Regression Modeling of Spatial Relationships

ECON6027 Spatial Econometrics & Data Analysis, 7b, theory

Spatial Linear Regression Models & Testing

Outline

Outline

The general static model

- This chapter considers different specifications of linear spatial econometric models that can be considered when we have reasons to believe that spatial autocorrelation exists in some form among the areal units.
- The general linear model that one can consider is,

$$\begin{aligned} Y_n &= \rho W_n Y_n + X_n \beta + u_n \qquad |\rho| < 1 \qquad (1) \\ u_n &= \lambda W_n u_n + \epsilon_n \qquad |\lambda| < 1 \qquad (2) \end{aligned}$$

  where X_n is a matrix of non-stochastic regressors, W_n is an exogenous weights matrix, $\epsilon_n | X_n \sim iidN(0, \sigma_\epsilon^2 I_n)$ and $\theta = (\beta_1', \beta_2', \rho, \lambda, \sigma_\epsilon^2)'$ is the vector of parameters to be estimated.
- W_n is generally row-normalised and this ensures the restriction on λ and ρ to hold.

The general model (contd.)

- Equation (1) considers the spatially lagged dependent variable $(W_n Y_n)$ with spatial parameter ρ.
- Equation (2) considers a spatial model of the stochastic disturbances with spatial parameter λ.
- in principle the weights matrix W_n can be the same or different. However, in practice there is no good reason for the W_n to be different.
- This model is known in the literature as the **SARAR(1,1) model**.
- If in addition the model also has a lagged independent variable, $W_n X_n \gamma$, then we call this the spatial *Durbin* model. We will not consider the Durbin specification in this lesson.

Nested spatial models

Nested within SARAR(1,1) model are three very interesting sub-models:

1. Pure spatial autoregressive model: $\beta = \lambda = 0$
2. Spatial lag model: $\lambda = 0$
3. Spatial error model: $\rho = 0$

Spatial parameter space

Equations (1) and (2) can also be expressed as,

$$
\begin{aligned}
Y_n &= (I_n - \rho W_n)^{-1}(X_n\beta + u_n) \\
u_n &= (I_n - \lambda W_n)^{-1}\epsilon_n
\end{aligned}
$$

When the weights matrix W_n is row standardised, the two inverse matrices exists when $|\lambda| < 1$ and $|\rho| < 1$.[1]

[1] In theory, one can impose much more general conditions involving the eigenvalues of W_n, that allow the parameter space to be larger than the specified space. However, in most practical applications, the given conditions are generally sufficient.

Outline

The pure spatial autoregressive (SAR) model

The pure spatial autoregressive (SAR) model is given as,

$$Y_n = \rho W_n Y_n + \epsilon_n, \qquad |\lambda| < 1 \tag{3}$$

The SAR model parameters $\theta = (\rho, \sigma_\epsilon^2)$ can be estimated using the maximum likelihood method. Note that the model can also be represented as,

$$Y_n = (I_n - \rho W_n)^{-1} \epsilon_n$$

Nore that in this case, $\mathrm{E}(Y_n) = 0$ and
$\mathrm{E}(Y_n Y_n') = \sigma_\epsilon^2 (I_n - \rho W_n)^{-1} (I_n - \rho W_n')^{-1}$

Maximum likelihood estimation of the SAR model

Under the assumption of $\epsilon_n \sim i.i.d.N(0, \sigma_\epsilon^2 I_n)$, the n areal units have a joint pdf of,

$$f(\epsilon_{1,\dots,n}|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}$$

The *likelihood function* has the same form except that it is a function of the parameters $\theta = (\rho, \sigma_\epsilon^2)$,

$$L(\theta|\epsilon_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}$$

If we only have $\epsilon_n \sim i.i.d(0, \sigma_\epsilon^2 I_n)$, then this is called a *quasi* likelihood function.

Log-likelihood function

Let $A_n(\rho) = (I_n - \rho W_n)$. The (quasi) log-likelihood function is,

$$
\begin{aligned}
\ell_n(\theta) &= \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{\epsilon_i^2}{2\sigma_\epsilon^2}\right\}\right) \\
&= -\frac{n}{2}\ln(2\pi\sigma_\epsilon^2) + \ln\left(\exp\left\{-\frac{Y_n' A_n'(\rho) A_n(\rho) Y_n}{2\sigma_\epsilon^2}\right\} \cdot |A_n(\rho)|\right) \\
&= -\frac{n}{2}\ln(2\pi\sigma_\epsilon^2) + \ln|A_n(\rho)| - \frac{1}{2\sigma_\epsilon^2} Y_n' A_n'(\rho) A_n(\rho) Y_n
\end{aligned}
$$

Note, $\sum_{i=1}^n e_i^2 = Y_n' A_n'(\rho) A_n(\rho) Y_n$. The transformation from ϵ_i to Y_i is via $\epsilon_n = A_n(\rho) Y_n$, thus, the Jacobean of the transformation is, $|\partial\epsilon_n / \partial Y_n| = |A_n(\rho)|$.

ML Estimators

The first order condition with respect to σ_ϵ^2 is,

$$\frac{\partial}{\partial \sigma_\epsilon^2} \ell_n(\theta) = -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} Y_n' A_n'(\rho) A_n(\rho) Y_n = 0$$

Thus the ML estimator for σ_ϵ^2 is,

$$\hat{\sigma}^2_{\epsilon,ML}(\rho) = \frac{1}{n} Y_n' A_n'(\rho) A_n(\rho) Y_n$$

Plugging $\hat{\sigma}^2_{\epsilon,ML}(\rho)$ back in the log-likelihood function gives us the *concentrated* log-likelihood function:

$$\ell_n^c(\rho) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2} \ln\left(\hat{\sigma}^2_{\epsilon,ML}(\rho)\right) + \ln|A_n(\rho)|$$

Maximum likelihood estimation

- The concentrated log-likelihood function is non-linear with respect to ρ in the sense that it **does not give a closed form solution** for the ρ estimator.
- As such, estimation of ρ requires numerical maximisation which can be done using the R function, *spatialreg::spautolm()*.
- The main problem is maximising the log-likelihood function (or the concentrated version) is the term $\ln |A_n(\rho)|$.
- Numerical optimisation of the concentrated log-likelihood function requires repeated evaluation of the determinant of an $n \times n$ matrix.
- If n is large, this process can be a demanding job even for the most powerful computers.

Ord/Spectral decomposition

- A way out is to use the *Ord decomposition*,

$$\ln |I_n - \rho W_n| = \left| \prod_{i=1}^{n} (1 - \rho \tau_i) \right|$$

  where, τ_i is the ith eigenvector of the weights matrix W_n.
- This is the methodology applied in software packages and is known as *pseudo-likelihood* method.[2]
- Although the Ord decomposition simplifies the computation it is not so accurate for large n since the spectral decomposition is also approximated in very large n.

[2]The term "pseudo-likelihood" was in fact introduced by Julian Besag in the context of analysing data having spatial dependence. The practical use of this is that it can provide an approximation to the likelihood function of a set of observed data which may either provide a computationally simpler problem for estimation, or may provide a way of obtaining explicit estimates of model parameters.

Outline

Spatial lag model (SLM)

The spatial lag model (SLM) is given as,

$$Y_n = \rho W_n Y_n + X_n \beta + \epsilon_n, \qquad |\rho| < 1 \tag{4}$$

The SAR model parameter vector is now, $\theta = (\beta', \rho, \sigma_\epsilon^2)$. Note that the model can also be expressed as,

$$Y_n = (I_n - \rho W_n)^{-1}(X_n \beta + \epsilon_n) \tag{5}$$

Estimation of the SLM parameters

- This model suffers from the problem of endogeneity in the sense that the spatially lagged dependent variable, $W_n Y_n$ is correlated with the disturbances.

$$\begin{aligned} E\left[(W_n Y_n)\epsilon_n'\right] &= E\left([W_n(I_n - \rho W_n)^{-1}(X_n\beta + \epsilon_n)]\epsilon_n'\right) \\ &= W_n(I_n - \rho W_n)^{-1}E(\epsilon_n\epsilon_n') \\ &= \sigma_\epsilon^2 W_n(I_n - \rho W_n)^{-1} \neq 0 \end{aligned}$$

  and the OLS estimator loses it's optimality properties.
- Two possible estimation methodologies are,
  1. maximum likelihood
  2. two stage least squares
- The ML estimators are consistent and efficient however, biased in small samples. On the other hand 2SLS estimator is unbiased and consistent, however, not so efficient. The 2SLS estimator also has the advantage of being computationally faster.

1. Maximum likelihood estimation of SLM parameters

The (quasi) log-likelihood function is,

$$\begin{aligned} \ell_n(\theta) &= -\frac{n}{2}\ln(2\pi\sigma_\epsilon^2) + \ln|A_n(\rho)| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\left[A_n(\rho)Y_n - X_n\beta\right]'\left[A_n(\rho)Y_n - X_n\beta\right] \end{aligned} \tag{6}$$

where $A_n(\rho) = (I_n - \rho W_n)$. As before this expression is non-linear with respect to the spatial parameter ρ and requires numerical maximisation which can be done using the R function, *spatialreg::lagsarlm()*.

Try this!

Given $\epsilon_n = A_n(\rho)Y_n - X_n\beta$, setup the likelihood function and derive the log-likelihood given in (6). In particular, show how the Jacobean of transformation term is $|A_n(\rho)|$.

ML estimators of SLM parameters

Closed form ML estimator for the parameters β and σ_ϵ^2 can be given by taking the first order conditions from the log-likelihood function (6),

$$\begin{aligned}
\tfrac{\partial}{\partial \beta'} \ell_n(\theta) &= \tfrac{1}{\sigma_\epsilon^2} X_n' \left[A_n(\rho) Y_n - X_n \beta \right] \\
\tfrac{\partial}{\partial \sigma_\epsilon^2} \ell_n(\theta) &= -\tfrac{n}{2\sigma_\epsilon^2} + \tfrac{1}{2\sigma_\epsilon^4} \left[A_n(\rho) Y_n - X_n \beta \right]' \left[A_n(\rho) Y_n - X_n \beta \right]
\end{aligned}$$

Thus, the closed form solutions to the ML estimators are:

$$\begin{aligned}
\hat{\beta}_{ML}(\rho) &= \left(X_n' X_n \right)^{-1} X_n' A_n(\rho) Y_n \\
\hat{\sigma}_{\epsilon,ML}^2(\rho) &= \frac{1}{n} \left[A_n(\rho) Y_n - X_n \hat{\beta} \right]' \left[A_n(\rho) Y_n - X_n \hat{\beta} \right]
\end{aligned}$$

ML estimators of SLM parameters (contd.)

- Once we plug these solutions back in the log-likelihood function (6), we get the *concentrated* log-likelihood function:

$$\ell_n^c(\rho) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2} \ln\left(\hat{\sigma}^2_{\epsilon,ML}(\rho)\right) + \ln|A_n(\rho)|$$

- In practice, it is this concentrated version of the likelihood function that is numerically optimised to find solutions for $\hat{\rho}$ which in turn is plugged into the closed form estimators $\hat{\beta}_{ML}(\rho)$ and $\hat{\sigma}^2_{\epsilon,ML}(\rho)$ to give ML estimates $\hat{\beta}_{ML}(\hat{\rho})$ and $\hat{\sigma}^2_{\epsilon,ML}(\hat{\rho})$. So the final vector of ML estimates are given as: $\hat{\theta} = \left(\hat{\beta}'_{ML}(\hat{\rho}), \hat{\rho}, \hat{\sigma}^2_{\epsilon,ML}(\hat{\rho})\right)'$.
- Note that the inability to give a closed form expression for the ρ estimator is what makes this likelihood function *non-linear* with respect to ρ and ultimately the source of the bias in finite samples.
- The maximum likelihood estimator, however, is consistent.

2. Two stage least squares estimation of SLM parameters

- In order to get rid of the endogeneity problem, we can also consider a 2SLS estimator.
- To implement the 2SLS procedure, we need valid instruments that are correlated with $W_n Y_n$ (instrument relevance), and uncorrelated with the disturbance term (instrument exogeneity).
- Consider the Taylor expansion,

$$(I_n - \rho W_n)^{-1} = I_n + \rho W_n + \rho^2 W_n^2 + \rho^3 W_n^3 + \dots$$

- Given the alternative expression of the SLM given in (5), we have the following moment condition,

$$E(Y_n) \quad = \quad (I_n - \rho W_n)^{-1} X_n \beta$$

2SLS estimation of SLM parameters (contd.)

- Using this moment condition, we have

$$\begin{aligned} E(Y_n) &= (I_n + \rho W_n + \rho^2 W_n^2 + \rho^3 W_n^3 + \ldots)X_n\beta \\ &= X_n\beta + \rho W_n X_n\beta + \rho^2 W_n^2 X_n\beta + \rho^3 W_n^3 X_n\beta + \ldots \end{aligned}$$

so that $E(Y_n)$ is a linear function of $X_n, W_nX_n, W_n^2X_n, \ldots$

- Now, consider the set of instruments,

$$H_n = \left[X_n, W_nX_n, W_n^2X_n\right]$$

- Express the model (4) as

$$Y_n = Z_n\gamma + \epsilon_n \tag{7}$$

where, $Z_n = [W_nY_n + X_n]$ and $\gamma = (\rho, \beta')'$.

2SLS estimation of SLM parameters (contd.)

Stage 1: Regress the independent variable Z_n on the instruments H_n using OLS,

$$\begin{aligned} Z_n &= H_n\delta + \eta_n \\ \hat{\delta}_{OLS} &= \left(H_n' H_n\right)^{-1} H_n' Z_n \\ \hat{Z}_n &= H_n \hat{\delta}_{OLS} = H_n \left(H_n' H_n\right)^{-1} H_n' Z_n \end{aligned}$$

Stage 2: Estimate the (7) using OLS,

$$\begin{aligned} Y_n &= \hat{Z}_n \gamma + \epsilon_n \\ \hat{\gamma}_{OLS} &= \left(\hat{Z}_n' \hat{Z}_n\right)^{-1} \hat{Z}_n' Y_n \end{aligned}$$

2SLS estimator can be implemented in R using the function, *spatialreg::stsls()*.

Outline

Spatial error model (SEM)

The spatial error model (SEM) is given as,

$$
\begin{aligned}
Y_n &= X_n\beta + u_n && (8)\\
u_n &= \lambda W_n u_n + \epsilon_n \qquad |\lambda| < 1 && (9)
\end{aligned}
$$

The SAR model parameter vector is now, $\theta = (\beta', \lambda, \sigma_\epsilon^2)$.

SEM moment conditions

- The moment conditions based on the disturbances are,

$$\begin{aligned} E(u_n) &= 0 \\ E(u_n u_n') &= \sigma_\epsilon^2 (I_n - \lambda W_n)^{-1} (I_n - \lambda W_n')^{-1} = \sigma_\epsilon^2 \Omega_n \end{aligned}$$

- Thus the disturbances in the SEM considers both heteroskedastic and autocorrelated structures.
- In this case, the usual GLS procedure described in 7a can only be applied if the parameter λ is known as a priori (which is unlikely the case).

Estimation of SEM

Given $u_n = (I_n - \lambda W_n)^{-1}\epsilon_n$ the equation (9) can also be written as,

$$\begin{aligned} Y_n &= X_n\beta + (I_n - \lambda W_n)^{-1}\epsilon_n \\ Y_n &= \lambda W_n Y_n + (I_n - \lambda W_n)X_n\beta + \epsilon_n \end{aligned}$$

Thus, the covariance between the lagged variable $W_n Y_n$ and disturbances is,

$$\begin{aligned} E\left[(W_n Y_n)\epsilon_n'\right] &= E\left(W_n[X_n\beta + (I_n - \lambda W_n)^{-1}\epsilon_n]\epsilon_n'\right) \\ &= E[W_n(I_n - \lambda W_n)^{-1}\epsilon_n\epsilon_n'] \\ &= \sigma_\epsilon^2 W_n(I_n - \lambda W_n)^{-1} \neq 0 \end{aligned}$$

As such the disturbance term is endogenous in that it is correlated with the spatially lagged variable $W_n Y_n$ and the OLS estimator loses it's optimality properties.

Estimation of SEM (contd.)

- Unlike the spatial lag model, it is not possible to apply a 2SLS procedure for estimation to the spatial error model since it is **not possible to identify instruments** for $W_n Y_n$ in this model that are linearly independent to the other regressors X_n and $W_n X_n$.
- Two possible estimation methodologies are available,
  1. maximum likelihood
  2. Feasible GLS (FGLS)
- The ML estimators are consistent and efficient however, biased in small samples. On the other hand FGLS estimators are unbiased and consistent, however, not as efficient as the ML estimators. In addition, the FGLS estimator has the advantage of being computationally faster.

1. Maximum likelihood estimation of SEM parameters

The (quasi) log-likelihood function is,

$$
\begin{aligned}
\ell_n(\theta) &= -\frac{n}{2}\ln(2\pi\sigma_\epsilon^2) - \frac{1}{2}\ln|\Omega_n| \\
&\quad -\frac{1}{2\sigma_\epsilon^2}(Y_n - X_n\beta)'\,\Omega_n^{-1}(Y_n - X_n\beta)
\end{aligned}
\tag{10}
$$

where, $\Omega_n = (I_n - \lambda W_n)^{-1}(I_n - \lambda W_n')^{-1}$ so that,

$$
\begin{aligned}
-\frac{1}{2}\ln|\Omega_n| &= -\frac{1}{2}\ln|(I_n - \lambda W_n)^{-1}(I_n - \lambda W_n')^{-1}| \\
&= \ln|I_n - \lambda W_n|
\end{aligned}
$$

Try this!

Given $\epsilon_n = B_n(\lambda)(Y_n - X_n\beta)$, where $B_n(\lambda) = I_n - \lambda W_n$, setup the likelihood function and derive the log-likelihood given in (10). In particular, show how the Jacobean of transformation term is $|B_n(\lambda)|$.

ML estimators of SEM parameters

- Equation (10) corresponds to deriving the log-likelihood function of the Cochrane–Orcutt transformed model,

$$Y_n^* = X_n^* \beta + \epsilon$$

where $Y_n^* = (I_n - \lambda W_n) Y_n$ and $X_n^* = (I_n - \lambda W_n) X_n$.

- This expression is non-linear with respect to the spatial parameter λ and requires numerical maximisation which can be done using the R function, *spatialreg::errorsarlm()*.

ML estimators of SEM parameters (contd.)

The first order conditions on (10) gives us the closed form expressions for ML estimators of β and σ^2_ϵ as,

$$\begin{aligned}\hat{\beta}_{ML}(\lambda) &= \left(X_n'\Omega_n^{-1}X_n\right)^{-1}X_n'\Omega_n^{-1}Y_n \\ \hat{\sigma}^2_{\epsilon,ML}(\lambda) &= \frac{1}{n}\left[Y_n - X_n\hat{\beta}\right]'\Omega_n^{-1}\left[Y_n - X_n\hat{\beta}\right]\end{aligned}$$

Once we plug these solutions back in the log-likelihood function (10), we get the *concentrated* log-likelihood function:

$$\ell_n^c(\lambda) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln\left(\hat{\sigma}^2_{\epsilon,ML}(\lambda)\right) + \ln|B_n(\lambda)|$$

where $B_n(\lambda) = I_n - \lambda W_n$.

ML estimators of SEM parameters (contd.)

- Once the concentrated log-likelihood function is numerically optimised to find solutions for $\hat{\lambda}$ we get ML estimates $\hat{\beta}_{ML}(\hat{\lambda})$ and $\hat{\sigma}^2_{\epsilon,ML}(\hat{\lambda})$. So the final vector of ML estimates are given as: $\hat{\theta} = \left(\hat{\beta}'_{ML}(\hat{\lambda}), \hat{\lambda}, \hat{\sigma}^2_{\epsilon,ML}(\hat{\lambda})\right)'$.
- Note that the inability to give a closed form expression for the λ estimator is what makes this likelihood function *non-linear* with respect to λ and ultimately the source of the bias in finite samples.
- The maximum likelihood estimator, however, is consistent.

2. Feasible GLS estimation of SEM parameters

Consider the model given in (9). A feasible GLS (FGLS) procedure can be obtained along the following steps:

Step **1:** Obtain a consistent estimate of β. As a consistent estimator for β, consider the OLS estimator,[3]

$$\tilde{\beta} = \left(X_n' X_n\right)^{-1} X_n' Y_n$$

Step **2:** Use $\tilde{\beta}$ to obtain an estimate of u_n

$$\hat{u}_n = Y_n - X_n \tilde{\beta}$$

Step **3:** Use $\hat{u}_n$ to estimate λ, say $\hat{\lambda}$ (details can be found in pages 58-62 of the reference).

[3] One can also consider the consistent maximum likelihood estimator for β from the previous slide.

Step **4:** Use $\hat{\lambda}$ to transform the model (9) as,

$$(I_n - \hat{\lambda} W_n) Y_n = (I_n - \hat{\lambda} W_n) X_n \beta + \epsilon$$

The VC matrix of this transformed model, $\Omega_n = (I_n - \lambda W_n)^{-1}(I_n - \lambda W_n')^{-1}$ can be estimated by plugging-in $\hat{\lambda}$.

Step **5:** Estimate the regression parameter β via GLS by substituting the VC matrix thus obtaining,

$$\hat{\beta} = \left(X_n' \hat{\Omega}_n^{-1} X_n \right)^{-1} X_n' \hat{\Omega}_n^{-1} Y_n$$

where $\hat{\Omega}_n$ the plug-in estimator from Step 4. This operation corresponds to applying the OLS method to the transformed model in Step 4.

The FGLS estimator can be implemented in R using the function, *spatialreg::GMerrorsar()*

Outline

The SARAR(1,1) model

The SARAR(1,1) model as given in equation (1) and (2) is,

$$\begin{aligned} Y_n &= \rho W_n Y_n + X_n\beta + u_n \qquad |\rho| < 1 \\ u_n &= \lambda W_n u_n + \epsilon_n \qquad\qquad |\lambda| < 1 \end{aligned}$$

where $\theta = (\beta_1', \beta_2', \rho, \lambda, \sigma_\epsilon^2)'$ is the vector of parameters to be estimated. This model can also be represented as,

$$Y_n = (I_n - \rho W_n)^{-1}[X_n\beta + (I_n - \lambda W_n)^{-1}\epsilon_n]$$

Estimation of SARAR(1,1)

- As with the nested models, the problem of endogeneity is still present where the covariance between the lagged variable $W_n Y_n$ and disturbances is non-zero.
- As such the OLS estimator loses it's optimality properties.
- The two main methods of estimating the model parameters in the literature are,
  1. Maximum likelihood estimator
  2. Generalised spatial two stage least squares estimator (GS2SLS)
- The ML estimators are consistent and efficient however, biased in small samples. On the other hand GS2SLS estimators are unbiased and consistent, however, not as efficient as the ML estimators. In addition, the GS2SLS estimator has the advantage of being computationally faster.

1. Maximum likelihood estimation of SARAR(1,1)

- Let $A_n(\rho) = (I_n - \rho W_n)$ and $B_n(\lambda) = (I_n - \lambda W_n)$. The (quasi) log-likelihood function is,

$$\begin{aligned} \ell_n(\theta) &= -\frac{n}{2}\ln(2\pi\sigma_\epsilon^2) + \ln|A_n(\rho)| + \ln|B_n(\lambda)| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\left[A_n(\rho)Y_n - X_n\beta\right]' B_n'(\lambda)B_n(\lambda)\left[A_n(\lambda)Y_n - X_n\beta\right] \end{aligned} \tag{11}$$

- This expression is non-linear with respect to the spatial parameters λ and ρ and requires numerical maximisation which can be done using the R function, *spatialreg::sacsarlm()*.
- As before, this expression corresponds to deriving the log-likelihood function of the Cochrane–Orcutt transformed model,

$$Y_n^* = X_n^*\beta + \epsilon$$

where $Y_n^* = (I_n - \lambda W_n)(I_n - \rho W_n)Y_n$ and $X_n^* = (I_n - \lambda W_n)X_n$.

Try this!

Given $\epsilon_n = B_n(\lambda)\left[A_n(\rho)Y_n - X_n\beta\right]$, setup the likelihood function and derive the log-likelihood given in (11). In particular, show how the Jacobean of transformation term is $|B_n(\lambda)A_n(\rho)|$.

GS2SLS estimation of SARAR(1,1)

- The complete details of the generalised spatial two stage least squares estimator, omitted here, can be found in pages 78-79 of the reference.
- The GS2SLS estimator can be implemented in R using the function, *spatialreg::gstsls()*

Outline

Lagrange multiplier test

- A pitfall of the Moran's test is that it does not consider a specific alternative in terms of regression modeling. i.e., if the null of no spatial autocorrelation among the residuals is rejected, the test does not give a *specific* alternative model to consider.
- On the other hand a Lagrange Multiplier (LM) test (a.k.a. Rao's Score test, or simply the score test) allows us to consider explicit alternative hypothesis tests either in the form of SLM or SEM.

LM test statistic

- Consider the general form of the LM test,

$$LM = s(\theta_0)' I(\theta_0) s(\theta_0) \tag{12}$$

  where θ_0 is a vector of parameters, $s(\theta_0) = \frac{\partial}{\partial \theta_0} \ell(\theta_0)$ is the score function corresponding to the first order conditions, and $I(\theta_0) = E\left[\frac{\partial^2}{\partial \theta \partial \theta'} \ell(\theta)\right]$ is the Fisher's Information matrix associated with the likelihood function under the null of no spatial autocorrelation (i.e., $Y_n = X_n \beta + \epsilon_n$).
- LM test has the advantage of requiring only the null model be estimated!

LM test with SEM alternative

- When the alternative hypothesis is specified as a SEM, then the log-likelihood function takes the form given in equation (10). As a result, the LM test statistic given in (12) takes the form,

$$LM_{SEM} = \frac{n^2}{T_n}\left[\frac{e_n' W_n e_n}{e_n' e_n}\right]^2$$

  where, $T_n = tr(W_n' W_n + W_n^2)$ e_n is the vector of residuals under the null model $Y_n = X_n\beta + \epsilon_n$.
- LM_{SEM} is asymptotically a χ_1^2 under the null model.
- Can you notice that LM_{SEM} is the square of the Moran's I test? (Given in 9a theory)
- Thus both LM_{SEM} and Moran's I test on OLS residuals will lead to the same inferential conclusion.

LM test with SLM alternative

- When the alternative hypothesis is specified as a SLM, then the log-likelihood function takes the form given in equation (6). As a result, the LM test statistic given in (12) takes the form,

$$LM_{SLM} = \frac{n^2}{Q_n} \left[\frac{e_n' W_n Y_n}{e_n' e_n} \right]^2$$

where $Q_n = \frac{1}{\hat{\sigma}_\epsilon^2} \left(W_n X_n \hat{\beta} \right)' M_n \left(W_n X_n \hat{\beta} \right) + T_n$,
$M_n = I_n - X_n \left(X_n' X_n \right)^{-1} X_n$, $T_n = tr(W_n' W_n + W_n^2)$, $\hat{\beta}$ and $\hat{\sigma}_\epsilon^2$ are the ML estimates for β and σ_ϵ^2 respectively, and e_n is the vector of residuals under the null model $Y_n = X_n \beta + \epsilon_n$.

- LM_{SLM} is asymptotically a χ_1^2 under the null model.
- Both LM tests can be implemented in R using the function *spdep::lm.LMtests()*

Robust LM tests

- The two tests, LM_{SEM} and LM_{SLM} are not independent so that one can only apply the alternative SEM, assuming there is no spatial lag component and vise versa.
- Due to this reason a robust version of both tests are considered in the literature as,

$$
\begin{aligned}
RLM_{SEM} &= \frac{n^2}{T_n(1 - T_n Q_n)} \left[\frac{e_n' W_n e_n}{e_n' e_n} - T_n Q_n^{-1} \frac{e_n' W_n Y_n}{e_n' e_n} \right]^2 \\
RLM_{SLM} &= \frac{n^2}{Q_n - T_n} \left[\frac{e_n' W_n e_n}{e_n' e_n} - \frac{e_n' W_n Y_n}{e_n' e_n} \right]^2
\end{aligned}
$$

- RLM_{SEM} is the robust LM tests for error dependence in the possible presence of a missing lagged dependent variable.
- RLM_{SLM} is the robust LM tests for lagged dependent variable in the possible presence of a missing error dependent variable.

Likelihood ratio test

- The likelihood ratio test statistic is often expressed as a difference between the log-likelihoods

$$\lambda_{LR} = -2\left[\ell(\theta_0) - \ell(\theta_1)\right]$$

  where $\ell(\theta_0)$ is the value of the log-likelihood function under the null model and $\ell(\theta_1)$ is the value of the log-likelihood function under the alternative model, both evaluated using ML estimates of the respective models.
- Multiplying by -2 ensures that (by Wilks' theorem) λ_{LR} converges asymptotically to a χ^2 distribution under the null hypothesis.

LR test implementation

- The likelihood-ratio test requires that the models be nested – i.e. the more complex model can be transformed into the simpler model by imposing constraints on the former's parameters with degrees of freedom equal to the difference in dimensionality of θ_0 and θ_1.
- When faced with model selection, we can use the LR tests to compare,
  - SARAR(1,1) vs. SEM
  - SARAR(1,1) vs. SLM
  - SLM vs. SAR
- The LR cannot be applied to compare the SEM and the SLM model since one is not nested within the other. One can use the robust LM tests for this purpose.
- LR tests can be implemented in R using the function *spatialreg::LR.Sarlm()*

Summary

- Pure spatial autoregressive model
  - maximum likelihood estimation
- Spatial lagged dependent model
  - maximum likelihood estimation
  - two stage least squares
- Spatial error dependent model
  - maximum likelihood estimation
  - feasible generalised least squares estimator
- SARAR(1,1) model
  - maximum likelihood estimation
  - generalised spatial two stage least squares estimator
- Testing and model selection
  - LM test
  - LR test

References I

G. Arbia.
A primer for spatial econometrics.
Palgrave Macmillan