

GMM Estimation of Spatial Error Autocorrelation with and without Heteroskedasticity

Luc Anselin

July 14, 2011

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 (2010), Arraiz et al. (2010) and Drukker et al. (2010, 2011) (jointly referred to in what follows as K-P-D). 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 Anselin-Rey 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 $\mathbf{W}\mathbf{y}$ as the spatial lag term with spatial autoregressive parameter ρ , $\mathbf{X}$ as an $n \times k$ matrix of observations on exogenous explanatory variables with $k \times 1$ coefficient vector β , and a $n \times 1$ vector of errors $\mathbf{u}$. The specification can be made more general, by including additional endogenous variables:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{Y}\gamma + \mathbf{u},$$

where $\mathbf{Y}$ is a $n \times s$ matrix of observations on endogenous variables other than the spatially lagged dependent variable, with associated coefficient vector γ .

When endogenous variables are included (either a spatial lag, other endogenous variables, or both), a $n \times p$ matrix of instruments $\mathbf{H}$ will be needed.

An alternative way to express the model is as:

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u}, \quad (1)$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}, \mathbf{W}\mathbf{y}]$ and the $k + s \times 1$ coefficient vector is rearranged as $\delta = [\beta' \gamma' \rho']'$ (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 $E[\varepsilon_i^2] = \sigma_i^2$. All other assumptions are standard.

Note that, in contrast to the general case outlined by K-P-D, we take the weights matrix in the spatial lag and in the spatial error part to be the same ($\mathbf{W}$). In K-P-D, 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-D 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., $\mathbf{W}\mathbf{y}$ 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-D 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:

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

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.

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-D 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-D 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.

3.1 Selection of Instruments

Instruments are needed whenever endogenous variables are present in the regression specification and 2SLS is carried out. We consider the most general case, where $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}, \mathbf{W}\mathbf{y}]$, with $\mathbf{Y}$ as a $n \times s$ matrix of observations on endogenous variables other than the spatially lagged dependent variable.

We need instruments for the spatially lagged dependent variable and for the endogenous variables. For the spatial lag, a number of papers have discussed the use of optimal instruments (e.g., Lee 2003, Das et al. 2003, Kelejian et al. 2004, Lee 2007). The basis for this follows from the reduced form of the spatial lag specification:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\beta + \mathbf{u}),$$

so that the optimal instruments for $\mathbf{W}\mathbf{y}$ consist of $\mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta$. In practice, the instruments consist of the exogenous variables and spatial lags of these, e.g., $\mathbf{H} = [\mathbf{X}, \mathbf{X}_L, \mathbf{X}_{LL}, \dots]$.

When additional endogenous variables are included in the regression specification, instruments must be included for those as well. We refer to this instruments matrix as $\mathbf{Q}$, for which the usual identification restrictions need to be satisfied, such that there are at least as many instruments as endogenous variables. This is elaborated upon for a fully specified system of simultaneous equations in Kelejian and Prucha (2004). For our purposes, consider the simple case of one additional endogenous variable $\mathbf{y}_2$, which itself is determined by a number of additional exogenous variables $\mathbf{X}^*$ in a linear specification for its reduced form:

$$\mathbf{y}_2 = \mathbf{X}^*\beta^* + \mathbf{v},$$

with β^* as the coefficient vector and $\mathbf{v}$ as an idiosyncratic error. Substituting the result back into the reduced form for $\mathbf{y}$ yields:¹

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\beta + \mathbf{X}^*\beta^* + \mathbf{v} + \mathbf{u}).$$

This would suggest that $\mathbf{X}^*$ and its spatial lags should be included in the instrument matrix $\mathbf{H}$. In general, in the absence of a full systems specification,

¹In general, the second endogenous variable may also depend on the spatial lag, in which case the simple expression for the reduced form given here does not hold.

it is not possible to obtain optimal instruments. In the most general case, the instrument matrix $\mathbf{H}$ should consist of both $\mathbf{X}$ and its spatial lags as well as $\mathbf{Q}$ and its spatial lags:

$$\mathbf{H} = [\mathbf{X}, \mathbf{X}_L, \mathbf{X}_{LL}, \dots, \mathbf{Q}, \mathbf{Q}_L, \mathbf{Q}_{LL}, \dots]$$

The effect of the selection of instruments on the efficiency of the estimators remains to be further investigated.

4 General Moment Equations

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

$$n^{-1}\mathbf{E}[\varepsilon'_L \varepsilon_L] = n^{-1}\text{tr}[\mathbf{W}\text{diag}[E(\varepsilon_i^2)]\mathbf{W}'] \quad (2)$$

$$n^{-1}\mathbf{E}[\varepsilon'_L \varepsilon] = 0, \quad (3)$$

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 $\text{diag}[E(\varepsilon_i^2)]$ in Equation 2. In the absence of heteroskedasticity, the RHS of the first condition simplifies to $\sigma^2 n^{-1}\text{tr}[\mathbf{W}\mathbf{W}']$. With σ^2 replaced by $\mathbf{E}[(n^{-1})\varepsilon'\varepsilon]$, the first moment condition then becomes:

$$\begin{aligned} n^{-1}\mathbf{E}[\varepsilon'_L \varepsilon_L] &= \mathbf{E}[n^{-1}\varepsilon'\varepsilon](n^{-1})\text{tr}[\mathbf{W}\mathbf{W}'] \\ &= n^{-1}\mathbf{E}[\varepsilon'(n^{-1})\text{tr}[\mathbf{W}\mathbf{W}']\mathbf{I}\varepsilon] \end{aligned}$$

4.1 Simplifying Notation

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

$$\mathbf{A}_1 = \mathbf{W}'\mathbf{W} - \text{diag}(\mathbf{w}'_i \mathbf{w}_i) \quad (4)$$

$$\mathbf{A}_2 = \mathbf{W}, \quad (5)$$

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 $\text{diag}(\mathbf{w}'_i \mathbf{w}_i)$ 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:

$$n^{-1}\mathbf{E}[\varepsilon'\mathbf{A}_1\varepsilon] = 0$$

$$n^{-1}\mathbf{E}[\varepsilon'\mathbf{A}_2\varepsilon] = 0$$

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:

$$n^{-1}E[\mathbf{u}'_s \mathbf{A}_1 \mathbf{u}_s] = 0 \quad (6)$$

$$n^{-1}E[\mathbf{u}'_s \mathbf{A}_2 \mathbf{u}_s] = 0 \quad (7)$$

4.2 Nonlinear Least Squares

The initial consistent estimate for λ is obtained by solving the moment conditions in Equations 6 and 7 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-D reorganize the equations as explicit functions of λ and λ^2 . This takes on the general form:

$$\mathbf{m} = \mathbf{g} - \mathbf{G} \begin{bmatrix} \lambda \\ \lambda^2 \end{bmatrix} = 0, \quad (8)$$

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

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

$$\mathbf{g}_1 = n^{-1} \mathbf{u}' \mathbf{A}_1 \mathbf{u} \quad (9)$$

$$\mathbf{g}_2 = n^{-1} \mathbf{u}' \mathbf{A}_2 \mathbf{u}, \quad (10)$$

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:

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

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

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

$$\mathbf{G}_{11} = 2n^{-1} \mathbf{u}' \mathbf{W}' \mathbf{A}_1 \mathbf{u} \quad (11)$$

$$\mathbf{G}_{12} = -n^{-1} \mathbf{u}' \mathbf{W}' \mathbf{A}_1 \mathbf{W} \mathbf{u} \quad (12)$$

$$\mathbf{G}_{21} = n^{-1} \mathbf{u}' \mathbf{W}' (\mathbf{A}_2 + \mathbf{A}'_2) \mathbf{u} \quad (13)$$

$$\mathbf{G}_{22} = -n^{-1} \mathbf{u}' \mathbf{W}' \mathbf{A}_2 \mathbf{W} \mathbf{u} \quad (14)$$

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

Considering each expression in turn, we find:

$$\begin{aligned}\mathbf{G}_{11} &= 2n^{-1}\mathbf{u}'_L\mathbf{A}_1\mathbf{u} \\ &= 2n^{-1}[\mathbf{u}'_{LL}\mathbf{u}_L - \mathbf{u}'_L\mathbf{D}\mathbf{u}]\end{aligned}\quad (15)$$

$$\begin{aligned}\mathbf{G}_{12} &= -n^{-1}\mathbf{u}'_L\mathbf{A}_1\mathbf{u}_L \\ &= -n^{-1}[\mathbf{u}'_{LL}\mathbf{u}_{LL} - \mathbf{u}'_L\mathbf{D}\mathbf{u}_L]\end{aligned}\quad (16)$$

$$\begin{aligned}\mathbf{G}_{21} &= n^{-1}\mathbf{u}'_L(\mathbf{W} + \mathbf{W}')\mathbf{u} \\ &= n^{-1}[\mathbf{u}'_L\mathbf{u}_L + \mathbf{u}'_{LL}\mathbf{u}]\end{aligned}\quad (17)$$

$$\begin{aligned}\mathbf{G}_{22} &= -n^{-1}\mathbf{u}'_L\mathbf{W}\mathbf{u}_L \\ &= -n^{-1}\mathbf{u}'_L\mathbf{u}_{LL}\end{aligned}\quad (18)$$

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). Alternatively, if the matrix $\mathbf{A}_1$ is stored efficiently in sparse form, we can use the cross products $\mathbf{u}'\mathbf{A}_1\mathbf{u}$, $\mathbf{u}'_L\mathbf{A}_1\mathbf{u}$ 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-D show the general expression for the elements of the 2×2 matrix Ψ to be of the form:

$$\psi_{q,r} = (2n)^{-1}\text{tr}[(\mathbf{A}_q + \mathbf{A}'_q)\Sigma(\mathbf{A}_r + \mathbf{A}'_r)\Sigma] + n^{-1}\mathbf{a}'_q\Sigma\mathbf{a}_r, \quad (19)$$

for $q, r = 1, 2$ and with Σ as a diagonal matrix with as elements $(u_i - \lambda u_{Li})^2 = u_{s_i}^2$, 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 second 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 $E[\mathbf{u}] = 0$.

The fundamental result from which the form of the variance-covariance matrix in Equation 19 follows was established in the central limit theorem for a linear quadratic form in Kelejian and Prucha (2001, pp. 226–227). In general, the quadratic form is specified as:

$$\begin{aligned} f &= \varepsilon' \mathbf{A} \varepsilon + \mathbf{b}' \varepsilon \\ &= \sum_i \sum_j a_{ij} \varepsilon_i \varepsilon_j + \sum_i b_i \varepsilon_i \end{aligned}$$

where the ε_i are independent error terms and the matrix $\mathbf{A}$ is taken to be symmetric (for the full set of assumptions, see Kelejian and Prucha 2001).² Under the maintained set of assumptions, the mean and variance of this quadratic form are:

$$\begin{aligned} \mu(f) &= \sum_{i=1}^n a_{ii} \sigma_i^2 \\ \sigma^2(f) &= 4 \sum_{i=1}^n \sum_{j=1}^{i-1} a_{ij}^2 \sigma_i^2 \sigma_j^2 + \sum_{i=1}^n b_i^2 \sigma_i^2 \\ &\quad + \sum_{i=1}^n a_{ii}^2 (\mu_i^{(4)} - \sigma_i^4) + 2 \sum_{i=1}^n b_i a_{ii} \mu_i^{(3)}, \end{aligned}$$

with $\sigma_i^2 = E[\varepsilon_i^2]$, $\mu_i^{(3)} = E[\varepsilon_i^3]$ and $\mu_i^{(4)} = E[\varepsilon_i^4]$. Note that when the diagonal elements of the matrix $\mathbf{A}$ are all zero, then $\mu_q = 0$ and the last two terms in the expression for the variance disappear. The variance-covariance matrix $\mathbf{\Psi}$ is constructed from the variances and covariances between $(n^{-1})E[\varepsilon' \mathbf{A}_1 \varepsilon]$ and $(n^{-1})E[\varepsilon' \mathbf{A}_2 \varepsilon]$. These conform to the general structure $f = \varepsilon' \mathbf{A}_{\mathbf{q}, \mathbf{r}} \varepsilon + \mathbf{b}_{\mathbf{q}, \mathbf{r}}' \varepsilon$ of Lemma A.1. in Kelejian and Prucha (2010, pp. 62–63) with mean and covariance as:

$$\begin{aligned} \mu_{(1,2)}(f) &= \sum_{i=1}^n a_{ii, (1,2)} \sigma_i^2 \\ \sigma_{q,r}(f) &= 2 \sum_{i=1}^n \sum_{j=1}^n a_{ij,q} a_{ij,r} \sigma_i^2 \sigma_j^2 + \sum_{i=1}^n b_{i,q} b_{i,r} \sigma_i^2 \\ &\quad + \sum_{i=1}^n a_{ii,q} a_{ii,r} [\mu_i^{(4)} - 3\sigma_i^4] + \sum_{i=1}^n (b_{i,q} a_{ii,r} + b_{i,r} a_{ii,q}) \mu_i^{(3)} \end{aligned}$$

²As argued in Kelejian and Prucha (2001, footnote 10), when $\mathbf{A}$ is not symmetric, it can be readily replaced by $(1/2)(\mathbf{A} + \mathbf{A}')$.

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:

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

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 $\mathbf{\Sigma}$ is a diagonal matrix, such that post-multiplying a matrix by $\mathbf{\Sigma}$ amounts to re-scaling the columns of that matrix by the elements on the diagonal of $\mathbf{\Sigma}$.

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-D as follows:

$$\mathbf{a}_r = (\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{H}\mathbf{P}\alpha_r, \quad (20)$$

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

$$\mathbf{P} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})[(n^{-1}\mathbf{Z}'\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})]^{-1}, \quad (21)$$

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). Pre-multiplying this result with the matrix $\mathbf{H}$ results in a vector of dimension $n \times 1$. We will refer to this as vector $\mathbf{v}_r$ (with $r = 1, 2$).

At first sight, the presence of the inverse matrix $(\mathbf{I} - \lambda\mathbf{W}')^{-1}$ in the expression for $\mathbf{a}_r$ in Equation 20 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:

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

As such, this does not help in computation, since the weights matrices involved are still of dimension $n \times n$. Also, the expression in Equation 20 pertains to the transpose of $\mathbf{W}$, i.e., $(\mathbf{I} - \lambda\mathbf{W}')^{-1}$. In order to apply the power expansion, it is easier to consider the transpose of $\mathbf{a}_r$:

$$\mathbf{a}'_r = \alpha'_r\mathbf{P}'\mathbf{H}'(\mathbf{I} - \lambda\mathbf{W})^{-1} = \mathbf{v}'_r(\mathbf{I} - \lambda\mathbf{W})^{-1}, \quad (22)$$

where $\mathbf{v}'_r$ is a vector of dimension $1 \times n$.

Using the power expansion yields:

$$\begin{aligned}\mathbf{a}'_r &= \mathbf{v}'_r[\mathbf{I} + \lambda\mathbf{W} + \lambda^2\mathbf{W}\mathbf{W} + \dots] \\ &= \mathbf{v}'_r + \lambda\mathbf{v}'_r\mathbf{W} + \lambda^2\mathbf{v}'_r\mathbf{W}\mathbf{W} + \dots\end{aligned}$$

This operation is relatively easy to implement computationally, since once $\lambda\mathbf{v}'_r\mathbf{W}$ is obtained (a vector of dimension $1 \times n$), all higher order terms consist of post-multiplying this vector with the (sparse) weights matrix $\mathbf{W}$ and the scalar λ . 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.³

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

³For 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.

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 20 for $\mathbf{a}_r$, which now becomes:

$$\mathbf{a}_r = \mathbf{H}\mathbf{P}\alpha_r \quad (23)$$

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:

$$\mathbf{P} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}_s)[(n^{-1}\mathbf{Z}_s'\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}_s)]^{-1} \quad (24)$$

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:

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

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, $\mathbf{Z} = [\mathbf{X}, \mathbf{W}\mathbf{y}]$, 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 $\hat{\mathbf{Z}} = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}$, or, in one expression, as:

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

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 19 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 $H_0 : \lambda = 0$.

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:

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

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:

$$\delta_{2,SWLS} = (\mathbf{Z}'_s \mathbf{Z}_s)^{-1} \mathbf{Z}'_s \mathbf{y}_s. \quad (25)$$

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:

$$\delta_{2,GS2SLS} = [\mathbf{Z}'_s \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1} \mathbf{H}'\mathbf{Z}_s]^{-1} \mathbf{Z}'_s \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1} \mathbf{H}'\mathbf{y}_s. \quad (26)$$

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

The asymptotic variance-covariance matrix Ω is for both the estimates of δ and the estimate of λ . For the spatial lag model, δ includes the spatial autoregressive coefficient ρ (for ease of notation, included as the last element in the vector). We distinguish between two major situations. In one, the estimates are based

on Steps 1–3. In other words, δ are the initial estimates of the regression model, and λ is the consistent and efficient GMM estimate. In the second case, the estimates are based on a spatially weighted estimation, including Steps 4 and 5. For each of these cases, we further distinguish between a model where only exogenous variables are present (estimated by OLS and SWLS), and a model where endogenous variables are included (estimated by 2SLS and GS2SLS). The latter encompasses the spatial lag model as a special case. In order to carry out inference, $n^{-1}\Omega$ is used instead of Ω as such.

To facilitate the expression for the asymptotic variance-covariance matrix, K-P-D introduce some additional notation. First, an auxiliary 2×1 vector $\mathbf{J}$ is defined, constructed as:

$$\mathbf{J} = \mathbf{G} \begin{bmatrix} 1 \\ 2\lambda \end{bmatrix}, \quad (27)$$

where the elements of the 2×2 matrix $\mathbf{G}$ are from Equations 15–18. K-P-D further define the overall weighting matrix Ψ_o as consisting of four submatrices, as in:

$$\Psi_o = \begin{bmatrix} \Psi_{\delta\delta} & \Psi_{\delta\lambda} \\ \Psi'_{\delta\lambda} & \Psi \end{bmatrix},$$

where Ψ is the weighting matrix of Equation 19, and the other submatrices differ between the standard estimation (Section 6.1) and the spatially weighted estimation (Section 6.2). Note that the dimension of the matrix Ψ_o is $(p+2) \times (p+2)$, where p is the number of instruments, and Ψ is of dimension 2×2 .

The estimator for the variance-covariance matrix takes on the general form:

$$\Omega = \begin{bmatrix} \mathbf{P}' & 0 \\ 0 & (\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1}\mathbf{J}'\Psi^{-1} \end{bmatrix} \Psi_o \begin{bmatrix} \mathbf{P} & 0 \\ 0 & \Psi^{-1}\mathbf{J}(\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1} \end{bmatrix}, \quad (28)$$

where $\mathbf{P}$, of dimension $p \times k$, is constructed differently for each model. Note that the dimension of the first matrix expression is $(k+1) \times (p+2)$, with the matrix in the lower right hand corner being of dimension 1×2 . The product of the three matrices yields a result of dimension $(k+1) \times (k+1)$.

6.1 First Step Estimation

The submatrices of Ψ_o appropriate for the case of estimation in the first step of the procedure are:

$$\Psi_{\delta\delta} = n^{-1}\mathbf{H}'(\mathbf{I} - \lambda\mathbf{W})^{-1}\Sigma(\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{H} \quad (29)$$

and

$$\Psi_{\delta\lambda} = n^{-1}\mathbf{H}'(\mathbf{I} - \lambda\mathbf{W})^{-1}\Sigma[\mathbf{a}_1, \mathbf{a}_2]. \quad (30)$$

In these expressions,⁴ the residuals from the initial estimation are used, i.e., $\mathbf{u}_1$, together with the efficient and consistent estimate λ_2 . The diagonal elements of Σ are the squares of the spatially filtered residuals, $u_{1is}^2 = (u_{1i} - \lambda_2 u_{1iL})^2$.

⁴Note that there is a typo in the expression in Equation (B2) on p. 612 of Arraiz et al. (2010). The correct expression for $\Psi_{\delta\delta}$ and for $\Psi_{\delta\rho}$ (in their notation) should include the matrix inverse operation (see A4 on p. 611).

6.1.1 OLS

In the special case of OLS estimation, $\mathbf{H} = \mathbf{Z}$ (or, $\mathbf{X}$) and the expression for $\mathbf{P}$ simplifies to $(n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$. In addition, we have shown that in this case both $\mathbf{a}_1$ and $\mathbf{a}_2$ are zero. As a result, the weighting matrix Ψ_o is block-diagonal

After some matrix algebra, the variance-covariance matrix follows as:

$$\Omega_{OLS} = \begin{bmatrix} n(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I} - \lambda\mathbf{W})^{-1}\Sigma(\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} & 0 \\ 0 & (\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1} \end{bmatrix}$$

Note that the main computational issue in $\mathbf{Z}'(\mathbf{I} - \lambda\mathbf{W})^{-1}$ is post-multiplying the $p \times n$ matrix $\mathbf{Z}'$ with an inverse matrix of dimension $n \times n$. In practice, we will again exploit the power expansion that was described in Section 4.3.2. Once this matrix expression is computed, its transpose is $(\mathbf{I} - \lambda\mathbf{W}')^{-1}\mathbf{Z}$.

6.1.2 2SLS

The 2SLS case follows as a straightforward implementation of Equation 28, with $\mathbf{P}$ as in Equation 21 (repeated here for convenience):

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

The submatrices of Ψ_o are as in Equations 29 and 30. The residuals are $\mathbf{u}_1$ from the 2SLS regression, and the spatial parameter is again λ_2 .

Note that in practice, these asymptotic variance-covariance matrices will seldom be used. They are appropriate for the case where the initial estimates from respectively OLS or 2SLS are not further improved upon. In practice, this is rarely of interest, since the whole purpose is to obtain more efficient estimates for the model parameters from the spatial Cochrane-Orcutt procedure.

6.2 Spatially Weighted Estimation

In the typical case where the estimates for δ are derived from a spatially weighted regression (yielding δ_2) in a second estimation step, the relevant submatrices of Ψ_o take the form:

$$\Psi_{\delta\delta} = n^{-1}\mathbf{H}'\Sigma\mathbf{H} \quad (31)$$

and

$$\Psi_{\delta\lambda} = n^{-1}\mathbf{H}'\Sigma[\mathbf{a}_1, \mathbf{a}_2]. \quad (32)$$

The expressions for $\mathbf{a}_1$ and $\mathbf{a}_2$ are as in Equation 23, using $\mathbf{u}_2$ as the residual vector and λ_3 for the autoregressive parameter. The diagonal elements of Σ are the squares of the spatially filtered residuals, $u_{2is}^2 = (u_{2i} - \lambda_3 u_{2iL})^2$. Note that it is no longer necessary to compute the inverse matrix $(\mathbf{I} - \lambda\mathbf{W})^{-1}$.

6.2.1 SWLS

When there are no endogenous variables present in $\mathbf{Z}$, the matrix Ψ_o is again block-diagonal, since $\Psi_{\delta\lambda} = 0$. Also, the expression for $\mathbf{P}$ simplifies to $(n^{-1}\mathbf{Z}'_s\mathbf{Z}_s)^{-1}$, a cross-product in the spatially filtered explanatory variables (using λ_3). Similarly, Equation 31 becomes $\Psi_{\delta\delta} = n^{-1}\mathbf{Z}'_s\mathbf{\Sigma}\mathbf{Z}_s$.

After some matrix algebra, the variance-covariance matrix follows as:

$$\Omega_{SWLS} = \begin{bmatrix} n(\mathbf{Z}'_s\mathbf{Z}_s)^{-1}\mathbf{Z}'_s\mathbf{\Sigma}\mathbf{Z}_s(\mathbf{Z}'_s\mathbf{Z}_s)^{-1} & 0 \\ 0 & (\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1} \end{bmatrix}.$$

6.2.2 GS2SLS

The relevant expression for $\mathbf{P}$ is as in Equation 24, repeated here for convenience:

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

The submatrices of Ψ_o are as in Equations 31 and 32. The residuals are $\mathbf{u}_2$ from the GS2SLS regression, and the spatial parameter is λ_3 .

7 Accounting for Heteroskedasticity in FGLS

The estimation equations for the weighted spatial regression in Kelejian and Prucha (2010) and Arraiz et al. (2010) implement the spatial Cochrane-Orcutt transformation, but do not correct for the presence of unspecified heteroskedasticity. As shown in Anselin (2006, p. 931), among others, the latter could be implemented by exploiting the classic result of White (1980) and including a consistent estimate for $\mathbf{X}'\mathbf{\Sigma}\mathbf{X}$ or $\mathbf{H}'\mathbf{\Sigma}\mathbf{H}$ in the second stage estimation in Step 4. Note that the role of these cross products is explicit in the estimates for the variance-covariance matrix.

7.1 SWLS with Heteroskedasticity

Equation 25 is adjusted to account for unspecified heteroskedasticity and becomes:

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

with the spatially filtered variables and the spatially filtered residuals based on λ_2 and $\mathbf{u}_1$.

7.2 GS2SLS with Heteroskedasticity

Similarly, Equation 26 is adjusted to account for unspecified heteroskedasticity and becomes:

$$\delta_{2,HS2SLS} = [(\mathbf{Z}'_s\mathbf{H})(\mathbf{H}'\mathbf{\Sigma}\mathbf{H})^{-1}(\mathbf{H}'\mathbf{Z}_s)]^{-1}(\mathbf{Z}'_s\mathbf{H})(\mathbf{H}'\mathbf{\Sigma}\mathbf{H})^{-1}\mathbf{H}'\mathbf{y}_s,$$

again, with the spatially filtered variables and the spatially filtered residuals based on λ_2 and $\mathbf{u}_1$.

8 General Two-Step Estimation in the Absence of Heteroskedasticity

In the recent working papers by Drukker et al. (2010, 2011), a general framework is presented for a two-step GMM estimation of the parameters in a SAR-SAR specification, allowing for endogenous variables other than the spatial lag. The moment conditions spelled out in the paper are more general than in previous work. Specifically, up to S weighting matrices $\mathbf{A}_s$ are allowed, which do not need to have zero diagonal, but only $\text{tr}\mathbf{A}_s = 0$ (zero diagonals being a special case). The moment conditions take the same form as in Section 4.1:

$$n^{-1}\mathbb{E}[\varepsilon'\mathbf{A}_s\varepsilon] = 0, \quad (33)$$

for $s = 1, \dots, S$.

The two step estimator actually consists of four steps, each of the two major steps consisting of two sub-steps. We consider each in turn.

8.1 Step One

The first step is identical to Step 1 in Section 5.1 and consists of either OLS or 2SLS estimation of the model to obtain an initial set of coefficient estimates δ_1 and residuals, $\mathbf{u}_1$.

8.1.1 Moment Conditions

The residuals $\mathbf{u}_1$ are used to solve the set of moment equations given in Equation 33. As in Section 4.2, the equations are re-organized in the same fashion as in Equation 8, using spatially filtered residuals instead of the error terms. In general, the $S \times 1$ vector $\mathbf{m}$ consists of the elements $\mathbf{m}_s = (n^{-1})(\mathbf{u} - \lambda\mathbf{u}_L)'\mathbf{A}_s(\mathbf{u} - \lambda\mathbf{u}_L)$. After some algebra, it is easy to see that this consists of a constant term, two terms in λ and one term in λ^2 :

$$\mathbf{m}_s = (n^{-1})\mathbf{u}'\mathbf{A}_s\mathbf{u} - (n^{-1})(\mathbf{u}'\mathbf{A}_s\mathbf{u}_L + \mathbf{u}_L'\mathbf{A}_s\mathbf{u})\lambda - (n^{-1})(-\mathbf{u}_L'\mathbf{A}_s\mathbf{u}_L)\lambda^2,$$

or

$$\mathbf{m}_s = (n^{-1})\mathbf{u}'\mathbf{A}_s\mathbf{u} - (n^{-1}) \begin{bmatrix} \mathbf{u}'\mathbf{A}_s\mathbf{u}_L + \mathbf{u}_L'\mathbf{A}_s\mathbf{u} & -\mathbf{u}_L'\mathbf{A}_s\mathbf{u}_L \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda^2 \end{bmatrix}. \quad (34)$$

The first term in Equation 34 is the element of the s -th row of $\mathbf{g}$, whereas the two elements in the row vector form the s -th row of the $S \times 2$ matrix $\mathbf{G}$. Note that Drukker et al. (2010, p. 7) assume that the matrix $\mathbf{A}_s$ is symmetric, in which case:

$$\mathbf{u}'\mathbf{A}_s\mathbf{u}_L + \mathbf{u}_L'\mathbf{A}_s\mathbf{u} = \mathbf{u}_L'\mathbf{A}_s'\mathbf{u} + \mathbf{u}_L'\mathbf{A}_s\mathbf{u} \quad (35)$$

$$= \mathbf{u}_L'(\mathbf{A}_s + \mathbf{A}_s')\mathbf{u} \quad (36)$$

$$= 2\mathbf{u}_L'\mathbf{A}_s\mathbf{u}. \quad (37)$$

However, in general, this is not the case. For example, for $\mathbf{A}_2 = \mathbf{W}$, this will not hold unless the weights matrix is symmetric (which is typically not the case). This can be remedied by setting $\mathbf{A}_2 = (1/2)(\mathbf{W} + \mathbf{W}')$.

For the homoskedastic case, the first moment equation results in a matrix $\mathbf{A}_1$ of the form (see, e.g., Kelejian and Prucha 2010, footnote 7):⁵

$$\mathbf{A}_1 = \mathbf{W}'\mathbf{W} - (n^{-1})tr(\mathbf{W}'\mathbf{W})\mathbf{I}.$$

Note that $\mathbf{A}_1$ is symmetric with $tr\mathbf{A}_1 = 0$, but its diagonal elements are non-zero.

The second moment equation is the same as in the heteroskedastic case, Equation 5, with $\mathbf{A}_2 = \mathbf{W}$, or, if the general expressions as in Equation 37 are used, with $\mathbf{A}_2 = (1/2)(\mathbf{W} + \mathbf{W}')$. Both $tr\mathbf{A}_2 = 0$ and its diagonal elements are zero.

To recap, the moment conditions take on the general form:

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

with:

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

and the 2×2 matrix $\mathbf{G}$ with the following elements:

$$\mathbf{G}_{11,21} = 2n^{-1}\mathbf{u}'_L\mathbf{A}_{1,2}\mathbf{u} \quad (38)$$

$$\mathbf{G}_{12,22} = -n^{-1}\mathbf{u}'_L\mathbf{A}_{1,2}\mathbf{u}_L \quad (39)$$

An initial consistent estimate of λ , say λ_1 is obtained as a nonlinear least squares solution to $\text{argmin}_\lambda(\mathbf{m}'\mathbf{m})$. Note that unlike the heteroskedastic approach in Kelejian and Prucha (2010) and Arraiz et al. (2010), this is not followed by an efficient GMM estimation step at this stage, i.e., using a weighting matrix.

8.2 Step Two – Consistent and Efficient Estimation of λ

The first part of the second step consists of a spatial Cochrane-Orcutt estimation, identical to the procedure outlined in Section 5.4. In the absence of spatially lagged and endogenous variables, this boils down to OLS on the spatially filtered variables:

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

⁵Kelejian and Prucha (2010, Equation 9) present the derivation in the context of a weighted nonlinear least squares estimator with a moments weighting matrix. The element corresponding to the first moment condition is $\nu = 1/[1 + [(n^{-1}tr(\mathbf{W}\mathbf{W}'))^2]]$. In Drukker et al. (2010, 2011) this scalar is included in $\mathbf{A}_1$. However, since the first solution is intended to produce a consistent estimate for λ , it is not clear why the scalar should be included.

When either spatially lagged or endogenous variables are present (or both), then the generalized spatial two stage least squares estimation is appropriate (we repeat the expression here for convenience):

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

Note that the instrument matrix $\mathbf{H}$ is the same as in Step 1. Specifically, these instruments are NOT spatially filtered.

We obtain a new set of residuals as $\mathbf{u}_2 = \mathbf{y} - \mathbf{Z}\delta_2$. More importantly, we also construct the spatially filtered residuals $\mathbf{u}_{2s}$, using the value λ_1 . Both $\mathbf{Z}_s$ and $\mathbf{u}_s$ are needed to carry out efficient GMM estimation of λ , as the result of $\text{argmin}_\lambda \mathbf{m}'\mathbf{\Psi}^{-1}\mathbf{m}$, where $\mathbf{\Psi}$ is a weighting matrix. The optimal weights correspond to the inverse variance of the moment conditions.

The results are very similar to those in Section 4.3, except that heteroskedasticity is not allowed for. In general, the elements of the 2×2 matrix $\mathbf{\Psi}$ are:

$$\psi_{q,r} = \hat{\sigma}^4(2n)^{-1}\text{tr}[(\mathbf{A}_q + \mathbf{A}'_q)(\mathbf{A}_r + \mathbf{A}'_r)] + \hat{\sigma}^2(n^{-1})\mathbf{a}'_q\mathbf{a}_r \quad (40)$$

$$+ (n^{-1})(\hat{\mu}^{(4)} - 3\hat{\sigma}^4)\text{vec}_D(\mathbf{A}_q)'\text{vec}_D(\mathbf{A}_r) \quad (41)$$

$$+ (n^{-1})\hat{\mu}^{(3)}[\mathbf{a}'_q\text{vec}_D(\mathbf{A}_r) + \mathbf{a}'_r\text{vec}_D(\mathbf{A}_q)] \quad (42)$$

with $q, r = 1, 2$, $\hat{\sigma}^2 = \mathbf{u}'_s\mathbf{u}_s/n$ as the estimate of the error variance obtained from the spatially filtered residuals, and $\hat{\mu}^{(3)}$ and $\hat{\mu}^{(4)}$ are similarly the third and fourth moments obtained from the spatially filtered residuals. Note that When the diagonal elements of the weighting matrix $\mathbf{A}$ are zero, then $\text{vec}_D(\mathbf{A}) = 0$. As pointed out earlier, this is the case for $\mathbf{A}_2$.

8.2.1 OLS estimation

As before, when no spatial lag or endogenous variables are included in the model specification, $\mathbf{a}_{q,r} = 0$ and the expression for $\psi_{q,r}$ simplifies greatly. Specifically, the second term (in Equation 40) and the fourth term (Equation 42) equal zero. Moreover, $\mathbf{A}_1$ is symmetric, so that $\mathbf{A}_1 + \mathbf{A}'_1 = 2\mathbf{A}_1$. Also, $\mathbf{A}_2 = (1/2)(\mathbf{W} + \mathbf{W}')$, which we use instead of $\mathbf{A}_2$, and $\text{vec}_D(\mathbf{A}_2) = 0$.

Consequently, we obtain the following results:

$$\psi_{11} = \hat{\sigma}^4(2n)^{-1}\text{tr}[(2\mathbf{A}_1)(2\mathbf{A}_1)] + (n^{-1})(\hat{\mu}^{(4)} - 3\hat{\sigma}^4)\text{vec}_D(\mathbf{A}_1)'\text{vec}_D(\mathbf{A}_1)$$

$$= 2\hat{\sigma}^4(n^{-1})\text{tr}[\mathbf{A}_1\mathbf{A}_1] + (n^{-1})(\hat{\mu}^{(4)} - 3\hat{\sigma}^4)\text{vec}_D(\mathbf{A}_1)'\text{vec}_D(\mathbf{A}_1),$$

$$\psi_{12} = \hat{\sigma}^4(2n)^{-1}\text{tr}[(2\mathbf{A}_1)(\mathbf{W} + \mathbf{W}')] = \hat{\sigma}^4(n^{-1})\text{tr}[\mathbf{A}_1(\mathbf{W} + \mathbf{W}')],$$

$$\psi_{21} = \psi_{12},$$

$$\psi_{22} = \hat{\sigma}^4(2n)^{-1}\text{tr}[(\mathbf{W} + \mathbf{W}')(\mathbf{W} + \mathbf{W}')] = \hat{\sigma}^4(n^{-1})\text{tr}[\mathbf{W}\mathbf{W}']$$

8.2.2 GS2SLS Estimation

In the general case, the terms $\mathbf{a}_{q,r}$, $q, r = 1, 2$ need to be computed.. However, the properties of $\mathbf{A}_1$ and $\mathbf{A}_2$ can be exploited to simplify the expressions,

as in the OLS case. The expressions for $\mathbf{a}_{q,r}$, $q, r = 1, 2$ are the same as in Section 4.3.3, Equation 23:

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

with $\mathbf{P}$ as in Equation 24:

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

and $\mathbf{H}$ is the matrix of instruments. $\mathbf{Z}_s$ is spatially filtered using λ_1 .

The expressions for $\alpha_{q,r}$, $q, r = 1, 2$ are as in Section 4.3.2, using the spatially filtered $\mathbf{Z}_s'$ and $\mathbf{u}_s$.

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:

$$\alpha_1 = -2n^{-1}[\mathbf{Z}_s'\mathbf{A}_1\mathbf{u}_s],$$

For α_2 , the corresponding expression is:

$$\alpha_2 = -n^{-1}[\mathbf{Z}_s'(\mathbf{W} + \mathbf{W}')\mathbf{u}_s]$$

Taking all of this together, the elements of the weighting matrix are:

$$\begin{aligned}\psi_{11} &= 2\hat{\sigma}^4(n^{-1})\text{tr}[\mathbf{A}_1\mathbf{A}_1] + \hat{\sigma}^2(n^{-1})\mathbf{a}_1'\mathbf{a}_1 \\ &\quad + (n^{-1})(\hat{\mu}^{(4)} - 3\hat{\sigma}^4)\text{vec}_D(\mathbf{A}_1)'\text{vec}_D(\mathbf{A}_1) \\ &\quad + 2(n^{-1})\hat{\mu}^{(3)}[\mathbf{a}_1'\text{vec}_D(\mathbf{A}_1)] \\ \psi_{12} &= \hat{\sigma}^4(n^{-1})\text{tr}[\mathbf{A}_1(\mathbf{W} + \mathbf{W}')] + \hat{\sigma}^2(n^{-1})\mathbf{a}_1'\mathbf{a}_2 \\ &\quad + (n^{-1})\hat{\mu}^{(3)}[\mathbf{a}_2'\text{vec}_D(\mathbf{A}_1)] \\ \psi_{21} &= \psi_{12}, \\ \psi_{22} &= \hat{\sigma}^4(2n)^{-1}\text{tr}[(\mathbf{W} + \mathbf{W}')(\mathbf{W} + \mathbf{W}')] + \hat{\sigma}^2(n^{-1})\mathbf{a}_2'\mathbf{a}_2\end{aligned}$$

8.3 Step Two – Variance-Covariance Matrix

The minimization of $\text{argmin}_\lambda \mathbf{m}'\Psi^{-1}\mathbf{m}$ yields a consistent and efficient estimate for λ , say λ_2 . Although it is only necessary to use a consistent estimate for λ to construct the appropriate variance-covariance matrix, it makes sense to use λ_2 rather than λ_1 in the expressions that follow. This also implies that spatially filtered residuals $\mathbf{u}_s$ and the spatially filtered $\mathbf{Z}_s$ need to be recomputed with the new value for λ_2 . In addition, all the relevant parts of the variance-covariance matrix that depend on λ need to be updated, such as the moments of the spatially filtered residuals, i.e., $\hat{\sigma}^2$, $\hat{\mu}^{(3)}$ and $\hat{\mu}^{(4)}$, among others.

The general expression for the variance-covariance matrix is the same as for the heteroskedastic case (Equation 28), which we repeat here for convenience:

$$\Omega = \begin{bmatrix} \mathbf{P}' & 0 \\ 0 & (\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1}\mathbf{J}'\Psi^{-1} \end{bmatrix} \Psi_o \begin{bmatrix} \mathbf{P} & 0 \\ 0 & \Psi^{-1}\mathbf{J}(\mathbf{J}'\Psi^{-1}\mathbf{J})^{-1} \end{bmatrix},$$

where, as in the heteroskedastic case, the submatrix Ψ_o consists of four submatrices:

$$\Psi_o = \begin{bmatrix} \Psi_{\delta\delta} & \Psi_{\delta\lambda} \\ \Psi'_{\delta\lambda} & \Psi \end{bmatrix}.$$

Again, as in the heteroskedastic case, an auxiliary 2×1 vector $\mathbf{J}$ is defined, constructed as:

$$\mathbf{J} = \mathbf{G} \begin{bmatrix} 1 \\ 2\lambda \end{bmatrix}, \quad (43)$$

where the elements of the 2×2 matrix $\mathbf{G}$ are from Equations 38 and 39.

The elements of Ψ are as in Section 8.2. Note that the new $\mathbf{Z}_s$ and $\mathbf{u}_s$ are needed to compute the $\alpha_{1,2}$, and the new $\mathbf{Z}_s$ enters into the computation of $\mathbf{P}$, both necessary to calculate the terms $\mathbf{a}_{1,2}$ in the GS2SLS case. In addition, the moments of the $\mathbf{u}_s$ residuals are required to compute the elements of Ψ for both OLS and GS2SLS cases.

The other submatrices of Ψ are:

$$\begin{aligned} \Psi_{\delta\delta} &= (n^{-1})\hat{\sigma}^2\mathbf{H}'\mathbf{H} \\ \Psi_{\delta\lambda} &= (n^{-1})\hat{\sigma}^2\mathbf{H}'[\mathbf{a}_1, \mathbf{a}_2] + (n^{-1})\hat{\mu}^{(3)}\mathbf{H}'[vec_D(\mathbf{A}_1), \mathbf{0}] \end{aligned}$$

Here, again, we need the residual moments and $[\mathbf{a}_1, \mathbf{a}_2]$ using λ_2 . Note, that since $vec_D\mathbf{A}_2 = \mathbf{0}$ the last element in the expression for $\Psi_{\delta\lambda}$ is a $n \times 1$ vector of zeros.

The corresponding elements of the variance-covariance matrix are then:

$$\begin{aligned} \Omega_{\delta\delta} &= \hat{\sigma}^2[\mathbf{Z}'_s\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}_s]^{-1} \\ \Omega_{\delta\lambda} &= \mathbf{P}'\Psi_{\delta\lambda}\Psi^{-1}\mathbf{J}[\mathbf{J}'\Psi^{-1}\mathbf{J}]^{-1} \\ \Omega_{\lambda\lambda} &= [\mathbf{J}'\Psi^{-1}\mathbf{J}]^{-1} \end{aligned}$$

Note that there may be a problem with the matrix $\Omega_{\delta\lambda}$ in the standard regression model. The classic result is that the variance-covariance matrix must be block-diagonal between the model coefficients (δ) and the error parameters (λ). In general, this will not be the case, since the second term in $\Psi_{\delta\lambda}$ does not disappear unless the diagonal elements of $\mathbf{A}_1$ are zero or the error distribution is symmetric or normal (in which case $\mu^{(3)} = 0$). This is not the case for the form of $\mathbf{A}_1$ used in Drukker et al. (2010, 2011). However, this is the case if we use

$$\mathbf{A}_1 = \mathbf{W}'\mathbf{W} - \text{diag}(\mathbf{w}'_i\mathbf{w}_i).$$

The properties of the corresponding estimator need to be further investigated.

References

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

- Anselin, L. (2006). Spatial econometrics. In Mills, T. and Patterson, K., editors, *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*, pages 901–969. Palgrave Macmillan, Basingstoke.
- 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.
- Drukker, D. M., Egger, P., and Prucha, I. R. (2010). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Working paper, Department of Economics, University of Maryland, College Park, MD.
- Drukker, D. M., Prucha, I. R., and Raciborski, R. (2011). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. *The Stata Journal*, 1:1–13.
- 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. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2):219–257.
- Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118:27–50.
- Kelejian, H. H. and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157:53–67.
- 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.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838.