GMM Estimation of Spatial Error Autocorrelation with Heteroskedasticity

Luc Anselin

July 29, 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) (jointly 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:

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

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:

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u}, \tag{1}$$

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:

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

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:

$$\begin{aligned}
\mathbf{y}_s &= \mathbf{y} - \lambda\mathbf{Wy} \\
&= \mathbf{y} - \lambda\mathbf{y}_L \\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{y} \\
\mathbf{X}_s &= \mathbf{X} - \lambda\mathbf{WX} \\
&= \mathbf{X} - \lambda\mathbf{X}_L \\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{X} \\
\mathbf{Wy}_s &= \mathbf{Wy} - \lambda\mathbf{WWy} \\
&= \mathbf{y_L} - \lambda\mathbf{y}_{LL} \\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{Wy} \\
\mathbf{Z}_s &= \mathbf{Z} - \lambda\mathbf{WZ} \\
&= \mathbf{Z} - \lambda\mathbf{Z}_L \\
&= (\mathbf{I} - \lambda\mathbf{W})\mathbf{Z}
\end{aligned}$$

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 insturments (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 in a feasible generalized least squares procedure (FGLS) 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:

$$n^{-1}\mathrm{E}[\varepsilon_L'\varepsilon_L] = n^{-1}\mathrm{tr}[\mathbf{W}\mathrm{diag}[E(\varepsilon_i^2)]\mathbf{W}'] \tag{2}$$

$$n^{-1}\mathrm{E}[\varepsilon_L'\varepsilon] = 0, \tag{3}$$

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 Equation 2. In the absence of heteroskedasticity, the RHS of the first condition 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

$$\mathbf{A}_1 = \mathbf{W}'\mathbf{W} - \mathrm{diag}(\mathbf{w}_{.i}'\mathbf{w}_{.i})$$
$$\mathbf{A}_2 = \mathbf{W},$$

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:

$$n^{-1}\mathrm{E}[\varepsilon'\mathbf{A_1}\varepsilon] = 0$$
$$n^{-1}\mathrm{E}[\varepsilon'\mathbf{A_2}\varepsilon] = 0$$

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:

$$n^{-1}\mathrm{E}[\mathbf{u}_s'\mathbf{A_1}\mathbf{u}_s] = 0 \tag{4}$$

$$n^{-1}\mathrm{E}[\mathbf{u}_s'\mathbf{A_2}\mathbf{u}_s] = 0 \tag{5}$$

4.2 Nonlinear Least Squares

The initial consistent estimate for λ is obtained by solving the moment conditions in Equations 4 and 5 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:

$$\mathbf{m} = \mathbf{g} - \mathbf{G} \begin{bmatrix} \lambda \\ \lambda^2 \end{bmatrix} = 0,$$

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:

$$\begin{aligned} \mathbf{g}_1 &= n^{-1}\mathbf{u'A}_1\mathbf{u} \\ \mathbf{g}_2 &= n^{-1}\mathbf{u'A}_2\mathbf{u}, \end{aligned}$$

with the $\mathbf{A}_{1,2}$ as defined above, and $\mathbf{u}$ is a vector of residuals. Note that there is actually no gain in using the notation $\mathbf{A}_2$, since it is the same as $\mathbf{W}$.

For computational purposes, it is useful to work out these expressions and express them as cross products of the residuals and their spatial lags. This yields:

$$\begin{aligned} \mathbf{g}_1 &= n^{-1}[\mathbf{u}'_L\mathbf{u}_L - \mathbf{u'Du}] \\ \mathbf{g}_2 &= n^{-1}\mathbf{u'u}_L, \end{aligned}$$

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

$$\begin{aligned} \mathbf{G}_{11} &= 2n^{-1}\mathbf{u'W'A}_1\mathbf{u} \\ \mathbf{G}_{12} &= -n^{-1}\mathbf{u'W'A}_1\mathbf{Wu} \\ \mathbf{G}_{21} &= n^{-1}\mathbf{u'W'}(\mathbf{A}_2 + \mathbf{A}'_2)\mathbf{u} \\ \mathbf{G}_{22} &= -n^{-1}\mathbf{u'W'A}_2\mathbf{Wu} \end{aligned}$$

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:

$$\begin{aligned} \mathbf{G}_{11} &= 2n^{-1}\mathbf{u'}_L\mathbf{A}_1\mathbf{u} \\ &= 2n^{-1}[\mathbf{u'}_{LL}\mathbf{u}_L - \mathbf{u'}_L\mathbf{Du}] \end{aligned}$$

$$\begin{aligned} \mathbf{G}_{12} &= -n^{-1}\mathbf{u'}_L\mathbf{A_1}\mathbf{u}_L \\ &= -n^{-1}[\mathbf{u'}_{LL}\mathbf{u}_{LL} - \mathbf{u'}_L\mathbf{Du}_L] \end{aligned}$$

$$\begin{aligned} \mathbf{G}_{21} &= n^{-1}\mathbf{u'}_L(\mathbf{W} + \mathbf{W'})\mathbf{u} \\ &= n^{-1}[\mathbf{u'}_L\mathbf{u}_L + \mathbf{u'}_{LL}\mathbf{u}] \end{aligned}$$

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

So, in summary, in order to compute the six elements of $\mathbf{g}$ and $\mathbf{G}$, we need five cross product terms: $\mathbf{u'u}_L$, $\mathbf{u'u}_{LL}$, $\mathbf{u'}_L\mathbf{u}_L$, $\mathbf{u'}_L\mathbf{u}_{LL}$, and $\mathbf{u'}_{LL}\mathbf{u}_{LL}$. In addition, we need three weighted cross products: $\mathbf{u'Du}$, $\mathbf{u'}_L\mathbf{Du}$, and $\mathbf{u'}_L\mathbf{Du}_L$ (note that $\mathbf{Du}$ only needs to be computed once). Alternatively, if the matrix $\mathbf{A}_1$ is stored efficiently in sparse form, we can use the cross products $\mathbf{u'A}_1\mathbf{u}$, $\mathbf{u'}_L\mathbf{A}_1\mathbf{u}$ and $\mathbf{u'}_L\mathbf{A}_1\mathbf{u}_L$.

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:

$$
\psi_{q,r} = (2n)^{-1}\text{tr}[(\mathbf{A}_q + \mathbf{A'}_q)\Sigma(\mathbf{A}_r + \mathbf{A'}_r)\Sigma] + n^{-1}\mathbf{a'}_q\Sigma\mathbf{a}_r, \tag{6}
$$

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 in Sections 4.3.2 and 4.3.3 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:

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

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:

$$\mathbf{a}_r = (\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{HP}\alpha_r, \tag{7}$$

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

$$\mathbf{P} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})[(n^{-1}\mathbf{Z}'\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})]^{-1}, \tag{8}$$

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

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

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:

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

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:

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

For easy computation, we will need $\mathbf{Z}_s$ and $\mathbf{u}_s$, as well as their spatial lags, $\mathbf{Z}_{s_L}$ and $\mathbf{u}_{s_L}$, and the three cross products $\mathbf{Z}'_s\mathbf{A}_1\mathbf{u}_s$, $\mathbf{Z}'_s\mathbf{u}_{s_L}$ and $\mathbf{Z}'_{s_L}\mathbf{u}_s$.

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:

$$(\mathbf{I} - \lambda\mathbf{W})^{-1} = \mathbf{I} + \lambda\mathbf{W} + \lambda^2\mathbf{W}\mathbf{W} + \ldots$$

As such, this does not help in computation, 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 $\lambda\mathbf{W}\mathbf{H}$, or $\lambda\mathbf{H}_L$, $\lambda^2\mathbf{W}(\mathbf{W}\mathbf{H})$, or $\lambda^2\mathbf{H}_{LL}$, 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

When the residuals used in the GMM estimation for λ are not the result of a standard procedure (such as OLS or 2SLS), but instead of a spatially weighted regression (such as SWLS or GS2SLS), the expressions for the optimal weighting matrix are different in two respects. The main difference is that the inverse term is no longer present in Equation 7 for $\mathbf{a}_r$, which now becomes:

$$\mathbf{a}_r = \mathbf{H}\mathbf{P}\alpha_r$$

The second difference is that the spatially filtered $\mathbf{Z}_s$ are used in the expression for $\mathbf{P}$ instead of $\mathbf{Z}$. The expression for $\mathbf{P}$ thus becomes:

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

5 Estimation Steps

For the purposes of this discussion, we will express the model in a generic form as in Equation 1, which we repeat here:

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u}.$$

[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.

This encompasses the two main cases. In the first, no endogenous variables are present (and thus also no spatially lagged dependent variable) and $\mathbf{Z} = \mathbf{X}$ in the usual notation. In the second case, endogenous variables are present. By convention, we will sort the variables such that the exogenous variables come first and the endogenous second. In the special case of a mixed regressive spatial autoregressive model, $\mathbf{Z} = [\mathbf{X}, \mathbf{Wy}]$, and $\delta = [\beta', \rho]'$.

The actual estimation proceeds in several steps, which are detailed in what follows.

5.1 Step 1 – Initial Estimates

The initial set of estimates, which we will denote as δ_1, are obtained from either OLS or 2SLS estimation of the model. In case of OLS, this yields:

$$\delta_{1,OLS} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y},$$

When endogenous variables are present (including the case of a spatially lagged dependent variable), a matrix of instruments $\mathbf{H}$ is needed and estimation follows from:

$$\delta_{1,2SLS} = (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{y},$$

with $\hat{\mathbf{Z}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}$, or, in one expression, as:

$$\delta_{1,2SLS} = [\mathbf{Z}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{y}.$$

In the presence of a spatially lagged dependent variable, the instruments should include multiple orders of spatial lags of the exogenous explanatory variables. In practice, up to two orders may be sufficient, such that $\mathbf{H} = [\mathbf{X}, \mathbf{X}_L, \mathbf{X}_{LL}]$. As always, care must be taken to avoid multicollinear instruments. For examples, this may be a problem when indicator variables are included in the model.

The estimates δ_1 yield an initial vector of residuals, $\mathbf{u}_1$ as:

$$\mathbf{u}_1 = \mathbf{y} - \mathbf{Z}\delta_1.$$

5.2 Step 2 – Consistent Estimation of λ

A first consistent estimate for λ, say λ_1 is obtained by substituting $\mathbf{u}_1$ into the moment equations of Section 4.2 and finding a solution by means of nonlinear least squares.

5.3 Step 3 – Efficient and Consistent Estimation of λ

An efficient estimate of λ is obtained by substituting the values of λ_1 and $\mathbf{u}_1$ into the elements of Equation 6 as specified in Section 4.3.2. This yields the weighting matrix $\Psi(\lambda_1)$, which then allows for a weighted nonlinear least squares solution to the moments equations. This results in the estimate λ_2.

At this point, we could stop and move to the construction of the asymptotic variance matrix of the estimates, as outlined in Section 6.1. For example, this would be relevant if we were only interested in testing the null hypothesis $\mathrm{H}_0 : \lambda = 0$.

Typically, however, one does not stop here and moves on to a spatially weighted estimation of the regression coefficients, which takes into account the consistent and efficient estimate λ_2 of the nuisance parameter. Note that only consistency of λ is required to obtain consistent estimates of the δ coefficients. The use of the more efficient λ_2 should result in more efficient estimates of δ as well, but consistency is not affected by this.

5.4 Step 4 – Spatially Weighted Estimation

The rationale behind spatially weighted estimation is that a simple transformation (the so-called spatial Cochrane-Orcutt) removes the spatial dependence from the error term in the regression equation:

$$\begin{aligned}(\mathbf{I} - \lambda \mathbf{W})y &= (\mathbf{I} - \lambda \mathbf{W})\mathbf{Z}\delta + (\mathbf{I} - \lambda \mathbf{W})\mathbf{u} \\ \mathbf{y}_s &= \mathbf{Z}_s\delta + \varepsilon,\end{aligned}$$

where $\mathbf{y}_s$ and $\mathbf{Z}_s$ are filtered variables, and ε is a heteroskedastic, but not spatially correlated innovation term.

We distinguish between two situations. In one, there are only exogenous variables in the regression, so that Spatially Weighted Least Squares (SWLS) is an appropriate estimation method. In the other, the presence of endogenous variables requires the use of 2SLS. The special case of a regression with a spatially lagged dependent variable also falls in this category.

5.4.1 Spatially Weighted Least Squares – SWLS

Spatially Weighted Least Squares is simply OLS applied to the spatially filtered variables:

$$\delta_{2,SWLS} = (\mathbf{Z'}_s\mathbf{Z}_s)^{-1}\mathbf{Z'}_s\mathbf{y}_s.$$

5.4.2 Generalized Spatial Two Stage Least Squares - GS2SLS

Similarly, Generalized Spatial Two Stage Least Squares is 2SLS applied to the spatially filtered variables:

$$\delta_{2,GS2SLS} = [\mathbf{Z'}_s\mathbf{H}(\mathbf{H'H})^{-1}\mathbf{H'Z}_s]^{-1}\mathbf{Z'}_s\mathbf{H}(\mathbf{H'H})^{-1}\mathbf{y}_s.$$

Note that the instrument matrix $\mathbf{H}$ is the same as in Step 1.

5.4.3 Residuals

The new estimate coefficient vector δ_2 yields an updated vector of residuals as:

$$\mathbf{u}_2 = \mathbf{y} - \mathbf{Z}\delta_2.$$

Note that the residuals are computed using the original variables and not the spatially filtered variables.

5.5 Step 5 – Iteration

The updated residual vector $\mathbf{u}_2$ can now be used to obtain a new estimate for λ. Since this is based on a spatially weighted regression, the proper elements of the weighting matrix Ψ are given in Section 4.3.3, with λ_2 and $\mathbf{u}_2$ substituted in the expressions. The solution by means of weighted nonlinear least squares yields the consistent and efficient estimate λ_3.

At this point, the value of λ_3 can be used together with δ_2 to construct an asymptotic variance matrix, as outlined in Section 6.2. This allows for joint inference on the coefficients δ and the spatial autoregressive term λ.

Alternatively, the new value of λ_3 could be used in an updated spatially weighted estimation to yield a new set of estimates for δ and an associated residual vector $\mathbf{u}$. These can then be substituted in the moment equations and in Ψ to result in a new estimate for λ. This iteration can be continued until a suitable stopping criterion is met. To date, there is no evidence as to the benefits of iteration beyond λ_3, but this remains a subject for further investigation.

6 Asymptotic Variance Matrix

6.1 Standard Estimation

6.2 Spatially Weighted Estimation

7 Accounting for Heteroskedasticity in FGLS

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.

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.