GMM Estimation of Spatial Error Autocorrelation with Heteroskedasticity

Luc Anselin

July 28, 2010

1 Background

This note documents the steps needed for an efficient GMM estimation of the regression parameters and autoregressive parameters using the moment conditions spelled out in Kelejian and Prucha (2009) and Arraiz et al. (2010) (referred to in what follows as K-P). Theoretical details are provided in those articles. The focus here is on the practical steps to carry out estimation in a number of special cases. I will be using "we" below since all this should eventually be moved into the spatial econometrics text.

2 Model Specification and Notation

The model we consider is the mixed regressive spatial autoregressive model with a spatial autoregressive error term, or so-called SAR-SAR model. Using the notation from Anselin (1988), this is expressed as:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u}. \tag{1}$$

The notation is standard, with $\mathbf{y}$ as a $n \times 1$ vector of observations on the dependent variable, $\mathbf{W}$ as a $n \times n$ spatial lag operator and $\mathbf{Wy}$ as the spatial lag term with spatial autoregressive parameter ρ, $\mathbf{X}$ as an $n \times k$ matrix of observations on explanatory variables with $k \times 1$ coefficient vector β, and a $n \times 1$ vector of errors $\mathbf{u}$. The specification is general, in the sense that $\mathbf{X}$ can contain both exogenous as well as endogenous variables. In the latter case, a $n \times p$ matrix of instruments $\mathbf{H}$ will be needed. An alternative way to express the model is as:

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u}, \tag{2}$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{Wy}]$ and the $k \times 1$ coefficient vector is rearranged as $\delta = [\beta' \rho]'$ (i.e., a column vector).

The error vector $\mathbf{u}$ follows a spatial autoregressive process:

$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \varepsilon \tag{3}$$

where λ is the spatial autoregressive parameter, and with heteroskedastic innovations, such that $\mathrm{E}[\varepsilon_i^2] = \sigma_i^2$. All other assumptions are standard.

Note that, in contrast to the general case outlined by K-P, we take the weights matrix in the spatial lag and in the spatial error part to be the same ($\mathbf{W}$). In K-P, the weights matrix for the error term is denoted as $\mathbf{M}$. While this is more general, it is hardly ever used in practice, hence we limit our treatment to the simpler case.

2.1 Spatially Lagged Variables

Spatially lagged variables play an important part in the GMM estimation procedure. In the original K-P papers, these are denoted by "bar" superscripts. Instead, we will use the L subscript throughout. In other words, a first order spatial lag of $\mathbf{y}$, i.e., $\mathbf{Wy}$ is denoted by $\mathbf{y}_L$, and similarly for spatially lagged explanatory variables, $\mathbf{X}_L$, and for $\mathbf{Z}_L$. Higher order spatial lags are symbolized by adding additional L subscripts. For example, a second order spatial lag of the error $\mathbf{u}$ would be $\mathbf{u}_{LL}$.

2.2 Spatial Cochrane-Orcutt Transformation

An important aspect of the estimation is the use of a set of spatially filtered variables in a spatially weighted regression. K-P refer to this as a spatial Cochrane-Orcutt transformation. The spatial filter is based on the weights matrix and the spatial autoregressive parameter for the error process. Since there is no distinction between the two weights here, the matrix $\mathbf{W}$ is used throughout. In the notation of what follows, we will use a subscript s for the spatially filtered variables. Also, to keep the notation simple, we will not distinguish between the notation for a parameter and its estimate. In practice, an estimate is always used, since the true parameter value is unknown.

The spatially filtered variables are then:

$$
\begin{aligned}
\mathbf{y}_s &= \mathbf{y} - \lambda\mathbf{W}\mathbf{y} && (4)\\
&= \mathbf{y} - \lambda\mathbf{y}_L && (5)\\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{y} && (6)\\
\mathbf{X}_s &= \mathbf{X} - \lambda\mathbf{W}\mathbf{X} && (7)\\
&= \mathbf{X} - \lambda\mathbf{X}_L && (8)\\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{X} && (9)\\
\mathbf{W}\mathbf{y}_s &= \mathbf{W}\mathbf{y} - \lambda\mathbf{W}\mathbf{W}\mathbf{y} && (10)\\
&= \mathbf{y_L} - \lambda\mathbf{y}_{LL} && (11)\\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{W}\mathbf{y} && (12)\\
\mathbf{Z}_s &= \mathbf{Z} - \lambda\mathbf{W}\mathbf{Z} && (13)\\
&= \mathbf{Z} - \lambda\mathbf{Z}_L && (14)\\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{Z} && (15)
\end{aligned}
$$

3 Outline of the GMM Estimation Procedure

The GMM estimation is carried out in multiple steps. The basic rationale is the following. First, an initial estimation yields a set of consistent (but not efficient) estimates for the model coefficients. For example, in a model with only exogenous explanatory variables, this would be based on ordinary least squares (OLS). In the presence of endogenous explanatory variables, two stage least squares (2SLS) would be necessary. This is also the case when a spatially lagged dependent variable is present. In the latter case, a number of papers have discussed the use of optimal weights (e.g., Lee 2003, Das et al. 2003, Kelejian et al. 2004, Lee 2007). In practice, the instruments consist of the exogenous variables and spatial lags of these, e.g., $\mathbf{H} = [\mathbf{X}, \mathbf{X}_L, \mathbf{X}_{LL}, \dots]$.

The initial consistent estimates provide the basis for the computation of a vector of residuals, say $\mathbf{u}$ (here, we do not use separate notation to distinguish the residuals from the error terms, since we always need residuals in practice). The residuals are used in a system of moment equations to provide a first consistent (but not efficient) estimate for the error autoregressive coefficient λ. The consistent estimate for λ is used to construct a weighting matrix that is necessary to obtain the optimal (consistent and efficient) GMM estimate of λ in a second iteration.

A third step then consists of estimating the regression coefficients (β and ρ, if appropriate) in a spatially weighted regression, using spatially filtered variables that incorporate the optimal GMM estimate of λ.

At this point, we could stop the estimation procedure and use the values of the regression coefficients, the corresponding residuals, and λ to construct a joint asymptotic variance-covariance matrix for all the coefficients (both regression and λ). Alternatively, we could go through another round of estimation using the updated residuals in the moment equations and potentially going through one more spatially weighted regression. While asymptotically, there are no grounds to prefer one over the other, in practice there may be efficiency gains from further iterations.

Finally, the estimation procedure as outlined in K-P only corrects for the presence of spatial autoregressive errors, but does not exploit the general structure of the heteroskedasticity in the estimation of the regression coefficients. The main contribution of K-P is to derive the moment equation such that the estimate for λ is consistent in the presence of general heteroskedasticity. The initial GM estimator presented in Kelejian and Prucha (1998, 1999) is only consistent under the assumption of absence of heteroskedasticity. We will need to further consider if the incorporation of both spatial autoregressive and heteroskedastic structures for the error variance improves the efficiency of the regression coefficients.

4 General Moment Equations

The point of departure for K-P's estimation procedure are two moment conditions, expressed as functions of the innovation terms ε. They are:

$$
\begin{aligned}
n^{-1}\mathrm{E}[\varepsilon_L' \varepsilon_L] &= n^{-1}\mathrm{tr}[\mathbf{W}\mathrm{diag}[E(\varepsilon_i^2)]\mathbf{W}'] && (16)\\
n^{-1}\mathrm{E}[\varepsilon_L' \varepsilon] &= 0, && (17)
\end{aligned}
$$

where ε_L is the spatially lagged innovation vector and tr stands for the matrix trace operator. The main difference with the moment equations in Kelejian and Prucha (1999) is that the innovation vector is allowed to be heteroskedastic of general form, hence the inclusion of the term $\mathrm{diag}[E(\varepsilon_i^2)]$. In the absence of heteroskedasticity, this term simplifies to $\sigma^2 n^{-1}\mathrm{tr}[\mathbf{W}\mathbf{W}']$. However, to carry out the GM estimation in this case, an additional equation is needed for σ^2.

4.1 Simplifying Notation

K-P introduce a number of simplifying notations that allow the moment conditions to be written in a very concise form. Specifically, they define

$$
\begin{aligned}
\mathbf{A}_1 &= \mathbf{W}'\mathbf{W} - \mathrm{diag}(\mathbf{w}_{.i}'\mathbf{w}_{.i}) && (18)\\
\mathbf{A}_2 &= \mathbf{W}, && (19)
\end{aligned}
$$

where $\mathbf{w_{.i}}$ is the i-th column of the weights matrix $\mathbf{W}$. Upon further inspection, we see that each element i of the diagonal matrix $\mathrm{diag}(\mathbf{w}_{.i}'\mathbf{w}_{.i})$ consists of the sum of squares of the weights in the i-th column. In what follows, we will designate this diagonal matrix as $\mathbf{D}$.

Using the new notation, the moment conditions become:

$$
\begin{aligned}
n^{-1}\mathrm{E}[\varepsilon'\mathbf{A_1}\varepsilon] &= 0 && (20)\\
n^{-1}\mathrm{E}[\varepsilon'\mathbf{A_2}\varepsilon] &= 0 && (21)
\end{aligned}
$$

In order to operationalize these equations, the (unobservable) innovation terms ε are replaced by their counterpart expressed as a function of regression residuals. Since $\mathbf{u} = \lambda\mathbf{u}_L + \varepsilon$, it follows that $\varepsilon = \mathbf{u} - \lambda\mathbf{u}_L = \mathbf{u}_s$, the spatially filtered residuals. The operational form of the moment conditions is then:

$$
\begin{aligned}
n^{-1}\mathrm{E}[\mathbf{u}_s'\mathbf{A_1}\mathbf{u}_s] &= 0 && (22)\\
n^{-1}\mathrm{E}[\mathbf{u}_s'\mathbf{A_2}\mathbf{u}_s] &= 0 && (23)
\end{aligned}
$$

4.2 Nonlinear Least Squares

The initial consistent estimate for λ is obtained by solving the moment conditions for this parameter, which is contained within $\mathbf{u}_s$. The parameter enters both linearly and in a quadratic form. Since there are two equations, there is no solution that actually sets the results to zero for both equations. Consequently, we try to get as close to this as possible and use a least squares rationale. In other words, if we consider the LHS of the equation as a deviation from zero, we try to minimize the sum of squared deviations.

In order to implement this in practice, K-P reorganize the equations as explicit functions of λ and λ^2. This takes on the general form:

$$\mathbf{m} = \mathbf{g} - \mathbf{G}\begin{bmatrix}\lambda \\ \lambda^2\end{bmatrix}, \tag{24}$$

such that an initial estimate of λ is obtained as a nonlinear least squares solution to these equations, $\text{argmin}_\lambda(\mathbf{m'm})$.

The vector $\mathbf{g}$ is a 2×1 vector with the following elements, as given by K-P:

$$\mathbf{g}_1 = n^{-1}\mathbf{u'A}_1\mathbf{u} \tag{25}$$

$$\mathbf{g}_2 = n^{-1}\mathbf{u'A}_2\mathbf{u}, \tag{26}$$

with the $\mathbf{A}_{1,2}$ as defined above, and $\mathbf{u}$ is a vector of residuals.

For computational purposes, it is actually more effective to work out these expressions and turn them into cross products of the residuals and their spatial lags. This yields:

$$\mathbf{g}_1 = n^{-1}[\mathbf{u}'_L\mathbf{u}_L - \mathbf{u'Du}] \tag{27}$$

$$\mathbf{g}_2 = n^{-1}\mathbf{u'u}_L, \tag{28}$$

The matrix $\mathbf{G}$ is a 2×2 matrix with the following elements, as given by K-P:

$$\mathbf{G}_{11} = 2n^{-1}\mathbf{u'W'A}_1\mathbf{u} \tag{29}$$

$$\mathbf{G}_{12} = -n^{-1}\mathbf{u'W'A}_1\mathbf{Wu} \tag{30}$$

$$\mathbf{G}_{21} = n^{-1}\mathbf{u'W'}(\mathbf{A}_2 + \mathbf{A}'_2)\mathbf{u} \tag{31}$$

$$\mathbf{G}_{22} = -n^{-1}\mathbf{u'W'A}_2\mathbf{Wu} \tag{32}$$

As before, this simplifies into a number of expressions consisting of cross products of the residuals and their spatial lags. Specifically, note that $\mathbf{Wu} = \mathbf{u}_L$, and $\mathbf{WWu} = \mathbf{u}_{LL}$, and, similarly, $\mathbf{u'W'} = \mathbf{u'}_L$ and $\mathbf{u'W'W'} = \mathbf{u'}_{LL}$.

Considering each expression in turn, we find:

$$\mathbf{G}_{11} = 2n^{-1}\mathbf{u'}_L\mathbf{A}_1\mathbf{u} \tag{33}$$

$$= 2n^{-1}[\mathbf{u'}_{LL}\mathbf{u}_L - \mathbf{u'}_L\mathbf{Du}] \tag{34}$$

$$\mathbf{G}_{12} = -n^{-1}\mathbf{u'}_L\mathbf{A_1u}_L \tag{35}$$

$$= -n^{-1}[\mathbf{u'}_{LL}\mathbf{u}_{LL} - \mathbf{u'}_L\mathbf{Du}_L] \tag{36}$$

$$\mathbf{G}_{21} = n^{-1}\mathbf{u'}_L(\mathbf{W} + \mathbf{W'})\mathbf{u} \tag{37}$$

$$= n^{-1}[\mathbf{u'}_L\mathbf{u}_L + \mathbf{u'}_{LL}\mathbf{u}] \tag{38}$$

$$\begin{aligned}
\mathbf{G}_{22} &= -n^{-1}\mathbf{u}'_L\mathbf{W}\mathbf{u}_L & (39)\\
&= -n^{-1}\mathbf{u}'_L\mathbf{u}_{LL} & (40)
\end{aligned}$$

So, in summary, in order to compute the six elements of $\mathbf{g}$ and $\mathbf{G}$, we need five cross product terms: $\mathbf{u}'\mathbf{u}_L$, $\mathbf{u}'\mathbf{u}_{LL}$, $\mathbf{u}'_L\mathbf{u}_L$, $\mathbf{u}'_L\mathbf{u}_{LL}$, and $\mathbf{u}'_{LL}\mathbf{u}_{LL}$. In addition, we need three weighted cross products: $\mathbf{u}'\mathbf{D}\mathbf{u}$, $\mathbf{u}'_L\mathbf{D}\mathbf{u}$, and $\mathbf{u}'_L\mathbf{D}\mathbf{u}_L$ (note that $\mathbf{D}\mathbf{u}$ only needs to be computed once).

Given a vector of residuals (from OLS, 2SLS or even Generalized Spatial 2SLS), the expression for $\mathbf{g}$ and $\mathbf{G}$ give us a way to obtain a consistent estimate for λ.

4.3 Weighted Nonlinear Least Squares

The estimates for λ obtained from the nonlinear least squares are consistent, but not efficient. Optimal estimates are found from a weighted nonlinear least squares procedure, or, $\text{argmin}_\lambda \mathbf{m}'\Psi^{-1}\mathbf{m}$, where Ψ is a weighting matrix. The optimal weights correspond to the inverse variance of the moment conditions.

K-P show the general expression for the elements of the 2×2 matrix Ψ to be of the form:

$$\psi_{q,r} = (2n)^{-1}\text{tr}[(\mathbf{A}_q + \mathbf{A}'_q)\Sigma(\mathbf{A}_r + \mathbf{A}'_r)\Sigma] + n^{-1}\mathbf{a}'_q\Sigma\mathbf{a}_r, \qquad (41)$$

for $q, r = 1, 2$ and with Σ as a diagonal matrix with as elements $(u_i - \lambda u_{L_i})^2 = u_{s_i}^2$, i.e., the squares of the spatially filtered residuals. The second term in this expression is quite complex, and will be examined more closely below. However, it is important to note that this term becomes zero when there are only exogenous explanatory variables in the model (i.e., when OLS is applicable). The term derives from the expected value of a cross product of expressions in the $\mathbf{Z}$ matrix and the error term $\mathbf{u}$. Hence, when no endogenous variables are included in $\mathbf{Z}$, the expected value of this cross product amounts to $\text{E}[\mathbf{u}] = 0$.

4.3.1 OLS Estimation

In the simplest case when no endogenous variables are present in the model, we only need to consider the trace term to obtain the elements of $\psi_{q,r}$. Note that $\mathbf{A}_1$ is symmetric, so that $\mathbf{A}_1 + \mathbf{A}'_1 = 2\mathbf{A}_1$. Also, $\mathbf{A}_2 = \mathbf{W}$ so that we don't need the extra notation at this point.

Consequently, we obtain the following results:

$$\begin{aligned}
\psi_{11} &= (2n)^{-1}\text{tr}[(2\mathbf{A}_1)\Sigma(2\mathbf{A}_1)\Sigma] & (42)\\
&= 2n^{-1}\text{tr}[\mathbf{A}_1\Sigma\mathbf{A}_1\Sigma], & (43)\\
\psi_{12} &= (2n)^{-1}\text{tr}[(2\mathbf{A}_1)\Sigma(\mathbf{W} + \mathbf{W}')\Sigma] & (44)\\
&= n^{-1}\text{tr}[\mathbf{A}_1\Sigma(\mathbf{W} + \mathbf{W}')\Sigma], & (45)\\
\psi_{21} &= \psi_{12}, & (46)\\
\psi_{22} &= (2n)^{-1}\text{tr}[(\mathbf{W} + \mathbf{W}')\Sigma(\mathbf{W} + \mathbf{W}')\Sigma & (47)
\end{aligned}$$

For computational purposes, it is important to keep in mind that while the matrices $\mathbf{A}_{1,2}$ are of dimension $n \times n$, they are typically very sparse. Furthermore, the matrix Σ is a diagonal matrix, such that post-multiplying a matrix by Σ amounts to re-scaling the columns of that matrix by the elements on the diagonal of Σ.

4.3.2 Standard 2SLS Estimation

The expression for $\mathbf{a}_r$, with $r = 1, 2$, for the case where the estimates are obtained from 2SLS is given in K-P as follows:

$$\mathbf{a}_r = (\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{HP}\alpha_r, \tag{48}$$

with $\mathbf{H}$ as a $n \times p$ matrix of instruments,

$$\mathbf{P} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})[(n^{-1}\mathbf{Z}'\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})]^{-1}, \tag{49}$$

as a matrix of dimension $p \times k$, and

$$\alpha_r = -n^{-1}[\mathbf{Z}'(\mathbf{I} - \lambda\mathbf{W}')(\mathbf{A}_r + \mathbf{A}'_r)(\mathbf{I} - \lambda\mathbf{W})\mathbf{u}], \tag{50}$$

where α_r is a vector of dimension $k \times 1$. As a result, $\mathbf{a}_r$ is of dimension $n \times 1$.

First, let's take a closer look at α_r, for $r = 1, 2$. Note that $\mathbf{Z}'(\mathbf{I} - \lambda\mathbf{W}')$ can be written in a simpler form as $\mathbf{Z}'_s$. Similarly, $(\mathbf{I}-\lambda\mathbf{W})\mathbf{u}$ is $\mathbf{u}_s$. For both filtered variables, we use the value of λ from the unweighted nonlinear least squares.

For α_1, since $\mathbf{A}_1$ is symmetric, $\mathbf{A}_1 + \mathbf{A}'_1 = 2\mathbf{A}_1$, and the corresponding expression can be written as:

$$\begin{aligned} \alpha_1 &= -2n^{-1}[\mathbf{Z}'_s\mathbf{A}_1\mathbf{u}_s] & (51) \\ &= -2n^{-1}[\mathbf{Z}'_{s_L}\mathbf{u}_{s_L} - \mathbf{Z}'_s\mathbf{D}\mathbf{u}_s], & (52) \end{aligned}$$

where $\mathbf{Z}'_{s_L}$ and $\mathbf{u}_{s_L}$ are the spatial lags (using weights matrix $\mathbf{W}$) of respectively the spatially filtered $\mathbf{Z}$ and the spatially filtered $\mathbf{u}$. For α_2, the expression is:

$$\begin{aligned} \alpha_2 &= -n^{-1}[\mathbf{Z}'_s(\mathbf{W} + \mathbf{W}')\mathbf{u}_s] & (53) \\ &= -n^{-1}[\mathbf{Z}'_s\mathbf{u}_{s_L} + \mathbf{Z}'_{s_L}\mathbf{u}_s] & (54) \end{aligned}$$

Pre-multiplying the respective expressions for α_1 and α_2 with the matrix $\mathbf{P}$ yields a vector of dimension $p \times 1$ (p is the number of instruments). At first sight, the presence of the inverse matrix $(\mathbf{I}-\lambda\mathbf{W})^{-1}$ in the expression for $\mathbf{a}_r$ for $r = 1, 2$ would seem to preclude large data analysis. However, we can exploit the approach outlined in Smirnov (2005). The typical power expansion (Leontief expansion) of the inverse matrix yields:

$$(\mathbf{I} - \lambda\mathbf{W})^{-1} = \mathbf{I} + \lambda\mathbf{W} + \lambda^2\mathbf{WW} + \dots \tag{55}$$

As such, this does not help the computational aspect, since the weights matrices involved are still of dimension $n \times n$. However, since $(\mathbf{I} - \lambda\mathbf{W})^{-1}$ pre-multiplies

$\mathbf{H}$, we see that the multiplication of each term in the expansion with $\mathbf{H}$ amounts to no more than a series of spatial lag operations, such as $\lambda\mathbf{WH}$, $\lambda^2\mathbf{W}(\mathbf{WH})$, etc. Depending on the value of λ, a reasonable approximation is readily obtained for a relatively low power in the expansion. For example, a value of λ of 0.5 (which is relatively large in practice) reduces to 0.00098 after a tenth power. Given that this is multiplied with the elements of a spatial weights matrix, the effective lag operator is even smaller. In our example, assuming five neighbors on average, this would yield a multiplier in the lag operator of $0.00098/5 = 0.0002$.[1]

With these elements in hand, we obtain the terms $\mathbf{a}_1$ and $\mathbf{a}_2$ needed in the expression $\mathbf{a'}_q\Sigma\mathbf{a}_r$, which, together with the trace terms, yield the four elements of the matrix Ψ.

4.3.3 Spatially Weighted Estimation

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Arraiz, I., Drukker, D. M., Kelejian, H. H., and Prucha, I. R. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science*, 50:592–614.

Das, D., Kelejian, H. H., and Prucha, I. R. (2003). Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, 82:1–27.

Kelejian, H. H. and Prucha, I. (1998). A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

Kelejian, H. H. and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.

Kelejian, H. H. and Prucha, I. R. (2009). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*. forthcoming.

Kelejian, H. H., Prucha, I. R., and Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In LeSage, J. P. and Pace, R. K., editors, *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, pages 163–198. Elsevier Science Ltd., Oxford, UK.

[1] For high values of λ, much higher powers are needed in order to obtain a reasonable approximation. For example, 0.9 to the tenth power is still 0.35, but after fifty powers (i.e., fifty lag operations) it is 0.005.

Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22:307–335.

Lee, L.-F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137:489–514.

Smirnov, O. (2005). Computation of the information matrix for models with spatial interaction on a lattice. *Journal of Computational and Graphical Statistics*, 14:910–927.