Lecture 7: GMM Estimation of Spatial Models

Mauricio Sarrias

Universidad Católica del Norte

November 22, 2017

Mandatory Reading

- (AR)-Chapter 7, 9 and 11
- Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of environmental economics and management, 45(1), 24-39.
- Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review, 40(2), 509-533.
- Messner, S. F., & Anselin, L. (2004). Spatial Analyses of Homicide with Areal Data. Spatially integrated social science, 12.
- Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics, 17(1), 99-121.
- Arraiz, I., Drukker, D. M., Kelejian, H. H., & Prucha, I. R. (2010). A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results. Journal of Regional Science, 50(2), 592-614.

S2SLS

Recall that Spatial Lag Model (SLM) is given by:

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

A more concise way to express the model is as:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Z} = [\mathbf{X}, \mathbf{W}\mathbf{y}]$ and the $(K+1) \times 1$ coefficient column vector is rearranged as $\boldsymbol{\delta} = (\boldsymbol{\beta}^\top, \rho)^\top$.

As we know, the presence of the spatially lagged dependent variable on the RHS induces endogeneity or simultaneous equation bias.

S2SLS

- Instead of applying QML or ML, we might rely on the IV approach to deal with the endogeneity problem.
- Principle: Find a set of instruments $\mathbf{H}$ that are strongly correlated with $\mathbf{Z}$, by asymptotically uncorrelated with $\boldsymbol{\varepsilon}$.
- Important point: We just assume that $\mathbf{Wy}$ is the only endogenous variable.
  - $\mathbf{H}$ should contain al the predetermined variables, $\mathbf{X}$, and the instrument(s) for $\mathbf{Wy}$.
  - We will relax this later.

What is the best instrument(s) for $\mathbf{W}\mathbf{y}$?

Optimal instrumental variables: the 'best instruments' for the r.h.s variables are the conditional means. Thus, the ideal instruments are:

$$\begin{aligned}\mathbb{E}\left(\mathbf{Z}|\mathbf{X}\right) &= \left[\mathbb{E}\left(\mathbf{X}|\mathbf{X}\right), \mathbb{E}\left(\mathbf{W}\mathbf{y}|\mathbf{X}\right)\right] \\ &= \left[\mathbf{X}, \mathbf{W}\mathbb{E}\left(\mathbf{y}|\mathbf{X}\right)\right] \quad \text{since } \mathbf{W} \text{ is non-stochastic.}\end{aligned}$$

- The best instruments for $\mathbf{X}$ is $\mathbf{X}$.
- The best instruments for $\mathbf{W}\mathbf{y}$ is $\mathbf{W}\mathbb{E}\left(\mathbf{y}|\,\mathbf{X}\right)$.

Given that the roots of $\rho \mathbf{W}_n$ are less than one in absolute value, the conditional expectation can also be written as:

$$
\begin{aligned}
\mathbb{E}(\mathbf{W}\mathbf{y}|\mathbf{X}) &= \mathbf{W}\mathbb{E}\left(\mathbf{y}|\,\mathbf{X}\right) \\
&= \mathbf{W}\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{W}\left[\mathbf{I}_n + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3 + ...\right]\mathbf{X}\boldsymbol{\beta}_0 \\
&= \mathbf{W}\left[\sum_{l=1}^{\infty}\rho_0^l\mathbf{W}^l\right]\mathbf{X}\boldsymbol{\beta}_0 \\
&= \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^2\mathbf{X}(\rho\boldsymbol{\beta}) + \mathbf{W}^3\mathbf{X}(\rho^2\boldsymbol{\beta}) + \mathbf{W}^4\mathbf{X}(\rho^3\boldsymbol{\beta}) + ...
\end{aligned}
$$

This reveals that $\mathbb{E}(\mathbf{y}|\mathbf{X})$ is linear in $\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}$... This observation motivated Kelejian and Prucha (1998) to select a set of instruments based on, at least, the linearly independent columns of $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X})$.

- Computational issues with the inverse of the $n \times n$ matrix $(\mathbf{I}_n - \rho_0 \mathbf{W})$.
- Kelejian and Prucha (1998, 1999) suggest the use of an **approximation** of the best instruments:
  - They suggest $\mathbf{H}$ which contains, say, $\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, ..., \mathbf{W}^l\mathbf{X}$, and to compute approximations of the best instruments from a regression of the rhs variables against $\mathbf{H}$, where l is a pre-selected finite constant and is generally set to 2 in applied studies.

Thus, in general we can write the instruments as:

$$\mathbf{H} = (\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X})$$

The intuition behind the instruments is the following:

- Since $\mathbf{X}$ determines $\mathbf{y}$, then it must be true that $\mathbf{WX}, \mathbf{W}^2\mathbf{X}, ...$ determines $\mathbf{Wy}$.
- Furthermore, since $\mathbf{X}$ is uncorrelated with ε, then $\mathbf{WX}$ must be also uncorrelated with ε.

Using the conditional expectation, Lee (2003) suggested the instrument matrix:

$$\mathbf{H} = \left[\mathbf{X}, \mathbf{W}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right],$$

which requires the use of consistent first round estimates for ρ and $\boldsymbol{\beta}$. In Kelejian et al. (2004), a similar approach is outlined where the matrix inverse is replaced by the power expansion. This yield an instruments matrix as:

$$\mathbf{H} = \left[\mathbf{X}, \mathbf{W}\left(\sum_{l=1}^{\infty}\rho_0^l\mathbf{W}^l\right)\mathbf{X}\boldsymbol{\beta}\right].$$

In any practical implementation, the power expansion must be truncated at some point.

Heterokedastic Errors (Kelejian and Prucha, 2010)

The errors $\{\epsilon_{i,n}, 1 \leq i \leq n, n \geq 1\}$ satisfy $\mathbb{E}(\epsilon_{i,n}) = 0$, $\mathbb{E}(\epsilon_{i,n}^2) = \sigma_{i,n}^2$, with $0 < \underline{a}^{\sigma} \leq \sigma_{i,n}^2 \leq \overline{a}^{\sigma} < \infty$. Additionally the errors are assumed to possess fourth moments, that is $\sup_{1 \leq i \leq n, n \geq 1} \mathbb{E} \left|\epsilon_{i,n}\right|^{4+\eta}$ for some $\eta > 0$. Furthermore, for each $n \geq 1$ the random variables $\epsilon_{1,n}, ..., \epsilon_{n,n}$ are totally independent.

- Remark: Triangular arrays.
- Note: Kelejian and Prucha (1998) derived the asymptotic properties assuming that the errors are homokedastic. But, Kelejian and Prucha (2010) extend the model by assuming heteroskedasticity.

Now, we state some assumptions about the behavior of the spatial weight matrix $\mathbf{W}$.

Diagonal elements of $\mathbf{W}_n$ (Kelejian and Prucha, 1998)

All diagonal elements of the spatial weighting matrix $\mathbf{W}_n$ are zero

This assumption is a normalization of the model and it also implies that no spatial unit is viewed as its own neighbor.

Nonsingularity (Kelejian and Prucha, 1998)

The matrix $(\mathbf{I}_n - \rho_0 \mathbf{W}_n)$ is nonsingular with $|\rho_0| < 1$.

Under Nonsigularity Assumption, we can write the reduced form of the true model as:

$$\mathbf{y}_n = (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1}\mathbf{X}_n \boldsymbol{\beta}_0 + (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1}\boldsymbol{\varepsilon}_n.$$

Heterokedastic-Errors Assumption implies further that the population variance-covariance matrix of $\mathbf{y}_n$ is equal to

$$\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^\top) = \boldsymbol{\Omega}_{y_n} = (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1}\boldsymbol{\Sigma}_n(\mathbf{I}_n - \rho_0 \mathbf{W}_n^\top)^{-1}, \tag{1}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma^2_{i,n})$. If we assume homokedasticity, then the variance-covariance matrix of $\mathbf{y}$ reduces to:

$$\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^\top) = \boldsymbol{\Omega}_{y_n} = \sigma^2_\epsilon (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1}(\mathbf{I}_n - \rho_0 \mathbf{W}_n^\top)^{-1}.$$

Bounded matrices (Kelejian and Prucha, 1998)

The row and column sums of the matrices $\mathbf{W}_n$ and $(\mathbf{I}_n - \rho_0 \mathbf{W}_n)$ are bounded uniformly in absolute value.

- Thus, the variance of $\mathbf{y}_n$ in Equation (1), which depend on $\mathbf{W}_n$ and $(\mathbf{I}_n - \rho_0 \mathbf{W}_n)$, are uniformly bounded in absolute value as n goes to infinity, thus limiting the degree of correlation between, respectively, the elements of $\boldsymbol{\varepsilon}_n$ and $\mathbf{y}_n$.
- Applied to $\mathbf{W}_n$, this assumption means that each cross-sectional unit can only have a limited number of neighbors.
- Applied to $(\mathbf{I}_n - \rho \mathbf{W}_n)$ limits the degree of correlation.

No Perfect Multicolinearity (Kelejian and Prucha, 1998)

The regressor matrices $\mathbf{X}_n$ have full column rank (for n large enough). Furthermore, the elements of the matrices $\mathbf{X}_n$ are uniformly bounded in absolute value.

Now we state some assumptions about the instruments.

Rank Instruments, (Kelejian and Prucha, 1998)

The instrument matrices $\mathbf{H}_n$ have full column rank $L \geq K + 1$ for all n large enough. Furthermore, the elements of the matrices $\mathbf{H}_n$ are uniformly bounded in absolute value. They are composed of a subset of the linearly independent columns of $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, ...)$.

Limits of Instruments (Kelejian and Prucha, 1998)

Let $\mathbf{H}_n$ be a matrix of instruments, then:

1. $\lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top\mathbf{H}_n = \mathbf{Q}_{HH}$ where $\mathbf{Q}_{HH}$ is finite and nonsingular (full rank).
2. $\text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n^\top\mathbf{Z}_n = \mathbf{Q}_{HZ}$ where $\mathbf{Q}_{HZ}$ is finite and has full column rank.

- $\lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{H}_n = \mathbf{Q}_{HH}$ implies that $\mathbf{W}_n\mathbf{X}_n$ and $\mathbf{X}_n$ cannot be linearly dependent.
    - This condition would be violated if for example $\mathbf{W}_n\mathbf{X}_n$ include a spatial lag for the constant term or the model is the pure SLM.
- $\text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{Z}_n = \mathbf{Q}_{HZ}$ requires a non-null correlation between the instruments and the original variables.

To operationalize the S2SLS we first need the predicted values for $\mathbf{Z}_n$ based on the OLS regression of $\mathbf{Z}_n$ on $\mathbf{H}_n$ in the first stage.

First stage

Consider this first stage as the regression $\mathbf{Z}_n = \mathbf{H}_n\boldsymbol{\theta} + \boldsymbol{\xi}_n$, so that $\widehat{\boldsymbol{\theta}} = (\mathbf{H}_n^\top \mathbf{H}_n)^{-1}\mathbf{H}_n^\top \mathbf{Z}_n$

The predicted values are computed as:

$$\widehat{\mathbf{Z}}_n = \mathbf{H}_n\widehat{\boldsymbol{\theta}}_n = \mathbf{H}_n(\mathbf{H}_n^\top \mathbf{H}_n)^{-1}\mathbf{H}_n^\top \mathbf{Z}_n = \mathbf{P}_{H,n}\mathbf{Z}_n \tag{2}$$

where $\mathbf{P}_{H,n}$ is the projection matrix. The second stage uses the predicted values of $\mathbf{Z}_n$:

Second Stage

$$\widehat{\boldsymbol{\delta}}_{S2SLS} = \left(\widehat{\mathbf{Z}}_n^\top \widehat{\mathbf{Z}}_n\right)^{-1}\widehat{\mathbf{Z}}_n^\top \mathbf{y}_n$$

Definition (Spatial Two Stage Least Square Estimator)

Let $\mathbf{H}_n$ be the matrix $(n \times L)$ of instruments. Then the S2SLS is given by:

$$\widehat{\boldsymbol{\delta}}_{S2SLS} = \left(\widehat{\mathbf{Z}}_n^\top \widehat{\mathbf{Z}}_n\right)^{-1} \widehat{\mathbf{Z}}_n^\top \mathbf{y}_n \tag{3}$$

where:

$$\widehat{\mathbf{Z}}_n = \mathbf{H}_n \widehat{\boldsymbol{\theta}}_n = \mathbf{H}_n (\mathbf{H}_n^\top \mathbf{H}_n)^{-1} \mathbf{H}_n^\top \mathbf{Z}_n = \mathbf{P}_{H,n} \mathbf{Z}_n \tag{4}$$

Note also that $\mathbf{H}_n$ is a $n \times L$ matrix, which also includes the exogenous variables $\mathbf{X}_n$. It is also important to note that the projection matrix does not affect $\mathbf{X}_n$, but it does affect the endogenous variable $\mathbf{W}_n \mathbf{y}_n$:

$$\mathbf{P}_{H,n} \mathbf{Z}_n = [\mathbf{X}_n, \mathbf{P}_{H,n} \mathbf{W}_n \mathbf{y}_n] = \left[\mathbf{X}_n, \widehat{\mathbf{W}_n \mathbf{y}_n}\right] \tag{5}$$

S2SLS Estimator as GMM

Recall that the GMM estimator is defined as the solution of the minimization problem:

$$\underbrace{\widehat{\boldsymbol{\delta}}_{GMM}}_{K\times 1} = \underset{\boldsymbol{\beta}}{\arg\min}\left\{\underbrace{\mathbf{g}_n(\boldsymbol{\beta})^\top}_{1\times L}\underbrace{\boldsymbol{\Upsilon}_n^{-1}}_{L\times L}\underbrace{\mathbf{g}_n(\boldsymbol{\beta})}_{L\times 1}\right\},$$

where

$$\mathbf{g}_n = \frac{1}{n}\mathbf{H}^\top\boldsymbol{\varepsilon} = \frac{1}{n}\mathbf{H}^\top(\mathbf{y}-\mathbf{Z}\boldsymbol{\delta})$$

The matrix $\boldsymbol{\Upsilon}_n^{-1}$ is the optimal weight matrix, which correspond to the inverse of the covariance matrix of the sample moments:

$$\boldsymbol{\Upsilon} = \frac{1}{n}\widehat{\sigma}_\epsilon^2\mathbf{H}^\top\mathbf{H}$$

Then, the function to minimize is:

$$J = \frac{1}{n\widehat{\sigma}^2}\left\{\left[\mathbf{H}^\top\mathbf{y}-\mathbf{H}^\top\mathbf{Z}\boldsymbol{\delta}\right]^\top\left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\left[\mathbf{H}^\top\mathbf{y}-\mathbf{H}^\top\mathbf{Z}\boldsymbol{\delta}\right]\right\}$$

Obtaining the first order conditions and solving for $\boldsymbol{\delta}$, we obtain:

$$\widehat{\boldsymbol{\delta}}_{GMM} = \left(\mathbf{Z}^\top\mathbf{P}_H\mathbf{Z}\right)^{-1}\mathbf{Z}^\top\mathbf{P}_H\mathbf{y} \tag{6}$$

S2SLS Asymptotic distribution

Theorem (Spatial 2SLS Estimator for SLM)

Suppose that Assumptions hold. Then the S2SLS estimator defined as

$$\widehat{\boldsymbol{\delta}}_n = \left(\widehat{\mathbf{Z}}_n^\top \widehat{\mathbf{Z}}_n\right)^{-1} \widehat{\mathbf{Z}}_n^\top \mathbf{y}_n \tag{7}$$

*is **consistent**, and its asymptotic distribution is:*

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_n) \tag{8}$$

where

$$\boldsymbol{\Omega}_n = \mathbf{P}^\top \mathbf{H}^\top \boldsymbol{\Sigma} \mathbf{H} \mathbf{P} \tag{9}$$

Inference on δ is then based on the asymptotic variance-covariance matrix:

$$\begin{aligned}
\mathrm{Var}(\widehat{\delta}_{2SLS}) &= \left[\mathbf{Z}^\top \mathbf{H} \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top \mathbf{Z}\right]^{-1} \\
&\times \left[\mathbf{Z}^\top \mathbf{H} \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \left(\mathbf{H}^\top \boldsymbol{\Sigma} \mathbf{H}\right) \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top \mathbf{Z}\right] \\
&\times \left[\mathbf{Z}^\top \mathbf{H} \left(\mathbf{H}^\top \mathbf{H}\right)^{-1} \mathbf{H}^\top \mathbf{Z}\right]^{-1} \\
&= \left(\widehat{\mathbf{Z}}^\top \mathbf{Z}\right)^{-1} \left(\widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma} \widehat{\mathbf{Z}}\right) \left(\mathbf{Z}^\top \widehat{\mathbf{Z}}\right)^{-1}
\end{aligned} \tag{10}$$

Definition (Order of Magnitude in Probability)

1. The sequence $\{X_n\}$ is $O_p(n^k)$ iff one can always find a finite interval within which the outcome $(1/n^k)X_n$ will occur with probability arbitrary close to 1 for each term in the sequence.
2. If a random sequence is $O_p(1)$, the random sequence is said to be bounded in probability.
3. The sequence $\{X_n\}$ is $o_p(n^k)$ iff $(1/n^k)X_n$ converges in probability to zero.

- Any random variable X with cdf F is $O_p(1)$ (White, 2014, pag.28)

Example (Order of Magnitude in Probability)

Let $\{X_n\}$ be such that $X_i \sim \text{N}(0,1), \forall i$, with all the terms in the sequence being independent random variables. Define $\{Z_n\}$ as $Z_n = \sum_{i=1}^{n} X_i$. Then $\{X_n\}$ itself is $O_p(1)$, i.e., is bounded in probability, and $\{Z_n\}$ is $O_p(n^{1/2})$. Note that $n^{-1/2} Z_n = n^{-1/2} \sum_{i=1}^{n} X_i \sim \text{N}(0,1)$. Finally, in the sequence defined by $Y_n = n^{-1/2}(X_n + Z_n)$, note that $n^{-1/2} X_n$ is $o_p(1)$, while $n^{-1/2} Z_n$ is $O_p(1)$, implying that as $n \to \infty$, $n^{-1/2} Z_n$ is the dominant term in the definition of Y_n while $n^{-1/2} X_n$ is the stochastically irrelevant as $n \to \infty$.

Theorem (Sufficient Conditions for Consistency)

Chebyshev's inequality implies that a sufficient conditions for an estimator base don a sample of size n, say $\widehat{\theta}_n$, say to be consistent for θ are:

$$\begin{aligned} \mathbb{E}\left(\widehat{\theta}_n\right) &= \theta_0 \\ \lim_{n\to\infty} \operatorname{Var}\left(\widehat{\theta}_n\right) &= 0 \end{aligned} \tag{11}$$

Slightly weaker conditions for consistency are:

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}\left(\widehat{\theta}_n\right) &= \theta_0 \\ \lim_{n\to\infty} \operatorname{Var}\left(\widehat{\theta}_n\right) &= 0 \end{aligned} \tag{12}$$

Theorem (CLT for Vectors of Linear Quadratic Forms with Heterokedastic Innovations)

Assume the following:

1. *For* $r = 1, ..., m$ *let* $\mathbf{A}_{r,n}$ *with elements* $(a_{ijr})_{i,j=1,...,n}$ *be an* $n \times n$ *non-stochastic symmetric real matrix with* $\sup_{1 \le j \le n, n \ge 1} \sum_{i=1}^{n} |a_{ijr}| < \infty$,
2. *and let* $\mathbf{a}_r = (a_{ir}, ..., a_{nr})^\top$ *be a* $n \times 1$ *non-stochastic real vector with* $\sup_n \frac{\sum_{i=1}^{n} |a_{ir}|^{\delta_1}}{n} < \infty$ *for some* $\delta_1 > 2$.
3. *Let* $\boldsymbol{\varepsilon} = (\epsilon_1, ..., \epsilon_n)^\top$ *be an* $n \times 1$ *random vector with the* ϵ_i *distributed totally independent with* $\mathbb{E}[\epsilon_i] = 0, \mathbb{E}\left[\epsilon_i^2\right]$, *and* $\sup_{1 \le i \le n, n \ge 1} \mathbb{E}|\epsilon_i|^{\delta_2} < \infty$ *for some* $\delta_2 > 4$.

Consider the $m \times 1$ *vector of linear quadratic forms* $\mathbf{v}_n = [Q_{1n}, ..., Q_{mn}]'$ *with:*

$$Q_{rn} = \boldsymbol{\varepsilon}' \mathbf{A}_r \boldsymbol{\varepsilon} + \mathbf{a}_r' \boldsymbol{\varepsilon} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ijr} \epsilon_i \epsilon_j + \sum_{i=1}^{n} a_{ir} \epsilon_i. \tag{13}$$

Theorem (CLT for Vectors of Linear Quadratic Forms with Heterokedastic Innovations, cont..)

Let $\mu_{\mathbf{v}} = \mathbf{E}\left[\mathbf{v}_n\right] = \left[\mu_{Q_1}, ..., \mu_{Q_2}\right]^\top$ *and* $\boldsymbol{\Sigma}_{\mathbf{v}_n} = \left[\sigma_{Q_{rs}}\right]_{r,s=1,\dots,m}$ *denote the mean and VC matrix of* $\mathbf{v}_n$*, respectively, then:*

$$\mu_{Q_r} = \sum_{i=1}^{n} a_{iir}\sigma_i^2$$

$$\sigma_{Q_{rs}} = 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ijr}a_{ijs}\sigma_i^2\sigma_j^2 + \sum_{i=1}^{n} a_{ir}a_{is}\sigma_i^2 + \sum_{i=1}^{n} a_{iir}a_{iis}\left[\mu_i^{(4)} - 3\mu_i^4\right] + \sum_{i=1}^{n}(a_{ir}a_{iis} + a_{is}a_{iir})\mu_i^{(3)}$$

with $\mu_i^{(3)} = \mathbf{E}(\epsilon_i^3)$ *and* $\mu_i^{(4)} = \mathbf{E}(\epsilon_i^4)$*. Furthermore, given that* $n^{-1}\lambda_{min}(\boldsymbol{\Sigma}_{\mathbf{v}_n}) \geq c$ *for some* $c > 0$*, then*

$$\boldsymbol{\Sigma}_{\mathbf{v}_n}^{-1/2}(\mathbf{v}_n - \mu_{\mathbf{v}_n}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}_m)$$

and thus:

$$n^{-1/2}(\mathbf{v}_n - \mu_{\mathbf{v}_n}) \overset{a}{\sim} \mathrm{N}(\mathbf{0}, n^{-1}\boldsymbol{\Sigma}_{\mathbf{v}_n})$$

Asymptotic distribution

Sketch of Proof.

As usual, we first write the estimator in terms of the population error term (for notational convenience we drop the sub indices):

$$
\begin{aligned}
\widehat{\boldsymbol{\delta}}_n &= \boldsymbol{\delta}_0 + \left(\widehat{\mathbf{Z}}^\top \widehat{\mathbf{Z}}\right)^{-1} \widehat{\mathbf{Z}}^\top \boldsymbol{\varepsilon} \\
&= \boldsymbol{\delta}_0 + \left[\left(\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right)^\top \left(\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right)\right]^{-1} \left(\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right)^\top \boldsymbol{\varepsilon} \\
&= \boldsymbol{\delta}_0 + \left[\mathbf{Z}^\top\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right]^{-1} \mathbf{Z}^\top\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\boldsymbol{\varepsilon}
\end{aligned}
$$

where we used **Rank of Instruments Assumption**. Solving for $\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0$ and multiplying by $\sqrt{n}$ we obtain:

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) &= \left[\left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)^\top \left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1} \left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)\right]^{-1} \left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)^\top \left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1} \frac{1}{\sqrt{n}}\mathbf{H}^\top\boldsymbol{\varepsilon} \\
&= \frac{1}{\sqrt{n}}\widetilde{\mathbf{P}}^\top\mathbf{H}^\top\boldsymbol{\varepsilon},
\end{aligned}
\tag{14}
$$

where

$$
\widetilde{\mathbf{P}} = \left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1} \left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right) \left[\left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)^\top \left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1} \left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)\right]^{-1}.
$$

Asymptotic distribution

Sketch of Proof.

From **Limits of Instruments Assumption**, we know that:

$$\begin{aligned} \lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{H}_n &= \mathbf{Q}_{HH} \\ \text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{Z}_n &= \mathbf{Q}_{HZ}. \end{aligned}$$

Therefore, $\widetilde{\mathbf{P}} \xrightarrow{p} \mathbf{P}$, where $\mathbf{P}$ is a finite matrix. Thus,

$$\widetilde{\mathbf{P}} - \mathbf{P} = o_p(1) \implies \widetilde{\mathbf{P}} = \mathbf{P} + o_p(1). \tag{15}$$

Inserting (15) into (14), we get:

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) &= \frac{1}{\sqrt{n}}\left[\mathbf{P} + o_p(1)\right]^\top \mathbf{H}^\top \boldsymbol{\varepsilon} \\ &= \mathbf{P}^\top \frac{1}{\sqrt{n}} \mathbf{H}^\top \boldsymbol{\varepsilon} + o_p(1) \end{aligned}$$

□

Asymptotic distribution

Sketch of Proof.

By the **Rank of Instruments Assumption** $\mathbf{H}$ is uniformly bounded in absolute value. The **Heterokedastic Errors Assumption** implies that $\epsilon_{i,n}$ forms a **triangular array** of identically distributed random variables. Furthermore, we know from that assumption that $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \text{diag}(\sigma^2_{i,n})$. By **Limits of Instruments Assumption**, we know that $\lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{H}_n = \mathbf{Q}_{HH}$ is finite and is nonsingular. Thus,

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{n}}\mathbf{H}^\top \boldsymbol{\varepsilon}\right) &= \mathbf{0} \\
\text{Var}\left(\frac{1}{\sqrt{n}}\mathbf{H}^\top \boldsymbol{\varepsilon}\right) &= \frac{1}{n}\mathbf{H}^\top \boldsymbol{\Sigma}\mathbf{H}
\end{aligned} \tag{16}
$$

Thus, by **Chebyshev's inequality** $n^{-1/2}\mathbf{P}\mathbf{H}^\top \boldsymbol{\varepsilon} = O_p(1)$ (that is it converges in distribution to something) and consequently:

$$
\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) = \mathbf{P}^\top \frac{1}{\sqrt{n}}\mathbf{H}\boldsymbol{\varepsilon} + o_p(1) = O_p(1) + o_p(1) = O_p(1) \tag{17}
$$

□

Asymptotic distribution

Sketch of Proof.

Therefore using CLT for Linear Quadratic Forms,

$$\frac{1}{\sqrt{n}}\mathbf{H}_n^{\top}\boldsymbol{\varepsilon}_n \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{H}_n^{\top}\boldsymbol{\Sigma}\mathbf{H}_n\right) \tag{18}$$

Finally :

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}) \tag{19}$$

where

$$\boldsymbol{\Omega} = \mathbf{P}^{\top}\mathbf{H}^{\top}\boldsymbol{\Sigma}\mathbf{H}\mathbf{P} \tag{20}$$

□

Some Remarks

- The estimator of $\boldsymbol{\Sigma}$ will based on HAC estimators.
- Under homokedasticity, the asymptotic variance-covariance matrix reduced to :

$$\text{Var}(\widehat{\delta}_{2SLS}) = \sigma^2_\epsilon \left(\mathbf{Q}^\top_{HZ}\mathbf{Q}^{-1}_{HH}\mathbf{Q}_{HZ}\right)^{-1} \tag{21}$$

A good estimator for the asymptotic variance will be:

$$\widehat{\text{Var}}(\widehat{\delta}_{2SLS}) = \widehat{\sigma}^2_\epsilon \left[\mathbf{Z}^\top\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{Z}\right]^{-1} \tag{22}$$

where:

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^\top\widehat{\boldsymbol{\varepsilon}}}{n}, \quad \widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \widehat{\mathbf{y}} \tag{23}$$

Preliminaries

- The SEM model is consistent but inefficient under OLS.
- ML approach estimate $\boldsymbol{\beta}$ and λ jointly by assuming the whole distribution of the error term.
- Then, we could in principle estimate this model using GLS since we know the exact form of the Heterogeneity.

Preliminaries

Recall that the SEM model is given by:

$$\begin{aligned} \mathbf{y}_n &= \mathbf{X}_n \boldsymbol{\beta}_0 + \mathbf{u}_n, \\ \mathbf{u}_n &= \lambda_0 \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n. \end{aligned} \tag{24}$$

Homokedastic Errors (Kelejian and Prucha, 1999)

The innovations $\{\epsilon_{i,n}, 1 \leq i \leq n, n \geq 1\}$ are independently and identically distributed for all n with zero mean and variance σ^2, where $0 < \sigma^2 < b$, with $b < \infty$. Additionally, the innovations are assumed to possess finite fourth moments.

Weight Matrix $\mathbf{M}_n$ (Kelejian and Prucha, 1999)

Assume the following:

1. All diagonal elements of the spatial weighting matrix $\mathbf{M}_n$ are zero.
2. The matrix $(\mathbf{I}_n - \lambda_0 \mathbf{M}_n)$ is nonsingular with $|\lambda_0| < 1$.

Given Equation (24), and Assumption (Weight Matrix $\mathbf{M}_n$), we can write $\mathbf{u}_n = (\mathbf{I}_n - \lambda \mathbf{M}_n)^{-1} \boldsymbol{\varepsilon}_n$. Therefore, the expectation and variance of $\mathbf{u}_n$ are $\mathbb{E}(\mathbf{u}_n) = 0$ and $\mathbb{E}(\mathbf{u}_n \mathbf{u}_n^\top) = \boldsymbol{\Omega}_n(\lambda_0)$, respectively, where:

$$\boldsymbol{\Omega}_n(\lambda_0) = \sigma_\epsilon^2 (\mathbf{I}_n - \lambda_0 \mathbf{M}_n)^{-1} (\mathbf{I}_n - \lambda_0 \mathbf{M}_n^\top)^{-1}.$$

Note that a row-standardized spatial weight matrix is typically not symmetric, such that $\mathbf{M}_n \neq \mathbf{M}_n^\top$ and thus $(\mathbf{I}_n - \lambda_0 \mathbf{M}_n)^{-1} \neq (\mathbf{I}_n - \lambda_0 \mathbf{M}_n^\top)^{-1}$.

SWLS

The spatially weighted least squares (SWLS) boils down to:

$$\widehat{\boldsymbol{\beta}}_{SWLS} = \left(\mathbf{X}_s^\top \mathbf{X}_s\right)^{-1} \mathbf{X}_s^\top \mathbf{y}_s, \tag{25}$$

were $\mathbf{X}_s = \mathbf{X} - \widehat{\lambda}\mathbf{W}\mathbf{X}$ and $\mathbf{y}_s = \mathbf{y} - \widehat{\lambda}\mathbf{M}\mathbf{y}$, using a consistent estimate $\widehat{\lambda}$ for the autoregressive parameter.

- Note that this model is basically and OLS applied to spatially filtered variables.
- Furthermore, it should be noted that the SWLS are nothing but a special case of Feasible Generalized Least Squares (FGLS).
- To note this consider the homoskedastic case, with $\mathbb{E}\left[\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}\right] = \sigma^2 \mathbf{I}$. Consequently:

$$\mathbb{E}\left[\mathbf{u}\mathbf{u}^\top\right] = \boldsymbol{\Omega} = \sigma^2 \left[\left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\right]^{-1}, \tag{26}$$

  and the corresponding Generalized Least Squares (GLS) estimator for $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left[\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X}\right]^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}.$$

The expression for the GLS estimator simplifies to:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{GLS} &= \left[\mathbf{X}^\top \frac{1}{\sigma^2}\left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{X}\right]^{-1} \mathbf{X}^\top \frac{1}{\sigma^2}\left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{y}, \\ &= \left[\mathbf{X}^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{X}\right]^{-1} \mathbf{X}^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{y}.\end{aligned}$$

The FGLS estimator substitutes a consistent estimate for λ into this expression, as:

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left[\mathbf{X}^\top \left(\mathbf{I}_n - \widehat{\lambda}\mathbf{M}\right)^\top \left(\mathbf{I}_n - \widehat{\lambda}\mathbf{M}\right)\mathbf{X}\right]^{-1} \mathbf{X}^\top \left(\mathbf{I}_n - \widehat{\lambda}\mathbf{M}\right)^\top \left(\mathbf{I}_n - \widehat{\lambda}\mathbf{M}\right)\mathbf{y},$$

which is the same as Equation (25).

Remarks

- So, I we had a consistent estimate of λ_0, we could use the FGLS.
- How to estimate λ_0 consistently? Kelejian and Prucha (1999) propose to estimate it by Method of Moments (MM).
- However, they do not provide the asymptotic distribution for $\widehat{\lambda}$: λ_0 is viewed purely as a nuisance parameter, whose only function is to aid in obtaining consistent estimates for $\boldsymbol{\beta}_0$.
- One advantage of the MM estimator (and of QML) is that they do not rely on the assumption of normality of the disturbances $\boldsymbol{\varepsilon}$. Nonetheless, both estimators assume that ϵ_i are independently and identically distributed for all i with zero mean and variance σ^2.

Moment Conditions

The basic idea behind a method of moments estimator is to find a set of population moments equations that provide a relationship between population moments and parameters.

Moment Conditions

Given the DGP in Equation (24), we can write:

$$\boldsymbol{\varepsilon} = \mathbf{u} - \lambda \mathbf{M}\mathbf{u},$$

where $\boldsymbol{\varepsilon}$ is the idiosyncratic error and $\mathbf{u}$ is the regression error. The MM estimation approach employs the following simple quadratic moment conditions:

$$\begin{aligned}
\mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] &= \sigma^2, \\
\mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon}\right] &= \frac{\sigma^2}{n}\mathbb{E}\left[\operatorname{tr}(\mathbf{M}^\top\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)\right], \\
\mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\boldsymbol{\varepsilon}\right] &= 0.
\end{aligned}$$

The Kelijian and Prucha (1999)'s MM estimator of λ is based on these three moments. The final value of $\mathbb{E}\left[\boldsymbol{\varepsilon}^\top\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon}\right]$ will depend on the assumption about the variance of $\boldsymbol{\varepsilon}$.

Moment Conditions

Definition (Moment Conditions)

Under **homoskedasticity** (Kelijian and Prucha, 1999) the moment conditions are:

$$\begin{aligned} \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] &= \sigma^2, \\ \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon}\right] &= \frac{\sigma^2}{n}\operatorname{tr}\left(\mathbf{M}^\top\mathbf{M}\right), \\ \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\boldsymbol{\varepsilon}\right] &= 0. \end{aligned} \tag{27}$$

Under **heterokedasticity** (Kelijian and Prucha, 2010) the moment conditions are:

$$\begin{aligned} \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] &= \sigma^2, \\ \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon}\right] &= n^{-1}\operatorname{tr}\left[\mathbf{W}\operatorname{diag}\left[\mathbb{E}(\epsilon_i^2)\right]\mathbf{W}^\top\right], \\ \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\mathbf{M}\boldsymbol{\varepsilon}\right] &= 0. \end{aligned} \tag{28}$$

Moment Conditions

In order to operationalize the moment conditions, we need to convert conditions on $\boldsymbol{\varepsilon}$ into conditions on $\mathbf{u}$ (since $\boldsymbol{\varepsilon}$ is not observed). Since $\mathbf{u} = \lambda \mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}$ if follows that $\epsilon = \mathbf{u} - \lambda \mathbf{M}\mathbf{u}$, i.e., the spatially filtered regression error terms.

$$
\begin{aligned}
\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} &= (\mathbf{u} - \lambda \mathbf{M}\mathbf{u})^\top (\mathbf{u} - \lambda \mathbf{M}\mathbf{u}) \\
&= \mathbf{u}^\top \mathbf{u} - 2\lambda \mathbf{u}^\top \mathbf{M}\mathbf{u} + \lambda^2 \mathbf{u}^\top \mathbf{M}^\top \mathbf{M}\mathbf{u} \qquad (29) \\
\boldsymbol{\varepsilon}^\top \mathbf{M}^\top \mathbf{M}\boldsymbol{\varepsilon} &= (\mathbf{u} - \lambda \mathbf{M}\mathbf{u})^\top \mathbf{M}^\top \mathbf{M}(\mathbf{u} - \lambda \mathbf{M}\mathbf{u}) \\
&= \mathbf{u}^\top \mathbf{M}^\top \mathbf{M}\mathbf{u} - 2\lambda \mathbf{u}^\top \mathbf{M}^\top \mathbf{M}\mathbf{M}\mathbf{u} + \lambda^2 \mathbf{u}^\top \mathbf{M}^\top \mathbf{M}\mathbf{M}^\top \mathbf{M}\mathbf{u} \qquad (30) \\
\boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon} &= (\mathbf{u} - \lambda \mathbf{M}\mathbf{u})^\top \mathbf{M}(\mathbf{u} - \lambda \mathbf{M}\mathbf{u}) \\
&= \mathbf{u}^\top \mathbf{M}\mathbf{u} - 2\lambda \mathbf{u}^\top \mathbf{M}\mathbf{M}\mathbf{u} + \lambda^2 \mathbf{u}^\top \mathbf{M}^\top \mathbf{M}\mathbf{M}\mathbf{u} \qquad (31)
\end{aligned}
$$

Let $\mathbf{u}_L = \mathbf{M}\mathbf{u}$, $\mathbf{u}_{LL} = \mathbf{M}\mathbf{M}\mathbf{u}$.

Moment Conditions

Taking the expectation over (29) and assuming Homokedasticity, we get:

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] &= \mathbb{E}\left[\mathbf{u}^\top\mathbf{u}\right] - 2\lambda\mathbb{E}\left[\mathbf{u}^\top\mathbf{M}\mathbf{u}\right] + \lambda^2\mathbb{E}\left[\mathbf{u}^\top\mathbf{M}^\top\mathbf{M}\mathbf{u}\right] \\
\sigma^2 &= \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}\right] - \lambda\frac{2}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}_L\right] + \lambda^2\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top\mathbf{u}_L\right] \quad \text{since } \mathbb{E}\left[n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] = \sigma^2 \\
0 &= \sigma^2 - \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}\right] + \lambda\frac{2}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}_L\right] - \lambda^2\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top\mathbf{u}_L\right] \\
0 &= \lambda\frac{2}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}_L\right] - \lambda^2\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top\mathbf{u}_L\right] + \frac{1}{n}\sigma^2 - \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}\right] \\
0 &= \begin{pmatrix} \frac{2}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}_L\right] & -\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top\mathbf{u}_L\right] & 1 \end{pmatrix}\begin{pmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{pmatrix} - \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top\mathbf{u}\right]
\end{aligned}
\tag{32}
$$

Moment Conditions

In similar fashion,

$$0 = \left(\tfrac{2}{n}\mathbb{E}\left[\mathbf{u}_{LL}^{\top}\mathbf{u}_{L}\right] \quad -\tfrac{1}{n}\mathbb{E}\left[\mathbf{u}_{LL}^{\top}\mathbf{u}_{LL}\right] \quad \tfrac{1}{n}\operatorname{tr}(\mathbf{M}^{\top}\mathbf{M}) \right) \begin{pmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{pmatrix} - \frac{1}{n}\mathbb{E}\left[\mathbf{u}_{L}^{\top}\mathbf{u}_{L}\right] \tag{33}$$

$$0 = \left(\tfrac{1}{n}\mathbb{E}\left[\mathbf{u}^{\top}\mathbf{u}_{LL} + \mathbf{u}_{L}^{\top}\mathbf{u}_{L}\right] \quad -\tfrac{1}{n}\mathbb{E}\left[\mathbf{u}_{L}^{\top}\mathbf{u}_{LL}\right] \quad 0 \right) \begin{pmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{pmatrix} - \frac{1}{n}\mathbb{E}\left[\mathbf{u}^{\top}\mathbf{u}_{L}\right] \tag{34}$$

At this point it is important to realized that **we have have three equations an three unknowns!**, λ, λ^2 and σ^2.

Moment Conditions

Consider the following three-equations system implied by Equations (32), (33) and (34)

$$\boldsymbol{\Gamma}_n \boldsymbol{\alpha} = \boldsymbol{\gamma}_n \tag{35}$$

where $\boldsymbol{\Gamma}_n$ is given in Equation (36), and $\boldsymbol{\alpha} = (\lambda, \lambda^2, \sigma^2)$. If $\boldsymbol{\Gamma}_n$ where known, Assumption (Identification) implies that Equation (35) determines $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha} = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n$ where:

$$\boldsymbol{\Gamma}_n = \begin{pmatrix} \frac{2}{n}\mathbb{E}\left[\mathbf{u}^\top \mathbf{u}_L\right] & -\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top \mathbf{u}_L\right] & 1 \\ \frac{2}{n}\mathbb{E}\left[\mathbf{u}_{LL}^\top \mathbf{u}_L\right] & -\frac{1}{n}\mathbb{E}\left[\mathbf{u}_{LL}^\top \mathbf{u}_{LL}\right] & \frac{1}{n}\operatorname{tr}(\mathbf{M}^\top \mathbf{M}) \\ \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top \mathbf{u}_{LL} + \mathbf{u}_L^\top \mathbf{u}_L\right] & -\frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top \mathbf{u}_{LL}\right] & 0 \end{pmatrix} \tag{36}$$

and

$$\boldsymbol{\gamma}_n = \begin{pmatrix} \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top \mathbf{u}\right] \\ \frac{1}{n}\mathbb{E}\left[\mathbf{u}_L^\top \mathbf{u}_L\right] \\ \frac{1}{n}\mathbb{E}\left[\mathbf{u}^\top \mathbf{u}_L\right] \end{pmatrix} \tag{37}$$

Moment Conditions

Now we express the moment conditions as sample averages in observables spatial lags of OLS residuals:

$$\mathbf{g}_n = \mathbf{G}_n \boldsymbol{\alpha} + \boldsymbol{\upsilon}_n(\lambda, \sigma^2) \tag{38}$$

which can be thought as a regression. Note also that

$$\mathbf{G}_n = \begin{pmatrix} \frac{2}{n}\widehat{\mathbf{u}}^\top \widehat{\mathbf{u}}_L & -\frac{1}{n}\widehat{\mathbf{u}}_L^\top \widehat{\mathbf{u}}_L & 1 \\ \frac{2}{n}\widehat{\mathbf{u}}_{LL}^\top \widehat{\mathbf{u}}_L & -\frac{1}{n}\widehat{\mathbf{u}}_{LL}^\top \widehat{\mathbf{u}}_{LL} & \frac{1}{n}\operatorname{tr}(\mathbf{M}^\top \mathbf{M}) \\ \frac{1}{n}\left[\widehat{\mathbf{u}}^\top \widehat{\mathbf{u}}_{LL} + \widehat{\mathbf{u}}_L^\top \widehat{\mathbf{u}}_L\right] & -\frac{1}{n}\widehat{\mathbf{u}}_L^\top \widehat{\mathbf{u}}_{LL} & 0 \end{pmatrix} \tag{39}$$

and

$$\mathbf{g}_n = \begin{pmatrix} \frac{1}{n}\widehat{\mathbf{u}}^\top \widehat{\mathbf{u}} \\ \frac{1}{n}\widehat{\mathbf{u}}_L^\top \widehat{\mathbf{u}}_L \\ \frac{1}{n}\widehat{\mathbf{u}}^\top \widehat{\mathbf{u}}_L \end{pmatrix} \tag{40}$$

where $\mathbf{G}_n$ is a 3×3 matrix, and where $\boldsymbol{\upsilon}_n(\lambda, \sigma^2)$ can be viewed as a vector of residuals.

Moment Conditions

We can define the GM estimator for λ and σ^2 as the Nonlinear Least Square (NLS) estimator corresponding to Equation (38):

$$(\widehat{\lambda}_{NLS,n}, \widehat{\sigma}^2_{NLS,N}) = \operatorname{argmin}\left\{\boldsymbol{\upsilon}_n(\lambda,\sigma^2)^\top \boldsymbol{\upsilon}_n(\lambda,\sigma^2) : \rho \in [-a,a], \sigma^2 \in [0,b]\right\} \tag{41}$$

Note that $(\widehat{\lambda}_{NLS,n}, \widehat{\sigma}^2_{NLS,N})$ are defined as the minimizers of

$$\left[\mathbf{g}_n - \mathbf{G}_n \begin{pmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{pmatrix}\right]^\top \left[\mathbf{g}_n - \mathbf{G}_n \begin{pmatrix} \lambda \\ \lambda^2 \\ \sigma^2 \end{pmatrix}\right]$$

Bounded Matrices (Kelejian and Prucha, 1999)

The row and column sums of the matrices $\mathbf{M}_n$ and $(\mathbf{I} - \lambda \mathbf{M}_n)$ are bounded uniformly in absolute value.

Residuals (Kelejian and Prucha, 1999)

Let $\widetilde{u}_{i,n}$ denote the i-th element of $\widetilde{\mathbf{u}}_n$. We then assume that

$$\widetilde{u}_{i,n} - u_{i,n} = \mathbf{d}_{i,n} \boldsymbol{\Delta}_n$$

where $\mathbf{d}_{i,n}$ and $\boldsymbol{\Delta}_n$ are $1 \times p$ and $p \times 1$ dimensional random vectors. Let $d_{ij,n}$ be the jth element of $\mathbf{d}_{i,n}$. Then, we assume that for some $\delta > 0$, $\mathbb{E}\left|d_{ij,n}\right|^{2+\delta} \leq c_d < \infty$, where c_d does not depend on n, and that

$$\sqrt{n}\left\|\boldsymbol{\Delta}_n\right\| = O_p(1). \qquad (42)$$

This assumption should be satisfied for most cases in which $\widetilde{\mathbf{u}}$ is based on $\sqrt{n}$-consistent estimators of the regression coefficients (non-linear OLS, linear OLS). Assumption Residuals (Kelejian and Prucha, 1999) comes from (Kelejian and Prucha, 2010) and is a bit stronger than the same assumption in (Kelejian and Prucha, 1999).

Assumptions

Identification (Kelejian and Prucha, 1999)

The smallest eigenvalues of $\boldsymbol{\Gamma}_n^\top \boldsymbol{\Gamma}_n$ is bounded away from zero, that is, $\omega_{min}(\boldsymbol{\Gamma}_n^\top \boldsymbol{\Gamma}_n) \geq \omega_* > 0$, where ω_* may depend on λ and σ^2

Theorem (Consistency)

Let $(\widehat{\lambda}_{NLS,n}, \widehat{\sigma}^2_{NLS,N})$ *given by:*

$$(\widehat{\lambda}_{NLS,n}, \widehat{\sigma}^2_{NLS,N}) = \operatorname{argmin}\left\{\boldsymbol{\upsilon}_n(\lambda, \sigma^2)^\top \boldsymbol{\upsilon}_n(\lambda, \sigma^2) : \rho \in [-a, a], \sigma^2 \in [0, b]\right\}$$

Then, given Assumptions (Heterokedastic errors), (Weight Matrix $\mathbf{M}_n$*), (Bounded Matrices), (Residuals), and (Identification),*

$$(\widehat{\lambda}_{NLS,n}, \widehat{\sigma}^2_{NLS,N}) \xrightarrow{p} (\lambda, \sigma^2) \quad as\ n \to \infty \tag{43}$$

An important remark is that this Theorem states only that the NLS estimates are consistent, but it does not tell us about the asymptotic distribution of $\widehat{\lambda}_{NLS,n}$.

Limiting Behavior

The elements of $\mathbf{X}$ are non-stochastic and bounded in absolute value by $c_X, 0 < c_X < \infty$. Also, $\mathbf{X}$ has full rank, and the matrix $\mathbf{Q}_X = \lim_{n\to\infty} n^{-1}\mathbf{X}^\top\mathbf{X}$ is finite and nonsingular. Furthermore, the matrices $\mathbf{Q}_X(\lambda) = \lim_{n\to\infty} n^{-1}\mathbf{X}^\top\boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}$ is finite and nonsingular for all $|\rho| < 1$

Asymptotics

Theorem (Asymptotic Properties of FGLS Estimator)

If assumptions (Homokedastic errors), (Weight Matrix $\mathbf{M}_n$*), (Bounded Matrices), and (Limiting Behavior) hold:*

1. *The true GLS estimator* $\widehat{\boldsymbol{\beta}}_{GLS}$ *is a consistent estimator for* $\boldsymbol{\beta}$*, and*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{Q}_X(\lambda)^{-1}\right) \tag{44}$$

2. *Let* $\widehat{\lambda}_n$ *be a consistent estimator for* λ*. Then the true GLS estimator* $\widehat{\boldsymbol{\beta}}_{GLS}$ *and the Feasible GLS estimator* $\widehat{\boldsymbol{\beta}}_{FGLS}$ *have the same asymptotic distribution.*
3. *Suppose further than* $\widehat{\sigma}_n^2$ *is a consistent estimator for* σ^2*. Then* $\widehat{\sigma}_n^2\left[n^{-1}\mathbf{X}^\top \boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1}\mathbf{X}\right]$ *is a consistent estimator for* $\sigma^2\mathbf{Q}_X(\lambda)^{-1}$*.*

Asymptotics

The Theorem assumes the existence of a consistent estimator of λ and σ^2. It can be shown that the OLS estimator:

$$\widehat{\boldsymbol{\beta}}_n = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}$$

is $\sqrt{n}$-consistent. Thus, the OLS residuals $\widetilde{u}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n$ satisfy Assumption 47 with $d_{i,n} = |\mathbf{x}_i|$ and $\Delta_n = \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$. Thus, OLS residuals can be used to obtain consistent estimators of ρ and σ^2.
Then, the FGLS estimator is given by

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left[\mathbf{X}^\top(\widetilde{\lambda})\mathbf{X}(\widetilde{\lambda})\right]^{-1} \mathbf{X}^\top(\widetilde{\lambda})\mathbf{y}(\widetilde{\lambda}) \tag{45}$$

where:

$$\begin{aligned} \mathbf{X}(\widetilde{\lambda}) &= (\mathbf{I} - \widetilde{\lambda}\mathbf{M})\mathbf{X} \\ \mathbf{y}(\widetilde{\lambda}) &= (\mathbf{I} - \widetilde{\lambda}\mathbf{M})\mathbf{y} \end{aligned} \tag{46}$$

Asymptotics

The variance covariance matrix of $\widehat{\boldsymbol{\beta}}_{FGLS}$ is estimated as:

$$\widehat{\mathrm{Var}}\left(\widehat{\boldsymbol{\beta}}_{FGLS}\right) = \widehat{\sigma}^2\left[\mathbf{X}^\top(\widetilde{\lambda})\mathbf{X}(\widetilde{\lambda})\right]^{-1}, \tag{47}$$

where:

$$\begin{aligned} \widehat{\sigma}^2 &= \widehat{\boldsymbol{\varepsilon}}^\top(\widetilde{\lambda})\widehat{\boldsymbol{\varepsilon}}(\widetilde{\lambda}) \\ \widehat{\boldsymbol{\varepsilon}}(\widetilde{\lambda}) &= \mathbf{y}(\widetilde{\lambda}) - \mathbf{X}(\widetilde{\lambda})\widehat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{I} - \widetilde{\lambda}\mathbf{M})\widehat{\mathbf{u}} \\ \widehat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{FGLS} \end{aligned} \tag{48}$$

Theorem (CLT for triangular arrays with homokedastic errors, Kelejian and Prucha (1999))

Let $\{v_{i,n}, 1 \leq i \leq n, n \geq 1\}$ be a triangular array of identically distributed random variables. Assume that the random variables $\{v_{i,n}, 1 \leq i \leq n\}$ are jointly independently distributed for each n with $\mathbb{E}(v_{i,n}) = 0$ and $\mathbb{E}(v_{i,n}^2) = \sigma^2 < \infty$. Let $\{a_{ij,n}, 1 \leq i \leq n, n \geq 1\}, j = 1, ..., k$ be triangular arrays of real numbers that are bounded in absolute value. Further let

$$\mathbf{v}_n = \begin{pmatrix} v_{1,n} \\ \vdots \\ v_{n,n} \end{pmatrix}, \quad \mathbf{A}_n = \begin{pmatrix} a_{11,n} & \dots & a_{1k,n} \\ \vdots & & \vdots \\ a_{n1,n} & \dots & a_{nk,n} \end{pmatrix}$$

Assume that $\lim_{n\to\infty} n^{-1}\mathbf{A}_n^\top \mathbf{A}_n = \mathbf{Q}_{AA}$ is finite and nonsingular matrix. Then

$$\frac{1}{\sqrt{n}}\mathbf{A}_n^\top \mathbf{v}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_{AA})$$

Sketch of Proof of Theorem.

We first prove part (a). Recall that the GLS and FGSL estimator are given by:

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left[\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{y}$$

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left[\mathbf{X}^\top \widehat{\boldsymbol{\Omega}}(\widehat{\lambda})^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^\top \widehat{\boldsymbol{\Omega}}(\widehat{\lambda})^{-1}\mathbf{y}$$

Since $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, the sampling error of $\widehat{\boldsymbol{\beta}}_{GLS}$ is,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left[\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{u}$$

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left[\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\left(\mathbf{I}_n - \lambda\mathbf{M}\right)^{-1}\boldsymbol{\varepsilon}$$

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left[\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^\top \left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top \boldsymbol{\varepsilon}$$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left[\frac{1}{n}\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}\frac{1}{\sqrt{n}}\mathbf{A}^\top \boldsymbol{\varepsilon}$$

where $\mathbf{A} = \left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{X}$. □

Sketch of Proof of Theorem.

By Assumption (Limiting Behavior):

$$\frac{1}{n}\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X} \to \mathbf{Q}_X(\lambda)$$

Since $\mathbf{Q}_X$ is not singular:

$$\left[\frac{1}{n}\mathbf{X}^\top \boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1} \to \mathbf{Q}_X^{-1}(\lambda)$$

Since $\mathbf{A}$ is bounded in abolute value, by Theorem it follows that:

$$\frac{1}{\sqrt{n}}\mathbf{A}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \lim_{n\to\infty} n^{-1}\sigma^2\mathbf{A}^\top\mathbf{A}\right) \tag{49}$$

where $\lim_{n\to\infty} n^{-1}\sigma^2\mathbf{A}^\top\mathbf{A} = \sigma^2 \lim_{n\to\infty} n^{-1}\mathbf{X}^\top\left(\mathbf{I}_n - \lambda\mathbf{M}\right)^\top\left(\mathbf{I}_n - \lambda\mathbf{M}\right)\mathbf{X} = \sigma^2\mathbf{Q}_X(\lambda)$. □

Sketch of Proof of Theorem.

Consequently:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \underbrace{\left[\frac{1}{n}\mathbf{X}^\top\boldsymbol{\Omega}(\lambda)^{-1}\mathbf{X}\right]^{-1}}_{\to\mathbf{Q}_X^{-1}(\lambda)} \quad \underbrace{\frac{1}{\sqrt{n}}\mathbf{A}^\top\boldsymbol{\varepsilon}}_{\xrightarrow{d}\mathrm{N}(\mathbf{0},\sigma^2\mathbf{Q}_X(\lambda))}$$

$$\overset{a}{\sim} \mathrm{N}\left[\mathbf{0}, \mathbf{Q}_X^{-1}(\lambda)\sigma^2\mathbf{Q}_X(\lambda)\mathbf{Q}_X^{-1}(\lambda)^\top)\right]$$

$$\overset{a}{\sim} \mathrm{N}\left[\mathbf{0}, \sigma^2\mathbf{Q}_X^{-1}(\lambda)\right]$$

This also implies that $\widehat{\boldsymbol{\beta}}_{GLS}$ is consistent. To show part (b), we can show that:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{GLS}-\widehat{\boldsymbol{\beta}}_{FGLS}) \xrightarrow{p} 0$$

Following Kelijian and Prucha (1999), if suffices to show that

$$\frac{1}{n}\mathbf{X}^\top\left[\boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1}-\boldsymbol{\Omega}(\lambda)^{-1}\right]\mathbf{X}\xrightarrow{p}\mathbf{0} \tag{50}$$

and

$$\frac{1}{n}\mathbf{X}^\top\left[\boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1}-\boldsymbol{\Omega}(\lambda)^{-1}\right]\mathbf{u}\xrightarrow{p}\mathbf{0}$$

Sketch of Proof of Theorem.

Since:

$$\boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1} - \boldsymbol{\Omega}(\lambda)^{-1} = (\lambda - \widehat{\lambda}_n)(\mathbf{M} + \mathbf{M}^\top) + (\lambda^2 - \widehat{\lambda}_n^2)\mathbf{M}^\top\mathbf{M}$$

Then using the fact the we have summable matrices,

$$\frac{1}{n}\mathbf{X}^\top\left[\boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1} - \boldsymbol{\Omega}(\lambda)^{-1}\right]\mathbf{X} = \underbrace{(\lambda - \widehat{\lambda}_n)}_{\xrightarrow{p} 0}\underbrace{n^{-1}\mathbf{X}^\top(\mathbf{M} + \mathbf{M}^\top)\mathbf{X}}_{O(1)} + \underbrace{(\lambda^2 - \widehat{\lambda}_n^2)}_{\xrightarrow{p} 0}\underbrace{n^{-1}\mathbf{X}^\top\mathbf{M}^\top\mathbf{M}\mathbf{X}}_{O(1)}$$

where $(\lambda - \widehat{\lambda}_n) \xrightarrow{p} 0$ since $\widehat{\lambda}_n$ is a consistent estimate of λ, and :

$$\begin{aligned}\frac{1}{n}\mathbf{X}^\top\left[\boldsymbol{\Omega}(\widehat{\lambda}_n)^{-1} - \boldsymbol{\Omega}(\lambda)^{-1}\right]\mathbf{u} &= \underbrace{(\lambda - \widehat{\lambda}_n)}_{\xrightarrow{p} 0}\underbrace{n^{-1/2}\mathbf{X}^\top(\mathbf{M} + \mathbf{M}^\top)\mathbf{u}}_{O_p(1)} + \underbrace{(\lambda^2 - \widehat{\lambda}_n^2)}_{\xrightarrow{p} 0}\underbrace{n^{-1/2}\mathbf{X}^\top\mathbf{M}^\top\mathbf{M}\mathbf{u}}_{O_p(1)} \\ &= o_p(1) * O_p(1) + o_p(1) * O_p(1) \\ &= o_p(1) + o_p(1) \\ &= o_p(1) \\ &\xrightarrow{p} 0\end{aligned} \tag{51}$$

□

Asymptotics

Sketch of Proof of Theorem.

To see that $n^{-1/2}\mathbf{X}^\top(\mathbf{M}+\mathbf{M}^\top)\mathbf{u} = O_p(1)$ note

$$\begin{aligned}
\mathbb{E}\left[n^{-1/2}\mathbf{X}^\top(\mathbf{M}+\mathbf{M}^\top)\mathbf{u}\right] &= 0 \\
\mathrm{Var}[n^{-1/2}\mathbf{X}^\top(\mathbf{M}+\mathbf{M}^\top)\mathbf{u}] &= n^{-1}\underbrace{\mathbf{X}^\top\underbrace{(\mathbf{M}+\mathbf{M}^\top)\boldsymbol{\Omega}(\mathbf{M}^\top+\mathbf{M})}_{\text{absolutely summable}}\mathbf{X}}_{O(n)} = O(1)
\end{aligned}$$

A similar result holds for $n^{-1/2}\mathbf{X}^\top\mathbf{M}^\top\mathbf{M}\mathbf{u}$.
Part 3 of the theorem follows from (50) and the fact that $\widehat{\sigma}^2$ is a consistent estimator for σ^2. □

A Feasible GLS (FGLS) can be obtained along with the following steps:

GLS (FGLS) Algorithm of SEM

The steps are the following:

1. First of all obtain a consistent estimate of $\boldsymbol{\beta}$, say $\widetilde{\boldsymbol{\beta}}$ using either OLS or NLS.
2. Use this estimate to obtain an estimate of $\mathbf{u}$, say $\widehat{\mathbf{u}}$,
3. Use $\widehat{\mathbf{u}}$, to estimate λ, say $\widehat{\lambda}$, using MM estimator (NLS estimator),
4. Estimate $\boldsymbol{\beta}$ using the FGLS estimator.

Consider the following SAC model:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \mathbf{u} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \\ \mathbf{u} &= \lambda\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \tag{52}$$

Applying the spatial Cochrane-Orcutt transformation to the first equation:

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + (\mathbf{I} - \lambda\mathbf{M})^{-1}\boldsymbol{\varepsilon} \\ (\mathbf{I} - \lambda\mathbf{M})\mathbf{y} &= (\mathbf{I} - \lambda\mathbf{M})\mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \\ \mathbf{y}_s(\lambda) &= \mathbf{Z}_s(\lambda)\boldsymbol{\delta} + \boldsymbol{\varepsilon} \end{aligned} \tag{53}$$

where the spatially filtered variables are given by:

$$\begin{aligned} \mathbf{y}_s(\lambda) &= \mathbf{y} - \lambda\mathbf{M}\mathbf{y} \\ &= \mathbf{y} - \lambda\mathbf{y}_L \\ &= (\mathbf{I} - \lambda\mathbf{M})\mathbf{y} \\ \mathbf{Z}_s(\lambda) &= \mathbf{Z} - \lambda\mathbf{M}\mathbf{Z} \\ &= \mathbf{Z} - \lambda\mathbf{Z}_L \\ &= (\mathbf{I} - \lambda\mathbf{M})\mathbf{Z} \end{aligned}$$

Intuition

If we knew λ, we would be able to apply an **IV approach on the transformed model**. For the discussion below, assume that we know λ. Note that the ideal instruments in this case will be:

$$\mathbb{E}(\mathbf{Z}) = \mathbb{E}[\mathbf{X}, \mathbf{W}\mathbb{E}(\mathbf{y})]$$
$$\mathbb{E}(\mathbf{MZ}) = \mathbb{E}[\mathbf{MX}, \mathbf{MW}\mathbb{E}(\mathbf{y})]$$

Given that all the columns of $\mathbb{E}(\mathbf{Z})$ and $\mathbb{E}(\mathbf{MZ})$ are linear in

$$\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, ..., \mathbf{MX}, \mathbf{MWX}, \mathbf{MW}^2\mathbf{X}, ... \tag{54}$$

Let the matrix of instruments, $\mathbf{H}$, be a subset of the columns in (54), for example

$$\mathbf{H} = \left[\mathbf{X}, \mathbf{WX}, ..., \mathbf{W}^l\mathbf{X}, \mathbf{MX}, \mathbf{MWX}, ..., \mathbf{MW}^l\mathbf{X}\right],$$

where typically, $l \leq 2$. Then:

$$\mathbf{PZ} = (\mathbf{X}, \mathbf{PWy})$$
$$\mathbf{PMZ} = (\mathbf{MX}, \mathbf{PMWy}).$$

Intuition

Since we have the instruments $\mathbf{H}$, and we have assumed that we have $\widehat{\lambda}$ such that $\widehat{\lambda} \xrightarrow{p} \lambda_0$ we might apply a GMM-type procedure using the following moment conditions:

$$\mathbf{m}(\lambda_0, \boldsymbol{\delta}_0) = \mathbb{E}\left[\frac{1}{n}\mathbf{H}^\top \mathbf{u}\right] = 0$$

Obviously, the corresponding GMM estimator is just the 2SLS estimator. Note that for the transformed model (53), the moment conditions would be

$$\mathbf{m}(\lambda_0, \boldsymbol{\delta}_0) = \mathbb{E}\left[\frac{1}{\sqrt{n}}\mathbf{H}^\top \boldsymbol{\varepsilon}\right] = 0$$

Now let $\widetilde{\lambda}$ some consistent estimator for λ_0 which can be obtained in a previous step, then the sample moment vector is:

$$\mathbf{m}^\delta(\widetilde{\lambda}, \boldsymbol{\delta}) = \frac{1}{\sqrt{n}}\mathbf{H}^\top \underbrace{\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]}_{\widetilde{\boldsymbol{\varepsilon}}},$$

where we explicitly state that the moments depends on $\boldsymbol{\delta}$—which will be estimated—and a consistent estimate of λ.

Intuition

Under **homoskedasticity** the variance-covariance matrix of the moment vector $\mathbf{g}(\lambda_0, \delta_0)$ is given by:

$$\text{Var}(\mathbf{m}(\lambda_0, \boldsymbol{\delta}_0)) = \mathbb{E}(\mathbf{m}(\lambda_0, \delta_0)\mathbf{m}(\lambda_0, \delta_0)^\top) = \sigma^2 n^{-1}\mathbf{H}^\top\mathbf{H},$$

which motivates the following two-step GMM estimator for $\boldsymbol{\delta}_0$:

$$\widehat{\boldsymbol{\delta}} = \underset{\boldsymbol{\delta}}{\text{argmin}} \quad \mathbf{g}_n^{\delta}(\widetilde{\lambda}, \boldsymbol{\delta})^\top \boldsymbol{\Upsilon}_n^{\delta\delta}\mathbf{g}_n^{\delta}(\widetilde{\lambda}, \boldsymbol{\delta})$$

with

$$\boldsymbol{\Upsilon}_n^{\delta\delta} = \left[\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right]^{-1}.$$

Intuition

Note that:

$$\begin{aligned}
J_n &= \left[\frac{1}{\sqrt{n}}\mathbf{H}^\top\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]\right]^\top \left[\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right]^{-1}\left[\frac{1}{\sqrt{n}}\mathbf{H}^\top\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]\right] \\
&= \frac{1}{n}\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]^\top \mathbf{H}\left[\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right]^{-1}\mathbf{H}^\top\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right] \\
&= \left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]^\top \mathbf{H}\left[\mathbf{H}^\top\mathbf{H}\right]^{-1}\mathbf{H}^\top\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right] \\
&= \left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]^\top \mathbf{P}_H\left[\mathbf{y}_s(\widetilde{\lambda}) - \mathbf{Z}_s(\widetilde{\lambda})\boldsymbol{\delta}\right]
\end{aligned}$$

Then, the estimator of $\boldsymbol{\delta}$ will be:

$$\widehat{\boldsymbol{\delta}} = \left[\widehat{\mathbf{Z}_s}^\top\mathbf{Z}_s\right]^{-1}\widehat{\mathbf{Z}_s}^\top\mathbf{y}_s$$

where $\widehat{\mathbf{Z}_s} = \mathbf{H}\left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}\mathbf{Z}_s$. This estimator has been called the Feasible Generalized Spatial Two-stage Least Squares (FGS2SLS) estimator (Kelijian and Prucha, 1998). However, this estimator is not fully efficient.

Intuation

The question is: How to obtain a consistent estimator of $\widehat{\lambda}$? As probably you can guess, this consistent estimator is obtained in a previous step by GM.

Moment conditions

Consider the **homokedastic** model and the following three moment conditions:

$$
\begin{aligned}
\mathbb{E}\left[\varepsilon^\top \varepsilon\right] &= \sigma^2 \\
\mathbb{E}\left[\varepsilon^\top \mathbf{M}\mathbf{M}\varepsilon\right] &= \sigma^2 \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M}\right) \\
\mathbb{E}\left[\varepsilon^\top \mathbf{M}\varepsilon\right] &= 0
\end{aligned}
$$

Substituting out σ^2 into the second moment equation yields:

$$
\begin{aligned}
\mathbb{E}\left[\varepsilon^\top \mathbf{M}\mathbf{M}\varepsilon\right] - \mathbb{E}\left[\varepsilon^\top \varepsilon\right] \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M}\right) &= 0 \\
\mathbb{E}\left[\varepsilon^\top \mathbf{M}\mathbf{M}\varepsilon - \varepsilon^\top \varepsilon \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M}\right)\right] &= 0 \\
\mathbb{E}\left[\varepsilon^\top \mathbf{M}\mathbf{M}\varepsilon - \varepsilon^\top \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M}\right)\varepsilon\right] &= 0 \\
\mathbb{E}\left[\varepsilon^\top \left(\mathbf{M}\mathbf{M} - \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M}\right)\mathbf{I}\right)\varepsilon\right] &= 0 \\
\mathbb{E}\left[\varepsilon^\top \mathbf{A}_1 \varepsilon\right] &= 0.
\end{aligned}
$$

Moment conditions

Generalizing this expression for the third moment we end up with two instead of three quadratic moment conditions:

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left[\boldsymbol{\varepsilon}^\top \mathbf{A}_1\boldsymbol{\varepsilon}\right] &= \mathbf{0}\\ \frac{1}{n}\mathbb{E}\left[\boldsymbol{\varepsilon}^\top \mathbf{A}_2\boldsymbol{\varepsilon}\right] &= \mathbf{0}\end{aligned} \tag{55}$$

with

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{M}\mathbf{M} - n^{-1}\operatorname{tr}\left(\mathbf{M}^\top\mathbf{M}\right)\mathbf{I}\\ \mathbf{A}_2 &= \mathbf{M}.\end{aligned}$$

Note that $\mathbf{A}_1$ is symmetric with $\operatorname{tr}(\mathbf{A}_1) = 0$ (you should be able to prove this), but its diagonal elements are non zero (In the heteroskedasticity case it is!). In Drukker et al. (2013), an additional scaling factor is included as:

$$\nu = 1/\left[1 + \left[(1/n)\operatorname{tr}\left(\mathbf{M}^\top\mathbf{M}\right)\right]^2\right].$$

Moment Conditions

Under this case the weighting matrices for quadratic moments are:

$$\mathbf{A}_1 = \nu \left[\mathbf{M}\mathbf{M} - n^{-1}\operatorname{tr}\left(\mathbf{M}^\top\mathbf{M}\right)\mathbf{I}\right]$$
$$\mathbf{A}_2 = \mathbf{M}_n.$$

If the errors are **heterokedastic**, then:

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{M}^\top\mathbf{M} - n^{-1}\operatorname{diag}\left(\mathbf{M}^\top\mathbf{M}\right) = \mathbf{M}^\top\mathbf{M} - n^{-1}\operatorname{diag}\left(\mathbf{m}_i^\top\mathbf{m}_i\right)\\ \mathbf{A}_2 &= \mathbf{M},\end{aligned}$$

where $\mathbf{m}_i$ is the ith column of the weights matrix $\mathbf{M}$. Note that $\operatorname{diag}\left(\mathbf{m}_i^\top\mathbf{m}_i\right)$ consists of the sum of the squares of the weight in the ith column. Denote this matrix as $\mathbf{D}$.

Moment Conditions

Since $\mathbf{u} = \lambda \mathbf{u}_L + \boldsymbol{\varepsilon}$, it follows that $\boldsymbol{\varepsilon} = \mathbf{u} - \lambda \mathbf{u}_L = \mathbf{u}_s$, the spatially filtered residuals. Then:

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left[\mathbf{u}_s^\top \mathbf{A}_1 \mathbf{u}_s\right] &= \mathbf{0} \\ \frac{1}{n}\mathbb{E}\left[\mathbf{u}_s^\top \mathbf{A}_2 \mathbf{u}_s\right] &= \mathbf{0} \end{aligned} \tag{56}$$

or more general

$$\mathbb{E}\left[\mathbf{u}^\top \left(\mathbf{I} - \lambda \mathbf{M}^\top\right) \mathbf{A}_q \left(\mathbf{I} - \lambda \mathbf{M}^\top\right) \mathbf{u}\right] = 0 \tag{57}$$

Now, we can express the **sample moment conditions** as :

$$\underset{2\times 1}{\mathbf{m}} = \underset{2\times 1}{\widehat{\mathbf{g}}} - \underset{2\times 2}{\widehat{\mathbf{G}}} \begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} = \mathbf{0}$$

The elements of $\widehat{\mathbf{g}}$ are the following (see Kelijian and Prucha, 2010, pag.56):

$$\mathbf{g}_1 = \frac{1}{n}\mathbf{u}^\top \mathbf{A}_1 \mathbf{u} = \frac{1}{n}\left[\mathbf{u}_L^\top \mathbf{u}_L - \mathbf{u}^\top \mathbf{D} \mathbf{u}\right] \tag{58}$$

$$\mathbf{g}_2 = \frac{1}{n}\mathbf{u}^\top \mathbf{A}_2 \mathbf{u} = \frac{1}{n}\mathbf{u}^\top \mathbf{u}_L \tag{59}$$

Moment Conditions

The $\widehat{\mathbf{G}}$ matrix is given by:

$$\mathbf{G}_{11} = 2n^{-1}\mathbf{u}^\top\mathbf{M}^\top\mathbf{A}_1\mathbf{u} \tag{60}$$

$$\mathbf{G}_{12} = -n^{-1}\mathbf{u}^\top\mathbf{M}^\top\mathbf{A}_1\mathbf{M}\mathbf{u} \tag{61}$$

$$\mathbf{G}_{21} = -n^{-1}\mathbf{u}^\top\mathbf{M}^\top\left(\mathbf{A}_2+\mathbf{A}_2^\top\right)\mathbf{u} \tag{62}$$

$$\mathbf{G}_{22} = -n^{-1}\mathbf{u}^\top\mathbf{M}\mathbf{A}_2\mathbf{M}\mathbf{u} \tag{63}$$

A more compact notation is:

$$\widetilde{\mathbf{G}} = \frac{1}{n}\begin{pmatrix} \widetilde{\mathbf{u}}^\top\left(\mathbf{A}_1+\mathbf{A}_1^\top\right)\widetilde{\mathbf{u}}_s & -\widetilde{\mathbf{u}}_s^\top\mathbf{A}_1\widetilde{\mathbf{u}}_s^\top \\ \vdots & \vdots \\ \widetilde{\mathbf{u}}^\top\left(\mathbf{A}_q+\mathbf{A}_q^\top\right)\widetilde{\mathbf{u}}_s & -\widetilde{\mathbf{u}}_s^\top\mathbf{A}_q\widetilde{\mathbf{u}}_s^\top \end{pmatrix}$$
$$\widetilde{\mathbf{g}} = \frac{1}{n}\begin{pmatrix} \widetilde{\mathbf{u}}^\top\mathbf{A}_1\widetilde{\mathbf{u}} \\ \vdots \\ \widetilde{\mathbf{u}}^\top\mathbf{A}_q\widetilde{\mathbf{u}} \end{pmatrix} \tag{64}$$

Assumptions

Now we will state the assumption for the SAC model under heteroskedasticity following Arraiz et al. (2010). The assumptions regarding the spatial weight matrix are the following:

Spatial Weights Matrices (Arraiz et al., 2010)

Assume the following:

(a) All diagonal elements $\mathbf{W}_n$ and $\mathbf{M}_n$ are zero.

(b) $\lambda \in (-1, 1)$, $\rho \in (-1, 1)$.

(c) The matrices $\mathbf{I}_n - \rho\mathbf{W}_n$ and $\mathbf{I}_n - \lambda\mathbf{M}_n$ are nonsingular for all $\lambda \in (-1, 1)$ and $\rho \in (-1, 1)$.

Part (a) is a normalization rule: a region cannot be a neighbor of itself. Part (b) has to do with the parameter space. This assumption is discussed by Kelejian and Prucha (2010, section 2.2). part (c) ensures that $\mathbf{y}$ and $\mathbf{u}$ are uniquely defined.

Thus, under these assumptions, we can write the model as:

$$\mathbf{y}_n = (\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} [\mathbf{X}_n \boldsymbol{\beta} + \mathbf{u}_n]$$
$$\mathbf{u}_n = (\mathbf{I}_n - \rho \mathbf{M}_n)^{-1} \boldsymbol{\varepsilon}_n.$$

The reduced form is:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} (\mathbf{I} - \lambda \mathbf{M})^{-1} \boldsymbol{\varepsilon}$$

The reduced form represents a system of n simultaneous equations. As in the standard spatial lag model, we can include endogenous explanatory variables on the right hand side of model specification. In this case:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{Y} \boldsymbol{\gamma} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \boldsymbol{\varepsilon}.$$

Assumptions

Heteroskedastic Errors (Arraiz et al., 2010)

The error term $\{\epsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ satisfy $\mathbb{E}(\epsilon_{i,n}) = 0$, $\mathbb{E}(\epsilon^2_{i,n}) = \sigma^2_{i,n}$, with $0 < \underline{a}^{\sigma} \leq \sigma^2_{i,n} \leq \overline{a}^{\sigma} < \infty$. Furthermore, for each $n \geq 1$ the random variables $\epsilon_{1,n}, ..., \epsilon_{n,n}$ are totally independent.

This assumption allows the innovations to be heteroskedastic with uniformly bounded variances. This assumption also allows for the innovations to depend on the sample size n, i.e., to form a triangular arrays.

Assumptions

Bounded Spatial Weight Matrices (Arraiz et al., 2010)

The row and column sums of the matrices $\mathbf{W}_n$ and $\mathbf{M}_n$ are bounded uniformly in absolute value, by , respectively, one and some finite constant, and the row and column sums of the matrices $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ and $(\mathbf{I} - \rho\mathbf{M}_n)^{-1}$ are bounded uniformly in absolute value by some finite constant.

This assumption limits the extent of spatial autocorrelation between $\mathbf{u}$ and $\mathbf{y}$. It ensures that the disturbance process and the process of the dependent variable exhibit a "fading" memory. Note that:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{u}_n\right] &= \mathbb{E}\left[(\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}\boldsymbol{\varepsilon}_n\right] \\
&= (\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}\mathbb{E}\left[\boldsymbol{\varepsilon}_n\right] \qquad (65) \\
&= \mathbf{0} \quad \text{by Heteroskedastic Errors}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\mathbf{u}_n\mathbf{u}_n^\top\right] &= \mathbb{E}\left[(\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^\top\left(\mathbf{I}_n - \lambda\mathbf{M}_n^\top\right)^{-1}\right] \\
&= (\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}\mathbb{E}\left[\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^\top\right]\left(\mathbf{I} - \lambda\mathbf{M}_n^\top\right)^{-1} \qquad (66) \\
&= (\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}\boldsymbol{\Sigma}\left(\mathbf{I}_n - \lambda\mathbf{M}_n^\top\right)^{-1}
\end{aligned}$$

where $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma^2_{i,n})$.

Assumptions

Regressors (Arraiz et al., 2010)

The regressor matrices $\mathbf{X}_n$ have full column rank (for n large enough). Furthermore, the elements of the matrices $\mathbf{X}_n$ are uniformly bounded in absolute value.

This assumption rules out multicollinearity problems, as well as unbounded exogenous variables.

Instruments I (Arraiz et al., 2010)

The instruments matrices $\mathbf{H}_n$ have full column rank $L \geq K + 1$ (for all n large enough). Furthermore, the elements of the matrices $\mathbf{H}_n$ are uniformly bounded in absolute value. Additionally, $\mathbf{H}_n$ is assumed to, at least, contain the linearly independent columns of $(\mathbf{X}_n, \mathbf{M}_n\mathbf{X}_n)$

There are some papers that discuss the use of optimal instruments for the spatial (see for example Lee, 2003; Das et al., 2003; Kelejian et al., 2004, Lee, 2007)

Assumptions

Instruments II (Identification) (Arraiz et al., 2010)

The instruments $\mathbf{H}_n$ satisfy furthermore:

(a) $\mathbf{Q}_{HH} = \lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{H}_n$ is finite and nonsingular.

(b) $\mathbf{Q}_{HZ} = \text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{Z}_n$ and $\mathbf{Q}_{HMZ} = \text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n^\top \mathbf{M}\mathbf{Z}_n$ are finite and have full column rank. Furthermore $\mathbf{Q}_{HZ,s}(\lambda) = \mathbf{Q}_{HZ} - \lambda\mathbf{Q}_{HMZ}$ has full column rank.

(c) $\mathbf{Q}_{H\Sigma H} = \lim_{n\to\infty} n^{-1}\mathbf{H}_n^\top \Sigma_n \mathbf{H}_n$ is finite and nonsingular, where $\boldsymbol{\Sigma}_n = \text{diag}(\sigma_{i,n}^2)$

In treating $\mathbf{X}_n$ and $\mathbf{H}_n$ as non-stochastic our analysis should be viewed as conditional on $\mathbf{X}_n$ and $\mathbf{H}_n$.

Procedure

Consider again the transformed model:

$$\mathbf{y}_s(\lambda_0) = \mathbf{Z}_s(\lambda_0)\boldsymbol{\delta}_0 + \epsilon$$

where $\mathbf{y}_s(\lambda_0) = \mathbf{y} - \lambda_0 \mathbf{M}\mathbf{y}$ and $\mathbf{Z}_s(\lambda_0) = \mathbf{Z} - \lambda_0 \mathbf{M}\mathbf{Z}$. If we would know λ_0, then we could apply the S2SLS to the transformed model. However, λ_0 is unknown and therefore we need to estimate it in a first place in order to estimate $\boldsymbol{\delta}$. The steps will be:

1. An initial IV estimator of δ leads to a set of consistent residuals.
2. With these residuals, derive the moment conditions that provide a consistent estimate of λ_0 using GMM Estimation procedure.
3. The estimate of λ_0 is then used to define a **weighting matrix** for the moment conditions in order to obtain a consistent and efficient estimator.
4. An estimate of δ_0 is obtained from the **transformed model**.
5. Finally, a **consistent and efficient** estimate of λ is based on GS2SLS residuals.

Now we will consider each step in detail

Step 1a: 2SLS estimator

In the first step, $\boldsymbol{\delta}$ is estimated by 2SLS applied to the untransformed model model $\mathbf{y}_n = \mathbf{Z}_n\boldsymbol{\delta} + \mathbf{u}_n$ using the instruments matrix $\mathbf{H}$. Then:

$$\widetilde{\delta}_{2SLS} = \left(\widetilde{\mathbf{Z}}^\top \mathbf{Z}\right)^{-1} \widetilde{\mathbf{Z}}^\top \mathbf{y} \tag{67}$$

where $\widetilde{\mathbf{Z}} = \mathbf{H}\left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top\mathbf{Z} = \mathbf{P}_H\mathbf{Z} = (\mathbf{X}, \widetilde{\mathbf{W}\mathbf{y}})$. The estimates $\widetilde{\delta}_{2SLS}$ yield an initial vector of residuals, $\mathbf{u}_{2SLS}$ as:

$$\widetilde{\mathbf{u}}_{2SLS} = \mathbf{y} - \mathbf{Z}\widetilde{\delta}_{2SLS} \tag{68}$$

Although $\widetilde{\delta}_{2SLS}$ is consistent, it does not utilize information relating to the spatial correlation error term. We therefore turn to the second step of the procedure.

Step 1b: Initial GMM estimator of λ based on 2SLS residuals

Using the consistent estimate $\mathbf{u}$ in the previous step, now we create the sample moments corresponding to Equation (57) for $q = 1, 2$ based on the estimated residuals, and $\widetilde{\mathbf{u}}_s = \mathbf{M}\widetilde{\mathbf{u}}$:

$$
\begin{aligned}
\mathbf{m}(\lambda, \widetilde{\boldsymbol{\delta}}_{2SLS}) &= \frac{1}{n}\begin{pmatrix} \widetilde{\mathbf{u}}_{2SLS}^{\top}\left(\mathbf{I}-\lambda\mathbf{M}^{\top}\right)\mathbf{A}_1\left(\mathbf{I}-\lambda\mathbf{M}\right)\widetilde{\mathbf{u}}_{2SLS} \\ \widetilde{\mathbf{u}}_{2SLS}^{\top}\left(\mathbf{I}-\lambda\mathbf{M}^{\top}\right)\mathbf{A}_2\left(\mathbf{I}-\lambda\mathbf{M}\right)\widetilde{\mathbf{u}}_{2SLS} \end{pmatrix} \\
&= \widetilde{\mathbf{G}}\begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} - \widetilde{\mathbf{g}}
\end{aligned} \tag{69}
$$

where,

$$
\begin{aligned}
\widetilde{\mathbf{G}} &= \frac{1}{n}\begin{pmatrix} \widetilde{\mathbf{u}}^{\top}\left(\mathbf{A}_1+\mathbf{A}_1^{\top}\right)\widetilde{\mathbf{u}}_s & -\widetilde{\mathbf{u}}_s^{\top}\mathbf{A}_1\widetilde{\mathbf{u}}_s^{\top} \\ \vdots & \vdots \\ \widetilde{\mathbf{u}}^{\top}\left(\mathbf{A}_q+\mathbf{A}_q^{\top}\right)\widetilde{\mathbf{u}}_s & -\widetilde{\mathbf{u}}_s^{\top}\mathbf{A}_q\widetilde{\mathbf{u}}_s^{\top} \end{pmatrix} \\
\widetilde{\mathbf{g}} &= \frac{1}{n}\begin{pmatrix} \widetilde{\mathbf{u}}^{\top}\mathbf{A}_1\widetilde{\mathbf{u}} \\ \vdots \\ \widetilde{\mathbf{u}}^{\top}\mathbf{A}_q\widetilde{\mathbf{u}} \end{pmatrix}
\end{aligned} \tag{70}
$$

Step 1b: Initial GMM estimator of λ based on 2SLS residuals

The initial GMM estimator for λ is then defined as

$$\breve{\lambda}_{gmm} = \underset{\lambda}{\operatorname{argmin}} \left\{ \left[\widetilde{\mathbf{G}} \begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} - \widetilde{\mathbf{g}} \right]^\top \left[\widetilde{\mathbf{G}} \begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} - \widetilde{\mathbf{g}} \right] \right\} \tag{71}$$

where $\boldsymbol{\Upsilon}^{\lambda\lambda} = \mathbf{I}$. This estimator is consistent but not efficient. For efficiency we need to replace $\boldsymbol{\Upsilon}^{\lambda\lambda}$ by the variance-covariance matrix of the sample moments. Furthermore, the expression above can be interpreted as a NLS system of equations. The initial estimate is thus obtained as a solution of the above system.

Now, we need to define the expression for the matrices $\mathbf{A}_s$. Drukker et al. (2013) suggest, for the homokedastic case, the following expressions:

$$\begin{aligned} \mathbf{A}_1 &= \upsilon \left[\mathbf{M}^\top \mathbf{M} - \frac{1}{n} \operatorname{tr}\left(\mathbf{M}^\top \mathbf{M} \right) \mathbf{I} \right] \\ \mathbf{A}_2 &= \mathbf{M} \end{aligned} \tag{72}$$

where υ is the scaling factor needed to obtain the same estimator of Kelejian and Prucha (1998, 1999).

Step 1b: Initial GMM estimator of λ based on 2SLS residuals

On the other hand, when heteroskedasticity is assumed, Kelejian and Prucha (2010) recommend the following expressions:

$$
\begin{aligned}
\mathbf{A}_1 &= \mathbf{M}^\top \mathbf{M} - \text{diag}(\mathbf{M}^\top \mathbf{M}) \\
\mathbf{A}_2 &= \mathbf{M}
\end{aligned}
\tag{73}
$$

Step 1c: Efficient GMM estimator of λ based on 2SLS residuals

The efficient GMM estimator of λ is a weighted NLS estimator. Specifically, this estimator is $\breve{\lambda}$ where:

$$\widetilde{\lambda}_{ogmm} = \underset{\lambda}{\operatorname{argmin}} \left[\mathbf{m}(\lambda, \widetilde{\boldsymbol{\delta}})^\top \widetilde{\boldsymbol{\Psi}}^{-1} \mathbf{m}(\lambda, \widetilde{\boldsymbol{\delta}})\right] \tag{74}$$

and where the weighting matrix is $\widetilde{\boldsymbol{\Psi}}_n^{-1}$. The matrix $\widetilde{\boldsymbol{\Psi}}_n^{-1} = \widetilde{\boldsymbol{\Psi}}_n^{-1}(\breve{\lambda}_{gmm})$ is defined as follows. Let $\widetilde{\boldsymbol{\Psi}} = \left[\widehat{\Psi}_{rs}\right]_{r,s=1,2}$ with

$$\widetilde{\Psi}_{rs} = (2n)^{-1} \operatorname{tr}\left[(\mathbf{A}_r + \mathbf{A}_r^\top)\widetilde{\boldsymbol{\Sigma}}(\mathbf{A}_s + \mathbf{A}_s^\top)\widetilde{\boldsymbol{\Sigma}}\right] + n^{-1}\widetilde{\mathbf{a}}_r^\top \widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{a}}_s, \tag{75}$$

where:

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}} &= \operatorname{diag}_{i=1,\dots,n}\left(\widetilde{\epsilon}_i^2\right) \\
\widetilde{\epsilon} &= \left(\mathbf{I} - \breve{\lambda}_{gmm}\mathbf{M}\right)\widetilde{\mathbf{u}} \\
\widetilde{\mathbf{a}}_r &= \left(\mathbf{I} - \breve{\lambda}_{gmm}\mathbf{M}\right)\mathbf{H}\widetilde{\mathbf{P}}\widetilde{\boldsymbol{\alpha}}_r \\
\widetilde{\boldsymbol{\alpha}}_r &= -n^{-1}\left[\mathbf{Z}^\top\left(\mathbf{I} - \breve{\lambda}_{gmm}\mathbf{M}\right)(\mathbf{A}_r + \mathbf{A}_r^\top)\left(\mathbf{I} - \breve{\lambda}_{gmm}\mathbf{M}\right)\widetilde{\mathbf{u}}\right] \\
\widetilde{\mathbf{P}} &= \left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1}\left(\frac{1}{n}\mathbf{H}_n^\top\mathbf{Z}_n\right)\left[\left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)^\top\left(\frac{1}{n}\mathbf{H}^\top\mathbf{H}\right)^{-1}\left(\frac{1}{n}\mathbf{H}^\top\mathbf{Z}\right)\right]^{-1}
\end{aligned} \tag{76}$$

Step 2a: G2SLS Estimator

Using $\breve{\lambda}_{ogmm}$ from step 1c (or the consistent estimator from step 1b) in the transformed model we have:

$$\widehat{\boldsymbol{\delta}}_n(\widetilde{\lambda}_{ogmm}) = \left[\widehat{\mathbf{Z}}_s^\top(\widetilde{\lambda}_{ogmm})\mathbf{Z}(\widetilde{\lambda}_{ogmm})\right]^{-1}\widehat{\mathbf{Z}}_s^\top(\widetilde{\lambda}_{ogmm})\mathbf{y}_s(\widetilde{\lambda}_{ogmm}) \tag{77}$$

where

$$\begin{aligned}
\mathbf{y}_s &= \mathbf{y} - \widetilde{\lambda}_{ogmm}\mathbf{M}\mathbf{y} \\
\mathbf{Z}_s &= \mathbf{Z} - \widetilde{\lambda}_{ogmm}\mathbf{M}\mathbf{Z} \\
\widehat{\mathbf{Z}}_s &= \mathbf{P}_H\mathbf{Z}_s \\
\mathbf{P}_H &= \mathbf{H}\left(\mathbf{H}^\top\mathbf{H}\right)^{-1}\mathbf{H}^\top
\end{aligned} \tag{78}$$

Step 2b: Efficient GMM estimator of λ using GS2SLS residual

An **efficient** GMM estimator of λ based on the GS2SLS residuals is obtained by minimizing the following expression:

$$\widehat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \left\{ \left[\widehat{\mathbf{G}} \begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} - \widehat{\mathbf{g}} \right]^{\top} (\widehat{\boldsymbol{\Psi}}^{\widehat{\lambda\lambda}})^{-1} \left[\widehat{\mathbf{G}} \begin{pmatrix} \lambda \\ \lambda^2 \end{pmatrix} - \widehat{\mathbf{g}} \right] \right\} \tag{79}$$

where $\widehat{\boldsymbol{\Psi}}^{\widehat{\lambda\lambda}}$ is an estimator for the variance-covariance matrix of the (normalized) sample moment vector based on the GS2SLS residuals. This estimator differs for the cases of homoskedastic and heteroskedastic errors.

Step 2b: Efficient GMM estimator of λ using GS2SLS residual

For the **homoskedastic** case the r, s (with $r, s = 1, 2$) element of $\widehat{\boldsymbol{\Psi}}^{\widehat{\lambda\lambda}}$ is given by:

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}_{rs}^{\widehat{\lambda\lambda}} = &\left[\widetilde{\sigma}^2\right]^2 (2n)^{-1} \operatorname{tr}\left[\left(\mathbf{A}_r + \mathbf{A}_r^\top\right)\left(\mathbf{A}_s + \mathbf{A}_s^\top\right)\right] \\ &+ \widetilde{\sigma}^2 n^{-1} \widetilde{\mathbf{a}}_r^\top \widetilde{\mathbf{a}}_s^\top \\ &+ n^{-1}\left(\widetilde{\mu}^{(4)} - 3\left[\widetilde{\sigma}^2\right]^2\right) \operatorname{vec}_D\left(\mathbf{A}_r\right)^\top \operatorname{vec}_D\left(\mathbf{A}_s\right) \\ &+ n^{-1}\widetilde{\mu}^{(3)}\left[\widetilde{\mathbf{a}}_r^\top \operatorname{vec}_D\left(\mathbf{A}_s\right) + \widetilde{\mathbf{a}}_s^\top \operatorname{vec}_D\left(\mathbf{A}_r\right)\right],\end{aligned} \tag{80}$$

Step 2b: Efficient GMM estimator of λ using GS2SLS residual

where

$$
\begin{aligned}
\widetilde{\mathbf{a}}_r &= \widehat{\mathbf{T}}\widetilde{\alpha}_r \\
\widehat{\mathbf{T}} &= \mathbf{H}\widehat{\mathbf{P}}, \\
\widehat{\mathbf{P}} &= \widehat{\mathbf{Q}}_{HH}^{-1}\widehat{\mathbf{Q}}_{HZ}\left[\widehat{\mathbf{Q}}_{HZ}^{\top}\widehat{\mathbf{Q}}_{HH}^{-1}\widehat{\mathbf{Q}}_{HZ}^{\top}\right]^{-1} \\
\widehat{\mathbf{Q}}_{HH}^{-1} &= \left(n^{-1}\mathbf{H}^{\top}\mathbf{H}\right), \\
\widehat{\mathbf{Q}}_{HZ} &= \left(n^{-1}\mathbf{H}^{\top}\mathbf{Z}\right), \\
\mathbf{Z} &= \left(\mathbf{I}-\widetilde{\lambda}\mathbf{M}\right)\mathbf{Z}, \\
\widetilde{\alpha}_r &= -n^{-1}\left[\mathbf{Z}^{\top}\left(\mathbf{A}_r+\mathbf{A}_r^{\top}\right)\widehat{\boldsymbol{\varepsilon}}\right] \\
\widehat{\sigma}^2 &= n^{-1}\widehat{\boldsymbol{\varepsilon}}\widehat{\boldsymbol{\varepsilon}}, \\
\widehat{\mu}^{(3)} &= n^{-1}\sum_{i=1}^{n}\widehat{\epsilon}_i^3, \\
\widehat{\mu}^{(4)} &= n^{-1}\sum_{i=1}^{n}\widehat{\epsilon}_i^4.
\end{aligned}
\tag{81}
$$

Step 2b: Efficient GMM estimator of λ using GS2SLS residual

For the **heteroskedastic** case the r, s (with $r, s = 1, 2$) element of $\widehat{\boldsymbol{\Psi}}^{\widehat{\lambda\lambda}}$ is given by:

$$\widehat{\boldsymbol{\Psi}}_{rs}^{\widehat{\lambda\lambda}} = (2n)^{-1} \operatorname{tr}\left[\left(\mathbf{A}_r + \mathbf{A}_r^\top\right) \widehat{\boldsymbol{\Sigma}} \left(\mathbf{A}_s + \mathbf{A}_s^\top\right) \widehat{\boldsymbol{\Sigma}}\right] + n^{-1} \widetilde{\mathbf{a}}_r^\top \widehat{\boldsymbol{\Sigma}} \widetilde{\mathbf{a}}_s^\top, \tag{82}$$

where, $\widehat{\boldsymbol{\Sigma}}$ is a diagonal matrix whose ith diagonal element is $\widehat{\epsilon}_i^2$.