i.e.,* $\text{plim}_{n\to\infty} \widetilde{\delta}_n = \delta$.

Remark 1: The essence of Theorem 1 is that the 2SLS estimator which is formulated in terms of the instruments H_n is consistent. For purposes which are related to our second step, however, it is also important to note that the rate of convergence is $n^{-1/2}$.

Although $\widetilde{\delta}_n$ is consistent, it does not utilize information relating to the spatial correlation of the error term. We therefore turn to the second step of our procedure.

3.2 The Second Step of the Procedure

Let $u_{i,n}$, $\overline{u}_{i,n}$, and $\overline{\overline{u}}_{i,n}$ be, respectively, the i-th elements of u_n, $\overline{u}_n = M_n u_n$, and $\overline{\overline{u}}_n = M_n^2 u_n$. Similarly, let $\varepsilon_{i,n}$ and $\overline{\varepsilon}_{i,n}$ be the i-th elements of ε_n and $\overline{\varepsilon}_n = M_n \varepsilon_n$. Then, the spatial correlation model implies

$$u_{i,n} - \rho \overline{u}_{i,n} = \varepsilon_{i,n}, \; i = 1, ..., n \tag{16}$$

and

$$\overline{u}_{i,n} - \rho \overline{\overline{u}}_{i,n} = \overline{\varepsilon}_{i,n}, \; i = 1, ..., n. \tag{17}$$

The following three equation system is obtained by squaring (16) and then summing, squaring (17) and summing, multiplying (16) by (17) and summing, and finally by dividing all terms by the sample size n.[11]

$$\begin{aligned}
&2\rho n^{-1} \sum u_{i,n} \overline{u}_{i,n} - \rho^2 n^{-1} \sum \overline{u}_{i,n}^2 + n^{-1} \sum \varepsilon_{i,n}^2 = n^{-1} \sum u_{i,n}^2 \\
&2\rho n^{-1} \sum \overline{u}_{i,n} \overline{\overline{u}}_{i,n} - \rho^2 n^{-1} \sum \overline{\overline{u}}_{i,n}^2 + n^{-1} \sum \overline{\varepsilon}_{i,n}^2 = n^{-1} \sum \overline{u}_{i,n}^2 \\
&\rho n^{-1} \sum [u_{i,n} \overline{\overline{u}}_{i,n} + \overline{u}_{i,n}^2] - \rho^2 n^{-1} \sum \overline{u}_{i,n} \overline{\overline{u}}_{i,n} + n^{-1} \sum \varepsilon_{i,n} \overline{\varepsilon}_{i,n} = n^{-1} \sum u_{i,n} \overline{u}_{i,n}
\end{aligned} \tag{18}$$

Assumption 5 implies $E(n^{-1} \sum \varepsilon_{i,n}^2) = \sigma_\varepsilon^2$. Noting that $\sum \overline{\varepsilon}_{i,n}^2 = \varepsilon_n' M_n' M_n \varepsilon_n$, Assumption 5 also implies that

$$\begin{aligned}
E(n^{-1} \sum \overline{\varepsilon}_{i,n}^2) &= n^{-1} E[Tr(\varepsilon_n' M_n' M_n \varepsilon_n)] = n^{-1} Tr(E \varepsilon_n \varepsilon_n' M_n' M_n) \\
&= \sigma_\varepsilon^2 n^{-1} Tr(M_n' M_n)
\end{aligned}$$

[11] All sums are taken over $i = 1, \ldots, n$.

where $Tr(.)$ denotes the trace operator. Finally, using similar manipulations, it is not difficult to show that Assumptions 1 and 5 imply $E(\ n^{-1}\sum \varepsilon_{i,n}\overline{\varepsilon}_{i,n}) = 0$. Now let $\alpha = (\rho, \rho^2, \sigma_\varepsilon^2)'$ and $\gamma_n = n^{-1}\left(E(u_n' u_n), E(\overline{u}_n' \overline{u}_n), E(u_n' \overline{u}_n)\right)'$. Then, if expectations are taken across (18), the resulting system of three equations can be expressed as

$$\Gamma_n\ \alpha = \gamma_n \tag{19}$$

where Γ_n is defined in Assumption 8. If Γ_n and γ_n were known, Assumption 8 implies that (19) determines α as $\alpha = \Gamma_n^{-1}\gamma_n$.

Kelejian and Prucha (1995) suggested two estimators of ρ and σ_ε^2. Essentially, these estimators are based on estimated values of Γ_n and γ_n. To define those estimators for ρ and σ_ε^2 within the present context, let $\widetilde{u}_n = y_n - Z_n\widetilde{\delta}_n$, $\widetilde{\overline{u}}_n = M_n\widetilde{u}_n$, and $\widetilde{\overline{\overline{u}}}_n = M_n^2\widetilde{u}_n$, where $\widetilde{\delta}_n$ is the 2SLS estimator obtained in the first step, and denote their i-th elements, respectively, as $\widetilde{u}_{i,n}$, $\widetilde{\overline{u}}_{i,n}$, and $\widetilde{\overline{\overline{u}}}_{i,n}$. Now consider the following estimators for Γ_n and γ_n:

$$G_n = \frac{1}{n}\begin{bmatrix} 2\sum \widetilde{u}_{i,n}\widetilde{\overline{u}}_{i,n} & -\sum \widetilde{\overline{u}}_{i,n}^2 & 1 \\ 2\sum \widetilde{\overline{u}}_{i,n}\widetilde{\overline{\overline{u}}}_{i,n} & -\sum \widetilde{\overline{\overline{u}}}_{i,n}^2 & Tr(M_n' M_n) \\ \sum[\widetilde{u}_{i,n}\widetilde{\overline{\overline{u}}}_{i,n} + \widetilde{\overline{u}}_{i,n}^2] & -\sum \widetilde{\overline{u}}_{i,n}\widetilde{\overline{\overline{u}}}_{i,n} & 0 \end{bmatrix},\ g_n = \frac{1}{n}\begin{bmatrix} \sum \widetilde{u}_{i,n}^2 \\ \sum \widetilde{\overline{u}}_{i,n}^2 \\ \sum \widetilde{u}_{i,n}\widetilde{\overline{u}}_{i,n} \end{bmatrix}. \tag{20}$$

Then, the empirical form of the relationship $\gamma_n = \Gamma_n\ \alpha$ in (19) is

$$g_n = G_n\alpha + v_n \tag{21}$$

where v_n can be viewed as a vector of regression residuals. The simplest of the two estimators of ρ and σ_ε^2 considered by Kelejian and Prucha (1995) is given by the first and the third element of the ordinary least squares estimator $\widetilde{\alpha}_n$ for α obtained from regressing g_n against G_n. Since G_n is a square matrix

$$\widetilde{\alpha}_n = G_n^{-1}g_n. \tag{22}$$

Clearly, $\widetilde{\alpha}_n$ is based on an overparameterization in that it does not utilize the information that the second element of α is the square of the first. We will henceforth denote the estimators of ρ and σ_ε^2 which are based on $\widetilde{\alpha}_n$ as $\widetilde{\rho}_n$ and $\widetilde{\sigma}_{\varepsilon,n}^2$. The second set of estimators of ρ and σ_ε^2, say $\widetilde{\widetilde{\rho}}_n$ and $\widetilde{\widetilde{\sigma}}_{\varepsilon,n}^2$, considered by Kelejian and Prucha (1995) – and which turned out to be more efficient - are defined as the nonlinear least squares estimators based on (21). That is

$\tilde{\tilde{\rho}}_n$ and $\tilde{\tilde{\sigma}}^2_{\varepsilon,n}$ are defined as the minimizers of

$$\left[g_n - G_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma_\varepsilon^2 \end{bmatrix}\right]' \left[g_n - G_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma_\varepsilon^2 \end{bmatrix}\right]. \tag{23}$$

The basic results corresponding to the second step of our procedure are contained in the following theorem. The proof of the theorem is given in the appendix.

Theorem 2 *Suppose the setup and the assumptions of Section 2 hold. Then* $(\tilde{\rho}_n, \tilde{\sigma}^2_{\varepsilon,n})$ *and* $(\tilde{\tilde{\rho}}_n, \tilde{\tilde{\sigma}}^2_{\varepsilon,n})$ *are consistent estimators of* $(\rho, \sigma^2_\varepsilon)$*.*

Remark 2: The essence of Theorem 2 is that a consistent estimator of ρ can be obtained by a relatively simple procedure. The third step of our procedure can be based upon either $\tilde{\rho}_n$ or $\tilde{\tilde{\rho}}_n$. The large sample properties of the 2SLS estimator in the third step are the same whether it is based on $\tilde{\rho}_n$ or $\tilde{\tilde{\rho}}_n$. However, $\tilde{\tilde{\rho}}_n$ is more efficient than $\tilde{\rho}_n$ as an estimator for ρ, and hence its use in the third step may be preferred due to small sample considerations.

3.3 The Third Step of the Procedure

If ρ were known we could estimate the vector of regression parameters δ by 2SLS based on (14). As remarked above, in analogy to the generalized least squares estimator, we refer to this estimator, say $\widehat{\delta}_n$, as the generalized spatial 2SLS estimator, or for short as the GS2SLS estimator. This estimator is given by

$$\widehat{\delta}_n = [\widehat{Z}_{n*}(\rho)'\widehat{Z}_{n*}(\rho)]^{-1}\widehat{Z}_{n*}(\rho)'y_{n*}(\rho) \tag{24}$$

where $\widehat{Z}_{n*}(\rho) = P_{H_n} Z_{n*}(\rho)$. (Recall that $Z_{n*}(\rho) = Z_n - \rho M_n Z_n$, $y_{n*}(\rho) = y_n - \rho M_n y_n$, $Z_n = (X_n, W_n y_n)$ and $P_{H_n} = H_n(H_n' H_n)^{-1} H_n'$.) Because H_n includes the linearly independent columns of both X_n and $M_n X_n$ it should be clear that $\widehat{Z}_{n*}(\rho) = (X_n - \rho M_n X_n, \widehat{W_n y_n - \rho M_n W_n y_n})$ where

$$\widehat{W_n y_n - \rho M_n W_n y_n} = P_{H_n}(W_n y_n - \rho M_n W_n y_n)$$

are the predicted values of $(W_n y_n - \rho M_n W_n y_n)$ in terms of the least squares regression on the instruments H_n.

Of course, in practical applications ρ is typically not known. In this case we may replace ρ in the above expressions by some estimator, say $\widehat{\rho}_n$. The resulting estimator may be termed the feasible GS2SLS estimator and is given by

$$\widehat{\delta}_{F,n} = [\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'y_{n*}(\widehat{\rho}_n), \tag{25}$$

with $\widehat{Z}_{n*}(\widehat{\rho}_n) = P_{H_n}Z_{n*}(\widehat{\rho}_n)$, $Z_{n*}(\widehat{\rho}_n) = Z_n - \widehat{\rho}_n M_n Z_n$, $y_{n*}(\widehat{\rho}_n) = y_n - \widehat{\rho}_n M_n y_n$. By the same argument as above $\widehat{Z}_{n*}(\widehat{\rho}_n) = (X_n - \widehat{\rho}_n M_n X_n, W_n y_n - \widehat{\widehat{\rho}_n M_n W_n y_n})$ with

$$W_n y_n - \widehat{\widehat{\rho}_n M_n W_n y_n} = P_{H_n}(W_n y_n - \widehat{\rho}_n M_n W_n y_n).$$

The proof of the following theorem is given in the appendix.

Theorem 3 *Suppose the setup and the assumptions of Section 2 hold, and $\widehat{\rho}_n$ is a consistent estimator for ρ. (Thus, in particular $\widehat{\rho}_n$ may be taken to be equal to $\widetilde{\rho}_n$ or $\widetilde{\widetilde{\rho}}_n$ which are defined in the second step of the procedure.) Furthermore, let $\widehat{\varepsilon}_n = y_{n*}(\widehat{\rho}_n) - Z_{n*}(\widehat{\rho}_n)\widehat{\delta}_{F,n}$, and $\widehat{\sigma}^2_{\varepsilon,n} = \widehat{\varepsilon}_n'\widehat{\varepsilon}_n/n$. Then*

(a) $\sqrt{n}(\widehat{\delta}_{F,n} - \delta) \xrightarrow{D} N(0, \Phi)$ *with*

$$\begin{aligned}\Phi &= \sigma^2_\varepsilon [\operatorname*{plim}_{n\to\infty} n^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1} \\ &= \sigma^2_\varepsilon [\operatorname*{plim}_{n\to\infty} n^{-1}\widehat{Z}_{n*}(\rho)'\widehat{Z}_{n*}(\rho)]^{-1}.\end{aligned} \tag{26}$$

(b) $\operatorname*{plim}_{n\to\infty} \widehat{\sigma}^2_{\varepsilon,n} = \sigma^2_\varepsilon$.

Remark 3: Among other things, Theorem 3 implies that $\widehat{\delta}_{F,n}$ is consistent. In addition, it suggests that small sample inferences concerning δ can be based on the small sample approximation

$$\widehat{\delta}_{F,n} \dot{\sim} N\left[\delta,\ \widehat{\sigma}^2_{\varepsilon,n}[\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}\right]. \tag{27}$$

4 Concluding Remarks

In this paper we propose a feasible GS2LSL (generalized spatial two stage least squares) procedure to estimate the parameters of a linear regression model which has a spatially lagged dependent variable as well as a spatially autoregressive disturbance term. We demonstrate that our estimator is consistent and asymptotically normal, and we give its large sample distribution. We also demonstrate that the autoregressive parameter in the disturbance process, ρ, is a nuisance parameter in the sense that the large sample distribution of our feasible GS2LSL estimator, which is based upon a consistent estimator of ρ, is the same as that of the GS2LSL estimator which is based upon the true value of ρ. We note that our results are not based upon the assumption that the disturbance terms are normally distributed.

Our feasible GS2LSL estimator is conceptually simple in the sense that its rational is obvious. It is also computationally feasible even in large samples. This is important to note because, at present, the only alternative to our estimator is the maximum likelihood estimator which may not be feasible in large samples unless the weighting matrices involved have simplifying features, such as spareness, symmetry, etc.

The analysis of the feasible GS2SLS estimator given in this paper focuses on its large sample distribution. An obvious suggestions for further research, therefore, relates to corresponding small sample issues. In this regard, a Monte Carlo study focusing on both our suggested GS2SLS procedure as well as the maximum likelihood estimator should be of interest. Such a study could also shed light on how well the large sample distribution given in this paper approximates the actual small sample distribution under various conditions.

A Appendix

Proof of (9) and (12): Let $\psi_n = n^{-1}H_n' W_n y_n$. Then from (4)

$$\psi_n = n^{-1}H_n' W_n (I - \lambda W_n)^{-1}(X_n\beta + u_n). \tag{A.1}$$

Because H_n, W_n and X_n are nonstochastic matrices, Assumption 5 implies that the mean vector and variance covariance matrix of ψ_n are

$$\begin{aligned} E(\psi_n) &= n^{-1}H_n' W_n (I - \lambda W_n)^{-1} X_n\beta \qquad \text{(A.2)} \\ E(\psi_n - E\psi_n)(\psi_n - E\psi_n)' &= n^{-2}H_n' W_n (I - \lambda W_n)^{-1}\Omega_{u_n}(I - \lambda W_n')^{-1}W_n' H_n \\ &= n^{-2}H_n' A_n H_n \end{aligned}$$

where $A_n = W_n(I - \lambda W_n)^{-1}\Omega_{u_n}(I - \lambda W_n')^{-1}W_n'$ and were Ω_{u_n} is given in (5). Assumption 3 and footnote 7 imply that the row and column sums of A_n are uniformly bounded in absolute value. That is there exists some finite constant c_a such that $\sum_{r=1}^n |a_{rs,n}| \leq c_a$ and $\sum_{s=1}^n |a_{rs,n}| \leq c_a$. Observe also that in light of Assumptions 3 and 4 the elements of H_n are uniformly bounded in absolute value by some finite constant, say c_h. Now let the (i,j)-th element of $E(\psi_n - E\psi_n)(\psi_n - E\psi_n)'$ be $\Delta_{ij,n}$. Then

$$\begin{aligned} |\Delta_{ij,n}| &\leq n^{-2}\sum_{s=1}^n \sum_{r=1}^n |h_{ri,n}|\ |a_{rs,n}|\ |h_{sj,n}| \qquad \text{(A.3)} \\ &\leq n^{-2}c_h \sum_{s=1}^n |h_{sj,n}| \sum_{r=1}^n |a_{rs,n}| \\ &\leq n^{-1}c_h^2\ c_a \to 0. \end{aligned}$$

The result in (9) follows from (A.2), (A.3), and Chebyshev's inequality. Since $E(W_n y_n) = W_n(I - \lambda W_n)^{-1}X_n\beta$ the result in (9) can also be stated as

$$\operatorname*{plim}_{n\to\infty} n^{-1}H_n' W_n y_n = \lim_{n\to\infty} n^{-1}H_n' E(W_n y_n).$$

The result in (12) follows as a special case. ■

The proof of Theorem 1 is based upon a central limit theorem for triangular arrays. This theorem is, e.g., given in Kelejian and Prucha (1995), and is described here for the convenience of the reader.

Theorem A.1 *Let $\{v_{i,n}, 1 \leq i \leq n, n \geq 1\}$ be a triangular array of identically distributed random variables. Assume that the random variables $\{v_{i,n}, 1 \leq i \leq n\}$ are (jointly) independently distributed for each n with $E(v_{i,n}) = 0$ and $E(v_{i,n}^2) = \sigma^2 < \infty$. Let $\{a_{ij,n},\ 1 \leq i \leq n,\ n \geq 1\}$, $j = 1, \ldots, k$, be triangular arrays of real numbers that are bounded in absolute value. Further, let*

$$v_n = \begin{bmatrix} v_{1,n} \\ \vdots \\ v_{n,n} \end{bmatrix}, \qquad A_n = \begin{bmatrix} a_{11,n} & \ldots & a_{1k,n} \\ \vdots & & \vdots \\ a_{n1,n} & \ldots & a_{nk,n} \end{bmatrix}.$$

Assume that $\lim_{n\to\infty} n^{-1}A_n'A_n = Q_{AA}$ is a finite and nonsingular matrix. Then $n^{-1/2}A_n'v_n \xrightarrow{D} N(0, \sigma^2 Q_{AA})$.

Proof of Theorem 1: Recall that $\widehat{Z}_n = P_{H_n}Z_n$ with $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$. Hence clearly $\widehat{Z}_n'\widehat{Z}_n = \widehat{Z}_n'Z_n$. In light of this we have from (13) and (15) that

$$\begin{aligned} \widetilde{\delta}_n &= (\widehat{Z}_n'\widehat{Z}_n)^{-1}\widehat{Z}_n'y_n \qquad (A.4) \\ &= \delta + (\widehat{Z}_n'\widehat{Z}_n)^{-1}\widehat{Z}_n'u_n \\ &= \delta + (\widehat{Z}_n'\widehat{Z}_n)^{-1}\widehat{Z}_n'(I - \rho M_n)^{-1}\varepsilon_n \\ &= \delta + [Z_n'H_n(H_n'H_n)^{-1}H_n'Z_n]^{-1}Z_n'H_n(H_n'H_n)^{-1}H_n'(I - \rho M_n)^{-1}\varepsilon_n. \end{aligned}$$

Let $Q_{HH,n} = n^{-1}H_n'H_n$, $Q_{HZ,n} = n^{-1}H_n'Z_n$, $F_n' = H_n'(I - \rho M_n)^{-1}$, then

$$\sqrt{n}(\widetilde{\delta}_n - \delta) = [Q_{HZ,n}'Q_{HH,n}^{-1}Q_{HZ,n}]^{-1}Q_{HZ,n}'Q_{HH,n}^{-1}n^{-1/2}F_n'\varepsilon_n. \qquad (A.5)$$

Observe that, as remarked in the text, in light of Assumptions 3, 4 and 6 the elements of H_n are bounded in absolute value. Observe further that by Assumption 3 the row and column sums of $(I - \rho M_n)^{-1}$ are uniformly bounded in absolute value. Consequently the elements of F_n are bounded in absolute value. Since $\lim_{n\to\infty} n^{-1}F_n'F_n = \Phi$ is finite and nonsingular by Assumption 7(c) it follows from Theorem A.1 that $n^{-1/2}F_n'\varepsilon_n \xrightarrow{D} N(0, \sigma^2\Phi)$. Given Assumptions 7(a),(b) it then follows from (A.5) that

$$\sqrt{n}(\widetilde{\delta}_n - \delta) \xrightarrow{D} N(0, \Delta) \qquad (A.6)$$

where

$$\Delta = \sigma^2[Q_{HZ}'Q_{HH}^{-1}Q_{HZ}]^{-1}Q_{HZ}'Q_{HH}^{-1}\Phi Q_{HH}^{-1}Q_{HZ}[Q_{HZ}'Q_{HH}^{-1}Q_{HZ}]^{-1}.$$

The claims in Theorem 1 now follow trivially from (A.6). ■

In proving Theorem 2 we will use the following notation: Let A be some matrix or vector. Then the Euclidean or l_2 norm of A is $||A|| = [Tr(A'A)]^{1/2}$. This norm is sub-multiplicative, i.e., if B is a conformable matrix, then $||AB|| \leq ||A|| \; ||B||$. We will utilize the following simple lemma, which is proven here for the convenience of the reader.

Lemma A.2 *Let* $\{\xi_{i,n} : 1 \leq i \leq n, n \geq 1\}$ *with* $\xi_{i,n} = (\xi_{i1,n}, \ldots, \xi_{im,n})$ *be a triangular array of* $1 \times m$ *random vectors. Then a sufficient condition for*

$$n^{-1} \sum_{i=1}^{n} ||\xi_{i,n}||^s = O_p(1), \qquad s > 0, \tag{A.7}$$

is that the s-th absolute moments $E\left|\xi_{ij,n}\right|^s$ *are uniformly bounded, i.e., that there exists a finite nonnegative constant* c_ξ *such that for all* $1 \leq i \leq n$, $n \geq 1$, *and* $j = 1, \ldots, m$

$$E\left|\xi_{ij,n}\right|^s \leq c_\xi < \infty. \tag{A.8}$$

Proof: First observe that a sufficient condition for (A.7) is that there exists some finite nonnegative constant c_1 such that

$$E\left(n^{-1} \sum_{i=1}^{n} ||\xi_{i,n}||^s\right) \leq c_1 \tag{A.9}$$

for all $n \geq 1$. To see this consider some arbitrary $\eta > 0$ and define the constant $c_2 = c_1/\eta$. Then

$$P\left(n^{-1} \sum_{i=1}^{n} ||\xi_{i,n}||^s \geq c_2\right) \leq \frac{E\left(n^{-1} \sum_{i=1}^{n} ||\xi_{i,n}||^s\right)}{c_2} \leq \frac{c_1}{c_2} = \eta,$$

which satisfies the requirements of the definition of $O_p(1)$. The first of the above inequalities follows from Markov's inequality. Of course a sufficient condition for (A.9) is that for all $1 \leq i \leq n$ and $n \geq 1$

$$E||\xi_{i,n}||^s \leq c_1. \tag{A.10}$$

Given the definition of $||.||$ we have

$$E||\xi_{i,n}||^s = E\left[\sum_{j=1}^{m}\xi_{ij,n}^2\right]^{s/2} \leq m^{s/2}\sum_{j=1}^{m}E\left|\xi_{ij,n}\right|^s, \tag{A.11}$$

where the last step is based on an inequality given, e.g., in Bierens (1981, p. 16). Hence clearly, if (A.8) holds, then we can find a constant c_1 such that (A.10) and hence (A.7) holds. ■

Proof of Theorem 2: We prove the theorem by demonstrating that all of the conditions assumed by Kelejian and Prucha (1995), i.e., their Assumptions 1-5, are satisfied here. Theorem 2 then follows as a direct consequence of Theorem 1 in Kelejian and Prucha (1995). Assumptions 1-3 and 5 in Kelejian and Prucha (1995) are readily seen to hold by comparing them with the assumptions maintained here. We now show that Assumption 4 in Kelejian and Prucha (1995) also holds.

Recall $Z_n = (X_n, \overline{y}_n)$ with $\overline{y}_n = W_n y_n$, and let $z_{i.,n} = (x_{i1,n}, \ldots, x_{ik,n}, \overline{y}_{i,n})$ be the i-th row of Z_n. Then via (13) in the text, $\widetilde{u}_n = y_n - Z_n\widetilde{\delta}_n = u_n + Z_n(\delta - \widetilde{\delta}_n)$ and so

$$|u_{i,n} - \widetilde{u}_{i,n}| \leq ||z_{i.,n}||\ ||\delta - \widetilde{\delta}_n||. \tag{A.12}$$

Assumption 4 in Kelejian and Prucha (1995) now holds if we can demonstrate that $(\delta - \widetilde{\delta}_n) = O_p(n^{-1/2})$, and that for some $\zeta > 0$

$$n^{-1}\sum_{i=1}^{n}||z_{i.,n}||^{2+\zeta} = O_p(1). \tag{A.13}$$

The former condition was established by Theorem 1. We now establish that (A.13) holds in particular for $\zeta = 1$. By Lemma A.2 a sufficient condition for this is that there exists some finite constant c_z such that for all $1 \leq i \leq n$, $n \geq 1$ and $j = 1, \ldots, k+1$

$$E\left|z_{ij,n}\right|^3 \leq c_z. \tag{A.14}$$

For $j = 1, \ldots, k$ we have $z_{ij,n} = x_{ij,n}$. Since the $x_{ij,n}$'s are assumed to be uniformly bounded in absolute value (A.14) is trivially satisfied for those $z_{ij,n}$'s. For $j = k+1$ we have $z_{ij,n} = \overline{y}_{i,n}$. To complete the proof we now establish that

$$E\left|\overline{y}_{i,n}\right|^3 \leq c_z \tag{A.15}$$

for some finite constant c_z. From (1) or (4) we have

$$\overline{y}_n = W_n y_n = W_n(I - \lambda W_n)^{-1} X_n \beta + W_n(I - \lambda W_n)^{-1}(I - \rho M_n)^{-1}\varepsilon_n. \quad \text{(A.16)}$$

Assumptions 3 and 4 imply that the elements of $d_n = W_n(I - \lambda W_n)^{-1} X_n \beta$ are bounded in absolute value and that the row and column sums of $D_n = W_n(I - \lambda W_n)^{-1}(I - \rho M_n)^{-1}$ are bounded uniformly in absolute value – compare footnote 7. Let c_d denote the common upper bound. From (A.16) we have

$$\overline{y}_{i,n} = d_{i,n} + \sum_{j=1}^{n} d_{ij,n}\varepsilon_{j,n} \quad \text{(A.17)}$$

and hence

$$\begin{aligned} \overline{y}_{i,n}^3 &= d_{i,n}^3 + 3d_{i,n}^2 \sum_{j=1}^{n} d_{ij,n}\varepsilon_{j,n} + 3d_{i,n}\sum_{j=1}^{n}\sum_{l=1}^{n} d_{ij,n}d_{il,n}\varepsilon_{j,n}\varepsilon_{l,n} \quad \text{(A.18)} \\ &\quad + \sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} d_{ij,n}d_{il,n}d_{im,n}\varepsilon_{j,n}\varepsilon_{l,n}\varepsilon_{m,n}. \end{aligned}$$

By Assumption 5 the $\varepsilon_{i,n}$'s are distributed identically, and for each n (jointly) independently, with finite fourth moments. Hence there exists some finite constant c_ε such that for all indices i, j, l, m, and all $n \geq 1$: $E\left|\varepsilon_{i,n}\right| \leq c_\varepsilon$, $E\left|\varepsilon_{j,n}\varepsilon_{l,n}\right| \leq c_\varepsilon$, $E\left|\varepsilon_{j,n}\varepsilon_{l,n}\varepsilon_{m,n}\right| \leq c_\varepsilon$. It now follows from (A.18) and the triangle inequality that

$$\begin{aligned} E\left|\overline{y}_{i,n}\right|^3 &\leq \left|d_{i,n}\right|^3 + 3\left|d_{i,n}\right|^2 \sum_{j=1}^{n}\left|d_{ij,n}\right| E\left|\varepsilon_{j,n}\right| \\ &\quad +3\left|d_{i,n}\right| \sum_{j=1}^{n}\sum_{l=1}^{n}\left|d_{ij,n}\right|\left|d_{il,n}\right| E\left|\varepsilon_{j,n}\varepsilon_{l,n}\right| \\ &\quad +\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\left|d_{ij,n}\right|\left|d_{il,n}\right|\left|d_{im,n}\right| E\left|\varepsilon_{j,n}\varepsilon_{l,n}\varepsilon_{m,n}\right| \\ &\leq c_d^3 + 3c_d^2 c_\varepsilon \sum_{j=1}^{n}\left|d_{ij,n}\right| + 3c_d c_\varepsilon \sum_{j=1}^{n}\sum_{l=1}^{n}\left|d_{ij,n}\right|\left|d_{il,n}\right| \\ &\quad +c_\varepsilon \sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\left|d_{ij,n}\right|\left|d_{il,n}\right|\left|d_{im,n}\right| \\ &\leq c_d^3(1 + 7c_\varepsilon), \end{aligned}$$

observing that $|d_{i,n}| \leq c_d$ and $\sum_{j=1}^{n} |d_{ij,n}| \leq c_d$. This establishes (A.15), which completes the proof. ■

Proof of Theorem 3: Proof of part (a). Recall that $\widehat{Z}_{n*}(\widehat{\rho}_n) = P_{H_n}(Z_n - \widehat{\rho}_n M_n Z_n)$ and $\widehat{Z}_{n*}(\rho) = P_{H_n}(Z_n - \rho M_n Z_n)$ with $P_{H_n} = H_n(H_n' H_n)^{-1} H_n'$. We first establish the following preliminary results:

$$\operatorname*{plim}_{n\to\infty} n^{-1} \widehat{Z}_{n*}(\widehat{\rho}_n)' \widehat{Z}_{n*}(\widehat{\rho}_n) = \operatorname*{plim}_{n\to\infty} n^{-1} \widehat{Z}_{n*}(\rho)' \widehat{Z}_{n*}(\rho) = \overline{Q} \tag{A.19}$$

$$n^{-1/2} \widehat{Z}_{n*}(\widehat{\rho}_n)' \varepsilon_n \xrightarrow{D} N(0, \sigma_\varepsilon^2 \overline{Q}) \tag{A.20}$$

$$\operatorname*{plim}_{n\to\infty} (\widehat{\rho}_n - \rho) n^{-1/2} \widehat{Z}_{n*}(\widehat{\rho}_n)' M_n u_n = 0 \tag{A.21}$$

where

$$\overline{Q} = [Q_{HZ} - \rho Q_{HMZ}]' Q_{HH}^{-1} [Q_{HZ} - \rho Q_{HMZ}] \tag{A.22}$$

is finite and nonsingular.

The result (A.19) follows immediately from Assumption 7 and the consistency of $\widehat{\rho}_n$ observing that

$$\begin{aligned} n^{-1} \widehat{Z}_{n*}(\widehat{\rho}_n)' \widehat{Z}_{n*}(\widehat{\rho}_n) &= n^{-1}(Z_n - \widehat{\rho}_n M_n Z_n)' P_{H_n} (Z_n - \widehat{\rho}_n M_n Z_n) \\ &= (n^{-1} Z_n' H_n - \widehat{\rho}_n n^{-1} Z_n' M_n' H_n) \\ &\quad (n^{-1} H_n' H_n)^{-1} (n^{-1} H_n' Z_n - \widehat{\rho}_n n^{-1} H_n' M_n Z_n). \end{aligned} \tag{A.23}$$

To prove result (A.20) observe that

$$\begin{aligned} n^{-1/2} \widehat{Z}_{n*}(\widehat{\rho}_n)' \varepsilon_n &= n^{-1/2} (Z_n - \widehat{\rho}_n M_n Z_n)' P_{H_n} \varepsilon_n \\ &= (n^{-1} Z_n' H_n - \widehat{\rho}_n n^{-1} Z_n' M_n' H_n)(n^{-1} H_n' H_n)^{-1} n^{-1/2} H_n' \varepsilon_n. \end{aligned} \tag{A.24}$$

In light of Assumptions 3, 4 and 6 the elements of H_n are bounded in absolute value. Given this and Assumptions 5 and 7 we have from Theorem A.1 that

$$n^{-1/2} H_n' \varepsilon_n \xrightarrow{D} N(0, \sigma_\varepsilon^2 Q_{HH}). \tag{A.25}$$

The result (A.20) now follows from (A.24) and (A.25), Assumption 7 and the consistency of $\widehat{\rho}_n$.

To prove result (A.21) observe that

$$\begin{aligned}(\widehat{\rho}_n-\rho)n^{-1/2}\widehat{Z}_{n*}(\widehat{\rho}_n)'M_nu_n &= (\widehat{\rho}_n-\rho)n^{-1/2}(Z_n-\widehat{\rho}_nM_nZ_n)' \\ &\quad P_{H_n}M_nu_n \\ &= (\widehat{\rho}_n-\rho)(n^{-1}Z_n'H_n-\widehat{\rho}_nn^{-1}Z_n'M_n'H_n) \\ &\quad (n^{-1}H_n'H_n)^{-1}n^{-1/2}H_n'M_nu_n.\end{aligned} \tag{A.26}$$

Note that $E(n^{-1/2}H_n'M_nu_n)=0$ and $E(n^{-1}H_n'M_nu_nu_n'M_n'H_n)=$ $n^{-1}H_n'M_n\Omega_{u_n}M_n'H_n$, where Ω_{u_n} is given in (5). Assumptions 3, 4, and 6 imply that the elements of $n^{-1}H_n'M_n\Omega_{u_n}M_n'H_n$ are bounded in absolute value and hence $n^{-1/2}H_n'M_nu_n=O_p(1)$. Given this the result (A.21) now follows from (A.26), Assumption 7 and the consistency of $\widehat{\rho}_n$.

To prove part (a) of the theorem observe that $\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)=$ $\widehat{Z}_{n*}(\widehat{\rho}_n)'Z_{n*}(\widehat{\rho}_n)$ and hence

$$\begin{aligned}\widehat{\delta}_{F,n} &= [\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'y_{n*}(\widehat{\rho}_n) \\ &= \delta+[\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'u_{n*}(\widehat{\rho}_n)\end{aligned} \tag{A.27}$$

where

$$u_{n*}(\widehat{\rho}_n)=y_{n*}(\widehat{\rho}_n)-Z_{n*}(\widehat{\rho}_n)\delta=\varepsilon_n-(\widehat{\rho}_n-\rho)M_nu_n. \tag{A.28}$$

Consequently

$$\begin{aligned}\sqrt{n}(\widehat{\delta}_{F,n}-\delta) &= [n^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}n^{-1/2}\widehat{Z}_{n*}(\widehat{\rho}_n)'\varepsilon_n \\ &\quad -[n^{-1}\widehat{Z}_{n*}(\widehat{\rho}_n)'\widehat{Z}_{n*}(\widehat{\rho}_n)]^{-1}(\widehat{\rho}_n-\rho)n^{-1/2}\widehat{Z}_{n*}(\widehat{\rho}_n)'M_nu_n.\end{aligned} \tag{A.29}$$

The second term on the r.h.s. of (A.29) converges to zero in probability in light of (A.19) and (A.21). Applying (A.19) and (A.20) to the first part on the r.h.s. of (A.29) yields $\sqrt{n}(\widehat{\delta}_{F,n}-\delta)\xrightarrow{D}N(0,\Phi)$ with $\Phi=\sigma_\varepsilon^2\overline{Q}^{-1}$, which establishes part (a) of the theorem.

Proof of part (b). To prove part (b) of the theorem observe that

$$\begin{aligned}\widehat{\varepsilon}_n &= y_{n*}(\widehat{\rho}_n)-Z_{n*}(\widehat{\rho}_n)\widehat{\delta}_{F,n} \\ &= y_{n*}(\widehat{\rho}_n)-Z_{n*}(\widehat{\rho}_n)\delta-Z_{n*}(\widehat{\rho}_n)(\widehat{\delta}_{F,n}-\delta) \\ &= \varepsilon_n-(\widehat{\rho}_n-\rho)M_nu_n-Z_{n*}(\widehat{\rho}_n)(\widehat{\delta}_{F,n}-\delta).\end{aligned} \tag{A.30}$$

Consequently

$$\widehat{\sigma}_\varepsilon^2=n^{-1}\widehat{\varepsilon}_n'\widehat{\varepsilon}_n=n^{-1}\varepsilon_n'\varepsilon_n+\Delta_n^1+\Delta_n^2+\Delta_n^3+\Delta_n^4+\Delta_n^5 \tag{A.31}$$

where

$$\begin{aligned}
\Delta_n^1 &= -2(\widehat{\delta}_{F,n} - \delta)'[n^{-1}Z_{n*}(\widehat{\rho}_n)'\varepsilon_n], \qquad \text{(A.32)}\\
\Delta_n^2 &= (\widehat{\delta}_{F,n} - \delta)'[n^{-1}Z_{n*}(\widehat{\rho}_n)'Z_{n*}(\widehat{\rho}_n)](\widehat{\delta}_{F,n} - \delta),\\
\Delta_n^3 &= 2(\widehat{\delta}_{F,n} - \delta)'[n^{-1}Z_{n*}(\widehat{\rho}_n)'M_n u_n](\widehat{\rho}_n - \rho),\\
\Delta_n^4 &= -2(\widehat{\rho}_n - \rho)[n^{-1}\varepsilon_n' M_n u_n],\\
\Delta_n^5 &= (\widehat{\rho}_n - \rho)^2[n^{-1}u_n' M_n' M_n u_n].
\end{aligned}$$

Assumption 5 and Chebyshev's inequality imply $\operatorname*{plim}_{n\to\infty} n^{-1}\varepsilon_n'\varepsilon_n = \sigma_\varepsilon^2$. To prove that $\operatorname*{plim}_{n\to\infty} \widehat{\sigma}_{\varepsilon,n}^2 = \sigma_\varepsilon^2$ we now demonstrate that $\operatorname*{plim}_{n\to\infty} \Delta_n^j = 0$ for $j = 1,\ldots,5$. Since $\operatorname*{plim}_{n\to\infty} \widehat{\delta}_{F,n} = \delta$ by part (a) of the theorem, and $\operatorname*{plim}_{n\to\infty} \widehat{\rho}_n = \rho$ by assumption, it suffices to show that each of the terms in square brackets on the r.h.s. of (A.32) is $O_p(1)$. By definition $Z_{n*}(\widehat{\rho}_n) = [Z_n - \widehat{\rho}_n M_n Z_n] = [X_n, W_n y_n] - \widehat{\rho}_n[M_n X_n, M_n W_n y_n]$ and thus it suffices to demonstrate that

$$\begin{aligned}
n^{-1}Z_n'\varepsilon_n &= \begin{bmatrix} n^{-1}X_n'\varepsilon_n \\ n^{-1}y_n'W_n'\varepsilon_n \end{bmatrix} = O_p(1), \qquad \text{(A.33)}\\
n^{-1}Z_n'M_n'\varepsilon_n &= \begin{bmatrix} n^{-1}X_n'M_n'\varepsilon_n \\ n^{-1}y_n'W_n'M_n'\varepsilon_n \end{bmatrix} = O_p(1),\\
n^{-1}Z_n'Z_n &= \begin{bmatrix} n^{-1}X_n'X_n & n^{-1}X_n'W_n y_n \\ n^{-1}y_n'W_n'X_n & n^{-1}y_n'W_n'W_n y_n \end{bmatrix} = O_p(1),\\
n^{-1}Z_n'M_n'M_n Z_n &= \begin{bmatrix} n^{-1}X_n'M_n'M_n X_n & n^{-1}X_n'M_n'M_n W_n y_n \\ n^{-1}y_n'W_n'M_n'M_n X_n & n^{-1}y_n'W_n'M_n'M_n W_n y_n \end{bmatrix}\\
&= O_p(1),\\
n^{-1}Z_n'M_n Z_n &= \begin{bmatrix} n^{-1}X_n'M_n X_n & n^{-1}X_n'M_n W_n y_n \\ n^{-1}y_n'W_n'M_n X_n & n^{-1}y_n'W_n'M_n W_n y_n \end{bmatrix}\\
&= O_p(1),\\
n^{-1}Z_n'u_n &= \begin{bmatrix} n^{-1}X_n'u_n \\ n^{-1}y_n'W_n'u_n \end{bmatrix} = O_p(1),\\
n^{-1}Z_n'M_n'u_n &= \begin{bmatrix} n^{-1}X_n'M_n'u_n \\ n^{-1}y_n'W_n'M_n'u_n \end{bmatrix} = O_p(1),\\
n^{-1}\varepsilon_n'M_n u_n &= O_p(1),\\
n^{-1}u_n'M_n'M_n u_n &= O_p(1).
\end{aligned}$$

Recall from (4) that $y_n = (I - \lambda W_n)^{-1} X_n \beta + (I - \lambda W_n)^{-1}(I - \rho M_n)^{-1}\varepsilon_n$ and $u_n = (I - \rho M_n)^{-1}\varepsilon_n$. Upon substitution of those expressions for y_n and u_n in (A.33) we see that the respective components are composed of three types of expressions. Those expressions are of the form, $n^{-1}A_n$, $n^{-1}B_n\varepsilon_n$ or $n^{-1}\varepsilon_n' C_n \varepsilon_n$, where A_n is a vector or matrix of nonstochastic elements, and B_n and C_n are matrices of nonstochastic elements. Given Assumptions 3 and 4 it is readily seen that the elements of expressions of the form $n^{-1}A_n$ are bounded in absolute value, i.e., $n^{-1}A_n = O(1)$. Furthermore it is seen that for expressions of the form $n^{-1}B_n\varepsilon_n$ and $n^{-1}\varepsilon_n' C_n \varepsilon_n$ the elements of the matrices B_n are bounded uniformly in absolute value, and the row and column sums of the matrices C_n are bounded uniformly in absolute value - compare footnote 7. Now let $c_b < \infty$ denote the bound for the absolute values of the elements of B_n, then we have

$$
\begin{aligned}
E\left|n^{-1}B_n\varepsilon_n\right| &= E\left|\left[\begin{array}{c} \vdots \\ n^{-1}\sum_{i=1}^n b_{ji,n}\varepsilon_{i,n} \\ \vdots \end{array}\right]\right| \qquad \text{(A.34)} \\
&\leq \left[\begin{array}{c} \vdots \\ n^{-1}\sum_{i=1}^n |b_{ji,n}|\, E\,|\varepsilon_{i,n}| \\ \vdots \end{array}\right] \leq \left[\begin{array}{c} \vdots \\ c_b E\,|\varepsilon_{1,n}| \\ \vdots \end{array}\right] < \infty.
\end{aligned}
$$

Similarly, let $c_c < \infty$ be the bound for the row and column sums of the absolute elements of C_n, then

$$
\begin{aligned}
E\left|n^{-1}\varepsilon_n' C_n \varepsilon_n\right| &= E\left|n^{-1}\sum_{i=1}^n\sum_{j=1}^n c_{ij,n}\varepsilon_{i,n}\varepsilon_{j,n}\right| \qquad \text{(A.35)} \\
&\leq n^{-1}\sum_{i=1}^n\sum_{j=1}^n |c_{ij,n}|\, E\,|\varepsilon_{i,n}|\,|\varepsilon_{j,n}| \leq \sigma_\varepsilon^2 c_c < \infty
\end{aligned}
$$

where we have also used the Cauchy-Schwartz inequality. Using Markov's inequality it now follows from (A.34) and (A.35) that $n^{-1}B_n\varepsilon_n = O_p(1)$ and $n^{-1}\varepsilon_n' C_n\varepsilon_n = O_p(1)$. We have thus established that all expression in (A.33) are $O_p(1)$, which complete the proof of part (b) of the theorem. ■

References

[1] Amemiya, T. *Advanced Econometrics,* Cambridge: Harvard University Press, 1985.

[2] Anselin, L. A. "New Look at Statistical Model Identification." *IEEE Transactions on Automatic Control AC* 19 (1982), 716-723.

[3] Anselin, L. *Spatial Econometrics: Methods and Models*, Boston: Kluwer Academic Publishers, 1988.

[4] Anselin, L. "Some Robust Approaches to Testing and Estimation in Spatial Econometrics." *Regional Science and Urban Economics* 20 (1990), 141-163.

[5] Anselin, L., A. Bera, R. Florax, and M. Yoon. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26 (1990), 77-104.

[6] Anselin, L. and S. Rey. "Properties of Tests for Spatial Dependence in Linear Regression Models." *Geographical Analysis* 23 (1991), 110-131.

[7] Bell, K. and N. Bockstael. "Applying the Generalized Method of Moments Approach to Spatial Problems Involving Micro-Level Data." Department of Agricultural and Resource Economics Working Paper 97-03, University of Maryland, 1997.

[8] Bierens, H.J. *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, Lecture Notes in Economics and Mathematical Systems 192, New York: Springer Verlag, 1981.

[9] Blommestein, H. "Specification and Estimation of Spatial Econometric Models." *Regional Science and Urban Economics* 13 (1983), 251-270.

[10] Case, A. "Spatial Patterns in Household Demand." *Econometrica* 59 (1991), 953-966.

[11] Case, A. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics* 22 (1992), 491-508.

[12] Case, A., Hines Jr., J., and H. Rosen. "Budget Spillovers and Fiscal Policy Independence: Evidence from the States." *Journal of Public Economics* 52 (1993), 285-307.

[13] Horn, R. and C. Johnson. *Matrix Analysis*, New York: Cambridge University Press, 1985.

[14] Kelejian, H. and D. Robinson. "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model." *Papers in Regional Science* 72 (1993), 297-312.

[15] Kelejian, H.H. and I.R. Prucha. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." Department of Economics, University of Maryland, Working Paper 95-03, 1995.

[16] Ord, J. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70 (1975), 120-126.

[17] Pace, R. and R. Barry. "Sparse Spatial Autoregressions." *Statistics and Probability Letters* 2158 (1996), 1-7.

[18] Pötscher, B.M. and I.R. Prucha. *Dynamic Nonlinear Econometric Models, Asymptotic Theory*, New York: Springer Verlag, 1997.

[19] Schmidt, P. *Econometrics*, New York: Marcel Dekker, 1976.

[20] Whittle, P. "On Stationary Processes in the Plane." *Biometrica* 41 (1954), 434-449.