2007

Modeling and Estimation Issues in Spatial Simultaneous Equations Models

Gebremeskel H. Gebremariam

Modeling and Estimation Issues in Spatial Simultaneous Equations Models

Gebremeskel H. Gebremariam, Post Doctoral Fellow[1]

RESEARCH PAPER 2007-13

Paper presented at the 54th Annual North American Meetings of the Regional Science Association International (RSAI) in Savannah, Georgia, November 7-10, 2007

Abstract: Spatial dependence is one of the main problems in stochastic processes and can be caused by a variety of measurement problems that are associated with the arbitrary delineation of spatial units of observation (such as counties boundaries, census tracts), problems of spatial aggregation, and the presence of spatial externalities and spillover effects. The existence of spatial dependence would then mean that the observations contain less information than if there had been spatial independence. Consequently, hypothesis tests and the statistical properties for estimators in the standard econometric approach will not hold. Thus, in order to obtain approximately the same information as in the case of spatial independence, the spatial dependence needs to be explicitly quantified and modeled. Although advances in spatial econometrics provide researchers with new avenues to address regression problems that are associated with the existence of spatial dependence in regional data sets, most of the applications have, however, been in single-equation frame-works. Yet, for many economic problems there are both multiple endogenous variables and data on observations that interact across space. Therefore, researchers have been in the undesirable position of having to choose between modeling spatial interactions in a single equation frame-work, or using multiple equations but losing the advantage of a spatial econometric approach. In an attempt to address this undesirable position, this research work deals with the modeling and estimation issues in spatial simultaneous equations models. The first part discusses modeling issues in multi-equation *Spatial Lag*, *Spatial Error*, and *Spatial Autoregressive* Models in both cross sectional and panel data sets. Whereas, the second part deals with estimation issues in spatial simultaneous equations models in both cross sectional and panel data sets. Finally, issues related specification tests in spatial simultaneous equations models are discussed.

1. *A joint appointment by the Regional Research Institute and Division of Resource Management, West Virginia University, Morgantown, WV 26506*

Modeling and Estimation Issues in Spatial Simultaneous Equations Models

1. Introduction

The investigation of regression residuals in the search of signs of a spatial structure is the first step in the analysis of spatial data. The usual graphical analysis tools and the residual mapping can provide the first indication that the observed values are more correlated than would be expected under random assignment. In this case, the presence of spatial clustering can be tested by using the spatial correlation tests such as Moran, Geary or G statistic, on the residuals. Although such tests can detect the presence of spatial clustering, however, they do not explain why such clustering occurs, nor do they explain which factors determine its shape and strength. In other words, the alternative hypothesis of spatial autocorrelation is too vague to be useful in the construction of theory. Rather, spatial autoregressive process, the process that expresses how observations at each location depend on values at neighboring locations –the spatial lag, is the relevant concept that formalizes the way in which the spatial association is generated (Anselin, 1992).

Spatial dependence is one of the main problems in stochastic processes and can be caused by a variety of measurement problems that are associated with the arbitrary delineation of spatial units of observation (such as counties boundaries, census tracts), problems of spatial aggregation, and the presence of spatial externalities and spillover effects. The existence of spatial dependence would then mean that the observations contain less information than if there had been spatial independence. Consequently, hypothesis tests and the statistical properties for estimators in the standard econometric approach will not hold. Thus, in order to obtain approximately the same information as in the case of spatial independence, the spatial dependence needs to be explicitly quantified and modeled.

2. Spatial Regression Models

The explicit inclusion of spatial dependence in regression models can be done in different ways. Anselin (2003), for example, attempts to extend the earlier work on spatial dependence and he notice that "the standard taxonomy of spatial autoregressive lag and error models commonly applied in spatial econometrics (Anselin, 1988) is perhaps too simplistic and leaves out other interesting possibilities for mechanisms through which phenomena or actions at a given location affects actors or properties at other locations". In this extension, he makes a distinction between global and local range of dependence which have implications for the econometric specifications of spatially lagged dependent variable (**Wy**), spatially lagged explanatory variables (**WX**) and spatially lagged error terms (**Wu**). The distinction between the global and local effects models depends upon the assumption on the underlying spatial process. The spatial regression models with global effects are based on the principle that the underlying spatial process on the analyzed data is stationary. This means that the spatial autocorrelation patterns of the data sets can be captured in one parameter only. The spatial regression models with local spatial effects, however, are based on the principle that the underlying spatial process on the analyzed data is non-stationary and hence spatial autocorrelation patterns of the data cannot be captured by one parameter only. Thus, when the spatial process is non-stationary, the coefficients of regression need to reflect the spatial heterogeneity.

The conditional autoregressive (CAR) model and the simultaneous models, spatial autoregressive (SAR) and the spatial moving average (SMA), models are the most commonly employed types of models in spatial statistics and spatial econometrics. In the conditional model, a random variable at a location is conditioned on the observations on that random variable at neighboring locations. The latter are treated as exogenous and can be exploited to construct optimal prediction for the random variable at unobserved locations (Anselin, 2003). The inverse covariance matrix for this model is constructed by $(\mathbf{I}-\rho\mathbf{D})/\sigma^{2}$ where $\mathbf{D}$ is a binary spatial weights matrix. This type of model is appropriate for studies involving first-order dependency which are most common in spatial statistics. In the simultaneous models (SAR and SMA models), however, the focus is on the explanation of the complete interactions between all observations or locations observed simultaneously. The covariance structure in such models is compatible with the spatial ordering and the inverse covariance matrix is constructed by $(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})/\sigma^{2}$ where $\mathbf{W}$ is a row-standardized spatial weights matrix. The simultaneity in these models follows from the nature of dependence in space which is two-directional. As a result, each location is in turn a neighbor for its neighbors, so that the effect of the neighbors has to be treated as endogenous (Anselin, 2003). The SAR model is appropriate for studies involving first-order as well as second-order dependency which are most common in regional studies. This research also follows the regional studies tradition. In the next subsection, the spatial dependence (spatial global effects) will be discussed, first in the context of cross-sectional setting and then the extension to panel data will follow.

2.1 Spatial Dependence in Cross-Sectional Models

There are two distinct ways of incorporating spatial dependence into the standard linear regression models: as an additional explanatory variable in the form of spatially lagged dependent variables ($\mathbf{Wy}_j$) (spatial lag), or in the error structure $E\left[\mathbf{u}_i\mathbf{u}_j\right] \neq \mathbf{0}$ (spatial error) (Anselin, 2001).

Spatial Lag Model

The spatial lag model combines the spatial dependence in the form of a spatial lag term with the usual linear explanation of a dependent variable by a set of explanatory variables. It is similar to the inclusion of a serially autoregressive term for the dependent variable in a time series context (Anselin and Bera, 1998; LeSage, 1999). This model is more appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction. Anselin (1993) referred this model as the spatial autoregressive model with substantive spatial dependence.

Formally, a spatial lag model, in the context of single equation and in a cross-sectional setting, is expressed as:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u} \tag{1.1}$$

where $\mathbf{y}$ is an n by 1 vector of observations on the dependent variable, $\mathbf{Wy}$ is the corresponding spatial lagged dependent variable for weights matrix $\mathbf{W}$, $\mathbf{X}$ is n by K matrix of observations on the explanatory variables, $\mathbf{u}$ is an n by 1 vector of error terms, ρ is the spatial autoregressive parameter and $\boldsymbol{\gamma}$ is a K by 1 vector of regression coefficients. The parameter ρ measures the degree of spatial dependence inherent in the data. As this model combines the standard regression model with a spatially lagged dependent variable, it is also called a mixed regressive-spatial autoregressive model (Anselin, 1998).

The spatial single equation model in equation (1.1) can be extended to a system of spatially interrelated cross sectional equations corresponding to n cross sectional units. But, first note that a standard G system of equations can be written as:

(1.2) $$\mathbf{Y}=\mathbf{Y}\mathbf{B}+\mathbf{X}\mathbf{\Gamma}+\mathbf{U}$$

with

$$\mathbf{Y} = \left(\mathbf{y}_1,\ldots,\mathbf{y}_G\right) \qquad \mathbf{X} = \left(\mathbf{x}_1,\ldots,\mathbf{x}_K\right) \qquad \mathbf{U} = \left(\mathbf{u}_1,\ldots,\mathbf{u}_G\right)$$

where $\mathbf{y}_j$ is the n by 1 vector of cross sectional observations on the dependent variable in jth equation, $\mathbf{x}_l$ is an n by 1 vector of cross sectional observations on the lth exogenous variable, $\mathbf{u}_j$ is an n by 1 vector of error terms in the jth equation, and $\mathbf{B}$ and $\mathbf{\Gamma}$ are correspondingly defined parameter matrices of dimension G by G and K by G, respectively. **B** is a diagonal matrix. Following Kelejian and Prucha (2004), the spatial lag dependent variables can be incorporated in equation (1.2) as follows:

(1.3) $$\mathbf{Y}=\mathbf{Y}\mathbf{B}+\mathbf{X}\mathbf{\Gamma}+\mathbf{W}\mathbf{Y}\mathbf{\Lambda}+\mathbf{U}$$

where **W** is an n by n weights matrix of known constants, and $\mathbf{\Lambda}$ is a G by G matrix of parameters. Note that the spatial global spillover effects in the endogenous variables is modeled via **WY,** with $\mathbf{W}\mathbf{y}_j$ representing the spatial lag in the jth equation for j = 1, … , G. The ith element of the vector of the spatial lag, $\mathbf{W}\mathbf{y}_j$, can be computed as:

(1.4) $$\mathbf{W}y_{ij} = \sum_{r=1}^{n} w_{ir} y_{rj}$$

where

$$w_{ir} = \begin{cases} \text{1 when i and r are neighbors (adjacent)} \\ \text{0 otherwise} \end{cases}.$$

Note that the spatial interactions in the system are determined by the nature of the $\mathbf{\Lambda}$ matrix. Specifying $\mathbf{\Lambda}$ as not a diagonal matrix of parameters allows the jth endogenous variable to depend on its spatial lag as well as on the spatial lags of the other endogenous variables in the model. If, however, there is a theoretical reason to believe that the jth endogenous variable depends either only on the spatial lags in the other endogenous variables in the model or only on its own spatial lag, then $\mathbf{\Lambda}$ should be specified as a diagonal matrix or as an identity matrix, respectively.

The system in equation (1.3) can be expressed in a form where its solution for the endogenous variables is clearly revealed. First, consider the following vector transformations:

$$
\begin{aligned}
vec(\mathbf{Y}) &= vec(\mathbf{YB}) + vec(\mathbf{X\Gamma}) + vec(\mathbf{WY\Lambda}) + vec(\mathbf{U}) \\
&= (\mathbf{B}'\otimes\mathbf{I})\, vec(\mathbf{Y}) + (\mathbf{\Gamma}'\otimes\mathbf{I})\, vec(\mathbf{X}) + (\mathbf{\Lambda}'\otimes\mathbf{W})\, vec(\mathbf{Y}) + \text{vec}(\mathbf{U})
\end{aligned}
\quad (1.5)
$$

Letting $\mathbf{y} = vec(\mathbf{Y})$, $\mathbf{x} = \text{vec}(\mathbf{X})$, and $\mathbf{u} = \text{vec}(\mathbf{U})$, it follows from equation (1.5) that

$$
\begin{aligned}
\mathbf{y} &= (\mathbf{B}'\otimes\mathbf{I})\mathbf{y} + (\mathbf{\Gamma}'\otimes\mathbf{I})\mathbf{x} + (\mathbf{\Lambda}'\otimes\mathbf{W})\mathbf{y} + \mathbf{u} \\
&= \left[(\mathbf{B}'\otimes\mathbf{I}) + (\mathbf{\Lambda}'\otimes\mathbf{W})\right]\mathbf{y} + (\mathbf{\Gamma}'\otimes\mathbf{I})\mathbf{x} + \mathbf{u}
\end{aligned}
\quad (1.6)
$$

The mixed regressive-spatial autoregressive specification given above can be interpreted in three different ways. First, in the specification given in equation (1.6), the interest is in finding out how each of variables in **y** relate to their values in the surrounding locations (spatial own lags), the values of the other endogenous variables in the surrounding locations (cross spatial lags) and the values of the other endogenous variables in the respective location, while controlling for the influence of other predetermined (exogenous) variables. The second perspective is when the interest is to detect the relations between the dependent variables y and the predetermined (exogenous) variables x, after all the spatial effects and the other endogenous variables effects are controlled for or filtered out. Formally, this can be expressed as:

$$
\begin{aligned}
\mathbf{y} - \left[(\mathbf{B}'\otimes\mathbf{I}) + (\mathbf{\Lambda}'\otimes\mathbf{W})\right]\mathbf{y} &= (\mathbf{\Gamma}'\otimes\mathbf{I})\mathbf{x} + \mathbf{u} \\
\left(\text{I-}\left[(\mathbf{B}'\otimes\mathbf{I}) + (\mathbf{\Lambda}'\otimes\mathbf{W})\right]\right)\mathbf{y} &= (\mathbf{\Gamma}'\otimes\mathbf{I})\mathbf{x} + \mathbf{u}
\end{aligned}
\quad (1.7)
$$

The third perspective is the interpretation of the model in its reduced form. The reduced form is nonlinear and it clearly illustrates how the expected value of the dependent variables at each location depend not only on the predetermined (exogenous) variables at the respective locations but also on the predetermined (exogenous) variables at all other locations. The reduced form is given by:

(1.8) $$\mathbf{y} = \text{ I-}\left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right]^{-1} \boldsymbol{\Gamma}'\otimes\mathbf{I} \;\mathbf{x} + \text{ I-}\left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right]^{-1}\mathbf{u}.$$

The expected or mean value can be computed by taking expectations on both sides of equation (1.8) as follows:

(1.9) $$\begin{aligned} E\;\mathbf{y} &= E\left[\text{ I-}\left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right]^{-1} \boldsymbol{\Gamma}'\otimes\mathbf{I}\;\mathbf{x}\right] \\ &+ E\left[\text{ I-}\left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right]^{-1}\mathbf{u}\right] \end{aligned}.$$

Since the mean of the error term is assumed to be zero, this gives:

(1.10) $$E\;\mathbf{y} = E\left[\text{ I-}\left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right]^{-1} \boldsymbol{\Gamma}'\otimes\mathbf{I}\;\mathbf{x}\right].$$

To continue with formulating the model in form more convenient to reveal its solution for the endogenous variables, consider equation (1.6). Let:

$$\mathbf{B}^{*} = \left[\mathbf{B}'\otimes\mathbf{I} + \boldsymbol{\Lambda}'\otimes\mathbf{W} \right] \; \textit{and} \; \boldsymbol{\Gamma}^{*} = \boldsymbol{\Gamma}'\otimes\mathbf{I} \;.$$

Then, equation (1.6) can be rewritten as:

(1.11) $$\begin{aligned} &\mathbf{y} = \mathbf{B}^{*}\mathbf{y} + \boldsymbol{\Gamma}^{*}\mathbf{x} + \mathbf{u} \\ &\mathbf{y} = \mathbf{I}_n - \mathbf{B}^{*}\;{}^{-1}\boldsymbol{\Gamma}^{*}\mathbf{x} + \mathbf{I}_n - \mathbf{B}^{*}\;{}^{-1}\mathbf{u} \quad \text{in reduced-form} \end{aligned}.$$

From this general form of the spatial econometric model, various specifications can be generated. By imposing zero restrictions on various model parameters, Rey and Boarnet (2004), for example, have identified 35 different specification cases from their two-equation spatial econometric model. In order to structure the taxonomy of the spatial econometric model, they considered three dimensions of simultaneity: feedback simultaneity; spatial autoregressive lag simultaneity; and spatial cross-regressive lag simultaneity. Depending on the underlying theoretical arguments, each equation of the model contains either all or some or none of these dimensions.

The system in equation (1.3) can also be expressed more compactly by imposing exclusion restriction on the parameters of the model. Particularly, let the vectors of nonzero elements of the jth column of $\mathbf{B}, \mathbf{\Gamma}$, and $\mathbf{\Lambda}$ be $\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}_j$, and $\boldsymbol{\lambda}_j$ respectively. Again, let the corresponding matrices of observations on the endogenous variables, exogenous variables, and the spatially lagged endogenous variables that appear in the jth equation be $\mathbf{Y}_j$, $\mathbf{X}_j$, and $\mathbf{WY}_j$ respectively. Then equation (1.3) can be written as:

(1.12)
$$\mathbf{y}_j = \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j$$

where

$$\mathbf{Z}_j = \left(\mathbf{Y}_j, \mathbf{X}_j, \mathbf{WY}_j \right) \text{ and } \boldsymbol{\delta}_j = \left(\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j, \boldsymbol{\lambda}'_j \right)'.$$

Spatial Error Models

A second way to incorporate spatial autocorrelation in a regression model is to specify a spatial process for the disturbance term. The disturbance terms in a regression model can be considered to contain all ignored elements, and when spatial dependence is present in the disturbance term, the spatial effects are assumed to be a noise, or perturbation, that is, a factor that needs to be removed (Anselin, 2001). Such spatial pattern in the residuals of the regression model may lead to the discovery of additional variables that should be included in the model. Spatial dependence in the disturbance term also violates the basic OLS estimation assumption of uncorrelated errors. Hence, when the spatial dependence is ignored, OLS estimates will be inefficient, though unbiased, the student t- and F-statistics for tests of significance will be biased, the R^2 measure will be misleading, which in turn lead to a wrong statistical interpretation of the regression mode (Anselin, 1996). More efficient estimators can be obtained by taking advantage of the particular structure of the error covariance implied by the spatial process. The disturbance term is non-spherical where the off-diagonal elements of the associated covariance matrix express the structure of spatial dependence.

A spatial dependence model is more common in social science applications using cross sectional data due to the predominance of spatial interaction and spatial externalities as well as due to the poor choice of spatial units in such applications (Anselin, 1992). The dependence in the disturbance term can be expressed either as spatial autoregressive or as a spatial moving average spatial process. The most common specification, however, is the spatial autoregressive spatial process, although most tests for spatial error autocorrelation are the same for either form (Anselin, 1992). The spatial dependence in the disturbance term, thus, can be expressed using matrix notation as:

$$\text{(1.13)} \qquad \mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}$$

with

$$\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$$

where $\mathbf{u}$ is assumed to follow a spatial autoregressive process, with ρ as the spatial autoregressive coefficient for the error lag **Wu**, and $\boldsymbol{\varepsilon}$ n by 1 vector of innovations or white noise error, and the other notations as defined before. Equation (1.13) is the structural form of the SAR model which expresses global spatial effects. The corresponding reduced form of the model can be specified as:

$$\text{(1.14)} \qquad \mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

with the corresponding error covariance matrix given as:

$$\text{(1.15)} \qquad E(\mathbf{u}\mathbf{u}') = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \rho\mathbf{W}')^{-1} = \sigma^2[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}.$$

The structure in equation (1.15) shows that the spatial error process leads to a non-zero error covariance between every pair of observation, but decreasing in magnitude with the order of contiguity. Note also that heteroskedasticity is induced in **u**, irrespective of the heteroskedasticity of $\boldsymbol{\varepsilon}$, because the inverse matrices in equation (1.15) yields non-constant diagonal element in the error covariance matrix.

An alternative structural form, the so-called spatial Durbin or common factor model, can be generated by pre-multiplying equation (1.14) by $\mathbf{I}-\rho\mathbf{W}$ and moving the spatial lag term to the right-hand side as:

(1.16) $$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} - \rho\mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

This spatial model has spatially lagged exogenous variables (**WX**) in addition to the spatially lagged dependent variable (**Wy**) and a well-behaved disturbance term $\boldsymbol{\varepsilon}$. Equation (1.16), however, becomes a proper spatial error model only if a set of K nonlinear constraints on the parameters, the so-called common factor constraints, $\rho\,\gamma = -\rho\gamma$ (the product of the spatial autoregressive coefficient ρ with the regression coefficient $\boldsymbol{\gamma}$ should equal the negative of the coefficient of spatially lagged exogenous variables (**WX**), $\rho\gamma$), are satisfied. The spatial error model can also be expressed in terms of spatially filtered variables as:

(1.17) $$\mathbf{I}-\rho\mathbf{W}\ \ \mathbf{y} = \ \mathbf{I}-\rho\mathbf{W}\ \ \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

The single equation spatial error model developed above can easily be extended to a system of spatially interrelated cross sectional equations corresponding to n cross sectional units. Assuming the spatial dependence in the error term, the system of simultaneous equations given in equation (1.2) can be expressed as:

(1.18) $$\mathbf{Y}=\mathbf{Y}\mathbf{B}+\mathbf{X}\boldsymbol{\Gamma}+\mathbf{U}$$

with

$$\mathbf{U} = \mathbf{WUC} + \mathbf{E} \tag{1.19}$$

where $\mathbf{WU} = (\mathbf{Wu}_1, \ldots, \mathbf{Wu}_G)$, $\mathbf{C} = \text{diag}_{j=1}^{G}(\rho_j)$, $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_G)$ and the other notations as defined before. Note that ρ_j denotes the spatial autoregressive parameter in the jth equation and since $\mathbf{C}$ is taken to be diagonal, the specification relates the disturbance vector in the jth equation only to its own spatial lag. Since it is assumed that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$, the disturbances, however, will be spatially correlated across units and across equations.

The system in equation (1.18) and (1.19) can be expressed in a form where its solution for the endogenous variables is clearly revealed. But, first consider the following vector transformations:

$$\begin{aligned} vec(\mathbf{Y}) &= vec(\mathbf{YB}) + vec(\mathbf{X\Gamma}) + vec(\mathbf{U}) \\ vec(\mathbf{Y}) &= vec(\mathbf{YB}) + vec(\mathbf{X\Gamma}) + vec(\mathbf{UWC} + \mathbf{E}) \\ &= (\mathbf{B}' \otimes \mathbf{I})\, vec(\mathbf{Y}) + (\mathbf{\Gamma}' \otimes \mathbf{I})\, vec(\mathbf{X}) + (\mathbf{C}' \otimes \mathbf{W})\, \text{vec}\mathbf{U} + vec\mathbf{E} \end{aligned} \tag{1.20}$$

Letting $\mathbf{y} = vec(\mathbf{Y})$, $\mathbf{x} = \text{vec}(\mathbf{X})$, $\mathbf{u} = \text{vec}(\mathbf{U})$, and $\boldsymbol{\varepsilon} = \text{vec}(\mathbf{E})$ it follows from equation (1.20) that

$$\begin{aligned} &\mathbf{y} = (\mathbf{B}' \otimes \mathbf{I})\mathbf{y} + (\mathbf{\Gamma}' \otimes \mathbf{I})\mathbf{x} + (\mathbf{C}' \otimes \mathbf{W})\mathbf{u} + \boldsymbol{\varepsilon} \\ &\text{or} \\ &\mathbf{y} = (\mathbf{B}' \otimes \mathbf{I})\mathbf{y} + (\mathbf{\Gamma}' \otimes \mathbf{I})\mathbf{x} + \mathbf{u}, \\ &\mathbf{u} = (\mathbf{C}' \otimes \mathbf{W})\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \tag{1.21}$$

The system in equation (1.21) can also be rewritten more compactly in a form that can reveal its solution for the endogenous variables as follows:

$$\begin{aligned} \mathbf{y} &= \mathbf{B}^{**}\mathbf{y} + \mathbf{\Gamma}^{*}\mathbf{x} + \mathbf{u}, \\ \mathbf{u} &= \mathbf{C}^{*}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \tag{1.22}$$

where $\mathbf{B}^{**} = (\mathbf{B}' \otimes \mathbf{I}_n)$, $\mathbf{C}^* = \mathbf{C}' \otimes \mathbf{W} = diag_{j=1}^{G}(\rho_j \mathbf{W})$, and the other notations as before. Furthermore, by imposing exclusion restriction on the system in equation (1.22), it can be expressed as:

$$\begin{aligned} \mathbf{y}_j &= \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j, \\ \mathbf{u}_j &= \rho_j \mathbf{W} \mathbf{u}_j + \boldsymbol{\varepsilon}_j, \quad j = 1, \ldots, G \end{aligned} \tag{1.23}$$

where

$$\mathbf{Z}_j = (\mathbf{Y}_j, \mathbf{X}_j) \text{ and } \boldsymbol{\delta}_j = (\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j)'.$$

Spatial Autoregressive Model

When there are no strong a priori theoretical reasons to believe that interdependences between spatial units arises either due to the spatial lags of the dependent variables or due to spatially autoregressive error terms, the standard approach is to model the system with both effects included (Anselin, 2003).The spatial lag model and the spatial error model are discussed in the above two subsections separately. By combining these two models, the spatial autoregressive model with both the spatial lag and spatial error effects can be expressed as (all notations are as defined before):

$$\mathbf{Y} = \mathbf{YB} + \mathbf{X\Gamma} + \mathbf{WY\Lambda} + \mathbf{U} \tag{1.24}$$

with

$$\mathbf{U} = \mathbf{WUC} + \mathbf{E}.$$

Combining equation (1.11) and (1.22) gives the system in its more compact form as (all notations are as expressed before):

$$\begin{aligned} \mathbf{y} &= \mathbf{B}^* \mathbf{y} + \boldsymbol{\Gamma}^* \mathbf{x} + \mathbf{u}, \\ \mathbf{u} &= \mathbf{C}^* \mathbf{u} + \boldsymbol{\varepsilon} \end{aligned}. \tag{1.25}$$

Assuming that $\mathbf{I}_{nG}-\mathbf{B}^{*}$ and $\mathbf{I}_{nG}-\mathbf{C}^{*}$ are nonsingular matrices with $\left|\rho_j\right|<1,\ \mathrm{j} = 1,...,\mathrm{G}$, the system in equation (1.25) can be expressed in its reduced form as:

$$\begin{aligned}\mathbf{y} &= \left(\mathbf{I}_{nG}-\mathbf{B}^{*}\right)^{-1}\left(\boldsymbol{\Gamma}^{*}\mathbf{x}+\mathbf{u}\right), \\ \mathbf{u} &= \left(\mathbf{I}_{nG}-\mathbf{C}^{*}\right)^{-1}\boldsymbol{\varepsilon}\end{aligned} \qquad (1.26)$$

Since the innovations are assumed to be independently and identically distributed, that is, $E\left(\boldsymbol{\varepsilon}\right)=\mathbf{0}$ and $E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\boldsymbol{\Sigma}\otimes\mathbf{I}_n$, the means and variance covariance matrices of the disturbance terms **u**, and the endogenous variables **y**, are given, respectively, as follows:

$$\begin{aligned}&E\left(\mathbf{u}\right)=0; \quad E\left(\mathbf{u}\mathbf{u}'\right)=\boldsymbol{\Omega}_{\mathbf{u}}=\left(\mathbf{I}_{nG}-\mathbf{C}^{*}\right)^{-1}\left(\boldsymbol{\Sigma}\otimes\mathbf{I}_n\right)\left(\mathbf{I}_{nG}-\mathbf{C}^{*\prime}\right)^{-1} \\ &E\left(\mathbf{y}\right)=\left(\mathbf{I}_{nG}-\mathbf{B}^{*}\right)^{-1}\boldsymbol{\Gamma}^{*}\mathbf{x}\ ; \quad E\left(\mathbf{y}\mathbf{y}'\right)=\boldsymbol{\Omega}_{\mathbf{y}}=\left(\mathbf{I}_{nG}-\mathbf{B}^{*}\right)^{-1}\boldsymbol{\Omega}_{\mathbf{u}}\left(\mathbf{I}_{nG}-\mathbf{B}^{*\prime}\right)^{-1}\end{aligned} \qquad (1.27)$$

The endogenous variables as well as the disturbances are, therefore, seen to be correlated both spatially and across equation, and furthermore will generally be hetroskedastic. In this study, the spatial units are counties and each county has only a small number of neighbors and, in turn, it is only a neighbor to a small number of counties. The weights matrix **W** is a row standardized sparse matrix and hence the row and column sums of the weights matrix is bounded in absolute values. It is also assumed that $\left(\mathbf{I}_n-\rho_j\mathbf{W}\right)$, j = 1,...,G and $\left(\mathbf{I}_n-\mathbf{B}^{*}\right)^{-1}$ are bounded uniformly in absolute values, which imply that $\boldsymbol{\Omega}_{\mathbf{u}}$ and $\boldsymbol{\Omega}_{\mathbf{y}}$ are also bounded uniformly as it can easily be seen from the relations in equation (1.27). Thus, the degree of correlation between the elements of **u** and **y** are limited, which is a necessary condition for all large sample analysis (see Kelejian and Prucha, 1998, 2004).

By imposing exclusion restrictions on the system in equation (1.24), the spatial autoregressive model can also be reformulated as follows (all notations are as defined before):

(1.28)
$$\mathbf{y}_j = \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j, \\ \mathbf{u}_j = \rho_j \mathbf{W}\mathbf{u}_j + \boldsymbol{\varepsilon}_j, \quad j = 1,...,G$$

where

$$\mathbf{Z}_j = (\mathbf{Y}_j, \mathbf{X}_j, \mathbf{W}\mathbf{Y}_j) \text{ and } \boldsymbol{\delta}_j = (\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j, \boldsymbol{\lambda}'_j)'.$$

Following Kelejian and Prucha (2004), a set of instruments are utilized to estimate the spatial models in equations (1.11), (1.23) and (1.28) using the instrumental variable techniques. Let **N** denote the n by p matrix of those instruments and as suggested by Kelejian and Prucha (2004), **N** will be chosen as a subset of the linearly independent columns of (**X, WX, …,** $\mathbf{W}^s$**X**), where s is an integer such that $1 \le s \le 2$. It is assumed that the elements of **N** are uniformly bounded in absolute value. Besides, **N** is full column rank non-stochastic instrument matrix with the following properties:

1) $\lim_{n\to\infty} \frac{\mathbf{N}'\mathbf{N}}{n} = \mathbf{Q}_{NN}$, where $\mathbf{Q}_{NN}$ is a finite and nonsingular matrix

2) $\lim_{n\to\infty} \frac{\mathbf{N}'E(\mathbf{Z}_j)}{n} = \mathbf{Q}_{NZ_j}$, where $\mathbf{Q}_{NZ_j}$ is a finite matrix which has full column rank, j = 1,...,G

3) $\lim_{n\to\infty} \frac{\mathbf{N}'\mathbf{W}E(\mathbf{Z}_j)}{n} = \mathbf{Q}_{NWZ_j}$, where $\mathbf{Q}_{NWZ_j}$ is a finite matrix which has full column rank, j = 1,...,G

4) $\mathbf{Q}_{NZ_j} - \rho_j \mathbf{Q}_{NWZ_j} = \lim_{n\to\infty} \frac{\mathbf{N}'(\mathbf{I}_n - \rho_j \mathbf{W})\mathbf{Z}_j}{n}$ has a full column rank, j = 1,...,G

5) $\lim_{n\to\infty} \frac{\mathbf{N}'(\mathbf{I}_n - \rho_j \mathbf{W})^{-1}(\mathbf{I}_n - \rho_j \mathbf{W}')^{-1}\mathbf{N}}{n} = \boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ is a finite and nonsingular matrix, j = 1,...,G

Assuming that the matrix of exogenous (nonstochastic) variables **X** has full column rank, properties 1 and 2 are important to ensure the consistency of the initial two stage least squares estimators. Property 2 also ensures that the instruments **N** allow the identification of the regression parameters $\boldsymbol{\delta}_j$ in equations (1.11), (1.23) and (1.28). Note that the 2SLS estimator for the parameters of the models in each of these equations is a generalized moments estimator corresponding to the moment conditions $E(\mathbf{N}'\mathbf{u}_j) = \mathbf{0}$. Let $\mathbf{u}_j(\underline{\boldsymbol{\delta}}_j) = \mathbf{y}_j - \mathbf{Z}_j\underline{\boldsymbol{\delta}}_j = \mathbf{u}_j + \mathbf{Z}_j(\boldsymbol{\delta}_j - \underline{\boldsymbol{\delta}}_j)$, then the condition that $\mathbf{Q}_{NZ_j}$ has full column rank implies that $\lim_{n\to\infty} \frac{E(\mathbf{N}'\mathbf{u}_j(\underline{\boldsymbol{\delta}}_j))}{n} = \left[\lim_{n\to\infty} \frac{\mathbf{N}'E(\mathbf{Z}_j)}{n}\right](\boldsymbol{\delta}_j - \underline{\boldsymbol{\delta}}_j)$ is zero if and only if $\underline{\boldsymbol{\delta}}_j = \boldsymbol{\delta}_j$. Thus fulfillment of the rank condition for $\mathbf{Q}_{NZ_j}$ ensures that the instruments **N** identify the true parameter vector $\boldsymbol{\delta}_j$, j = 1,…,G, and the objective function is uniquely maximized at $\underline{\boldsymbol{\delta}}_j = \boldsymbol{\delta}_j$, at least in the limit.

Properties 3 and 4 are important in ensuring the consistency of the generalized two and three stage estimators, which are based on a Cochrane-Orcutt-type transformation of the models. Property 5 is used in deriving the limiting distribution of the initial two-stage least squares estimator from the untransformed model (see Kelejian and Prucha (2004) for details and proofs).

2.2 Spatial Dependence in Simultaneous-Equations Panel Data Models

When data is available across space and over time, spatial dependence can be incorporated into the standard simultaneous equations panel data models in a straightforward way. Spatial lag dependent variables, for example, can be written as follows (all notations are as given before):

(1.29) $$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

where

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \vdots\\ \mathbf{y}_G\end{bmatrix};\quad \mathbf{Z}=\text{diag}\left(\mathbf{Z}_j\right)=\begin{bmatrix}\mathbf{Z}_1 & \cdots & & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0}\\ \vdots & & \ddots & \vdots\\ \mathbf{0} & \cdots & & \mathbf{Z}_G\end{bmatrix};\quad \boldsymbol{\delta}=\begin{bmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\delta}_2\\ \vdots\\ \boldsymbol{\delta}_G\end{bmatrix};\quad \textit{and}\quad \mathbf{u}=\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_G\end{bmatrix}.$$

Note that $\mathbf{y}_j$ is nT x 1 vector of observations on the endogenous variable in the jth equation, $\mathbf{Z}_j = \left(\mathbf{Y}_j, \mathbf{X}_j, \mathbf{WY}_j\right)$ a matrix of dimension nT by $\left(G1_j\text{-}1+K1_j+G1_j\right)$, $\boldsymbol{\delta}_j = \left(\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j, \boldsymbol{\lambda}'_j\right)'$ where $\boldsymbol{\beta}_j$ is $G1_j\text{-}1$ by 1, $\boldsymbol{\gamma}_j$ is $K1_j$ by 1, and $\boldsymbol{\lambda}_j$ is $G1_j$ by 1, and $\mathbf{u}_j$ is nT by 1 vector of disturbance in the jth equation, for j = 1, 2, ..., G .For the one-way error component model, the disturbance of the jth equation $\mathbf{u}_j$ is given by:

(1.30) $$\mathbf{u}_j = \mathbf{Z}_\mu \boldsymbol{\mu}_j + \boldsymbol{\omega}_j$$

where

$$\mathbf{Z}_\mu = \left(\mathbf{I}_n \otimes \boldsymbol{\iota}_T\right),\ \boldsymbol{\mu}'_j = \left(\mu_{1j}, \mu_{2j}, \ldots, \mu_{nj}\right),\ \text{and } \boldsymbol{\omega}'_j = \left(\omega_{11j}, \omega_{11j}, \ldots, \omega_{1Tj}, \ldots, \omega_{n1j}, \ldots, \omega_{nTj}\right).$$

Thus,

(1.31) $$\boldsymbol{\Omega}_{jl} = E\left(\mathbf{u}_j \mathbf{u}'_l\right) = \boldsymbol{\sigma}^2_{\mu_{jl}}\left(\mathbf{I}_n \otimes \mathbf{J}_T\right) + \boldsymbol{\sigma}^2_{\omega_{jl}}\left(\mathbf{I}_n \otimes \mathbf{I}_T\right)$$

where $\mathbf{I}_T$ and $\mathbf{I}_n$ are identity matrices with dimensions T and n, respectively, $\boldsymbol{\iota}_T$ is a vector of ones of dimension T, $\mathbf{J}_T$ is a matrix of ones of dimension T and $\otimes$ denotes Kronecker product.

In this case, the covariance matrix between the disturbances of different equations has the same one-way error component form. But, now there are additional cross equation variances components to be estimated. When one considers the whole model, the variance-covariance matrix for the set of G structural equations is given by:

(1.32) $$\boldsymbol{\Omega} = E\left(\mathbf{u}\mathbf{u}'\right) = \boldsymbol{\Sigma}_\mu \otimes \left(\mathbf{I}_n \otimes \mathbf{J}_T\right) + \boldsymbol{\Sigma}_\omega \otimes \left(\mathbf{I}_n \otimes \mathbf{I}_T\right)$$

where $\boldsymbol{\Sigma}_\mu = \left[\boldsymbol{\sigma}^2_{\mu_{jl}}\right]$ and $\boldsymbol{\Sigma}_\omega = \left[\boldsymbol{\sigma}^2_{\omega_{jl}}\right]$ are both G x G matrices, and $\mathbf{u}' = (\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_G)$ is a 1 x nGT vector of disturbances with $\mathbf{u}_j$ defined in equation (1.30) for j = 1,2,…G. Before proceeding, it is helpful to define two matrices, **P** and **H,** which are useful in transforming the structural equations. Let **P** be the matrix which averages the observations across time for each individual and **H** be the matrix which obtains the deviations from individual means. Thus,

$$\begin{bmatrix} \mathbf{P} = \mathbf{Z}_\mu (\mathbf{Z}'_\mu \mathbf{Z}_\mu)^{-1} \mathbf{Z}'_\mu = \mathbf{I}_n \otimes \bar{\mathbf{J}}_T, \text{ where } \bar{\mathbf{J}}_T = \mathbf{J}_T / T \\ \mathbf{H} = \mathbf{I}_{nT} - \mathbf{P} = \left(I_T - \frac{J_T}{T} \right) \otimes I_n \end{bmatrix}^1 .$$

Now it is possible to transform the stacked system of equations in equation (1.29) by $(\mathbf{I}_G \otimes \mathbf{H})$ and $(\mathbf{I}_G \otimes \mathbf{P})$ to get, respectively,

$$\hat{\mathbf{y}} = \hat{\mathbf{Z}}\boldsymbol{\delta} + \hat{\mathbf{u}} \text{ and } \breve{\mathbf{y}} = \breve{\mathbf{Z}}\boldsymbol{\delta} + \breve{\mathbf{u}} \tag{1.33}$$

where $\hat{\mathbf{y}} = (\mathbf{I}_G \otimes \mathbf{H})\mathbf{y}$, $\hat{\mathbf{Z}} = (\mathbf{I}_G \otimes \mathbf{H})\mathbf{Z}$, and $\hat{\mathbf{u}} = (\mathbf{I}_G \otimes \mathbf{H})\mathbf{u}$; and $\breve{\mathbf{y}} = (\mathbf{I}_G \otimes \mathbf{P})\mathbf{y}$, $\breve{\mathbf{Z}} = (\mathbf{I}_G \otimes \mathbf{P})\mathbf{Z}$, and $\breve{\mathbf{u}} = (\mathbf{I}_G \otimes \mathbf{P})\mathbf{u}$. The W3SLS and B3SLS estimators can be obtained by performing 3SLS on these transformed equations using, respectively, $\mathbf{I}_G \otimes \hat{\mathbf{N}}$ and $\mathbf{I}_G \otimes \breve{\mathbf{N}}$ as sets of instruments, where $\hat{\mathbf{N}} = \mathbf{HN}$ and $\breve{\mathbf{N}} = \mathbf{PN}$.

Similarly, spatial dependence in the errors can be written as follows:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \tag{1.34}$$

$$\mathbf{u} = \rho (\mathbf{I}_T \otimes \mathbf{W})\mathbf{u} + (\mathbf{I}_n \otimes \boldsymbol{\iota}_T)\mu + \boldsymbol{\omega}$$

where

[1] **P** and **H** are idempotent, orthogonal and sum to the identity matrix.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{bmatrix}; \quad \mathbf{Z} = \text{diag}\left(\mathbf{Z}_j\right) = \begin{bmatrix} \mathbf{Z}_1 & \cdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & & \mathbf{Z}_G \end{bmatrix}; \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_G \end{bmatrix}; \quad \textit{and} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_G \end{bmatrix}$$

with $\mathbf{Z}_j = \left(\mathbf{Y}_j, \mathbf{X}_j\right)$, $\boldsymbol{\delta}_j = \left(\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j\right)'$ for j = 1,…,G and the other notations as given in equation (1.29).

In order to facilitate the modeling of spatial error dependence in the context of panel data, the data is arranged time wise. Apart from this difference in the arrangement of the data, the format is similar to equation (1.29) above. Now consider the jth equation of the system in equation (1.34):

(1.35)
$$\mathbf{y}_j = \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j$$

$$\mathbf{u}_j = \rho_j \left(\mathbf{I}_T \otimes \mathbf{W}\right) \mathbf{u}_j + \boldsymbol{\upsilon}_j$$

where

(1.36)
$$\boldsymbol{\upsilon}_j = \left(\mathbf{I}_n \otimes \boldsymbol{\iota}_T\right) \mu_j + \boldsymbol{\omega}_j .$$

The mean and the covariance of the innovation vector $\boldsymbol{\upsilon}_j$ can be given by:

(1.37)
$$\begin{aligned} &E\left(\boldsymbol{\upsilon}_j\right) = 0 \\ &E\left(\boldsymbol{\upsilon}_j \boldsymbol{\upsilon}'_j\right) = \boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j} = \sigma^2_{\mu_j}\left(\mathbf{J}_T \otimes \mathbf{I}_n\right) + \sigma^2_{\omega_j}\left(\mathbf{I}_n \otimes \mathbf{I}_n\right) \end{aligned} .$$

From the second part of equation (1.35) it follows that

(1.38)
$$\mathbf{u}_j = \left[\mathbf{I}_T \otimes \left(\mathbf{I}_n - \rho_j \mathbf{W}\right)^{-1}\right] \boldsymbol{\upsilon}_j .$$

Thus, the mean and the covariance of $\mathbf{u}_j$ can be given as follows:

$$
(1.39)\quad
\begin{aligned}
E(\mathbf{u}_j) &= 0\\
E(\mathbf{u}_j\mathbf{u}_j') &= \boldsymbol{\Omega}_{\mathbf{u}_j} = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W})^{-1}\right]\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W}')^{-1}\right]\\
&= \boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W})^{-1}\right]\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W}')^{-1}\right]\\
&= \boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W})^{-1}(\mathbf{I}_n - \rho_j\mathbf{W}')^{-1}\right]
\end{aligned}
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}$ is as given in equation (1.37). This can easily be extended to the whole model. But, first note that the variance-covariance of the innovations for the whole model can be given by:

$$
(1.40)\quad \boldsymbol{\Omega}_{\boldsymbol{\upsilon}} = \boldsymbol{\Sigma}_1 \otimes \mathbf{P} + \boldsymbol{\Sigma}_{\omega} \otimes \mathbf{H}
$$

where $\boldsymbol{\Sigma}_1 = T\boldsymbol{\Sigma}_{\mu} + \boldsymbol{\Sigma}_{\omega}$.
Thus, the variance-covariance matrix for the set of G structural equations is computed as:

$$
(1.41)\quad E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Omega}_{\mathbf{u}} = \boldsymbol{\Omega}_{\boldsymbol{\upsilon}}\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j\mathbf{W})^{-1}(\mathbf{I}_n - \rho_j\mathbf{W}')^{-1}\right],\ \text{for j = 1,...,G}.
$$

Both spatial lag dependence and spatial error dependence can also be incorporated into the standard simultaneous equations models in the context of panel data. Recalling equations (1.29) and (1.34) the spatial autoregressive panel data model with spatial autoregressive disturbances can be formulated as follows (all notations and definitions are as expressed before):

$$
(1.42)\quad
\begin{aligned}
\mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{u},\\
\mathbf{u} &= \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u} + \boldsymbol{\upsilon}
\end{aligned}
$$

with

$$
\boldsymbol{\upsilon} = (\mathbf{I}_n \otimes \boldsymbol{\iota}_T)\mu + \boldsymbol{\omega}
$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{bmatrix}; \quad \mathbf{Z} = \text{diag}\left(\mathbf{Z}_j\right) = \begin{bmatrix} \mathbf{Z}_1 & \cdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & & \mathbf{Z}_G \end{bmatrix}; \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_G \end{bmatrix}; \quad and \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_G \end{bmatrix}.$$

Note that $\mathbf{y}_j$ is nT x 1 vector of observations on the endogenous variable in the jth equation, $\mathbf{Z}_j = \left(\mathbf{Y}_j, \mathbf{X}_j, \mathbf{WY}_j\right)$ a matrix of dimension $\text{nT by } \left(\text{G1}_j\text{-1+K1}_j + \text{G1}_j\right)$, $\boldsymbol{\delta}_j = \left(\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j, \boldsymbol{\lambda}'_j\right)'$ where $\boldsymbol{\beta}_j$ is $\text{G1}_j\text{-1}$ by 1, $\boldsymbol{\gamma}_j$ is K1_j by 1, and $\boldsymbol{\lambda}_j$ is G1_j by 1, and $\mathbf{u}_j$ is nT by 1 vector of disturbance in the jth equation, for $j = 1, 2, ..., G$. The variance-covariance of the innovations and the disturbances for this model are the same to those given in equations (1.40) and (1.41), respectively.

3. Estimation Issues in Spatial Simultaneous-Equations Models

The presence of a combination of feedback simultaneity, spatial autoregressive lag simultaneity and spatial cross-regressive lag simultaneity in spatial simultaneous equations models creates a number of complications of which the questions of whether or not each equation of the model is identified, the choice of estimators and the treatment of instruments are the most important ones (Rey and Boarnet, 2004). The traditional rank and order conditions for identification, for example, are not applied if the system is expressed as (all notations are as defined before except now **B** is a matrix of coefficients whose diagonal elements are 1):

(1.43) $$\mathbf{YB} = \mathbf{X\Gamma} + \mathbf{WY\Lambda} + \mathbf{U}.$$

Pre- and post-multiplying the matrix of endogenous variables in equation (1.43) by two distinct coefficient matrices leads to two different reduced forms. Thus, this system as it stands does not lend itself to the application of the traditional rank and order conditions for checking identification. If the models, however, are viewed as special cases of simultaneous equations that are nonlinear in endogenous variables, identification can be checked by checking for the following necessary conditions:

1. All the endogenous variables in the model can be expressed in terms of the exogenous variables and the disturbance terms. This condition is fulfilled if matrices $\mathbf{I}_{nG} - \mathbf{B}^*$ and $\mathbf{I}_n - \rho_j \mathbf{W}$ are nonsingular with $|\rho_j| < 1$, for j = 1, …, G.
2. The solution of the model for the endogenous variables in terms of the exogenous variables and the disturbance terms is unique. This condition is fulfilled if the instruments matrix **N** is selected in such a way that $\lim_{n \to \infty} \frac{\mathbf{N}' E \mathbf{Z}_j}{n}$ is a finite matrix which has full column rank for j = 1, .., G.
3. The number of endogenous variables appearing on the right hand side of an equation must be less than or equal to the number of exogenous and lagged endogenous variables appearing in the model but not in that equations

It is well known that the presence of endogenous variables on the right hand side of an equation in the simultaneous equations system violates the assumption of zero correlation between the regressors and the disturbance term upon which the unbiasedness or consistency of ordinary least squares (OLS) estimators are based. For the same reason, the presence of the endogenous variables in their lagged form on the right hand side of an equation leads to biased and inconsistent OLS estimates. Besides, the existence of a spatially autoregressive error term in an equation leads to the inconsistency of OLS. Hence, an alternative estimation method must be used in order to obtain unbiased and consistent estimator for the parameters of the spatial simultaneous equations models. One such an approach is the instrumental variables procedure suggested in Kelejian and Prucha (2004).

3.1. Cross-Sectional Data Setting

Kelejian and Prucha (2004) suggest limited and full information instrumental variable estimator for the parameters of a spatial simultaneous equation model and derive the limiting distribution of those estimators. In the case of limited information (single equation) estimation, they proposed a three step generalized spatial two-stage least squares (GS2SLS) procedure to estimate the unknown parameters in the jth equation of the model in equation (1.24). The first step consists of the estimation of the model parameter vector $\boldsymbol{\delta}_j$ in equation (1.28) by two-stage least squares (2SLS) using the instruments **N**, where **N** is defined in reference to equation (1.28) above. The resulting 2SLS estimator is given by:

(1.44) $$\hat{\boldsymbol{\delta}}_j = \left(\bar{\mathbf{Z}}_j'\bar{\mathbf{Z}}_j\right)^{-1}\bar{\mathbf{Z}}_j'\mathbf{y}_j$$

where $\bar{\mathbf{Z}}_j = \mathbf{P}_N\mathbf{Z}_j = \left(\bar{\mathbf{Y}}_j, \mathbf{X}_j, \overline{\mathbf{WY}}\right)$ with $\bar{\mathbf{Y}}_j = \mathbf{P}_N\mathbf{Y}_j, \overline{\mathbf{WY}}_j = \mathbf{P}_N\mathbf{WY}_j$, and $\mathbf{P}_N = \mathbf{N}\left(\mathbf{N}'\mathbf{N}\right)^{-1}\mathbf{N}'$ is a projection matrix. Although $\hat{\boldsymbol{\delta}}_j$ is consistent estimator of $\boldsymbol{\delta}_j$,it does not utilize information relating to the spatial correlation of the disturbance terms. These 2SLS estimates are used to compute estimates for the disturbances $\mathbf{u}_j$ which in turn are used to estimate the autoregressive parameter ρ_j in the second step of the procedure. The resulting 2SLS residuals are hence given by:

(1.45) $$\tilde{\mathbf{u}}_j = \mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j .$$

In the second step, Kelejian and Prucha (2004) used the generalized moments procedure to estimate the spatial autoregressive parameter of the disturbances of the jth equation, for j =

1,…, G, of the model in equation (1.28). Note that from the relation in equation (1.28) we have:

(1.46) $$\mathbf{u}_j - \rho_j \mathbf{W}\mathbf{u}_j = \boldsymbol{\varepsilon}_j$$

and pre-multiplication by the weights matrix **W** gives:

(1.47) $$\mathbf{W}\mathbf{u}_j - \rho_j \mathbf{W}^2\mathbf{u}_j = \mathbf{W}\boldsymbol{\varepsilon}_j .$$

The following three-equation system is obtained from the relationships between equations (1.46) and (1.47):

(1.48) $$\begin{aligned}
\frac{\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{u}_j'\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} - 2\rho_j \frac{\mathbf{u}_j'\ \mathbf{W}\mathbf{u}_j}{n} \\
\frac{\mathbf{W}\boldsymbol{\varepsilon}_j{}'\ \mathbf{W}\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}^2\mathbf{u}_j{}'\ \mathbf{W}^2\mathbf{u}_j}{n} - 2\rho_j \frac{\mathbf{W}^2\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} \\
\frac{\boldsymbol{\varepsilon}_j'\mathbf{W}\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{u}_j'\ \mathbf{W}\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}^2\mathbf{u}_j}{n} - \rho_j \frac{\left(\mathbf{u}_j'\ \mathbf{W}^2\mathbf{u}_j + \mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j\right)}{n}
\end{aligned}$$

Taking expectations across equation (1.48):

(1.49) $$E\left[\begin{aligned}
\frac{\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{u}_j'\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} - 2\rho_j \frac{\mathbf{u}_j'\ \mathbf{W}\mathbf{u}_j}{n} \\
\frac{\mathbf{W}\boldsymbol{\varepsilon}_j{}'\ \mathbf{W}\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}^2\mathbf{u}_j{}'\ \mathbf{W}^2\mathbf{u}_j}{n} - 2\rho_j \frac{\mathbf{W}^2\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j}{n} \\
\frac{\boldsymbol{\varepsilon}_j'\mathbf{W}\boldsymbol{\varepsilon}_j}{n} &= \frac{\mathbf{u}_j'\ \mathbf{W}\mathbf{u}_j}{n} + \rho_j^2 \frac{\mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}^2\mathbf{u}_j}{n} - \rho_j \frac{\left(\mathbf{u}_j'\ \mathbf{W}^2\mathbf{u}_j + \mathbf{W}\mathbf{u}_j{}'\ \mathbf{W}\mathbf{u}_j\right)}{n}
\end{aligned}\right]$$

yields:

$$(1.50)\quad \begin{bmatrix} \sigma_j^2 \\ \sigma_j^2 \dfrac{tr\left(\mathbf{W}'\mathbf{W}\right)}{n} \\ \sigma_j^2 \dfrac{tr\left(\mathbf{W}\right)}{n} = 0 \end{bmatrix} = E \begin{bmatrix} \dfrac{\mathbf{u}_j'\mathbf{u}_j}{n} + \rho_j^2 \dfrac{\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)}{n} - 2\rho_j \dfrac{\mathbf{u}_j'\left(\mathbf{Wu}_j\right)}{n} \\ \dfrac{\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)}{n} + \rho_j^2 \dfrac{\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)}{n} - 2\rho_j \dfrac{\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{Wu}_j\right)}{n} \\ \dfrac{\mathbf{u}_j'\left(\mathbf{Wu}_j\right)}{n} + \rho_j^2 \dfrac{\left(\mathbf{Wu}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)}{n} - \rho_j \dfrac{\left(\mathbf{u}_j'\left(\mathbf{W}^2\mathbf{u}_j\right) + \left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} \end{bmatrix}$$

and after rearranging:

$$(1.51)\quad E\begin{bmatrix} \dfrac{\mathbf{u}_j'\mathbf{u}_j}{n} \\ \dfrac{\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)}{n} \\ \dfrac{\mathbf{u}_j'\left(\mathbf{Wu}_j\right)}{n} \end{bmatrix} = \begin{bmatrix} 2\rho_j \dfrac{E\left(\mathbf{u}_j'\left(\mathbf{Wu}_j\right)\right)}{n} & -\rho_j^2 \dfrac{E\left(\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & \sigma_j^2 \\ 2\rho_j \dfrac{E\left(\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & -\rho_j^2 \dfrac{E\left(\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)\right)}{n} & \sigma_j^2 \dfrac{tr\left(\mathbf{W}'\mathbf{W}\right)}{n} \\ \rho_j \dfrac{E\left(\mathbf{u}_j'\left(\mathbf{W}^2\mathbf{u}_j\right) + \left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & -\rho_j^2 \dfrac{E\left(\left(\mathbf{Wu}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)\right)}{n} & 0 \end{bmatrix}$$

$$(1.52)\quad \underbrace{E\begin{bmatrix} \dfrac{\mathbf{u}_j'\mathbf{u}_j}{n} \\ \dfrac{\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)}{n} \\ \dfrac{\mathbf{u}_j'\left(\mathbf{Wu}_j\right)}{n} \end{bmatrix}}_{\tau_j} = \underbrace{\begin{bmatrix} 2\dfrac{E\left(\mathbf{u}_j'\left(\mathbf{Wu}_j\right)\right)}{n} & -\dfrac{E\left(\left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & 1 \\ 2\dfrac{E\left(\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & -\dfrac{E\left(\left(\mathbf{W}^2\mathbf{u}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)\right)}{n} & \dfrac{tr\left(\mathbf{W}'\mathbf{W}\right)}{n} \\ \dfrac{E\left(\mathbf{u}_j'\left(\mathbf{W}^2\mathbf{u}_j\right) + \left(\mathbf{Wu}_j\right)'\left(\mathbf{Wu}_j\right)\right)}{n} & -\dfrac{E\left(\left(\mathbf{Wu}_j\right)'\left(\mathbf{W}^2\mathbf{u}_j\right)\right)}{n} & 0 \end{bmatrix}}_{\Upsilon_j} \underbrace{\begin{bmatrix} \rho_j \\ \rho_j^2 \\ \sigma_j^2 \end{bmatrix}}_{\alpha_j}$$

Thus, the system in equation (1.52) can be rewritten as (j = 1,…,G):

$$(1.53)\quad \boldsymbol{\tau}_j = \Upsilon_j \boldsymbol{\alpha}_j \rightarrow \boldsymbol{\alpha}_j = \Upsilon_j^{-1} \boldsymbol{\tau}_j$$

The parameter vector $\boldsymbol{\alpha}_j = (\rho_j, \rho_j^2, \sigma_j^2)'$ would be completely determined in terms of the relation in equation (1.53) if τ_j and Υ_j were known. Note that τ_j and Υ_j are not observable. Following the suggestions in Kelejian and Prucha (2004), however, the following estimators for τ_j and Υ_j in terms of sample moments can be defined as:

$$\mathbf{o}_j = \frac{\left[\tilde{\mathbf{u}}_j'\tilde{\mathbf{u}}_j,\ (\mathbf{W}\tilde{\mathbf{u}}_j)'(\mathbf{W}\tilde{\mathbf{u}}_j),\ \tilde{\mathbf{u}}_j'(\mathbf{W}\tilde{\mathbf{u}}_j)\right]}{n},$$

(1.54)
$$\mathbf{O}_j = \frac{1}{n}\begin{bmatrix} 2\tilde{\mathbf{u}}_j'(\mathbf{W}\tilde{\mathbf{u}}_j) & -(\mathbf{W}\tilde{\mathbf{u}}_j)'(\mathbf{W}\tilde{\mathbf{u}}_j) & \text{n} \\ 2(\mathbf{W}^2\tilde{\mathbf{u}}_j)'(\mathbf{W}\tilde{\mathbf{u}}_j) & -(\mathbf{W}^2\tilde{\mathbf{u}}_j)'(\mathbf{W}^2\tilde{\mathbf{u}}_j) & \text{tr}(\mathbf{W}'\mathbf{W}) \\ \left(\tilde{\mathbf{u}}_j'(\mathbf{W}^2\tilde{\mathbf{u}}_j) + (\mathbf{W}\tilde{\mathbf{u}}_j)'(\mathbf{W}\tilde{\mathbf{u}}_j)\right) & -(\mathbf{W}\tilde{\mathbf{u}}_j)'(\mathbf{W}^2\tilde{\mathbf{u}}_j) & 0 \end{bmatrix}.$$

Thus, given the estimates in equation (1.54), the empirical form of the relationship in equation (1.53) can be given by:

(1.55)
$$\mathbf{o}_j = \mathbf{O}_j\boldsymbol{\alpha}_j + \boldsymbol{\xi}_j.$$

Since $\mathbf{o}_j$ and $\mathbf{O}_j$ are observable and $\boldsymbol{\alpha}_j$ is vector of parameters to be estimated, $\boldsymbol{\xi}_j$ can be viewed as a vector of regression residuals. Thus, the second step estimators of ρ_j and σ_j^2 , say, $\hat{\rho}_j$ and $\hat{\sigma}_j^2$, are nonlinear least squares estimators defined as the minimizer of:

(1.56)
$$\left[\mathbf{o}_j - \mathbf{O}_j\begin{bmatrix}\rho_j \\ \rho_j^2 \\ \sigma_j^2\end{bmatrix}\right]'\left[\mathbf{o}_j - \mathbf{O}_j\begin{bmatrix}\rho_j \\ \rho_j^2 \\ \sigma_j^2\end{bmatrix}\right]$$

or

$$\hat{\rho}_j, \hat{\sigma}_j^2 = \text{argmin}\left[\mathbf{o}_j - \mathbf{O}_j \begin{bmatrix} \rho_j \\ \rho_j^2 \\ \sigma_j^2 \end{bmatrix}\right]' \left[\mathbf{o}_j - \mathbf{O}_j \begin{bmatrix} \rho_j \\ \rho_j^2 \\ \sigma_j^2 \end{bmatrix}\right].$$

In the third step of the procedure a Cochrane-Orcutt type transformation is applied to the model in equation (1.28). More specifically, let:

$$\mathbf{y}_j^*(\rho_j) = \mathbf{y}_j - \rho_j \mathbf{W}\mathbf{y}_j \text{ and } \mathbf{Z}_j^*(\rho_j) = \mathbf{Z}_j - \rho_j \mathbf{W}\mathbf{Z}_j .$$

Then, equation (1.28) becomes:

(1.57) $$\mathbf{y}_j^*(\rho_j) = \mathbf{Z}_j^*(\rho_j)\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j$$

If ρ_j were know we could perform 2SLS on equation (1.57) to obtain the generalized spatial two-stage least squares (GS2SLS) estimator for $\boldsymbol{\delta}_j$. That is:

(1.58) $$\hat{\boldsymbol{\delta}}_j = \bar{\mathbf{Z}}_j^*(\rho_j)'\bar{\mathbf{Z}}_j^*(\rho_j)^{-1} \bar{\mathbf{Z}}_j^*(\rho_j)'\mathbf{y}_j^*(\rho_j)$$

where $\bar{\mathbf{Z}}_j^*(\rho_j) = \mathbf{P}_N \mathbf{Z}_j^*(\rho_j)$ and $\mathbf{P}_N = \mathbf{N} \; \mathbf{N}'\mathbf{N}^{-1} \mathbf{N}'$ .But, since in practical applications ρ_j is not known, it is replaced with its estimate as defined in equation (1.56) and estimate the model in equation (1.58) using 2SLS. The resulting estimator is termed as the feasible GS2SLS and is given by:

(1.59) $$\hat{\boldsymbol{\delta}}_j^F = \bar{\mathbf{Z}}_j^*(\hat{\rho}_j)'\bar{\mathbf{Z}}_j^*(\hat{\rho}_j)^{-1} \bar{\mathbf{Z}}_j^*(\hat{\rho}_j)'\mathbf{y}_j^*(\hat{\rho}_j)$$

where $\bar{\mathbf{Z}}_j^*(\hat{\rho}_j) = \mathbf{P}_N [\mathbf{Z}_j - \hat{\rho}_j \mathbf{W}\mathbf{Z}_j]$ and $\mathbf{y}_j^*(\hat{\rho}_j) = \mathbf{y}_j - \hat{\rho}_j \mathbf{W}\mathbf{y}_j$.

The three step GS2SLS procedure is applied in estimating the parameters of spatial simultaneous equation models when the spatial dependence is either spatial error dependence or both spatial error dependence and spatial lag dependence. When the spatial dependence is only spatial lag type and if the disturbances are white noise, then, the second step and consequently the third step are not required.

One of the limitations of the limited information (single equation) estimation technique is that it does not take into account the information provided by the potential cross equation correlation in the innovation vectors $\boldsymbol{\varepsilon}_j$. In order to use the information from such cross equation correlations, it is important to stack the equations given in equation (1.56) as follows:

(1.60) $$\mathbf{y}^*(\rho) = \mathbf{Z}^*(\rho)\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

where

$\mathbf{y}^*(\rho) = \left(\mathbf{y}_1^*(\rho_1)', ..., \mathbf{y}_G^*(\rho_G)'\right)'$, $\mathbf{Z}^*(\rho) = diag_{j=1}^{G}\left(\mathbf{Z}_j^*(\rho_j)\right)$, $\rho = \left(\rho_1, ..., \rho_G\right)'$ *and* $\delta = \left(\delta_1, ..., \delta_G\right)'$

Recall from equation (1.26) that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$. Assuming that ρ and $\boldsymbol{\Sigma}$ were known, equation (1.60) could be estimated using the instrumental variable technique. In that case, the resulting systems instrumental variable estimator of $\boldsymbol{\delta}$ would be the generalized spatial three-stage least squares (GS3SLS) estimator which can be given by (all notations as defined before):

(1.61) $$\hat{\boldsymbol{\delta}} = \left[\overline{\mathbf{Z}}^*(\hat{\rho})'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n\right)\overline{\mathbf{Z}}^*(\hat{\rho})\right]^{-1}\overline{\mathbf{Z}}^*(\hat{\rho})'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n\right)\mathbf{y}^*(\hat{\rho}).$$

Since in practical applications ρ and $\boldsymbol{\Sigma}$ are not known, their estimators are required to obtain the feasible estimator for $\boldsymbol{\delta}$. The generalized moments estimators for ρ_j and σ_j^2 are defined in equation (1.56). Note that σ_j^2 is the jth diagonal element of $\boldsymbol{\Sigma}$. Besides, a consistent estimator for $\boldsymbol{\Sigma}$ can be derived by combining equations (1.57) and (1.59) as:

(1.62) $$\hat{\boldsymbol{\sigma}}_{jl}^2 = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}_j'\hat{\boldsymbol{\varepsilon}}_l,\ j, l = 1, ..., G$$

where $\hat{\boldsymbol{\varepsilon}}_j = \mathbf{y}_j^*(\hat{\rho}_j) - \mathbf{Z}_j^*(\hat{\rho}_j)\hat{\boldsymbol{\delta}}_j^F$. Then, the G by G matrix whose (j,l)th element is $\hat{\boldsymbol{\sigma}}_{jl}^2$ defines a consistent estimator for $\boldsymbol{\Sigma}$ denoted by $\hat{\boldsymbol{\Sigma}}$. Substituting $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}$ in equation (1.61) gives the feasible generalized spatial three-stage least squares (FGS3SLS) estimator for $\boldsymbol{\delta}$. That is:

(1.63) $$\hat{\boldsymbol{\delta}}^F = \left[\bar{\mathbf{Z}}^*(\hat{\rho})' \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n\right) \bar{\mathbf{Z}}^*(\hat{\rho})\right]^{-1} \bar{\mathbf{Z}}^*(\hat{\rho})' \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_n\right) \mathbf{y}^*(\hat{\rho}).$$

3.2.Panel Data Setting

The same procedure is also applicable to the panel data case with minor change in the arrangement of the data set and some changes in notations. Now, recall the spatial autoregressive panel data model with spatial autoregressive disturbances, from equation (1.42):

(1.64) $$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}, \\ \mathbf{u} &= \rho\left(\mathbf{I}_T \otimes \mathbf{W}\right)\mathbf{u} + \boldsymbol{\upsilon} \end{aligned}$$

where

$$\boldsymbol{\upsilon} = \left(\mathbf{I}_n \otimes \boldsymbol{\iota}_T\right)\mu + \boldsymbol{\omega}.$$

As it is evident from equations (1.37)-(1.41), the variance-covariance matrix $\boldsymbol{\Omega}_u$ depends on $\rho, \boldsymbol{\sigma}_\omega^2$ and $\boldsymbol{\sigma}_1^2$. Thus, a feasible estimator for the parameter of the model requires consistent estimators of $\rho, \boldsymbol{\sigma}_\omega^2$ and $\boldsymbol{\sigma}_1^2$. To this end, generalized moments estimators of $\rho, \boldsymbol{\sigma}_\omega^2$ and $\boldsymbol{\sigma}_1^2$ are defined in the following subsection. These generalized moments estimators generalize the generalized moments estimators given in Kelejian and Prucha (2004) for the case of a single cross section. First, the generalized moments estimators of $\rho, \boldsymbol{\sigma}_{\omega}^2$ and $\boldsymbol{\sigma}_1^2$ in context of single equation are defined and then the procedure is generalized to system of simultaneous equations model.

Consider the jth equation of the system in (1.64). The generalized moments estimators of $\rho_j, \boldsymbol{\sigma}_{\omega j}^2$ and $\boldsymbol{\sigma}_{1j}^2$ are defined in terms of six moments conditions. These six moments conditions, for $T \geq 2$, are given as follows:

(1.65)
$$
\begin{aligned}
&E\left[\frac{1}{n(T-1)}\boldsymbol{\upsilon}_j'\mathbf{H}\boldsymbol{\upsilon}_j\right] &&= \boldsymbol{\sigma}_{\omega j}^2 \\
&E\left[\frac{1}{n(T-1)}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)'\mathbf{H}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)\right] &&= \boldsymbol{\sigma}_{\omega j}^2 \frac{1}{n} tr\left(\mathbf{W}'\mathbf{W}\right) \\
&E\left[\frac{1}{n(T-1)}\boldsymbol{\upsilon}_j'\mathbf{H}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)\right] &&= \mathbf{0} \\
&E\left[\frac{1}{n}\boldsymbol{\upsilon}_j'\mathbf{P}\boldsymbol{\upsilon}_j\right] &&= \boldsymbol{\sigma}_{1j}^2 \\
&E\left[\frac{1}{n}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)'\mathbf{P}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)\right] &&= \boldsymbol{\sigma}_{1j}^2 \frac{1}{n} tr\left(\mathbf{W}'\mathbf{W}\right) \\
&E\left[\frac{1}{n}\boldsymbol{\upsilon}_j'\mathbf{P}\left((\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j\right)\right] &&= \mathbf{0}
\end{aligned}
.
$$

Note that since $\boldsymbol{\upsilon}_j$ are not observable, $E\left[\frac{1}{n(T-1)}\boldsymbol{\upsilon}_j'\mathbf{H}\boldsymbol{\upsilon}_j\right]$ and $E\left[\frac{1}{n}\boldsymbol{\upsilon}_j'\mathbf{P}\boldsymbol{\upsilon}_j\right]$ do not represent the unbiased analysis of variance estimators of $\boldsymbol{\sigma}_{\omega j}^2$ and $\boldsymbol{\sigma}_{1j}^2$, respectively. The innovations $\boldsymbol{\upsilon}_j$ can, however, be expressed in terms of the disturbances and the disturbances in turn can be substituted by their estimated values which are observable. Thus, using the relations in the jth equation of (1.64), let:

(1.66)
$$
\begin{aligned}
\boldsymbol{\upsilon}_j &= \mathbf{u}_j - \rho_j (\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j \\
(\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j &= (\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j - \rho_j (\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j
\end{aligned}
.
$$

Substituting the expressions in (1.66) for $\boldsymbol{\upsilon}_j$ and $(\mathbf{I}_T \otimes \mathbf{W})\boldsymbol{\upsilon}_j$ into (1.65) gives the moments conditions in (1.65) in terms of the disturbances. That is:

$$
(1.67)\qquad
\begin{aligned}
&E\frac{1}{n(T-1)}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{H}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right] &&=\boldsymbol{\sigma}^2_{\omega j}\\
&E\frac{1}{n(T-1)}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right]'\mathbf{H}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right] &&=\boldsymbol{\sigma}^2_{\omega j}\frac{1}{n}tr(\mathbf{W}'\mathbf{W})\\
&E\frac{1}{n(T-1)}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{H}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right] &&=\mathbf{0}\\
&E\frac{1}{n}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{P}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right] &&=\boldsymbol{\sigma}^2_{1j}\\
&E\frac{1}{n}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right]'\mathbf{P}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right] &&=\boldsymbol{\sigma}^2_{1j}\frac{1}{n}tr(\mathbf{W}'\mathbf{W})\\
&E\frac{1}{n}\left[\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{P}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j-\rho_j(\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{u}_j\right] &&=\mathbf{0}
\end{aligned}.
$$

After rearranging, this yields a system of six equations that can be expressed as:

$$
(1.68)\qquad \Upsilon_j\alpha_j-\tau_j=0
$$

where

$$
\tau_j=E\begin{bmatrix}
\dfrac{\mathbf{u}_j'\mathbf{H}\mathbf{u}_j}{n(T-1)}\\[2ex]
\dfrac{\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{H}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]}{n(T-1)}\\[2ex]
\dfrac{\mathbf{u}_j'\mathbf{H}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]}{n(T-1)}\\[2ex]
\dfrac{\mathbf{u}_j'\mathbf{P}\mathbf{u}_j}{n}\\[2ex]
\dfrac{\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]'\mathbf{P}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]}{n}\\[2ex]
\dfrac{\mathbf{u}_j'\mathbf{P}\left[(\mathbf{I}_T\otimes\mathbf{W})\mathbf{u}_j\right]}{n}
\end{bmatrix},
$$

$$
\Upsilon_j = \begin{bmatrix}
2\frac{E\left[\mathbf{u}_j'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n(T-1)} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n(T-1)} & 1 & 0 \\
2\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right)'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n(T-1)} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right)'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right]}{n(T-1)} & \frac{tr(\mathbf{W}'\mathbf{W})}{n(T-1)} & 0 \\
\frac{E\left[\mathbf{u}_j'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j + \left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n(T-1)} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right]}{n(T-1)} & 0 & 0 \\
2\frac{E\left[\mathbf{u}_j'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n} & 0 & 1 \\
2\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right)'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right)'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right]}{n} & 0 & \frac{tr(\mathbf{W}'\mathbf{W})}{n} \\
\frac{E\left[\mathbf{u}_j'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j + \left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right]}{n} & -\frac{E\left[\left((\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j\right)'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j\right]}{n} & 0 & 0
\end{bmatrix}
$$

, and

$$
\alpha_j = \begin{bmatrix} \rho_j \\ \rho_j^2 \\ \sigma_{\omega j}^2 \\ \sigma_{1j}^2 \end{bmatrix}.
$$

The system in (1.68) consists of six equations involving the second moments of $\mathbf{u}_j$, $(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j$ and $(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j$. However, since $\mathbf{u}_j$, $(\mathbf{I}_T \otimes \mathbf{W})\mathbf{u}_j$ and $(\mathbf{I}_T \otimes \mathbf{W}^2)\mathbf{u}_j$ are not observable, the generalized moments estimators are defined in terms of sample moments. These sample moments are obtained by replacing the $\mathbf{u}_j$'s in (1.68) by their estimated values ($\tilde{\mathbf{u}}_j$), where the estimated disturbances are computed as:

(1.69) $$\tilde{\mathbf{u}}_j = \mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j$$

where $\mathbf{y}_j$ and $\mathbf{Z}_j$ are defined in equation (1.42) and $\hat{\boldsymbol{\delta}}_j$ is an estimator of $\boldsymbol{\delta}_j$ obtained by estimating the regression model in equation (1.64) by two-stage least squares(2SLS) using the instruments **N.** After substituting $\mathbf{u}_j$ by $\tilde{\mathbf{u}}_j$ the system in (1.68) becomes the sample analogue to (1.68) and can be expressed as:

(1.70) $$\mathbf{O}_j\boldsymbol{\alpha}_j - \mathbf{o}_j = \xi_j(\boldsymbol{\alpha}_j)$$

where

$$\mathbf{o}_j = E\begin{bmatrix} \dfrac{\tilde{\mathbf{u}}_j'\mathbf{H}\tilde{\mathbf{u}}_j}{n(T-1)} \\ \dfrac{[(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j]'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j}{n(T-1)} \\ \dfrac{\tilde{\mathbf{u}}_j'\mathbf{H}(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j}{n(T-1)} \\ \dfrac{\tilde{\mathbf{u}}_j'\mathbf{P}\tilde{\mathbf{u}}_j}{n} \\ \dfrac{[(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j]'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j}{n} \\ \dfrac{\tilde{\mathbf{u}}_j'\mathbf{P}(\mathbf{I}_T \otimes \mathbf{W})\tilde{\mathbf{u}}_j}{n} \end{bmatrix},$$

$$
\mathbf{O}_j = \begin{bmatrix}
2\frac{E\left[\tilde{\mathbf{u}}_j'\mathbf{H}\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]}{n\left(T-1\right)} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n\left(T-1\right)} & 1 & 0 \\
2\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n\left(T-1\right)} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]\right)}{n\left(T-1\right)} & \frac{tr\left(\mathbf{W}'\mathbf{W}\right)}{n\left(T-1\right)} & 0 \\
\frac{E\left(\tilde{\mathbf{u}}_j'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]+\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n\left(T-1\right)} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{H}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]\right)}{n\left(T-1\right)} & 0 & 0 \\
2\frac{E\left(\tilde{\mathbf{u}}_j'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & 0 & 1 \\
2\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & 0 & \frac{tr\left(\mathbf{W}'\mathbf{W}\right)}{n} \\
\frac{E\left(\tilde{\mathbf{u}}_j'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]+\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & -\frac{E\left(\left[\left(\mathbf{I}_T\otimes\mathbf{W}\right)\tilde{\mathbf{u}}_j\right]'\mathbf{P}\left[\left(\mathbf{I}_T\otimes\mathbf{W}^2\right)\tilde{\mathbf{u}}_j\right]\right)}{n} & 0 & 0
\end{bmatrix},
$$

$\boldsymbol{\alpha}_j = \begin{bmatrix} \rho_j \\ \rho_j^2 \\ \sigma_{\omega j}^2 \\ \sigma_{1j}^2 \end{bmatrix}$, and $\xi_j\left(\boldsymbol{\alpha}_j\right)$ is a vector of residuals.

Now, it is possible to define the unweighted and weighted generalized moments estimators of $\rho_j, \boldsymbol{\sigma}_{\omega j}^2$ and $\boldsymbol{\sigma}_{1j}^2$. When equal weights are given to the moments conditions, the generalized moments estimators of $\rho_j, \boldsymbol{\sigma}_{\omega j}^2$ and $\boldsymbol{\sigma}_{1j}^2$, say, $\hat{\rho}_j, \hat{\boldsymbol{\sigma}}_{\omega j}^2$ and $\hat{\boldsymbol{\sigma}}_{1j}^2$ respectively, are defined as the unweighted nonlinear least squares estimators corresponding to (1.70). More formally, $\hat{\rho}_j, \hat{\boldsymbol{\sigma}}_{\omega j}^2$ and $\hat{\boldsymbol{\sigma}}_{1j}^2$ are defined as the nonlinear least squares estimators that minimize:

$$(1.71)\qquad \boldsymbol{\xi}_j\left(\rho_j,\sigma^2_{\omega j},\sigma^2_{1j}\right)'\boldsymbol{\xi}_j\left(\rho_j,\sigma^2_{\omega j},\sigma^2_{1j}\right)=\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]'\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]$$

or

$$\left(\hat{\rho}_j,\hat{\sigma}^2_{\omega j},\hat{\sigma}^2_{1j}\right)=\operatorname{argmin}\left\{\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]'\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right],\ \rho_j\in\left[-1,1\right],\sigma^2_{\omega j}\in\left[0,c_\omega\right],\sigma^2_{1j}\in\left[0,c_1\right]\right\}.$$

When the moments conditions are weighed by a weighing matrix, say, $\boldsymbol{\Phi}$, the generalized moments estimators of $\rho_j,\boldsymbol{\sigma}^2_{\omega j}$ and $\boldsymbol{\sigma}^2_{1j}$, however, are defined as the nonlinear least squares estimators that minimize:

$$(1.72)\qquad \boldsymbol{\xi}_j\left(\rho_j,\sigma^2_{\omega j},\sigma^2_{1j}\right)'\tilde{\boldsymbol{\Phi}}_j^{-1}\boldsymbol{\xi}_j\left(\rho_j,\sigma^2_{\omega j},\sigma^2_{1j}\right)=\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]'\tilde{\boldsymbol{\Phi}}_j^{-1}\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]$$

or

$$\left(\tilde{\rho}_j,\tilde{\sigma}^2_{\omega j},\tilde{\sigma}^2_{1j}\right)=\operatorname{argmin}\left\{\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right]'\tilde{\boldsymbol{\Phi}}_j^{-1}\left[\mathbf{o}_j-\mathbf{O}_j\begin{bmatrix}\rho_j\\ \sigma^2_{\omega j}\\ \sigma^2_{1j}\end{bmatrix}\right],\ \rho_j\in\left[-1,1\right],\sigma^2_{\omega j}\in\left[0,c_\omega\right],\sigma^2_{1j}\in\left[0,c_1\right]\right\}$$

where $\tilde{\rho}_j,\tilde{\sigma}^2_{\omega j}$ and $\tilde{\sigma}^2_{1j}$ are the weighted generalized moments estimators of $\rho_j,\boldsymbol{\sigma}^2_{\omega j}$ and $\boldsymbol{\sigma}^2_{1j}$ respectively and $\tilde{\boldsymbol{\Phi}}_j$ is the consistent estimator of $\boldsymbol{\Phi}_j$ with $\boldsymbol{\Phi}_j$ representing the variance covariance matrix of the sample moments at the true parameter values. Recall that in this research the sample moments at the true parameter values are given by the left hand side expressions on equation (1.65) with the expectations operator suppressed. Thus, $\mathbf{\Phi}_j$ can be expressed by:

$$(1.73) \qquad \mathbf{\Phi}_j = \begin{bmatrix} \dfrac{\boldsymbol{\sigma}^4_{\omega j}}{T-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}^4_{1j} \end{bmatrix} \otimes \begin{bmatrix} 2 & 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}}{n}\right) & \mathbf{0} \\ 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}}{n}\right) & 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}\mathbf{W}'\mathbf{W}}{n}\right) & \mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}\;\;\mathbf{W}'+\mathbf{W}}{n}\right) \\ \mathbf{0} & \mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}\;\;\mathbf{W}'+\mathbf{W}}{n}\right) & \mathrm{tr}\left(\dfrac{\mathbf{W}\mathbf{W}+\mathbf{W}'\mathbf{W}}{n}\right) \end{bmatrix}$$

where $\boldsymbol{\sigma}^4_{\omega j}$ and $\boldsymbol{\sigma}^4_{1j}$ are as defined in equation (1.71) .When $\boldsymbol{\sigma}^4_{\omega j}$ and $\boldsymbol{\sigma}^4_{1j}$ are replaced by their consistent estimators, say, $\tilde{\boldsymbol{\sigma}}^4_{\omega j}$ and $\tilde{\boldsymbol{\sigma}}^4_{1j}$ respectively, equation (1.73) becomes :

$$(1.74) \qquad \tilde{\mathbf{\Phi}}^{-1}_j = \begin{bmatrix} \dfrac{\tilde{\boldsymbol{\sigma}}^4_{\omega j}}{T-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\sigma}}^4_{1j} \end{bmatrix} \otimes \begin{bmatrix} 2 & 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}}{n}\right) & \mathbf{0} \\ 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}}{n}\right) & 2\mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}\mathbf{W}'\mathbf{W}}{n}\right) & \mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}\;\;\mathbf{W}'+\mathbf{W}}{n}\right) \\ \mathbf{0} & \mathrm{tr}\left(\dfrac{\mathbf{W}'\mathbf{W}(\mathbf{W}'+\mathbf{W})}{n}\right) & \mathrm{tr}\left(\dfrac{\mathbf{W}\mathbf{W}+\mathbf{W}'\mathbf{W}}{n}\right) \end{bmatrix}$$

$\tilde{\mathbf{\Phi}}_j$ is consistent estimator of $\mathbf{\Phi}_j$ provided the estimators for $\boldsymbol{\sigma}^2_{\omega \mathrm{j}}$ and $\boldsymbol{\sigma}^2_{1j}$ are consistent.

Now, consider once again the jth equation of the system given in (1.74) and recall (1.39). If $\rho_j, \boldsymbol{\sigma}^2_{\omega \mathrm{j}}$ and $\boldsymbol{\sigma}^2_{1j}$ were known, then the generalized spatial two-stage least squares (GS2SLS) estimator of $\boldsymbol{\delta}_j$ can be given by:

(1.75)
$$\begin{aligned}\hat{\boldsymbol{\delta}}_j &= \left[\bar{\mathbf{Z}}_j' \ \boldsymbol{\Omega}_{u_j}^{-1}(\rho_j, \boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2) \ \mathbf{Z}_j\right]^{-1} \bar{\mathbf{Z}}_j' \ \boldsymbol{\Omega}_{u_j}^{-1}(\rho_j, \boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2) \ \mathbf{y}_j \\ &= \left\{\bar{\mathbf{Z}}_j' \left[\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}(\boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2)\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j \mathbf{W})^{-1} (\mathbf{I}_n - \rho_j \mathbf{W}')^{-1}\right]\right]^{-1} \bar{\mathbf{Z}}_j\right\}^{-1} \\ &\quad \times \bar{\mathbf{Z}}_j' \left[\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}(\boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2)\left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j \mathbf{W})^{-1} (\mathbf{I}_n - \rho_j \mathbf{W}')^{-1}\right]\right]^{-1} \mathbf{y}_j \\ &= \left[\bar{\mathbf{Z}}_j^*(\rho_j)' \left(\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}^{-1}(\boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2)\right) \mathbf{Z}_j^*(\rho_j)\right]^{-1} \bar{\mathbf{Z}}_j^*(\rho_j)' \left(\boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}^{-1}(\boldsymbol{\sigma}_{\omega j}^2, \boldsymbol{\sigma}_{1j}^2)\right) \mathbf{y}_j^*(\rho_j)\end{aligned}$$

where $\bar{\mathbf{Z}}_j^*(\rho_j) = \mathbf{P}_N \mathbf{Z}_j^*(\rho_j)$ with $\mathbf{Z}_j^*(\rho_j) = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j \mathbf{W})\right]\mathbf{Z}_j$, $\mathbf{P}_N = \mathbf{N}(\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'$,and $\mathbf{y}_j^*(\rho_j) = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j \mathbf{W})\right]\mathbf{y}_j$. But, since in practical applications $\rho_j, \boldsymbol{\sigma}_{\omega j}^2$ and $\boldsymbol{\sigma}_{1j}^2$ are not known, their estimators as defined in (1.71) or (1.72) are used instead. The resulting estimator is termed as feasible generalized spatial two-stage least squares estimator and is given by:

(1.76)
$$\hat{\boldsymbol{\delta}}_j^F = \left[\bar{\mathbf{Z}}_j^*(\tilde{\rho}_j)' \ \boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}^{-1}(\tilde{\boldsymbol{\sigma}}_{\omega j}^2, \tilde{\boldsymbol{\sigma}}_{1j}^2) \ \bar{\mathbf{Z}}_j^*(\tilde{\rho}_j)\right]^{-1} \bar{\mathbf{Z}}_j^*(\tilde{\rho}_j)' \ \boldsymbol{\Omega}_{\boldsymbol{\upsilon}_j}^{-1}(\tilde{\boldsymbol{\sigma}}_{\omega j}^2, \tilde{\boldsymbol{\sigma}}_{1j}^2) \ \mathbf{y}_j^*(\tilde{\rho}_j) .$$

The variables $\mathbf{y}_j^*(\rho_j)$ and $\mathbf{Z}_j^*(\rho_j)$ can be viewed as the result of a spatial Cochrane-Orcutt type transformation of the jth equation of the model in (1.74). That is, pre-multiplication of the first and the second parts of the jth equation of (1.74) by $\mathbf{I}_T \otimes (\mathbf{I}_n - \rho_j \mathbf{W})$ gives (j = 1,…, G):

(1.77)
$$\mathbf{y}_j^*(\rho_j) = \mathbf{Z}_j^*(\rho_j)\boldsymbol{\delta}_j + \boldsymbol{\upsilon}_j .$$

One of the limitations of this estimator in (1.75), however, is that it does not take into account the information provided by the potential cross equation correlation in the innovation vectors $\boldsymbol{\upsilon}_j$. In order to use the information from such cross equation correlations, it is important to stack the equations given in (1.77) as follows:

(1.78)
$$\mathbf{y}^*(\rho) = \mathbf{Z}^*(\rho)\boldsymbol{\delta} + \boldsymbol{\upsilon}$$

where

$\mathbf{y}^*(\rho) = \left[\mathbf{y}_1^*(\rho_1)',...,\mathbf{y}_G^*(\rho_G)'\right]'$, $\mathbf{Z}^*(\rho) = diag_{j=1}^G\left(\mathbf{Z}_j^*(\rho_j)\right)$, $\rho = \left(\rho_1,...,\rho_G\right)'$ *and* $\delta = \left(\delta_1,...,\delta_G\right)'$

Note that the mean of the innovation vector $\boldsymbol{\upsilon}$ is zero. Recall also from (1.40) that $E\left(\boldsymbol{\upsilon\upsilon}'\right) = \boldsymbol{\Omega}_{\boldsymbol{\upsilon}} = \boldsymbol{\Sigma}_1 \otimes \mathbf{P} + \boldsymbol{\Sigma}_\varpi \otimes \mathbf{H}$, where $\boldsymbol{\Sigma}_1 = T\boldsymbol{\Sigma}_\mu + \boldsymbol{\Sigma}_\varpi$, *with* $\boldsymbol{\Sigma}_\mu = \left[\sigma^2_{\mu_{jl}}\right]$, $\boldsymbol{\Sigma}_\varpi = \left[\sigma^2_{\varpi_{jl}}\right]$ and hence $\boldsymbol{\Sigma}_1 = \left[\sigma^2_{1_{jl}} = T\sigma^2_{\mu_{jl}} + \sigma^2_{\varpi_{jl}}\right]$ all G by G matrices. Assuming that ρ and $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$ were known, the system in (1.78) could be estimated using instrumental variable estimation technique. The resulting systems instrumental variable estimator of $\boldsymbol{\delta}$ would be the generalized spatial three-stage least squares (GS3SLS) estimator and can be expressed as:

(1.79) $$\hat{\boldsymbol{\delta}} = \left[\bar{\mathbf{Z}}^*(\rho)'\left(\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}^{-1}(\boldsymbol{\sigma}^2_\varpi, \boldsymbol{\sigma}^2_1)\right)\bar{\mathbf{Z}}^*(\rho)\right]^{-1}\bar{\mathbf{Z}}^*(\rho)'\left(\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}^{-1}(\boldsymbol{\sigma}^2_\varpi, \boldsymbol{\sigma}^2_1)\right)\mathbf{y}^*(\rho).$$

However, since ρ and $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$ are not known in practical applications, their estimators are required to obtain a feasible estimator for $\boldsymbol{\delta}$. The generalized moments estimator for ρ can be obtained from (1.72) and since $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$ is composed of $\boldsymbol{\sigma}^2_\varpi$ and $\boldsymbol{\sigma}^2_1$ its estimator can also be obtained from (1.72). Besides, consistent estimators of $\boldsymbol{\sigma}^2_\varpi$ and $\boldsymbol{\sigma}^2_1$ and hence of $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$ can be derived by combining (1.76) and (1.77) as (j, l =1,…,G):

(1.80) $$\begin{aligned} \tilde{\boldsymbol{\sigma}}^2_{\varpi jl} &= \left[\mathbf{y}_j^*(\tilde{\rho}_j) - \mathbf{Z}_j^*(\tilde{\rho}_j)\hat{\boldsymbol{\delta}}_j^F\right]'\mathbf{H}\left[\mathbf{y}_l^*(\tilde{\rho}_l) - \mathbf{Z}_l^*(\tilde{\rho}_l)\boldsymbol{\delta}_l^F\right] \Big/ n(T-1) \\ \tilde{\boldsymbol{\sigma}}^2_{1jl} &= \left[\mathbf{y}_j^*(\tilde{\rho}_j) - \mathbf{Z}_j^*(\tilde{\rho}_j)\hat{\boldsymbol{\delta}}_j^F\right]'\mathbf{P}\left[\mathbf{y}_l^*(\tilde{\rho}_l) - \mathbf{Z}_l^*(\tilde{\rho}_l)\boldsymbol{\delta}_l^F\right] \Big/ n \end{aligned}.$$

Then, the G by G matrix whose (j,l)th element is $\tilde{\boldsymbol{\sigma}}^2_{\varpi jl}$ defines a consist estimator for $\boldsymbol{\Sigma}_\varpi$ denoted by $\tilde{\boldsymbol{\Sigma}}_\varpi$ and the G by G matrix whose (j,l)th element is $\tilde{\boldsymbol{\sigma}}^2_{1jl}$ defines a consistent estimator for $\boldsymbol{\Sigma}_1$ denoted by $\tilde{\boldsymbol{\Sigma}}_1$ .Thus, (1.81) $$\tilde{\mathbf{\Omega}}_{\mathbf{\upsilon}} = \tilde{\mathbf{\Sigma}}_{\varpi} \otimes \mathbf{H} + \tilde{\mathbf{\Sigma}}_{1} \otimes \mathbf{P}\ .$$

Replacing ρ and $\mathbf{\Omega}_{\mathbf{\upsilon}}$ by $\tilde{\rho}$ and $\tilde{\mathbf{\Omega}}_{\mathbf{\upsilon}} = \mathbf{\Omega}_{\mathbf{\upsilon}}(\tilde{\boldsymbol{\sigma}}_{\varpi}^{2}, \tilde{\boldsymbol{\sigma}}_{1}^{2})$ in (1.79), hence, yields the feasible generalized spatial three-stage least squares (FGS3SLS) estimator which can be expressed as:

(1.82) $$\hat{\boldsymbol{\delta}}^{F} = \left[\bar{\mathbf{Z}}^{*}(\tilde{\rho})' \ \mathbf{\Omega}_{\mathbf{\upsilon}}^{-1}(\tilde{\boldsymbol{\sigma}}_{\varpi}^{2}, \tilde{\boldsymbol{\sigma}}_{1}^{2}) \ \bar{\mathbf{Z}}^{*}(\tilde{\rho}) \right]^{-1} \bar{\mathbf{Z}}^{*}(\tilde{\rho})' \ \mathbf{\Omega}_{\mathbf{\upsilon}}^{-1}(\tilde{\boldsymbol{\sigma}}_{\varpi}^{2}, \tilde{\boldsymbol{\sigma}}_{1}^{2}) \ \mathbf{y}^{*}(\tilde{\rho})\ .$$

4. Specification Tests

Specification tests form one of the most important areas of research in econometrics (Hausman, 1978). Once the system of simultaneous equations is specified, there is an opportunity to test both coefficient restrictions and asymptotic orthogonality assumptions. When there are more instruments than needed to identify an equation, a test can be done to investigate whether the additional instruments are valid in the sense that they are uncorrelated with the error term. This is commonly known as test of the overidentifying restriction. To explain this, consider the jth equation of the system of simultaneous equations given in this study:

(1.83) $$\mathbf{y_j} = \mathbf{Y_j}\boldsymbol{\beta}_\mathbf{j} + \mathbf{X_j}\boldsymbol{\gamma}_\mathbf{j} + \mathbf{u_j}$$

where $\mathbf{Y_j}$ is a vector of G_j-1 included right-hand endogenous variables, $\mathbf{X_j}$ is a vector of k_j included predetermined variables, $\mathbf{u_j}$ is disturbance term, and the vector of K_j excluded predetermined variables is given by $\mathbf{X^*_j}$.

Anderson and Rubin (1950) was the first to develop the procedure for testing the overidentifying restriction based on the asymptotic distribution of the smallest characteristic root ($\boldsymbol{\lambda}_\mathbf{j}$) derived from LIML estimation. Their likelihood ratio test statistic is base on n($\boldsymbol{\lambda}_\mathbf{j}$-1) which under the null hypothesis is distributed as Chi-squared with (K_j-

(G_j-1)) degree of freedom which is equal to the number of overidentifying restrictions.. That is,

$$\mathbf{LR} = \chi^2\left[\mathbf{K_j} - \mathbf{G_j} - \mathbf{1}\right] = n\ \boldsymbol{\lambda}_\mathbf{j} - \mathbf{1} \tag{1.84}$$

where n is the sample size and all other notations are as defined above. A large value for LR is an indication that there are exogenous variables in the model that have been inappropriately omitted from the jth equation. This test statistic, however, is difficult to compute. Hausman (1983) proposed an alternative test statistic based on Lagrange multiplier principle which is asymptotically equivalent but easier to compute. This test statistic is obtained as nR_u^2, where n is the sample size and R_u^2 is the usual R-squared or the uncentered R-squared of the regression of residuals from the second-stage equation on all included and excluded instruments. In other words, simply estimate equation (1.83) by 2SLS, GMM, LIML or any efficient limited-information estimator and obtain the resulting residuals, $\hat{u}_j$. Then, regress these on all instruments and calculate nR_u^2. The statistic has a limiting chi-squared distribution with (K_j-(G_j-1)) degree of freedom which is equal to the number of overidentifying restrictions, under the assumed specification of the model.

One potential source of misspecification in spatial econometric models comes from spatial autocorrelation in the dependent variable or in the error term or in both. Anselin and Kelejian (1997) proposed Moran's I statistics based on residuals that are obtained from an instrumental variable (IV) procedure such as 2SLS in a general model that encompasses endogeneity due to feedback simultaneities as well as spatial autoregressive /or cross-regressive lag simultaneities. Following Anselin and Kelejian (1997), this statistic is specified as:

(1.85) $$I^{*} = n\left(\frac{\hat{u}'W\hat{u}}{S_0\hat{u}'\hat{u}}\right)$$

where n is the sample size, $\hat{u}$ is the IV residuals, W is the spatial weights matrix, and S_0 is the usual normalizing factor given by

$$S_0 = \sum_{i}^{n}\sum_{j}^{n} w_{ij} \ .$$

Note that for a row standardized spatial weights matrix, S_0 is equal to n because each row sums to one. Hence equation (1.85) is simplified to:

(1.86) $$I^{*} = \left(\frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}\right).$$

Anselin and Kelejian (1997) shows that

$$n^{1/2} I^{*} \xrightarrow{D} N\left[0, \phi^2\right]$$

where

(1.87) $$\phi^2 = \frac{S_2}{2S_1^2} + \left(\frac{4}{S_1^2\sigma_u^2}\right)A$$

with S_1 and S_2 finite constants such that

$$S_1 = \lim_{n\to\infty}\sum_{i}^{n}\sum_{j}^{n} w_{ij} \Big/ n \text{, and}$$

$$S_2 = \lim_{n\to\infty} tr\left[\ W + W' \quad W + W'\ \right]\Big/ n \text{,}$$

σ_u^2 is error variance, and

$$A = p\lim\left[\ n^{-1}u'WZ_j \quad n \ \ Z_j'P_X Z_j \ ^{-1} \quad n^{-1}Z_j W'u'\ \right].$$

Replacing S_1, S_2, A by their finite sample counterparts, respectively

$$\hat{S}_1 = \sum_{i}^{n}\sum_{j}^{n} w_{ij} \Big/ n$$

$$\hat{S}_2 = \ tr\left[\ W + W' \quad W + W'\ \right]\Big/ n$$

$$\hat{A} = \left[n^{-1}u'WZ_j \quad n \quad Z_j' P_X Z_j \quad ^{-1} \quad n^{-1} Z_j W' u' \right].$$

and σ_u^2 by its consistent estimator, $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/n$ in equation (1.87) would give a consistent estimator for ϕ^2, say $\hat{\phi}^2$.

With ϕ^2 replaced by its consistent estimator $\hat{\phi}^2$, an asymptotic test can be constructed such that the null hypothesis of no spatial autocorrelation may be rejected at the α level of significance if

$$\left| \frac{n^{1/2} I^*}{\hat{\phi}} \right| \succ z_\alpha$$

where z_α is the value of the standard normal variate corresponding to α.

References

Anderson, T. W. and H. Rubin, 1950, "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics,* **21**:570–582

Anselin, Luc and A. Bera, 1998, "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," *In Handbook of Applied Economic Statistics,* edited by A. Ullah and D. E. Giles, 237-289, New York: Marcel Dekker.

Anselin, Luc, 1988, *Spatial Econometrics: Methods, and Models,* Dordrecht: Kluwer Academic.

Anselin, Luc, 1992, "Space and Applied Econometrics: Introduction," Spatial Issue, *Regional Science and Urban Economics*, **22**: 307-316.

Anselin, Luc, 1993, "Discrete Space Autoregressive Models," in Michael F. Goodchild, Bradley O. Parks and T. Steyaert, eds., *Environmental Modeling with GIS,* pp. 454-469, Oxford: Oxford University Press.

Anselin, Luc, 1996, "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association," in M. Fischer, H. Scholten and D. Unwin, eds., *Spatial Analytical Perspective on GIS,* pp. 111-125, London: Taylor & Francis.

Anselin, Luc, 1998, "GIS Research Infrastructure for Spatial Analysis of Real Estate Markets," *Journal of Housing Research,* **9**: 113-133.

Anselin, Luc, 2001, "Spatial Econometrics," in B. Baltagi, (ed.), *A Companion to Theoretical Econometrics*, pp. 310-330, Oxford, England: Blackwell.

Anselin, Luc, 2003, "Spatial Externalities, Spatial Multipliers and Spatial Econometrics," *International Regional Science Review,* **26**(2): 153-166.

Anselin, Luc, and Harry H. Kelejian, 1997, "Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors," *International Regional Science Review*, **20**: 153-182.

Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica*, **46**: 1251-1271.

Hausman, J., 1983, "Specification and Estimation of Simultaneous Equations Models," in Z. Griliches and m. Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North Holland.

Kelejian, Harry H. and Ingram R. Prucha, 1998, "A Generalized Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Spatial Autoregressive Disturbances," *Journal of Real Estate, Finance and Economics,* **17**(1): 99-121.

Kelejian, Harry H. and Ingram R. Prucha, 2004, "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations," *Journal of Econometrics*, **118**:27-50.

LeSage, J.P., 1999, "Spatial Econometrics, University of Toledo, available at http://www.rri.wvu.edu/WebBook/LeSage/spatial/spatial.html.

Rey, Sergio J. anfd Marlon G. Boarnet, 2004, "A Taxonomy of Spatial Econometric Models for Simultaneous Equations Systems," in L. Anselin, R.J.G.M. Florax and S.J. Rey, *Advances in Spatial Econometrics: methodology, Tools and Applications,* Berlin: Springer.