Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations

Harry H. Kelejian and Ingmar R. Prucha[¤]
Department of Economics
University of Maryland
College Park, MD 20742

June 2001, Revised April 2002

Abstract

In this paper we consider a simultaneous system of spatially interrelated cross sectional equations. Our speci…cation incorporates spatial lags in the endogenous and exogenous variables. In modelling the disturbance process we allow for both spatial correlation as well as correlation across equations. The data set is taken to be a single cross section of observations. The model may be viewed as an extension of the widely used single equation Cli¤-Ord model. We suggest computationally simple limited and full information instrumental variable estimators for the parameters of the system and give formal large sample results.

JEL classi…cation: C31

Keywords: Spatial dependence; Simultaneous equation system; two-stage least squares, three-stage least squares; Generalized moments estimation

[¤]Corresponding author: Ingmar Prucha, Department of Economics, University of Maryland, College Park, MD 20742, Tel.: 301-405-3499, Fax: 301-405-3542, Email: prucha@econ.umd.edu

1 Introduction[¤]

Spatial models have attracted considerable interest in the recent economics and econometrics literature, both on an empirical and theoretical level.[1] One of the most widely used spatial models is the single equation model introduced by Cli¤ and Ord (1973, 1981). This model is a variant of the model introduced by Whittle (1954) and is sometimes referred to as a spatial autoregressive model; see, e.g., Anselin (1988). In this paper we consider an extension of the single equation Cli¤ and Ord model. In particular, we consider the estimation of a simultaneous system of cross sectional equations with spatial dependencies. The data set is assumed to be a single cross section of observations on the variables involved.[2] The spatial dependencies arise for two reasons. First, the error terms are assumed to be spatially correlated, as well as correlated across equations. Second, the value of the dependent variable in a given equation corresponding to a given cross sectional observation is assumed, in part, to depend upon a weighted sum of that dependent variable over "neighboring" cross sectional units. Such weighted sums over neighboring units are often described in the literature as spatial lags of the variables involved. Our equations may also contain spatial lags of the exogenous variables.

We introduce both limited and full information estimators for the model parameters that are in the spirit of the classical two and three stage least

---

[¤]We thank two anonymous referees and the participants of seminars at Monterey and Vienna for helpful comments. Financial support from the National Science Foundation through grant SES-0001780 is gratefully acknowledged.

[1]Recent empirical and theoretical papers include Delong and Summers (1991), Case (1991), Krugman (1991, 1995), Case, Hines, and Rosen (1993), Holtz-Eakin (1994), Shroder (1995), Anselin et al. (1996), Audretsch and Feldman (1996), Ausubel et al. (1997), Driscoll and Kraay (1998), Kelejian and Robinson (1997), Kelejian and Prucha (1998, 1999, 2001a,b,c), Pinkse and Slade (1998), Buettner (1999), Conley (1999), Pinkse (1999), Lee (1999a,b, 2001a,b, 2002), Rey and Boarnet (1998), Bell and Bockstael (2000), Baltagi, Song and Koh (2000), Baltagi and Li (2001a,b), and Giacomini and Granger (2001). For reviews and general discussions relating to spatial models see, e.g., Cli¤ and Ord (1973, 1981), Anselin (1988), and Cressie (1993).

[2]The force of our modelling and suggested estimation procedure is that it only requires a single cross section of data. Evident variations of our procedure could also be considered if panel data were available and the number of time periods, say T, were small relative to the number of cross sectional units, say n. In the panel data case, if T were large relative to n, a wide variety of models and estimation procedures would be available. For example, see Prucha (1985) and Baltagi (1995).

squares estimators. We give formal large sample results relating to our suggested estimators. Specifically, we demonstrate that our estimators are consistent and asymptotically normal. Our results therefore generalize those given by Kelejian and Prucha (1998) in a single equation framework. One step in our procedure is based on a generalized moments (GM) estimator of spatial autoregressive coefficients. The GM estimator was introduced by Kelejian and Prucha (1999).

It will become evident that our systems estimators are computationally simple even in large samples. One reason for this is that our procedure is based in part on our GM procedure rather than on a quasi maximum likelihood procedure, which is often considered in a single equation framework, e.g., see Anselin (1988), Case (1991), and Case, Hines, and Rosen (1993). Even in a single equation framework, such quasi maximum likelihood procedures are often infeasible in moderate to large size samples unless the weights matrix is of a special form, - see, e.g., the discussion and references in Kelejian and Prucha (1999).

The model is specified and interpreted in Section 2. The limited and full information estimators are defined, and their asymptotic properties are given in Section 3. Conclusions and suggestions for further work are given in Section 4. Proofs and other technical details are relegated to the appendix.

2 Model

In this section we specify the model along with a discussion of the maintained assumptions. It proves helpful to introduce the following notational conventions and definitions: Let $(A_n)_{n \in N}$ be some sequence of $np \times np$ matrices where $p \geq 1$ is some fixed positive integer. Then we denote its (i, j)-th element as $a_{ij,n}$. If A_n is a square nonsingular matrix, then A_n^{-1} denotes its inverse and a_n^{ij} denotes its (i, j)-th element; if A_n is singular, A_n^{-1} denotes the generalized inverse. If A_n is some vector or matrix, then $\|A_n\| = [tr(A_n' A_n)]^{1/2}$ where $tr(.)$ denotes the trace. Furthermore we say that the row and column sums of the sequence of matrices A_n are bounded uniformly in absolute value if there exits a positive finite constant c_A, independent of n, such that

$$\max_{1 \leq i \leq np} \sum_{j=1} |a_{ij,n}| \leq c_A \text{ and } \max_{1 \leq j \leq np} \sum_{i=1} |a_{ij,n}| \leq c_A$$

for all $n \in N$. We note for future reference that if the row and column sums of A_n and B_n are bounded uniformly in absolute value, then (assuming conformability for multiplication) so are the row and column sums of $C_n = A_n B_n$; see Remark A1 in the appendix.

2.1 Model Specification

The following spatial simultaneous equation model can be viewed as an extension of the widely used spatial single equation model introduced by Cliff and Ord (1973, 1981). In particular, we consider the following system of spatially interrelated cross sectional equations corresponding to n cross sectional units:

$$Y_n = Y_n B + X_n C + \overline{Y}_n \Lambda + U_n; \tag{1}$$

with

$$\begin{aligned} Y_n &= (y_{1,n}, \ldots, y_{m,n}), \; X_n = (x_{1,n}, \ldots, x_{k,n}), \; U_n = (u_{1,n}, \ldots, u_{m,n}), \\ \overline{Y}_n &= (\overline{y}_{1,n}, \ldots, \overline{y}_{m,n}), \; \overline{y}_{j,n} = W_n y_{j,n}, \quad j = 1, \ldots, m, \end{aligned}$$

where $y_{j,n}$ is the $n \times 1$ vector of cross sectional observations on the dependent variable in the j-th equation, $x_{l,n}$ is the $n \times 1$ vector of cross sectional observations on the l-th exogenous variable, $u_{j,n}$ is the $n \times 1$ disturbance vector in the j-th equation, W_n is an $n \times n$ weights matrix of known constants,[3] and B, C, and Λ are correspondingly defined parameter matrices of dimension $m \times m$, $k \times m$ and $m \times m$, respectively. In this model spatial spillovers in the endogenous variables are modeled via $\overline{y}_{j,n}$, $j = 1, \ldots, m$. The vector $\overline{y}_{j,n}$ is typically referred to as the spatial lag of $y_{j,n}$. The i-th element of $\overline{y}_{j,n}$ is given by

$$\overline{y}_{ij,n} = \sum_{r=1} w_{ir,n} y_{rj,n}.$$

The weights $w_{ir,n}$ are usually specified to be nonzero if cross sectional unit i relates to unit r in a meaningful way. In such cases, units i and r are said to be neighbors. Usually neighboring units are taken to be those units that are close in some dimension, such as geographic or technological. We note

[3] We are assuming that the system only involves one weights matrix. This assumption is made for ease of presentation, but also seems to be the typical specification in applied work. Our results can be generalized in a straight forward way to the case in which each spatially lagged variable depends upon a weights matrix which is unique to that variable.

that Λ is not assumed to be diagonal, and hence the specification allows for the j-th endogenous variable to depend on its own spatial lag as well as the spatial lags of other endogenous variables.

In addition to allowing for general spatial lags in the endogenous variables we also allow for spatial autocorrelation in the disturbances. In particular we assume that the disturbances are generated by the following spatially autoregressive process:

$$U_n = \overline{U}_n R + E_n, \tag{2}$$

with

$$\begin{aligned} E_n &= (\varepsilon_{1,n}, \ldots, \varepsilon_{m,n}), \quad R = \operatorname{diag}_{j=1}^{m}(\rho_j), \\ \overline{U}_n &= (\overline{u}_{1,n}, \ldots, \overline{u}_{m,n}), \; \overline{u}_{j,n} = W_n u_{j,n}, \quad j = 1, \ldots, m, \end{aligned}$$

where $\varepsilon_{j,n}$ denotes the $n \times 1$ vector of innovations and ρ_j denotes the spatial autoregressive parameter in the j-th equation. Analogous to the terminology used above, the vector $\overline{u}_{j,n}$ is typically referred to as the spatial lag of $u_{j,n}$. Since R is taken to be diagonal the specification relates the disturbance vector in the j-th equation only to its own spatial lag.[4] However, as will become evident below, the disturbances will be spatially correlated across units and across equations via our assumptions concerning the innovations $\varepsilon_{j,n}$, $j = 1, \ldots, m$.

For purposes of generality we have allowed for the elements of the weights matrices, the exogenous regressor matrices, the innovation vectors, and therefore, the endogenous variable matrices to depend on the sample size n, i.e., for the variables to form triangular arrays. We emphasize that by allowing the elements of the exogenous regressor matrices to depend on the sample size we implicitly also allow for spatial lags among the exogenous regressors, in addition to spatial lags in the endogenous variables and disturbances. At this point we also note that our analysis is conditionalized on the realized values of the exogenous variables and so we will henceforth view the matrix X_n as a matrix of constants.

We now express the model in (1) and (2) in a form that will more clearly

[4] Allowing for R to be non-diagonal would further complicate the analysis, and is beyond the scope of the present paper.

reveal its solution for the endogenous variables. Let

$$
\begin{aligned}
y_n &= \operatorname{vec}(Y_n); \overline{y}_n = \operatorname{vec}(\overline{Y}_n);\ x_n = \operatorname{vec}(X_n); \\
u_n &= \operatorname{vec}(U_n); \overline{u}_n = \operatorname{vec}(\overline{U}_n);\ \varepsilon_n = \operatorname{vec}(E_n).
\end{aligned}
$$

Noting that $\overline{y}_n = (I_m \otimes W_n)y_n$ and, if A_1 and A_2 are conformable matrices, that $\operatorname{vec}(A_1A_2) = (A_2' \otimes I)\operatorname{vec}(A_1)$, it follows from (1) and (2) that

$$
\begin{aligned}
y_n &= B_n^* y_n + C_n^* x_n + u_n; \qquad (3)\\
u_n &= R_n^* u_n + \varepsilon_n;
\end{aligned}
$$

where $B_n^* = [(B' \otimes I_n) + (\Lambda' \otimes W_n)]$; $C_n^* = (C' \otimes I_n)$; and $R_n^* = (R \otimes W_n) = \operatorname{diag}_{j=1}^m(\rho_j W_n)$, since R is a diagonal matrix.

Finally, we impose exclusion restrictions on the system in (1). Specifically, let β_j, γ_j, and λ_j be the vectors of nonzero elements of the j-th columns of, respectively, B, C, and Λ. Similarly, let $Y_{j,n}$, $X_{j,n}$, and $\overline{Y}_{j,n}$ be the corresponding matrices of observations on the endogenous variables, exogenous variables, and spatially lagged endogenous variables that appear in the j-th equation. Then, the system in (1) and (2) can be expressed as $(j = 1, \ldots, m)$

$$
\begin{aligned}
y_{j,n} &= Z_{j,n}\delta_j + u_{j,n}, \qquad (4)\\
u_{j,n} &= \rho_j W_n u_{j,n} + \varepsilon_{j,n};
\end{aligned}
$$

where $Z_{j,n} = (Y_{j,n}, X_{j,n}, \overline{Y}_{j,n})$ and $\delta_j = (\beta_j', \gamma_j', \lambda_j')'$.

We make the following assumptions.

Assumption 1 The diagonal elements of the spatial weights matrices W_n are zero.

Assumption 2 (a) The matrices $I_{mn} - B_n^*$ are nonsingular.
(b) The matrices $I_n - \rho_j W_n$ are nonsingular with $|\rho_j| < 1$, $j = 1, \ldots, m$.

Assumption 3 The row and column sums of the matrices W_n, $(I_{mn} - B_n^*)^{-1}$, and $(I_n - \rho_j W_n)^{-1}$, $j = 1, \ldots, m$, are bounded uniformly in absolute value.

Assumption 4 The matrix of exogenous (nonstochastic) regressors X_n has full column rank (for n sufficiently large). In addition, the elements of X_n are uniformly bounded in absolute value.

The next assumption defines the basic properties of the innovations process. In the following let $V_n = [v_{1,n}, \ldots, v_{m,n}]$ be an $n \times m$ matrix of basic innovations and let $v_n = vec(V_n)$.

Assumption 5 The innovations ε_n are generated as follows:

$$\varepsilon_n = (\Sigma_*' \otimes I_n)\, v_n$$

where Σ_* is a nonsingular $m \times m$ matrix and the random variables $\{v_{ij,n} : i = 1, \ldots, n; j = 1, \ldots, m\}$ are, for each n, identically and independently distributed with zero mean, variance one and finite fourth moments, and where the distribution does not depend on n. Furthermore, let $\Sigma = \Sigma_*' \Sigma_*$, then the diagonal elements of Σ are bounded by some constant $b < \infty$.

Let $\varepsilon_n(i)$ and $v_n(i)$ denote the i-th rows of, respectively, E_n and V_n. Then observing that $E_n = V_n \Sigma_*$, and hence $\varepsilon_n(i) = v_n(i)\Sigma_*$, it follows from Assumption 5 that the innovation vectors $\{\varepsilon_n(i) : 1 \le i \le n\}$ are distributed identically and independently with zero mean and variance covariance matrix Σ. Thus the innovations entering the disturbance process are spatially uncorrelated. However, analogous to the classical simultaneous equation model, the specification allows for the innovations corresponding to the same cross sectional unit to be correlated across equations. This is also seen observing that $E\varepsilon_n = 0$ and $E\varepsilon_n\varepsilon_n' = \Sigma \otimes I_n$.

Our suggested estimation procedures are instrumental variable techniques. Let H_n denote the $n \times p$ matrix of instruments utilized by these procedures. As discussed below, in practice H_n will frequently be chosen as a subset of the linearly independent columns of $(X_n, W_n X_n, \ldots, W_n^s X_n)$, where $s \ge 1$ is a finite integer which would typically be less than or equal to two. We maintain the following assumptions concerning the instruments:

Assumption 6 The (nonstochastic) instrument matrix H_n contains at least the linearly independent columns of $(X_n, W_n X_n)$. The elements of H_n are uniformly bounded in absolute value. Furthermore H_n has the following properties:

(a) $Q_{HH} = \lim_{n \to \infty} n^{-1} H_n' H_n$ is a finite nonsingular matrix;

(b) $Q_{HZ_j} = \lim_{n \to \infty} n^{-1} H_n' E(Z_{j,n})$ is a finite matrix which has full column rank, $j = 1, \ldots, m$;

(c) $Q_{HWZ_j} = \lim_{n \to \infty} n^{-1} H_n' W_n E(Z_{j,n})$ is a finite matrix which has full column rank, $j = 1, \ldots, m$;

(d) $Q_{HZ_j} - \rho_j Q_{HWZ_j}$ has full column rank, $j = 1, \ldots, m$;

(e) $\Psi_j = \lim_{n \to \infty} n^{-1} H_n' (I_n - \rho_j W_n)^{-1} (I_n - \rho_j W_n')^{-1} H_n$ is a finite nonsingular matrix, $j = 1, \ldots, m$.

Our next assumption ensures that the autoregressive parameters $\rho_1, \ldots, \rho_m$ are "identifiably unique" – see, e.g., Kelejian and Prucha (1999).

Assumption 7 For $j = 1, \ldots, m$, let

$$\Gamma_{j,n} = n^{-1} E \left\{ \begin{array}{ccc} 2u_{j,n}' \bar{u}_{j,n} & -\bar{u}_{j,n}' \bar{u}_{j,n} & n \\ 2\bar{\bar{u}}_{j,n}' \bar{u}_{j,n} & -\bar{\bar{u}}_{j,n}' \bar{\bar{u}}_{j,n} & tr(W_n' W_n) \\ (u_{j,n}' \bar{\bar{u}}_{j,n} + \bar{u}_{j,n}' \bar{u}_{j,n}) & -\bar{u}_{j,n}' \bar{\bar{u}}_{j,n} & 0 \end{array} \right\}, \tag{5}$$

where $\bar{u}_{j,n} = W_n u_{j,n}$ and $\bar{\bar{u}}_{j,n} = W_n \bar{u}_{j,n} = W_n^2 u_{j,n}$. Let $\lambda_{j,n}$ be the smallest eigenvalue of $\Gamma_{j,n}' \Gamma_{j,n}$. Then we assume that $\lambda_{j,n} \geq \lambda_* > 0$, i.e., the smallest eigenvalues are bounded away from zero.

For future reference, we define $\bar{\varepsilon}_{j,n}$ and $\bar{\bar{\varepsilon}}_{j,n}$ in a similar fashion, namely $\bar{\varepsilon}_{j,n} = W_n \varepsilon_{j,n}$ and $\bar{\bar{\varepsilon}}_{j,n} = W_n^2 \varepsilon_{j,n}$.

2.2 Model Implications

Assumption 1 is a normalization of the model; it also implies that no unit is viewed as its own neighbor. Assumption 2 implies that the system in (1) and (2), or in (3), is complete in that it defines the endogenous variables in terms of the exogenous variables and innovations. In particular, since $I_{mn} - R_n^* = diag_{j=1}^m (I_n - \rho_j W_n)$ it follows from (3) that

$$\begin{aligned} y_n &= (I_{mn} - B_n^*)^{-1} [C_n^* x_n + u_n] \\ u_n &= (I_{mn} - R_n^*)^{-1} \varepsilon_n. \end{aligned} \tag{6}$$

Since $E\varepsilon_n = 0$ by Assumption 5, we have $Eu_n = 0$ and $Ey_n = (I_{mn} - B_n^*)^{-1} C_n^* x_n$. Recalling that Assumption 5 implies $E\varepsilon_n \varepsilon_n' = \Sigma \otimes I_n$ we obtain from (6) the following expressions for the variance covariance matrix of u_n, say $\Omega_{u,n}$, and of y_n, say $\Omega_{y,n}$:

$$\begin{aligned} \Omega_{u,n} &= (I_{mn} - R_n^*)^{-1}(\Sigma \otimes I_n)(I_{mn} - R_n^{*\prime})^{-1}, \\ \Omega_{y,n} &= (I_{mn} - B_n^*)^{-1}\Omega_{u,n}(I_{mn} - B_n^{*\prime})^{-1}. \end{aligned} \tag{7}$$

The disturbances u_n and the endogenous variables y_n are thus seen to be correlated both spatially as well as across equations, and furthermore will generally be heteroskedastic.

Consider now Assumption 3 and its implications for $\Omega_{u,n}$ and $\Omega_{y,n}$. Since the row and column sums of products of matrices, whose row and column sums are bounded in absolute value, have the same property, Assumption 3 implies that the row and column sums of both $\Omega_{u,n}$ and $\Omega_{y,n}$ are bounded uniformly in absolute value. Therefore, this assumption limits the degree of correlation between the elements of u_n and of y_n. For perspective, we note that in virtually all large sample analyses it is necessary to restrict the degree of permissible correlation – see, e.g., Amemiya (1985, ch. 3,4) and Pötscher and Prucha (1997, ch. 5,6).

Now consider Assumption 3 as it relates to the row and column sums of W_n. In practice, it is often assumed that each cross sectional unit has only a finite, and typically, a small number of neighbors and, in turn, it is only a neighbor to a finite and typically small number of other units. It is also often assumed that the rows of the weights matrices are normalized to sum to unity - see, e.g., Case (1991) and Kelejian and Robinson (1995). Under such assumptions the row and column sums of the weights matrices would obviously be bounded in absolute value. In other cases the weights matrices may not be sparse, but the weights are specified to be proportional to the inverse of a distance measure - see, e.g., Dubin (1988), and DeLong and Summers (1991). Again, under reasonable conditions the row and column sums of the weights matrices would be bounded in absolute value, provided the weights decline sufficiently fast as the distances between units increases.

Assumption 4 and parts (a) and (b) of Assumption 6 are crucial in ensuring the consistency of our initial two stage least squares estimator. Parts (c) and (d) of Assumption 6 are analogous in that they are crucial in ensuring the consistency of our generalized two and three stage estimators, which are based on a Cochrane-Orcutt-type transformation of the model. Part (e) of Assumption 6 is used in deriving the limiting distribution of the initial two stage least squares estimator from the untransformed model.

For a further interpretation we note that part (b) of Assumption 6 is a high level condition used to ensure that the instruments H_n allow us to identify the regression parameters δ_j in (4), $j = 1, \ldots, m$. In particular, consider the 2SLS estimator for the parameters in (4), and observe that this estimator is a generalized moments estimator corresponding the the moment conditions

$$E(H_n' u_{j,n}) = 0.$$

Let $u_{j,n}(\underline{\delta}_j) = y_{j,n} - Z_{j,n}\underline{\delta}_j = u_{j,n} + Z_{j,n}(\delta_j - \underline{\delta}_j)$. The condition that $\lim_{n\to\infty} n^{-1} H_n' E(Z_{j,n})$ has full column rank, as maintained in part (b) of Assumption 6, implies that

$$\lim_{n\to\infty} n^{-1} E\left[H_n' u_{j,n}(\underline{\delta}_j)\right] = \left[\lim_{n\to\infty} n^{-1} H_n' E(Z_{j,n})\right](\delta_j - \underline{\delta}_j)$$

is zero if and only if $\underline{\delta}_j = \delta_j$. Thus the condition ensures that, at least asymptotically, the instruments H_n identify the true parameter vector δ_j, $j = 1, \ldots, m$. We note that a similar assumption was made in Kelejian and Prucha (1998), as well as in Lee (2001b), who also provides a discussion of certain cases in which this assumption is violated.[5] In terms of the objective function of the 2SLS estimator, i.e., $u_{j,n}(\underline{\delta}_j)' H_n (H_n' H_n)^{-1} H_n' u_{j,n}(\underline{\delta}_j)$, parts (a) and (b) of Assumption 6 ensure that in the limit the objective function is uniquely maximized at $\underline{\delta}_j = \delta_j$; compare also Amemiya (1985, p.246). Parts (c) and (d) of Assumption 6 play an analogous role in the identification of the model parameters after a Cochrane-Orcutt-type transformation of the model.

The optimal instruments for Y_n and $\overline{Y}_n = W_n Y_n$ are based on their (conditional) means. It is not difficult to see from (3) that if the largest eigenroot of $I_{mn} - B_n^*$ is less than one in absolute value $EY_n = \sum_{s=0}^{\infty} W_n^s X_n \Pi_s$, where Π_s are matrices whose elements are functions of the elements of B, C, and Λ. The instrumental variable estimators considered below are obtained by instrumenting Y_n and $\overline{Y}_n$ in terms of fitted values from regressions on H_n. Our recommendation for choosing H_n to be a subset of the linearly independent columns of $(X_n, W_n X_n, \ldots, W_n^s X_n)$, $s \geq 1$, may hence be viewed as being geared towards achieving a computationally simple approximation to the optimal instruments.[6]

[5] See also Kelejian and Prucha (2001c) for another case in which part (b) of Assumption 6 is violated.

[6] The basic computational operations needed to compute $W_n X_n$ are of the order n^2.

As indicated earlier, Assumption 7 is essentially an identifiability condition for the autoregressive parameters ρ_j. This will become clear from our results in the appendix.

It seems of interest to compare our model with the space-time simultaneous equation model mentioned by Anselin (1988, p.156). While a definite interpretation of the model is difficult because of typographical errors relating to the indices, and because of a lack of formal specifications, it appears that his model can be viewed as a classical simultaneous equation model. The model contains one equation for each cross sectional unit, the variables of which are assumed to be observed over T periods. The dependent variables of the model are simultaneously interrelated across units, and the number of time periods T is assumed to be large relative to the number of cross sectional units n. In contrast, our specification considers a system of equations corresponding to each cross sectional unit, the variables of which are observed for only one time period. Our specifications allow for simultaneity between the different variables corresponding to a particular unit as well as for simultaneity of these variables across units. As an illustration, Anselin's system could relate to time series observations on the demand for police expenditures in each state, which is determined in part by the demand for police expenditures in neighboring states. In contrast, our model could relate, in a given year, to the demand for police expenditures, as well as that for education, roads, parks, etc., for each state. These variables would interact simultaneously within a state, as well as between states.

3 Estimation

In the following we define limited and full information instrumental variable estimators for the parameters of the spatial simultaneous equation model specified above, and derive the limiting distribution of those estimators.

3.1 Limited Information Estimation: GS2SLS

In this section we introduce a generalized spatial two stage least squares procedure (GS2SLS) for the estimation of the parameters in the j-th equation.

We recommend to compute, e.g., $W_n^2 X_n$ recursively by multiplying $W_n X_n$ into W_n, which keeps the computational burden at the order n^2. This approach avoids the need to compute W_n^2, which requires computational operations of the order n^3.

This procedure generalizes the estimator considered in Kelejian and Prucha (1998) for a single equation spatially autoregressive model. The proposed GS2SLS estimation procedure consists of three steps. In the first step we estimate the model parameter vector δ_j in (4) by two stage least squares (2SLS) using H_n as the instrument matrix. Based on the 2SLS estimates of δ_j we compute estimates of the disturbances $u_{j,n}$. In the second step we use those estimated disturbances to estimate the autoregressive parameter ρ_j using the generalized moments procedure introduced in Kelejian and Prucha (1999). In the third step the estimate for ρ_j is used to account for the spatial autocorrelation in the disturbances $u_{j,n}$ using a Cochran-Orcutt-type transformation. The GS2SLS estimator for δ_j is obtained by estimating the transformed model by 2SLS using H_n as the instrument matrix.

3.1.1 Initial 2SLS Estimation

Consider the system in (4) and let $\hat{Z}_{j,n} = P_H Z_{j,n}$, where $P_H = H_n(H_n'H_n)^{-1}H_n'$ and H_n is defined in reference to Assumption 6. Given our assumptions concerning H_n, we have $\hat{Z}_{j,n} = (\hat{Y}_{j,n}, X_{j,n}, \hat{\overline{Y}}_{j,n})$, where $\hat{Y}_{j,n} = P_H Y_{j,n}$, and $\hat{\overline{Y}}_{j,n} = P_H \overline{Y}_{j,n}$. The 2SLS estimator of δ_j is then given by

$$\tilde{\delta}_{j,n} = (\hat{Z}_{j,n}' Z_{j,n})^{-1} \hat{Z}_{j,n}' y_{j,n}. \tag{8}$$

The 2SLS residuals are given by

$$\tilde{u}_{j,n} = y_{j,n} - Z_{j,n}\tilde{\delta}_{j,n}. \tag{9}$$

In the following let $u_{ij,n}$ and $\tilde{u}_{ij,n}$ denote the i-th element of $u_{j,n}$ and $\tilde{u}_{j,n}$, and let $z_{i.,j,n}$ denote the r-th row of $Z_{j,n}$. The proof of the following theorem is given in the appendix.

Theorem 1 Suppose Assumptions 1–6 hold. Then $\tilde{\delta}_{j,n} = \delta_j + O_p(n^{-1/2})$, and so $\tilde{\delta}_{j,n}$ is a $n^{1/2}$-consistent estimator for δ_j. Furthermore

$$|\tilde{u}_{ij,n} - u_{ij,n}| \le \|z_{i.,j,n}\| \, \|\delta_j - \tilde{\delta}_{j,n}\| ,$$

with $n^{-1}\sum_{i=1}^{n} \|z_{i.,j,n}\|^{2+\phi} = O_p(1)$ for some $\phi > 0$.

The theorem shows that the 2SLS residuals satisfy Assumption 4 maintained in Kelejian and Prucha (1999) in connection with the generalized moments estimator for the spatial autoregressive parameter of a disturbance process. This observation will be utilized in demonstrating the consistency of the generalized moments estimator for ρ_j discussed in the next step.

3.1.2 Estimation of the Spatial Autoregressive Parameter

As a second step we apply the generalized moments procedure introduced in Kelejian and Prucha (1999) to estimate the spatial autoregressive parameter of the disturbance process of each equation. To motivate the procedure observe that the relationship in (2) implies that $u_{j,n} - \rho_j \overline{u}_{j,n} = \varepsilon_{j,n}$. Premultiplication by W_n then yields $\overline{u}_{j,n} - \rho_j \overline{\overline{u}}_{j,n} = \overline{\varepsilon}_{j,n}$. These two relationships imply $(j = 1, \dots, m)$

$$
\begin{aligned}
n^{-1}\varepsilon_{j,n}'\varepsilon_{j,n} &= n^{-1}u_{j,n}'u_{j,n} + \rho_j^2 n^{-1}\overline{u}_{j,n}'\overline{u}_{j,n} - 2\rho_j n^{-1}u_{j,n}'\overline{u}_{j,n} \\
n^{-1}\overline{\varepsilon}_{j,n}'\overline{\varepsilon}_{j,n} &= n^{-1}\overline{u}_{j,n}'\overline{u}_{j,n} + \rho_j^2 n^{-1}\overline{\overline{u}}_{j,n}'\overline{\overline{u}}_{j,n} - 2\rho_j n^{-1}\overline{\overline{u}}_{j,n}'\overline{u}_{j,n} \\
n^{-1}\varepsilon_{j,n}'\overline{\varepsilon}_{j,n} &= n^{-1}u_{j,n}'\overline{u}_{j,n} + \rho_j^2 n^{-1}\overline{u}_{j,n}'\overline{\overline{u}}_{j,n} - \rho_j n^{-1}[u_{j,n}'\overline{\overline{u}}_{j,n} + \overline{u}_{j,n}'\overline{u}_{j,n}]
\end{aligned} \tag{11}
$$

Assumption 5 implies that $E(n^{-1}\varepsilon_{j,n}'\varepsilon_{j,n}) = \sigma_{jj}$, where σ_{jj} is the j-th diagonal element of Σ. Noting that $\overline{\varepsilon}_{j,n} = W_n \varepsilon_{j,n}$, it follows from Assumptions 1 and 5 that $E(n^{-1}\overline{\varepsilon}_{j,n}'\overline{\varepsilon}_{j,n}) = \sigma_{jj}\, n^{-1}\mathrm{tr}(W_n' W_n)$, and $E(n^{-1}\varepsilon_{j,n}'\overline{\varepsilon}_{j,n}) = \sigma_{jj} n^{-1}\mathrm{tr}(W_n) = 0$. Let $\alpha_j = (\rho_j, \rho_j^2, \sigma_{jj})'$, and $\gamma_{j,n} = n^{-1}[E(u_{j,n}'u_{j,n}), E(\overline{u}_{j,n}'\overline{u}_{j,n}), E(u_{j,n}'\overline{u}_{j,n})]'$. Then, if expectations are taken across (11) the resulting system can be expressed as $(j = 1, \dots, m)$

$$
\gamma_{j,n} = \Gamma_{j,n}\alpha_j \tag{12}
$$

where $\Gamma_{j,n}$ is defined in (5).

Clearly if $\Gamma_{j,n}$ and $\gamma_{j,n}$ were known, ρ_j and σ_{jj} would be perfectly determined in terms of the vector $\alpha_j = \Gamma_{j,n}^{-1}\gamma_{j,n}$. Following the general approach of Kelejian and Prucha (1999) we define the following estimators for $\Gamma_{j,n}$ and $\gamma_{j,n}$:

$$
\begin{aligned}
G_{j,n} &= n^{-1}\begin{bmatrix} 2\tilde{u}_{j,n}'\overline{\tilde{u}}_{j,n} & -\overline{\tilde{u}}_{j,n}'\overline{\tilde{u}}_{j,n} & n \\ 2\overline{\overline{\tilde{u}}}_{j,n}'\overline{\tilde{u}}_{j,n} & -\overline{\overline{\tilde{u}}}_{j,n}'\overline{\overline{\tilde{u}}}_{j,n} & \mathrm{tr}(W_n'W_n) \\ (\tilde{u}_{j,n}'\overline{\overline{\tilde{u}}}_{j,n} + \overline{\tilde{u}}_{j,n}'\overline{\tilde{u}}_{j,n}) & -\overline{\tilde{u}}_{j,n}'\overline{\overline{\tilde{u}}}_{j,n} & 0 \end{bmatrix}, \\
g_{j,n} &= n^{-1}\left[\tilde{u}_{j,n}'\tilde{u}_{j,n},\ \overline{\tilde{u}}_{j,n}'\overline{\tilde{u}}_{j,n},\ \tilde{u}_{j,n}'\overline{\tilde{u}}_{j,n}\right]',
\end{aligned} \tag{13}
$$

where $\tilde{u}_{j,n}$ denotes the 2SLS residuals defined in (9), $\overline{\tilde{u}}_{j,n} = W_n\tilde{u}_{j,n}$, and $\overline{\overline{\tilde{u}}}_{j,n} = W_n\overline{\tilde{u}}_{j,n} = W_n^2\tilde{u}_{j,n}$. Then, an empirical form of (12) is

$$
g_{j,n} = G_{j,n}\alpha_j + \zeta_{j,n} \tag{14}
$$

where $\zeta_{j,n}$ can be viewed as a vector of regression residuals. Our generalized moments estimator of (ρ_j, σ_{jj}), say $(\tilde{\rho}_j, \tilde{\sigma}_{jj})$, is defined as the nonlinear least squares estimator based on (14). That is [7]

$$(\tilde{\rho}_{j,n}, \tilde{\sigma}_{jj,n}) = \underset{\rho_j \in [-a,a], \sigma_{jj} \in [0,b]}{\arg\min} [g_{j,n} - G_{j,n}\alpha_j]' [g_{j,n} - G_{j,n}\alpha_j] \tag{15}$$

where $a > 1$ is a pre-selected constant. The next theorem establishes the consistency of $(\tilde{\rho}_{j,n}, \tilde{\sigma}_{jj,n})$.

Theorem 2 Suppose Assumptions 1–5 and 7 hold. Then $(\tilde{\rho}_{j,n}, \tilde{\sigma}_{jj,n}) - (\rho_j, \sigma_{jj}) \xrightarrow{p} 0$ as $n \to \infty$, $j = 1, \ldots, m$.

The above theorem establishes the consistency of the generalized moments estimator $\tilde{\rho}_{j,n}$. Using Monte Carlo simulations Kelejian and Prucha (1999) and Das, Kelejian and Prucha (2001) compare the small sample distribution of $\tilde{\rho}_{j,n}$ with that of the maximum likelihood estimator within the context of a single equation model. They found that the two estimators are very similar in small samples. Although such a study has not been performed for the model at hand, we conjecture that a similar finding also holds within the present context.

It is important to note that the optimization space for ρ_j described in (15) is a compact set containing the actual parameter space. The optimization space does not exclude values of ρ_j for which $I_n - \rho_j W_n$ is singular.

3.1.3 Generalized Spatial 2SLS Estimation

Let μ be a scalar and define $y^*_{j,n}(\mu) = y_{j,n} - \mu W_n y_{j,n}$ and $Z^*_{j,n}(\mu) = Z_{j,n} - \mu W_n Z_{j,n}$. Given this notation, we see that applying a Cochrane-Orcutt-type transformation to (4) yields $(j = 1, \ldots, m)$:

$$y^*_{j,n}(\rho_j) = Z^*_{j,n}(\rho_j)\delta_j + \varepsilon_{j,n}. \tag{16}$$

Assume for a moment that ρ_j is known. The generalized spatial two stage least squares (GS2SLS) estimator for δ_j, say $\hat{\delta}_{j,n}$, is then defined as the 2SLS estimator based on (16), i.e.,

$$\hat{\delta}_{j,n} = \left[\hat{Z}^*_{j,n}(\rho_j)' Z^*_{j,n}(\rho_j) \right]^{-1} \hat{Z}^*_{j,n}(\rho_j)' y^*_{j,n}(\rho_j); \tag{17}$$

[7] Following Kelejian and Prucha (1999) we could also define an estimator for ρ_j and σ_{jj} based on the ordinary least squares estimator for α_j from (14). We do not consider this estimator here since results given in Kelejian and Prucha (1999) suggest it is less efficient.

where $\hat{Z}^*_{j,n}(\rho_j) = P_H Z^*_{j,n}(\rho_j)$ with $P_H = H_n(H_n' H_n)^{-1}H_n'$. Our feasible generalized spatial two stage least squares (FGS2SLS) estimator for δ_j, say $\hat{\delta}^F_{j,n}$, is now defined by substituting the generalized moments estimator $\tilde{\rho}_{j,n}$ for ρ_j in the above expression, i.e.,

$$\hat{\delta}^F_{j,n} = \left[\hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' Z^*_{j,n}(\tilde{\rho}_{j,n})\right]^{-1} \hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' y^*_{j,n}(\tilde{\rho}_{j,n}). \tag{18}$$

We have the following theorem concerning the asymptotic distribution of $\hat{\delta}_{j,n}$ and $\hat{\delta}^F_{j,n}$.

Theorem 3 Suppose Assumptions 1–7 hold. Then for $j = 1, \ldots, m$ we have $n^{1/2}\left(\hat{\delta}^F_{j,n} - \hat{\delta}_{j,n}\right) \xrightarrow{P} 0$ as $n \to \infty$ and

$$n^{1/2}\left(\hat{\delta}^F_{j,n} - \delta_j\right) \xrightarrow{D} N(0, \Omega_j)$$

as $n \to \infty$ where

$$\begin{aligned} \Omega_j &= \sigma_{jj}\, \operatorname*{p\,lim}_{n\to\infty} \left[n^{-1}\hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' \hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})\right]^{-1} \\ &= \sigma_{jj}\, \operatorname*{p\,lim}_{n\to\infty} \left[n^{-1}\hat{Z}^*_{j,n}(\rho_j)' \hat{Z}^*_{j,n}(\rho_j)\right]^{-1} \\ &= \sigma_{jj} \left[(Q_{HZ_j} - \rho_j Q_{HWZ_j})' Q_{HH}^{-1} (Q_{HZ_j} - \rho_j Q_{HWZ_j})\right]^{-1}. \end{aligned} \tag{20}$$

The theorem shows that the true and feasible GS2SLS estimators have the same asymptotic distribution. The theorem also holds if $\tilde{\rho}_{j,n}$ is replaced by any other consistent estimator for ρ_j, and thus ρ_j is seen to be a nuisance parameter. A consistent estimator for Ω_j can be found in an obvious way from the first line of (20) by replacing σ_{jj} by $\tilde{\sigma}_{jj}$ or any other consistent estimator for σ_{jj}; see Lemma 1 below.

3.2 Full Information Estimation: GS3SLS

The GS2SLS estimator takes into account potential spatial correlation, but is limited in the information it utilizes in that it does not take into account potential cross equation correlation in the innovation vectors ε_j. To utilize the full system information it is helpful to stack the equations in (16) as

$$y^*_n(\rho) = Z^*_n(\rho)\delta + \varepsilon_n \tag{21}$$

where

$$\begin{aligned} y^*_n(\rho) &= (y^*_{1,n}(\rho_1)', \ldots, y^*_{m,n}(\rho_m)')', \\ Z^*_n(\rho) &= \operatorname{diag}_{j=1}^m(Z^*_{j,n}(\rho_j)), \end{aligned}$$

and $\rho = (\rho_1, \ldots, \rho_m)'$ and $\delta = (\delta_1', \ldots, \delta_m')'$. Recall that $E\varepsilon_n = 0$ and $E\varepsilon_n\varepsilon_n' = \Sigma \otimes I_n$. If ρ and Σ were known, a natural systems instrumental variable estimator of δ would be

$$\delta_n = [\hat{Z}^*_n(\rho)'(\Sigma^{-1} \otimes I_n)(Z^*_n(\rho)]^{-1}\hat{Z}^*_n(\rho)'(\Sigma^{-1} \otimes I_n)y^*_n(\rho) \tag{22}$$

where $\hat{Z}^*_n(\rho) = \operatorname{diag}_{j=1}^m(\hat{Z}^*_{j,n}(\rho_j))$, and as before $\hat{Z}^*_{j,n}(\rho_j) = P_H Z^*_{j,n}(\rho_j)$. Consistent with our terminology for limited information estimators we refer to this estimator as the generalized spatial three stage least squares (GS3SLS) estimator.

A feasible analog of δ_n requires estimators for ρ and Σ. As our estimator for ρ we take $\tilde{\rho}_n = (\tilde{\rho}_{1,n}, \ldots, \tilde{\rho}_{m,n})$ where $\tilde{\rho}_{j,n}$ denotes the generalized moments estimator for ρ_j defined in the previous section. We now suggest a consistent estimator of Σ. In light of (16) let $\tilde{\varepsilon}_{j,n} = y^*_{j,n}(\tilde{\rho}_{j,n}) - Z^*_{j,n}(\tilde{\rho}_{j,n})\hat{\delta}^F_{j,n}$, and define $\hat{\sigma}_{jl,n} = n^{-1}\tilde{\varepsilon}_{j,n}'\tilde{\varepsilon}_{l,n}$ for $j, l = 1, \ldots, m$. Furthermore, let $\hat{\Sigma}_n$ be the $m \times m$ matrix whose (j,l)-th element is $\hat{\sigma}_{jl,n}$. The following lemma establishes that $\hat{\Sigma}_n$ is a consistent estimator for Σ.

Lemma 1 Suppose Assumptions 1–7 hold. Then $\operatorname{plim}_{n\to\infty} \hat{\Sigma}_n = \Sigma$.

Corresponding to the GS3SLS estimator δ_n we now define a feasible generalized spatial three stage least squares (FGS3SLS) estimator as

$$\delta^F_n = [\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes I_n)(Z^*_n(\tilde{\rho}_n)]^{-1}\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes I_n)y^*_n(\tilde{\rho}_n). \tag{23}$$

The next theorem establishes the asymptotic distribution of δ_n and δ^F_n.

Theorem 4 Suppose Assumptions 1–7 hold. Then we have $n^{1/2}\left(\delta^F_n - \delta_n\right) \xrightarrow{P} 0$ as $n \to \infty$ and

$$n^{1/2}\left(\delta^F_n - \delta\right) \xrightarrow{D} N(0, \Omega) \tag{24}$$

as $n \to \infty$ where

$$\begin{aligned}
\Omega &= \left[p \lim_{n\to\infty} n^{-1} \hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes I_n)\hat{Z}_n^*(\tilde{\rho}_n) \right]^{-1} \qquad (25)\\
&= \left[p \lim_{n\to\infty} n^{-1} \hat{Z}_n^*(\rho)'(\Sigma^{-1} \otimes I_n)\hat{Z}_n^*(\rho) \right]^{-1} \\
&= \left\{ \operatorname{diag}_{j=1}^m (Q_{HZ_j} - \rho_j Q_{HWZ_j})'(\Sigma^{-1} \otimes Q_{HH}^{-1}) \operatorname{diag}_{l=1}^m (Q_{HZ_l} - \rho_l Q_{HWZ_l}) \right\}^{-1}.
\end{aligned}$$

The theorem shows that the true and feasible GS3SLS estimators have the same asymptotic distribution. We note that the theorem also holds if $\tilde{\rho}_n$ and $\hat{\Sigma}_n$ are replaced by any other consistent estimators, and thus ρ and Σ are nuisance parameters. A comparison of (20) and (25) shows, using arguments along the lines of, e.g., Schmidt (1976, pp. 209-211), that the GS3SLS estimator δ_n^F is efficient relative to GS2SLS estimator $\hat{\delta}_n^F$, as is expected. The theorem also suggests that the small sample distribution of δ_n^F can be approximated as follows

$$\delta_n^F \stackrel{\cdot}{\sim} N\left(\delta, \left[\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes I_n)\hat{Z}_n^*(\tilde{\rho}_n) \right]^{-1} \right).$$

Suppose we are interested in testing the hypothesis $H_0 : h(\delta) = 0$ versus $H_1 : h(\delta) \neq 0$, where h is some (possibly vector valued) differentiable function. Then the theorem can also be used to construct, in the usual way, Wald tests of that hypothesis. In particular, we can test in this way for the presence of spatial lags in the endogenous variables and/or exogenous variables. Kelejian and Prucha (2001a) give general results concerning the distribution of the Moran I test statistic. Those results can be used to test the hypothesis that the regression disturbances are not spatially correlated.

4 Conclusion

This paper develops estimation theory for a simultaneous system of spatially interrelated cross sectional equations. The model may be viewed as an extension of the widely used single equation model of Cliff and Ord (1973,1981). We introduce both a limited information estimator, termed the FGS2SLS estimator, and a full information estimator, termed the FGS3SLS estimator, and rigorously derive their asymptotic properties. These estimators are based on an approximation of the optimal instruments, and as a result these estimators are computationally simple even in large samples. In future research it should be of interest to explore the small sample properties of these estimators and compare them to those of the maximum likelihood estimator. Comparisons of this sort, within the context of a single equation spatial autoregressive model, have been considered by Das, Kelejian and Prucha (2001). They found that the maximum likelihood estimator and the FGS2SLS estimator exhibited very similar small sample properties, provided at least two spatial lags of the exogenous variables were included among the instruments. We conjecture that these finding will extend to the systems case. Das, Kelejian and Prucha (2001) also found minor differences in the small sample efficiencies of the maximum likelihood and generalized moments estimators of the spatial autoregressive coefficient in the disturbance process. Similar results are also reported in Kelejian and Prucha (1999).

In future research it should be of interest to extend the analysis of this paper to instrumental variable estimators that are based on asymptotically optimal instruments along the lines of Lee (1999a) and Kelejian and Prucha (2001b), who considered such optimal instruments in the context of a single equation spatial autoregressive model. In future research it would also be of interest to derive the limiting distribution of the maximum likelihood estimator in a systems framework under a reasonable set of low level assumptions. Another avenue of suggested research relates to the development of further tests of hypothesis in a spatial systems framework based on the Lagrange Multiplier and Likelihood Ratio testing principles. Such a development could in part expand on results by Baltagi, Song, and Koh (2000) and Baltagi and Li (2001b). Also, the central limit theorem for quadratic forms given in Kelejian and Prucha (2001a) should be helpful towards establishing the asymptotic distribution of those tests. Finally, it should be of interest to develop necessary conditions in the form of counting rules for the identification of the model parameters of systems such as (1). We conjecture that, given the spatial weights satisfy appropriate conditions, for the purpose of these counting rule the spatially lagged dependent variables can be treated as if they are predetermined, since their conditional means will in general differ from the exogenous variables appearing in the original system. For an analogous discussion of counting rules within the framework of a simultaneous equation system that is nonlinear in variables see, e.g., Kelejian and Oates (1981, pp. 288-299).

A Appendix

In this appendix we will repeatedly make use of the following observations.

Remark A1 Let A_n and B_n be $np \times np$ matrices ($p \geq 1$) whose row and column sums are bounded uniformly in absolute value by finite constants c_A and c_B, let S_n be some $np \times s$ matrix whose elements are bounded uniformly in absolute value by some finite constant c_S, and let ξ_n and η_n be $np \times 1$ vectors of uncorrelated random variables with zero mean and finite variances σ_ξ^2 and σ_η^2, i.e., $\xi_n \sim (0, \sigma_\xi^2 I_{np})$ and $\eta_n \sim (0, \sigma_\eta^2 I_{np})$. Then:

(i) The row and column sums of $C_n = A_n B_n$ are bounded uniformly in absolute value by $c_A c_B$.

(ii) The elements of $A_n S_n$ are bounded uniformly in absolute value by the constant $c_A c_S$.

(iii) The elements of $n^{-1} S_n' S_n$ are $O(1)$, the elements of $n^{-1/2} S_n' \xi_n$ are $O_p(1)$, and $n^{-1} \xi_n' A_n \eta_n$ is $O_p(1)$.

The above observations can be readily established: For part (i) see, e.g., Kelejian and Prucha (1999), p. 526. For the last observation in part (iii) note that $E|n^{-1}\xi_n' A_n \eta_n| \leq n^{-1} \sum_i \sum_j |a_{ij,n}| E\left[|\xi_{i,n}\eta_{j,n}|\right] \leq \sigma_\xi \sigma_\eta n^{-1} \sum_i \sum_j |a_{ij,n}| \leq \sigma_\xi \sigma_\eta c_A < \infty$. Also note that the statement allows for the case where $\xi_n = \eta_n$.

Lemma A1 Given Assumptions 1-5 hold, then for $j = 1, \ldots, m$:

$$\operatorname*{plim}_{n\to\infty} n^{-1} H_n' Z_{j,n} = \lim_{n\to\infty} n^{-1} H_n' E(Z_{j,n}) = Q_{HZ_j}; \tag{A.1}$$

$$\operatorname*{plim}_{n\to\infty} n^{-1} H_n' W_n Z_{j,n} = \lim_{n\to\infty} n^{-1} H_n' W_n E(Z_{j,n}) = Q_{HWZ_j}. \tag{A.2}$$

Proof: Recall that $Z_{j,n} = (Y_{j,n}, X_{j,n}, \overline{Y}_{j,n})$ and that $X_{j,n}$ is nonstochastic. Hence to prove (A.1) it suffices to show that

$$\operatorname*{plim}_{n\to\infty} n^{-1} H_n' Y_n = \lim_{n\to\infty} n^{-1} H_n' E(Y_n); \tag{A.3}$$

$$\operatorname*{plim}_{n\to\infty} n^{-1} H_n' \overline{Y}_n = \lim_{n\to\infty} n^{-1} H_n' E(\overline{Y}_n). \tag{A.4}$$

In light of (6) and Assumption 5 observe that $y_n = Ey_n + A_n v_n$ with $A_n = (I_{mn} - B_n^*)^{-1}(I_{mn} - R_n^*)^{-1}(\Sigma_*' \otimes I_n)$. Thus

$$\begin{aligned} n^{-1}\mathrm{vec}(H_n' Y_n) &= n^{-1}(I_m \otimes H_n')y_n \\ &= n^{-1}(I_m \otimes H_n')Ey_n + n^{-1}(I_m \otimes H_n')A_n v_n \\ &= n^{-1}\mathrm{vec}(H_n' EY_n) + n^{-1}(I_m \otimes H_n')A_n v_n. \end{aligned}$$

In light of Remark A1 the elements of $(I_m \otimes H_n')A_n$ are bounded uniformly in absolute value. Since v_n is, by Assumption 5, a vector of i.i.d. random variables with zero mean and variance one it follows from Remark A1 that $n^{-1}(I_m \otimes H_n')A_n v_n = o_p(1)$, which completes the demonstration of (A.3). The demonstration of (A.4) is similar. Analogous arguments can be used to prove (A.2). ∎

Proof of Theorem 1: Recall from (8) that $\tilde\delta_{j,n} = (\hat Z_{j,n}' Z_{j,n})^{-1}\hat Z_{j,n}' y_{j,n}$, where $\hat Z_{j,n} = P_H Z_{j,n}$ and $P_H = H_n(H_n' H_n)^{-1}H_n'$. In light of (4) it is readily seen that

$$\begin{aligned} n^{1/2}\left(\tilde\delta_{j,n} - \delta_j\right) = {} & \left[n^{-1}Z_{j,n}' H_n\left(n^{-1}H_n' H_n\right)^{-1}\left(n^{-1}H_n' Z_{j,n}\right)\right]^{-1} \\ & \left(n^{-1}Z_{j,n}' H_n\right)\left(n^{-1}H_n' H_n\right)^{-1} n^{-1/2}F_{j,n}' \varepsilon_{j,n}. \end{aligned}$$

where $F_{j,n}' = H_n'(I_n - \rho_j W_n)^{-1}$. By Lemma A1 and Assumption 6 plim $n^{-1}H_n' Z_{j,n} = Q_{HZ_j}$, which is finite and has full column rank, and plim $n^{-1}H_n' H_n = Q_{HH}$, which is finite and nonsingular. By Assumption 5 the elements of $\varepsilon_{j,n}$ are i.i.d. with finite variance σ_{jj}. Observe further that in light of Remark A1 and Assumptions 3 and 6, the elements of $F_{j,n}$ are bounded in absolute value and $\Xi_j = \lim_{n\to\infty} n^{-1}F_{j,n}' F_{j,n}$ is finite and nonsingular. Given Theorem A in Kelejian and Prucha (1999) it follows that $n^{-1/2}F_{j,n}' \varepsilon_{j,n} \xrightarrow{d} N(0, \sigma_{jj}\Xi_j)$. As a consequence we have $n^{1/2}\left(\tilde\delta_{j,n} - \delta_j\right) \xrightarrow{d} N(0, \Psi_j)$ with

$$\Psi_j = \left[Q_{HZ_j}' Q_{HH}^{-1} Q_{HZ_j}\right]^{-1} Q_{HZ_j}' Q_{HH}^{-1} \Xi_j Q_{HH}^{-1} Q_{HZ_j} \left[Q_{HZ_j}' Q_{HH}^{-1} Q_{HZ_j}\right]^{-1}.$$

Thus $n^{1/2}\left(\tilde\delta_{j,n} - \delta_j\right) = O_p(1)$, or equivalently $\tilde\delta_{j,n} = \delta_j + O_p(n^{-1/2})$, which completes the proof of the first part of the theorem.

We next prove the second part of the theorem. Clearly $|\hat u_{ij,n} - u_{ij,n}| \le \|z_{i.,j,n}\|\,\|\delta_j - \tilde\delta_{j,n}\|$ in light of (4) and (9), and since the norm $\|.\|$ is submultiplicative. A sufficient condition for $n^{-1}\sum_{i=1}^{n}\|z_{i.,j,n}\|^{2+\phi} = O_p(1)$, $\phi > 0$, is that the $(2+\delta)$-th absolute moment of the elements of $z_{i:j;n}$ are uniformly bounded. In the following we demonstrate that this is indeed the case for, say, $\delta = 1$. The vector $z_{i:j;n}$ may contain exogenous, endogenous, and spatially lagged endogenous variables, which will be considered in turn. By Assumption 4 the exogenous variables are bounded uniformly in absolute value, and thus so are their third moments. In light of (3) we have

$$y_n = d_y + D_y v_n; \qquad \overline{y}_n = d_{\overline{y}} + D_{\overline{y}} v_n; \tag{A.5}$$

where

$$\begin{aligned} d_y &= (I_{mn} - B_n^*)^{-1} C_n^* x_n; & d_{\overline{y}} &= (I_m \otimes W_n) d_y; \\ D_y &= (I_{mn} - B_n^*)^{-1} (I_{mn} - R_n^*)^{-1} (\Sigma_*' \otimes I_n); & D_{\overline{y}} &= (I_m \otimes W_n) D_y. \end{aligned} \tag{A.6}$$

Given Assumptions 3 and 4 it follows immediately from Remark A1 that the elements of d_y and $d_{\overline{y}}$ are bounded uniformly in absolute value, and that the row and column sums of D_y and $D_{\overline{y}}$ are bounded uniformly in absolute value. Assumption 5 implies that the elements of v_n are i.i.d. with finite 4-th moments. It is hence an immediately consequence of Lemma A2 below that $E\,|y_{ij;n}|^3 \le \text{const} < \infty$ and $E\,\left|\overline{y}_{ij;n}\right|^3 \le \text{const} < \infty$, where the constants do not depend on any of the indices. ∎

Lemma A2 Let $\varepsilon_n = (\varepsilon_{1;n}, \ldots, \varepsilon_{np;n})'$ be a $np \times 1$ random vector $(p \ge 1)$, where, for each n, the elements are identically and independently distributed with finite fourth moments. Let $d_n = (d_{1;n}, \ldots, d_{np;n})'$ be some nonstochastic $np \times 1$ vector whose elements are uniformly bounded in absolute value, and let $D_n = (d_{ij;n})$ be a nonstochastic $np \times np$ matrix whose row and column sums are uniformly bounded in absolute value by some (finite) constant c_d. Define

$$\xi_n = (\xi_{1;n}, \ldots, \xi_{np;n})' = d_n + D_n \varepsilon_n;$$

then $E\left|\xi_{i;n}\right|^3 \le c < \infty$, where c is a finite constant that does not depend on i and n.

Proof: Clearly $\xi_{i;n} = d_{i;n} + f_{i;n}$ where $f_{i;n} = \sum_j d_{ij;n} \varepsilon_{j;n}$. By Minkovski's inequality $\left[E\left|\xi_{i;n}\right|^3\right]^{1/3} \le \left[|d_{i;n}|^3\right]^{1/3} + \left[E\,|f_{i;n}|^3\right]^{1/3}$. Since the $d_{i;n}$'s are uniformly bounded in absolute value it suffices to show that moments $E\,|f_{i;n}|^3$

are uniformly bounded. By assumption the $\varepsilon_{i,n}$'s are identically distributed with finite fourth moments. Hence there exists some finite constant c_ε such that for all indices j,k,l and all $n \geq 1$: $E\left|\varepsilon_{j,n}\varepsilon_{k,n}\varepsilon_{l,n}\right| \leq c_\varepsilon$. Applying the triangle inequality yields

$$
\begin{aligned}
E|f_{i,n}|^3 &\leq \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} |d_{ij,n}||d_{ik,n}||d_{il,n}|\, E\left|\varepsilon_{j,n}\varepsilon_{k,n}\varepsilon_{l,n}\right| \\
&\leq c_\varepsilon \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} |d_{ij,n}||d_{ik,n}||d_{il,n}| \leq c_d^3 c_\varepsilon;
\end{aligned}
$$

observing $\sum_{j=1}^{n}|d_{ij,n}| \leq c_d$, which completes the proof. ■

.

Proof of Theorem 2: Recall from (4) that the disturbance process for the j-th equation is defined as $u_{j,n} = \rho_j W_n u_{j,n} + \varepsilon_{j,n}$. To prove the theorem we verify that all of the conditions assumed by Kelejian and Prucha (1999), i.e., their Assumptions 1-5, are satisfied here - with ρ_j, $u_{j,n}$, $\varepsilon_{j,n}$ and W_n corresponding to ρ, u_n, ε_n and M_n in the earlier paper. Assumptions 1-3 and 5 in Kelejian and Prucha (1999) are readily seen to hold by comparing them with Assumptions 1-3, and 7 maintained here. Assumption 4 in Kelejian and Prucha (1999) is satisfied in light of Theorem 1 above. Theorem 2 now follows as a direct consequence of Theorem 1 in Kelejian and Prucha (1999). ■

Proof of Theorem 3: Observe that substitution of (16), with ρ_j replaced by $\tilde{\rho}_{j,n}$, into (18) yields

$$
\hat{\delta}^F_{j,n} = \delta_j + \left[\hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' Z^*_{j,n}(\tilde{\rho}_{j,n})\right]^{-1} \hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' u^*_{j,n}(\tilde{\rho}_{j,n})
$$

where

$$
u^*_{j,n}(\tilde{\rho}_{j,n}) = y^*_{j,n}(\tilde{\rho}_{j,n}) - Z^*_{j,n}(\tilde{\rho}_{j,n})\delta_j = \varepsilon_{j,n} - (\tilde{\rho}_{j,n} - \rho_j) W_n (I_n - \rho_j W_n)^{-1}\varepsilon_{j,n}.
$$

Consequently

$$
\begin{aligned}
n^{1/2}(\hat{\delta}^F_{j,n} - \delta_j) = {} & \left[n^{-1}\hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})' Z^*_{j,n}(\tilde{\rho}_{j,n})\right]^{-1} \\
& \left[n^{-1/2}\hat{Z}^*_{j,n}(\tilde{\rho}_{j,n})'\varepsilon_{j,n} + \Delta_{j,n}\right].
\end{aligned}
\tag{A.7}
$$

where

$$\Delta_{j,n} = -(\tilde\rho_{j,n} - \rho_j) n^{-1/2} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' W_n (I_n - \rho_j W_n)^{-1} \varepsilon_{j,n}.$$

Clearly $\hat{Z}^*_{j,n}(\tilde\rho_{j,n})' Z^*_{j,n}(\tilde\rho_{j,n}) = \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' \hat{Z}^*_{j,n}(\tilde\rho_{j,n})$. To prove the theorem we proceed to establish the following results:

$$\begin{aligned} p\lim_{n\to\infty} n^{-1} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' Z^*_{j,n}(\tilde\rho_{j,n}) &= \overline{Q}_{jj}; \\ n^{-1/2} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' \varepsilon_{j,n} &\xrightarrow{d} N(0, \sigma_{jj} \overline{Q}_{jj}); \qquad \text{(A.8)} \\ n^{-1/2} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' W_n (I_n - \rho_j W_n)^{-1} \varepsilon_{j,n} &= O_p(1); \end{aligned}$$

where

$$\overline{Q}_{jj} = \left[Q_{HZ_j} - \rho_j Q_{HWZ_j} \right]' Q_{HH}^{-1} \left[Q_{HZ_j} - \rho_j Q_{HWZ_j} \right]. \qquad \text{(A.9)}$$

The matrix $\overline{Q}_{jj}$ is finite and nonsingular in light of Assumption 6. Given (A.8), the claim concerning the limiting distribution of $n^{1/2}(\hat\delta^F_{j,n} - \delta_j)$ is then readily seen to hold, observing that $\tilde\rho_{j,n} - \rho_j = o_p(1)$.

The first line in (A.8) follows immediately from Lemma A1, Assumption 6, and the consistency of $\tilde\rho_{j,n}$, observing that

$$\begin{aligned} n^{-1} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' Z^*_{j,n}(\tilde\rho_{j,n}) &= (n^{-1} Z'_{j,n} H_n - \tilde\rho_{j,n} n^{-1} Z'_{j,n} W'_n H_n) \qquad \text{(A.10)} \\ &\quad (n^{-1} H'_n H_n)^{-1} (n^{-1} H'_n Z_{j,n} - \tilde\rho_{j,n} n^{-1} H'_n W_n Z_{j,n}). \end{aligned}$$

Next observe that

$$\begin{aligned} n^{-1/2} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' \varepsilon_{j,n} &= (n^{-1} Z'_{j,n} H_n - \tilde\rho_{j,n} n^{-1} Z'_{j,n} W'_n H_n) \qquad \text{(A.11)} \\ &\quad (n^{-1} H'_n H_n)^{-1} (n^{-1/2} H'_n \varepsilon_{j,n}). \end{aligned}$$

In light of Assumption 5 the elements of $\varepsilon_{j,n}$ are i.i.d. with zero mean and finite variance σ_{jj}. Given Assumption 6 concerning the instruments H_n it then follows from Theorem A in Kelejian and Prucha (1999) that $n^{-1/2} H'_n \varepsilon_{j,n} \xrightarrow{d} N(0, \sigma_{jj} Q_{HH})$. The second line in (A.8) is now readily seen to hold, utilizing Lemma A1, Assumption 6, and the consistency of $\tilde\rho_{j,n}$.

Now observe that

$$\begin{aligned} n^{-1/2} \hat{Z}^*_{j,n}(\tilde\rho_{j,n})' W_n (I_n - \rho_j W_n)^{-1} \varepsilon_{j,n} &= (n^{-1} Z'_{j,n} H_n - \tilde\rho_{j,n} n^{-1} Z'_{j,n} W'_n H_n) \\ &\quad (n^{-1} H'_n H_n)^{-1} (n^{-1/2} F^{*\prime}_{j,n} \varepsilon_{j,n}) \end{aligned}$$

with $F_{j,n}^{*\prime} = H_n' W_n (I_n - \rho_j W_n)^{-1}$. Given Assumptions 3 and 6 it follows from part (i) and (ii) of Remark A1 that the elements of $F_{j,n}^{*\prime}$ are uniformly bounded in absolute value. As remarked, the elements of $\varepsilon_{j,n}$ are i.i.d. by Assumption 5. It hence follows from part (iii) of Remark A1 that $n^{-1/2} F_{j,n}^{*\prime} \varepsilon_{j,n} = O_p(1)$. The third line in (A.8) is now again readily seen to hold, utilizing Lemma A1, Assumption 6, and the consistency of $\tilde{\rho}_{j,n}$.

We note that in the above arguments we have only utilized the consistency of $\tilde{\rho}_{j,n}$. The expressions on the l.h.s. of (A.8) differ from the analogous expressions obtained by replacing $\tilde{\rho}_{j,n}$ by ρ_j only by terms of $o_p(1)$. Thus it is furthermore readily seen that $n^{1/2}\left(\hat{\delta}^F_{j,n} - \hat{\delta}_{j,n}\right) \xrightarrow{p} 0$ as $n \to \infty$. ■

We shall make use of the following lemma.

Lemma A3 Let A_n be some matrix whose row and column sums are bounded uniformly in absolute value. Then, given Assumptions 1-5 hold:

$$
\begin{aligned}
n^{-1}\varepsilon_{j,n}' A_n \varepsilon_{l,n} &= O_p(1); \\
n^{-1} Z_{j,n}' A_n \varepsilon_{l,n} &= O_p(1); \\
n^{-1} Z_{j,n}' A_n Z_{l,n} &= O_p(1);
\end{aligned}
\tag{A.12}
$$

for all $j = 1, \ldots, m$ and $l = 1, \ldots, m$.

Proof: Consider the expression for y_n in (A.5) and (A.6). Let $d_{r,n}$ denote the r-th subvector of d_n of dimension $n \times 1$, and let $D_{rs,n}$ denote the (r,s)-th submatrix of D_n of dimension $n \times n$, then

$$
y_{r,n} = d_{r,n} + \sum_{s=1}^{m} D_{rs,n} v_{s,n}. \tag{A.13}
$$

As remarked after (A.5) and (A.6), the elements of d_y, and hence the elements of $d_{r,n}$, are bounded uniformly in absolute value; furthermore, the row and column sums of D_y, and hence those of $D_{rs,n}$, are bounded uniformly in absolute value. Also observe from Assumption 5 that

$$
\varepsilon_{l,n} = \sum_{s=1}^{m} \sigma^*_{sl} v_{s,n}. \tag{A.14}
$$

By definition $Z_{j,n} = (Y_{j,n}, X_{j,n}, \overline{Y}_{j,n})$ and $\overline{Y}_n = W_n Y_n$. Upon substitution of expression (A.13) for the columns of Y_n and expression (A.14) for $\varepsilon_{l,n}$ into (A.12) we see that the elements of each term in (A.12) can be expressed as a finite sum of three basic types of expressions. Those expressions are of the form, $n^{-1}a_n$, $n^{-1}b_n'v_{s,n}$ or $n^{-1}v_{r,n}'C_nv_{s,n}$, where the a_n's are nonstochastic scalars, the b_n's are nonstochastic $n \times 1$ vectors and the C_n's are nonstochastic $n \times n$ matrices. Given Assumptions 1-5, the implied properties of d_n and D_n, and the assumption maintained for A_n it follows from Remark A1 that the 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 $n^{-1}b_n'v_{s,n}$ and $n^{-1}v_{r,n}'C_nv_{s,n}$ the elements of 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. Since the elements of v_n are i.i.d. with finite 4-th moments it follows furthermore from Remark A1 that $n^{-1}b_n'v_{s,n} = O_p(1)$ and $n^{-1}v_{r,n}'C_nv_{s,n} = O_p(1)$. Observing that finite sums of random variables of the order $O_p(1)$ are again $O_p(1)$ completes the proof. ■

Proof of Lemma 1: To prove the lemma observe that for $j = 1, \ldots, m$:

$$
\begin{aligned}
\tilde{u}_{j,n} &= y^*_{j,n}(\tilde\rho_{j,n}) - Z^*_{j,n}(\tilde\rho_{j,n})\hat{\delta}^F_{j,n} && \text{(A.15)}\\
&= y^*_{j,n}(\tilde\rho_{j,n}) - Z^*_{j,n}(\tilde\rho_{j,n})\delta_j - Z^*_{j,n}(\tilde\rho_{j,n})(\hat{\delta}^F_{j,n} - \delta_j)\\
&= \varepsilon_{j,n} - (\tilde\rho_{j,n} - \rho_j)W_n(I_n - \rho_j W_n)^{-1}\varepsilon_{j,n} - Z^*_{j,n}(\tilde\rho_{j,n})(\hat{\delta}^F_{j,n} - \delta_j).
\end{aligned}
$$

Consequently for $i, j = 1, \ldots, m$:

$$
\begin{aligned}
\tilde\sigma_{jl,n} = {} & n^{-1}\tilde{u}'_{j,n}\tilde{u}_{l,n} = n^{-1}\varepsilon'_{j,n}\varepsilon_{l,n} && \text{(A.16)}\\
& - (\tilde\rho_{j,n} - \rho_j)[n^{-1}\varepsilon'_{j,n}(I_n - \rho_j W_n')^{-1}W_n'\varepsilon_{l,n}]\\
& - (\hat{\delta}^F_{j,n} - \delta_j)'[n^{-1}Z^*_{j,n}(\tilde\rho_{j,n})'\varepsilon_{l,n}]\\
& - (\tilde\rho_{l,n} - \rho_l)[n^{-1}\varepsilon'_{j,n}W_n(I_n - \rho_l W_n)^{-1}\varepsilon_{l,n}]\\
& + (\tilde\rho_{j,n} - \rho_j)(\tilde\rho_{l,n} - \rho_l)[n^{-1}\varepsilon'_{j,n}(I_n - \rho_j W_n')^{-1}W_n'W_n(I_n - \rho_l W_n)^{-1}\varepsilon_{l,n}]\\
& + (\tilde\rho_{l,n} - \rho_l)(\hat{\delta}^F_{j,n} - \delta_j)'[n^{-1}Z^*_{j,n}(\tilde\rho_{j,n})'W_n(I_n - \rho_l W_n)^{-1}\varepsilon_{l,n}]\\
& - [n^{-1}\varepsilon'_{j,n}Z^*_{l,n}(\tilde\rho_{l,n})](\hat{\delta}^F_{l,n} - \delta_l)\\
& + (\tilde\rho_{j,n} - \rho_j)[n^{-1}\varepsilon'_{j,n}(I_n - \rho_j W_n')^{-1}W_n'Z^*_{l,n}(\tilde\rho_{l,n})](\hat{\delta}^F_{l,n} - \delta_l)\\
& + (\hat{\delta}^F_{j,n} - \delta_j)'[n^{-1}Z^*_{j,n}(\tilde\rho_{j,n})'Z^*_{l,n}(\tilde\rho_{l,n})](\hat{\delta}^F_{l,n} - \delta_l).
\end{aligned}
$$

Consider the first term on the r.h.s. of (A.16), i.e., $n^{-1}\varepsilon'_{j,n}\varepsilon_{l,n} = n^{-1}\sum_{i=1}^{n}\varepsilon_{ij,n}\varepsilon_{il,n}$. Assumption 5 implies that the products $\varepsilon_{ij,n}\varepsilon_{il,n}$, $i = 1,\dots,n$, are i.i.d. Kolmogorov's law of large numbers - see, e.g., Pötscher and Prucha (2001), p.217 - then implies that $\operatorname{plim}_{n\to\infty} n^{-1}\varepsilon'_{j,n}\varepsilon_{l,n} = \sigma_{jl}$. To prove the lemma we next show that all other terms on the r.h.s. of (A.16) are $o_p(1)$. Since $\tilde{\rho}_{j,n} - \rho_j = o_p(1)$ by Theorem 2 and $\hat{\delta}^F_{j,n} - \delta_j = o_p(1)$ by Theorem 3 it suffices to show that each of the terms in square brackets on the r.h.s. of (A.16) is $O_p(1)$. Substitution of $Z^*_{j,n}(\tilde{\rho}_{j,n}) = Z_{j,n} - \tilde{\rho}_{j,n}W_nZ_{j,n}$ into those terms shows all of them are composed of expressions of the three types considered in (A.12) of Lemma A3 above, possibly multiplied by $\tilde{\rho}_{j,n}$ and $\tilde{\rho}_{l,n}$. Given Assumption 3 it follows immediately from Remark A1 that the row and column sums of all matrices A_n appearing in those expressions are uniformly bounded in absolute value. Hence by Lemma A3 the terms in square brackets on the r.h.s. of (A.16) are seen to be indeed $O_p(1)$, which complete the proof of the lemma. We note that the proof only used the consistency of $\tilde{\rho}_{j,n}$ and $\hat{\delta}^F_{j,n}$, and not any other feature of those estimators. ∎

Proof of Theorem 4: Analogous to the proof of Theorem 3 observe that substitution of (21), with ρ_j replaced by $\tilde{\rho}_{j,n}$, into (23) yields

$$\hat{\delta}^F_n = \delta + \left[\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}^{-1}_n \otimes I_n)Z^*_n(\tilde{\rho}_n)\right]^{-1}\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}^{-1}_n \otimes I_n)u^*_n(\tilde{\rho}_n)$$

where

$$\begin{aligned} u^*_n(\tilde{\rho}_n) &= [u^*_{1,n}(\tilde{\rho}_{1,n})',\dots,u^*_{m,n}(\tilde{\rho}_{m,n})'] \\ &= \varepsilon_n - \left[\operatorname{diag}_{j=1}^{m}(\tilde{\rho}_{j,n} - \rho_j)W_n(I_n - \rho_jW_n)^{-1}\right]\varepsilon_n. \end{aligned}$$

Consequently

$$\begin{aligned} n^{1/2}(\hat{\delta}^F_n - \delta) = &\left[n^{-1}\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}^{-1}_n \otimes I_n)Z^*_n(\tilde{\rho}_n)\right]^{-1} \\ &\left[n^{-1/2}\hat{Z}^*_n(\tilde{\rho}_n)'(\hat{\Sigma}^{-1}_n \otimes I_n)\varepsilon_n + \Delta_n\right] \end{aligned} \tag{A.18}$$

where

$$
\begin{aligned}
\Delta_n &= -n^{-1/2}\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)\operatorname{diag}_{j=1}^m\left[(\tilde{\rho}_{j,n}-\rho_j)W_n(I_n-\rho_j W_n)^{-1}\right]\varepsilon_n \\
&= -n^{-1/2}\begin{bmatrix} \sum_{l=1}^m(\tilde{\rho}_{l,n}-\rho_l)\hat{\sigma}_n^{1l}\hat{Z}_{1,n}^*(\tilde{\rho}_{1,n})'W_n(I_n-\rho_l W_n)^{-1}\varepsilon_{l,n} \\ \vdots \\ \sum_{l=1}^m(\tilde{\rho}_{l,n}-\rho_l)\hat{\sigma}_n^{ml}\hat{Z}_{m,n}^*(\tilde{\rho}_{m,n})'W_n(I_n-\rho_l W_n)^{-1}\varepsilon_{l,n} \end{bmatrix}.
\end{aligned}
$$

Clearly $\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)Z_n^*(\tilde{\rho}_n) = \hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)\hat{Z}_n^*(\tilde{\rho}_n)$. To prove the theorem we demonstrate in the following that

$$
\begin{aligned}
&\operatorname*{p\,lim}_{n\to\infty} n^{-1}\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)Z_n^*(\tilde{\rho}_n) = \overline{Q}, \\
&n^{-1/2}\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)\varepsilon_n \xrightarrow{d} N(0,\overline{Q}), \qquad \text{(A.19)} \\
&n^{-1/2}\hat{Z}_{j,n}^*(\tilde{\rho}_{j,n})'W_n(I_n-\rho_l W_n)^{-1}\varepsilon_{l,n} = O_p(1), \quad j,l = 1,\ldots,m,
\end{aligned}
$$

where

$$
\overline{Q} = \operatorname{diag}_{j=1}^m(Q_{HZ_j}-\rho_j Q_{HWZ_j})'(\Sigma^{-1}\otimes Q_{HH}^{-1})\operatorname{diag}_{l=1}^m(Q_{HZ_l}-\rho_l Q_{HWZ_l}). \tag{A.20}
$$

The matrix $\overline{Q}$ is finite and nonsingular in light of Assumption 6. Given (A.19), the claim concerning the limiting distribution of $n^{1/2}(\hat{\delta}^F_{j,n}-\delta_j)$ is then readily seen to hold, observing that by Theorem 2 and Lemma 1, $\tilde{\rho}_{j,n}-\rho_j = o_p(1)$ and $\hat{\Sigma}_n - \Sigma = o_p(1)$, and thus $\Delta_n = o_p(1)$ provided the third line in (A.19) holds indeed.

The (j,l)-th block of the matrix on the l.h.s. of the first line of (A.19) is given by $\hat{\sigma}_n^{jl}n^{-1}\hat{Z}_{j,n}^*(\tilde{\rho}_{j,n})'Z_{l,n}^*(\tilde{\rho}_{l,n})$. Since Σ is nonsingular we have also $\hat{\Sigma}_n^{-1}-\Sigma^{-1} = o_p(1)$ and hence $\operatorname{p\,lim}_{n\to\infty}\hat{\sigma}_n^{jl} = \sigma^{jl}$. Furthermore, by arguments analogous to those used to prove the first line of (A.8) we have

$$
\operatorname*{p\,lim}_{n\to\infty} n^{-1}\hat{Z}_{j,n}^*(\tilde{\rho}_{j,n})'Z_{l,n}^*(\tilde{\rho}_{l,n}) = \left[Q_{HZ_j}-\rho_j Q_{HWZ_j}\right]'Q_{HH}^{-1}\left[Q_{HZ_l}-\rho_l Q_{HWZ_l}\right]
$$

for $j,l = 1,\ldots,m$. From this the first line in (A.19) is now readily seen to hold.

Next observe that utilizing Assumption 5

$$
\begin{aligned}
&n^{-1/2}\hat{Z}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1}\otimes I_n)\varepsilon_n \\
= \;&\operatorname{diag}_{j=1}^m(n^{-1}Z_{j,n}'H_n-\tilde{\rho}_{j,n}n^{-1}Z_{j,n}'W_n'H_n) \\
&\left[\hat{\Sigma}_n^{-1}\Sigma_*'\otimes(n^{-1}H_n'H_n)^{-1}\right]n^{-1/2}(I_m\otimes H_n')v_n.
\end{aligned}
$$

By Assumption 5 the elements of v_n are i.i.d. with zero mean and variance one. Given Assumption 6 concerning the instruments H_n it then follows from Theorem A in Kelejian and Prucha (1999) that $n^{-1/2}(I_m \otimes H_n')v_n \xrightarrow{d} N(0, I_m \otimes Q_{HH})$. Observing again that $\tilde{\rho}_{j,n} - \rho_j = o_p(1)$, $\hat{\Sigma}_n^{-1} - \Sigma^{-1} = o_p(1)$ and $\Sigma = \Sigma_*'\Sigma_*$ the second line in (A.19) is now readily seen to hold from arguments analogous to those used to prove the second line of (A.8)

Furthermore, using arguments analogous to those used to prove the third line in (A.8) also shows that the third line in (A.19) holds.

We note that in the above arguments we have only utilized the consistency of $\tilde{\rho}_{j,n}$ and $\hat{\Sigma}_n$. The expressions on the l.h.s. of (A.19) differ from the analogous expressions obtained by replacing $\tilde{\rho}_{j,n}$ by ρ_j only by terms of $o_p(1)$. Thus it is furthermore readily seen that $n^{1/2}(\delta_n^F - \delta_n) \xrightarrow{P} 0$ as $n \to \infty$. ■

References

[1] Amemiya, T., 1985, Advanced Econometrics (Harvard University Press, Cambridge).

[2] Anselin, L., 1988, Spatial Econometrics: Methods and Models (Kluwer Academic Publishers, Boston).

[3] Anselin, L., A. Bera, R. Florax and M. Yoon, 1996, Simple diagnostic tests for spatial dependence, Regional Science and Urban Economics 26, 77-104.

[4] Audretsch, D. and M. Feldmann, 1996, R&D spillovers and the geography of innovation and production, American Economic Review 86, 630-640.

[5] Ausubel, L., P. Cramton, R. McAfee and J. McMillan, 1997, Synergies in wireless telephony: Evidence from broadcast PCS, Journal of Economics and Management Strategy 6, 497-527.

[6] Baltagi, B.H., 1995, Econometric Analysis of Panel Data Wiley, New York).

[7] Baltagi, B.H. and D. Li, 2001a, Double Length Arti...cial Regressions For Testing Spatial Dependence, Econometric Reviews 20, 31-40.

[8] Baltagi, B.H. and D. Li, 2001b, LM Tests for Functional Form and Spatial Error Correlation, International Regional Science Review 24, 194-225.

[9] Baltagi, B.H., S.H. Song and W. Koh, 2000, Testing Panel Data Regression Models with Spatial Error Correlation, Department of Economics, Texas A & M University, College Station, mimeo.

[10] Bell, K.P. and N.E. Bockstael, 2000, Applying the generalized-moments estimation approach to spatial problems involving micro level data, Review of Economics and Statistics 82, 72-82.

[11] Buettner, T, 1999, The effect of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data, Papers in Regional Science 78, 47-67.

[12] Case, A., 1991, Spatial patterns in household demand, Econometrica 59, 953-966.

[13] Case, A., J. Hines Jr. and H. Rosen, 1993, Budget spillovers and fiscal policy independence: Evidence from the states, Journal of Public Economics 52, 285-307.

[14] Cliff, A. and J. Ord., 1973, Spatial Autocorrelation (Pion, London).

[15] Cliff, A. and J. Ord., 1981, Spatial Processes,Models and Applications (Pion, London).

[16] Conley, T., 1999, GMM estimation with cross sectional dependence, Journal of Econometrics 92, 1-45.

[17] Cressie, N., 1993, Statistics of Spatial Data (Wiley, New York).

[18] Das, D., H.H. Kelejian and I.R. Prucha, 2001, Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances, Papers in Regional Science, forthcoming.

[19] De Long, J. and L. Summers, 1991, Equipment investment and economic growth, Quarterly Journal of Economics 106, 445-502.

[20] Driscoll, J. and A. Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, The Review of Economics and Statistics LXXX, 549-560.

[21] Dubin, R., 1988, Estimation of regression coefficients in the presence of spatially autocorrelated error terms, Review of Economics and Statistics 70, 466-474.

[22] Giacomini, R. and C.W.J. Granger, 2001, Aggregation of space-time processes, Department of Economics, University of California, San Diego, mimeo.

[23] Holtz-Eakin, D., 1994, Public sector capital and the productivity puzzle, Review of Economics and Statistics 76, 12-21.

[24] Kelejian, H.H. and W.E. Oates, 1981, Introduction to Econometrics, Principles and Applications (Harper and Row, New York).

[25] Kelejian, H.H. and I.R. Prucha, 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, Journal of Real Estate Finance and Economics 17, 99-121.

[26] Kelejian, H.H. and I.R. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, International Economic Review 40, 509-533.

[27] Kelejian, H.H. and I.R. Prucha, 2001a, On the asymptotic distribution of the Moran I test statistic with applications, Journal of Econometrics 104, 219-257.

[28] Kelejian, H.H. and I.R. Prucha, 2001b, Best series instrumental variable estimator for a spatial autoregressive model with autoregressive disturbances, Department of Economics, University of Maryland, College Park, mimeo.

[29] Kelejian, H.H. and I.R. Prucha, 2001c, 2SLS and OLS in a spatial autoregressive model with equal spatial weights, Regional Science and Urban Economics, forthcoming.

[30] Kelejian, H.H. and D. Robinson, 1995, Spatial correlation: A suggested alternative to the autoregressive model. In L. Anselin and R. Florax, eds., New Directions in Spatial Econometrics (Springer, New York), 75-95.

[31] Kelejian, H.H. and D. Robinson, 1997, Infrastructure productivity estimation and its underlying econometric speci…cations: A sensitivity analysis, Papers in Regional Science 76, 115-131.

[32] Krugman, P., 1991, Geography and Trade (MIT Press, Cambridge).

[33] Krugman, P., 1995, Development, Geography, and Economic Theory (MIT Press, Cambridge).

[34] Lee, L.-F., 1999a, Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, Department of Economics, Hong Kong University of Science and Technology, mimeo.

[35] Lee, L.-F., 1999b, Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models, Department of Economics, Hong Kong University of Science and Technology, mimeo.

[36] Lee, L.-F., 2001a, Generalized Method of Moments Estimation of Spatial Autoregressive Processes, Department of Economics, Ohio State University, mimeo.

[37] Lee, L.-F., 2001b, GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, Department of Economics, Ohio State University, mimeo.

[38] Lee, L.-F., 2002, Consistency and e¢ciency of least squares estimation for mixed regressive, spatial autoregessive models, Econometric Theory 18, 252-277.

[39] Pinkse, J., 1999, Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models, Department of Economics, University of British Columbia and University College London, mimeo.

[40] Pinkse, J., and M. Slade, 1998, Contracting in space: An application of spatial statistics to discrete-choice models, Journal of Econometrics 85, 125-154.

[41] Pötscher, B.M. and I.R. Prucha, 1997, Dynamic Nonlinear Econometric Models, Asymptotic Theory (Springer Verlag, New York).

[42] Pötscher, B.M. and I.R. Prucha, 2001, Basic elements of asymptotic theory. In B.H. Baltagi, ed., A Companion to Theoretical Econometrics (Blackwell, New York), 201-229.

[43] Prucha, I.R., 1985, Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components, International Economic Review 26, 491-506.

[44] Rey, S. and M. Boarnet, 1999, A taxonomy of spatial econometric models for simultaneous systems. In L. Anselin and R. Florax, eds., Advances in Spatial Econometrics (Springer Verlag, New York),

[45] Schmidt, P., 1976, Econometrics (Marcel Dekker,New York).

[46] Shroder, M., 1995, Games the states don't play: Welfare bene…ts and the theory of …scal federalism, Review of Economics and Statistics 77, 183-191.

[47] Whittle, P., 1954, On stationary processes in the plane, Biometrica 41, 434-449.