Lecture 4: Maximum Likelihood Estimator of SLM and SEM

Mauricio Sarrias

Universidad Católica del Norte

October 16, 2017

- (LK)-Chapter 3.
- (AR)-Chapter 10.
- Anselin, L., & Bera, A. K. (1998).Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics.Statistics Textbooks and Monographs,155, 237-290.
- Ertur, C., & Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence.Journal of applied econometrics,22(6), 1033-1062.

Log-likelihood function

The SLM model is given by the following structural model:

$$\mathbf{y} = \rho_0 \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon},$$
$$\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}_n, \sigma_0^2 \mathbf{I}_n), \tag{1}$$

where:

- $\mathbf{y}$ is a vector $n \times 1$ that collects the dependent variable for each spatial units;
- $\mathbf{W}$ is a $n \times n$ spatial weight matrix;
- $\mathbf{X}$ is a $n \times K$ matrix of independent variables;
- $\boldsymbol{\beta}_0$ is a known $K \times 1$ vector of parameters;
- ρ_0 measures the degree of spatial correlation;
- and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of error terms.

Note that we are making the explicit assumption that the error terms follows a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_0^2 \mathbf{I}_n$. That is, we are assuming that all spatial units have the **same error variance**.

Remarks:

- Since we are assuming the distribution of $\boldsymbol{\varepsilon}$, then we will be able to use the MLE.
- MLE approach: $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^{\top}, \widehat{\rho}, \widehat{\sigma}^2)^{\top}$ are chosen so as to maximize the probability of generating or obtaining the observed sample.
- So, we need the joint density function $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$: the probability of observing $\mathbf{y}$ given $\mathbf{X}$ under the unknown parameter vector $\boldsymbol{\theta}$

Log-likelihood function

Using the Transformation Theorem, we need the following transformation:

$$f(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) = f(\boldsymbol{\varepsilon}(\mathbf{y})|\mathbf{X};\boldsymbol{\theta}) \left| \frac{\partial \boldsymbol{\varepsilon}}{\partial \mathbf{y}} \right|.$$

where $\frac{\partial \boldsymbol{\varepsilon}}{\partial \mathbf{y}}$ is known as the Jacobian; and the unobseved errors is a functional form of the observed $\mathbf{y}$.

The idea is the following:

Since $\mathbf{y}$ and not the $\boldsymbol{\varepsilon}$ are observed quantities, the parameters must be estimated by maximizing $L(\mathbf{y})$, not $L(\boldsymbol{\varepsilon})$

Log-likelihood function

To obtain $f(\boldsymbol{\varepsilon}(\mathbf{y})|\mathbf{X}; \boldsymbol{\theta})$, recall that $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, thus:

$$f(\boldsymbol{\varepsilon}|\mathbf{X}) = (2\pi \cdot \sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}\right].$$

But the model can be written as:

$$\boldsymbol{\varepsilon} = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \quad \text{where} \quad \mathbf{A} = \mathbf{I}_n - \rho\mathbf{W}$$

So,

$$f(\boldsymbol{\varepsilon}(\mathbf{y})|\mathbf{X}; \boldsymbol{\theta}) = (2\pi \cdot \sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \qquad (2)$$

Log-likelihood function

To move from the the distribution of the error term to the distribution for the observable random variable $\mathbf{y}$ we need the Jacobian transformation:

$$\det\left(\frac{\partial\boldsymbol{\varepsilon}}{\partial\mathbf{y}}\right) = \det\left(\mathbf{J}\right) = \det(\mathbf{I}_n - \rho\mathbf{W}),$$

where $\mathbf{J} = \left(\frac{\partial\boldsymbol{\varepsilon}}{\partial\mathbf{y}}\right)$ is the Jacobian matrix, and $\det(\mathbf{I}_n - \rho\mathbf{W})$ is the determinant of a $n \times n$ matrix.

Remarks:

- In contrast to the time-series case, the spatial Jacobian is not the determinant of a triangular matrix, but of a full matrix.
- This may complicate its computation considerably.
- Recall that this Jacobian reduces to a scalar 1 in the standard regression model, since the partial derivative becomes $|\partial(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\partial\mathbf{y}| = |\mathbf{I}_n| = 1$.

Log-likelihood function

Putting the pieces togheter, the Likelihood function is:

$$L(\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{X}; \boldsymbol{\theta}) = (2\pi \cdot \sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \det(\mathbf{A})$$

And the log-likelihood function is:

$$\log L(\boldsymbol{\theta}) = \log|\mathbf{A}| - \frac{n\log(2\pi)}{2} - \frac{n\log(\sigma^2)}{2} - \frac{1}{2\sigma^2}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (3)$$

- To find the ML estimates for the SLM model, we need to maximize Equation (3) with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \rho)^\top$.
- So, we need to find the first order condition (FOC) of this optimization problem (score function)

The score function

Definition (Some useful results on matrix calculus)

Some important results are the followings:

$$\frac{\partial(\rho\mathbf{W})}{\partial\rho} = \mathbf{W} \tag{4}$$

$$\begin{aligned}\frac{\partial\mathbf{A}}{\partial\rho} &= \frac{\partial(\mathbf{I}_n - \rho\mathbf{W})}{\partial\rho} \\ &= \frac{\partial\mathbf{I}_n}{\partial\rho} - \frac{\partial\rho\mathbf{W}}{\rho} \\ &= -\mathbf{W}\end{aligned} \tag{5}$$

$$\frac{\partial\log|\mathbf{A}|}{\partial\rho} = \mathrm{tr}(\mathbf{A}^{-1}\partial\mathbf{A}/\partial\rho) = \mathrm{tr}\left[\mathbf{A}^{-1}(-\mathbf{W})\right] \tag{6}$$

Let $\boldsymbol{\varepsilon} = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, then:

$$\frac{\partial\boldsymbol{\varepsilon}}{\partial\rho} = \frac{\partial(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\partial\rho} = -\mathbf{W}\mathbf{y} \tag{7}$$

$$\frac{\partial\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}}{\partial\rho} = \boldsymbol{\varepsilon}^\top(\partial\boldsymbol{\varepsilon}/\partial\rho) + (\partial\boldsymbol{\varepsilon}^\top/\partial\rho)\boldsymbol{\varepsilon} = 2\boldsymbol{\varepsilon}^\top(\partial\boldsymbol{\varepsilon}/\partial\rho) = 2\boldsymbol{\varepsilon}^\top(-\mathbf{W})\mathbf{y} \tag{8}$$

$$\frac{\partial\mathbf{A}^{-1}}{\partial\rho} = -\mathbf{A}^{-1}(\partial\mathbf{A}/\partial\rho)\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1} \tag{9}$$

$$\frac{\partial\,\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{W}\right)}{\partial\rho} = \mathrm{tr}\left(\partial\mathbf{A}^{-1}\mathbf{W}/\partial\rho\right) \tag{10}$$

Taking the derivative of Equation (3) respect to $\boldsymbol{\beta}$, we obtain

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2}\left[-2\left((\mathbf{Ay})^\top \mathbf{X}\right)^\top + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}\right] = \frac{1}{\sigma^2}\mathbf{X}^\top(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}), \quad (11)$$

and with respect to σ^2, we obtain

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}\right)^\top \left(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}\right). \quad (12)$$

Taking the derivative of Equation (3) respect to ρ, we obtain

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \rho} &= \left(\frac{\partial}{\partial \rho}\right)\log|\mathbf{A}| - \frac{1}{2\sigma^2}\left(\frac{\partial}{\partial \rho}\right)\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \\
&= -\operatorname{tr}(\mathbf{A}^{-1}\mathbf{W}) + \frac{1}{2\sigma^2}2\boldsymbol{\varepsilon}^\top \mathbf{Wy} \quad \text{Using (6) and (8)} \\
&= -\operatorname{tr}(\mathbf{A}^{-1}\mathbf{W}) + \frac{1}{2\sigma^2}2\boldsymbol{\varepsilon}^\top \mathbf{Wy} \\
&= -\operatorname{tr}(\mathbf{A}^{-1}\mathbf{W}) + \frac{1}{\sigma^2}\boldsymbol{\varepsilon}^\top \mathbf{Wy}
\end{aligned} \quad (13)$$

Thus the complete gradient is:

$$\nabla_{\boldsymbol{\theta}} = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma^2} \\ \frac{\partial \log L(\boldsymbol{\theta})}{\partial \rho} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \mathbf{X}^\top \boldsymbol{\varepsilon} \\ \frac{1}{2\sigma^4} (\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} - n\sigma^2) \\ -\operatorname{tr}(\mathbf{A}^{-1} \mathbf{W}) + \frac{1}{\sigma^2} \boldsymbol{\varepsilon}^\top \mathbf{W} \mathbf{y} \end{pmatrix}$$

Note that if we replace $\mathbf{y} = \mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{A}^{-1} \boldsymbol{\varepsilon}$ in Equation (13), we get:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \rho} = \frac{1}{\sigma^2} (\mathbf{C} \mathbf{X} \boldsymbol{\beta})^\top \boldsymbol{\varepsilon} + \frac{1}{\sigma^2} (\boldsymbol{\varepsilon}^\top \mathbf{C} \boldsymbol{\varepsilon} - \sigma^2 \operatorname{tr}(\mathbf{C})),$$

where:

$$\mathbf{C} = \mathbf{W} \mathbf{A}^{-1}.$$

- The $\widehat{\boldsymbol{\theta}}_{ML}$ is found by solving $\nabla_{\boldsymbol{\theta}} = 0$...
- ... however, solving for ρ is very tricky (highly non-linear).
- So, how do we proceed?

Note the following. Setting both (11) and (12) to 0 and solving, we obtain:

$$\widehat{\boldsymbol{\beta}}_{ML}(\rho) = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{A}(\rho)\mathbf{y} \tag{14}$$

$$\widehat{\sigma}^2_{ML}(\rho) = \frac{\left(\mathbf{A}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{ML}\right)^\top \left(\mathbf{A}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{ML}\right)}{n} \tag{15}$$

Note that conditional on ρ (assuming we know ρ), these estimates are simply OLS applied to the **spatial filtered** dependent variable $\mathbf{Ay}$ and the exploratory variables $\mathbf{X}$.

If we knew ρ, we could easily get $\widehat{\boldsymbol{\beta}}_{ML}(\rho)$ and $\widehat{\sigma}^2_{ML}(\rho)$

How can we estimate ρ?.

After some manipulation Equation (14) can be re-written as:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{ML}(\rho) &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y} - \rho \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y} \\
&= \widehat{\boldsymbol{\beta}}_0 - \rho \widehat{\boldsymbol{\beta}}_L
\end{aligned} \tag{16}
$$

Note that the first term in (16) is just the OLS regression of $\mathbf{y}$ on $\mathbf{X}$, whereas the second term is just ρ times the OLS regression of $\mathbf{W}\mathbf{y}$ on $\mathbf{X}$. Next, define the following:

$$
\mathbf{e}_0 = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_0 \text{ and } \mathbf{e}_L = \mathbf{W}\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}_L}. \tag{17}
$$

Then, plugging (16) into (15)

$$\sigma^2_{ML}\left[\boldsymbol{\beta}_{ML}(\rho), \rho\right] \quad = \quad \frac{(\mathbf{e}_0 - \rho\mathbf{e}_L)^{\top}(\mathbf{e}_0 - \rho\mathbf{e}_L)}{n}. \qquad (18)$$

Note that both (16) and (18) rely only on observables, except for ρ, and so are readily calculable given some estimate of ρ. Therefore, plugging (16) and (18) back into the likelihood (3) we obtain the **concentrated log-likelihood function**:

$$\log L(\rho)_c = -\frac{n}{2} - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left[\frac{(\mathbf{e}_0 - \rho\mathbf{e}_L)^{\top}(\mathbf{e}_0 - \rho\mathbf{e}_L)}{n}\right] + \log|\mathbf{I}_n - \rho\mathbf{W}|, \qquad (19)$$

which is a **nonlinear** function of a single parameter ρ. A ML estimate for ρ, $\widehat{\rho}_{ML}$, is obtained from a numerical optimization of the concentrated log-likelihood function (19). Once we obtain $\widehat{\rho}$, we can easily obtain $\widehat{\boldsymbol{\beta}}$.

The algorithm to perform the ML estimation of the SLM is the following:

1. Perform the two auxiliary regression of $\mathbf{y}$ and $\mathbf{Wy}$ on $\mathbf{X}$ to obtain $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\beta}}_L$ as in Equation (16).

2. Use $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\beta}}_L$ to compute the residuals in Equation (17).

3. Maximize the concentrated likelihood given in Equation (19) by numerical optimization to obtain an estimate of ρ.

4. Use the estimate of $\widehat{\rho}$ to plug it back in to the expression for $\boldsymbol{\beta}$ (Equation 14) and σ^2 (Equation 15).

Ord's Jacobian

- $|\mathbf{I}_n - \rho\mathbf{W}|$ is computationally this is burdensome
  - determining $\widehat{\rho}$ rests on the evaluation of $|\mathbf{I}_n - \rho\mathbf{W}|$ in each iteration.
- However, Ord (1975) note that

$$|\omega\mathbf{I}_n - \mathbf{W}| = \prod_{i=1}^{n}(\omega - \omega_i).$$

Therefore:

$$|\mathbf{I}_n - \rho\mathbf{W}| = \prod_{i=1}^{n}(1 - \rho\omega_i)$$

and the log-determinant term follows as

$$\log|\mathbf{I}_n - \rho\mathbf{W}| = \sum_{i=1}^{n}\log(1 - \rho\omega_i) \tag{20}$$

- The advantage of this approach is that the eigenvalues only need to be computed once.

Ord's Jacobian

- Domain of ρ: We need that $1 - \rho\omega_i \neq 0$, which occurs only if $1/\omega_{min} < \rho < 1/\omega_{max}$. For row-standardized matrix, the largest eigenvalues is 1.

- Finally, the new concentrated log-likelihood function is:

$$\log L(\rho)_c = -\frac{n}{2} - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left[\frac{(\mathbf{e}_0 - \rho\mathbf{e}_L)^\top (\mathbf{e}_0 - \rho\mathbf{e}_L)}{n}\right] +$$

$$+ \sum_{i=1}^{n}\log(1 - \rho\omega_i)$$

$$= \text{const} - \log\left[\frac{\mathbf{e}_0'\mathbf{e}_0 - 2\rho\mathbf{e}_L'\mathbf{e}_0 + \rho^2\mathbf{e}_L'\mathbf{e}_L}{n}\right] + \frac{2}{n}\sum_{i=1}^{n}\log(1 - \rho\omega_i)$$

$$(21)$$

Recall that:

$$\widehat{\boldsymbol{\theta}} \overset{a}{\sim} \mathrm{N}\left[\boldsymbol{\theta}_0, -\mathbb{E}\left[\mathbf{H}(\mathbf{w}_i; \boldsymbol{\theta}_0)\right]^{-1}\right]$$

where $\mathbf{H}(\mathbf{w}_i; \boldsymbol{\theta}_0)$ is the Hessian Matrix.

The Hessian is a $(K+2) \times (K+2)$ matrix of second derivatives given by :

$$\mathbf{H}(\boldsymbol{\beta}, \sigma^2, \rho) = \begin{pmatrix} \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \boldsymbol{\beta} \partial \sigma^2} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \boldsymbol{\beta} \partial \rho} \\ \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \sigma^2 \partial \boldsymbol{\beta}^\top} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial (\sigma^2)^2} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \sigma^2 \partial \rho} \\ \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \rho \partial \boldsymbol{\beta}^\top} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \rho \partial \sigma^2} & \frac{\ell(\boldsymbol{\beta}, \sigma^2, \rho)}{\partial \rho^2} \end{pmatrix}$$

The Hessian Matrix for the SLM is:

$$\mathbf{H}(\boldsymbol{\beta}, \sigma^2, \rho) = \begin{pmatrix} -\frac{1}{\sigma^2}(\mathbf{X}^\top \mathbf{X}) & -\frac{1}{(\sigma^2)^2}\mathbf{X}^\top \boldsymbol{\varepsilon} & -\frac{1}{\sigma^2}\mathbf{X}^\top \mathbf{W}\mathbf{y} \\ . & \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} & -\frac{\boldsymbol{\varepsilon}^\top \mathbf{W}\mathbf{y}}{\sigma^4} \\ . & . & -\operatorname{tr}\left[(\mathbf{W}\mathbf{A}^{-1})^2\right] - \frac{1}{\sigma^2}(\mathbf{y}^\top \mathbf{W}^\top \mathbf{W}\mathbf{y}) \end{pmatrix}$$

which is symmetric.

Proof on blackboard.

Thus, the expected value of the Hessian is:

$$-\mathbb{E}\left[\mathbf{H}(\boldsymbol{\beta}, \sigma^2, \rho)\right] = \begin{pmatrix} \frac{1}{\sigma^2}(\mathbf{X}^\top\mathbf{X}) & \mathbf{0}^\top & \frac{1}{\sigma^2}\mathbf{X}^\top(\mathbf{CX}\boldsymbol{\beta}) \\ & \frac{n}{2\sigma^4} & \frac{1}{\sigma^2}\operatorname{tr}(\mathbf{C}) \\ & & \operatorname{tr}(\mathbf{C}^s\mathbf{C}) + \frac{1}{\sigma^2}(\mathbf{CX}\boldsymbol{\beta})^\top(\mathbf{CX}\boldsymbol{\beta}) \end{pmatrix} \quad (22)$$

Proof on blackboard.

The asymptotic variance matrix follows as the inverse of the information matrix:

$$\text{Var}(\boldsymbol{\beta}, \sigma^2, \rho) = \begin{pmatrix} \frac{1}{\sigma^2}(\mathbf{X}^\top \mathbf{X}) & \mathbf{0}^\top & \frac{1}{\sigma^2}\mathbf{X}^\top(\mathbf{C}\mathbf{X}\boldsymbol{\beta}) \\ & \frac{n}{2\sigma^4} & \frac{1}{\sigma^2}\text{tr}(\mathbf{C}) \\ & & \text{tr}(\mathbf{C}^s\mathbf{C}) + \frac{1}{\sigma^2}(\mathbf{C}\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{C}\mathbf{X}\boldsymbol{\beta}) \end{pmatrix}^{-1} \quad (23)$$

An important feature is that the covariance between $\boldsymbol{\beta}$ and the error variance is zero, as in the standard regression model, this is not the case for ρ and the error variance. This lack of block diagonality in the information matrix for the spatial lag model will lead to some interesting results on the structure of specification test.

However, we can use the eigenvalues approximation. Recall that

$$\left(\frac{\partial}{\partial \rho}\right) \log |\mathbf{A}| = -\sum_{i=1}^{n} \frac{\omega_i}{1 - \rho \omega_i}, \tag{24}$$

so that,

$$\text{Var}(\boldsymbol{\beta}, \sigma^2, \rho) = \begin{pmatrix} \frac{1}{\sigma^2}(\mathbf{X}^\top \mathbf{X}) & \mathbf{0}' & \frac{1}{\sigma^2}\mathbf{X}^\top(\mathbf{CX}\boldsymbol{\beta}) \\ . & \frac{N}{2\sigma^4} & \frac{1}{\sigma^2}\text{tr}(\mathbf{C}) \\ . & . & \alpha + \text{tr}(\mathbf{CC}) + \frac{1}{\sigma^2}(\mathbf{CX}\boldsymbol{\beta})^\top(\mathbf{CX}\boldsymbol{\beta}) \end{pmatrix}^{-1}$$

where $\alpha = \sum_{i=1}^{n} \left(\frac{\omega_i^2}{(1 - \rho \omega_i)^2}\right)$. Note that while the covariance between $\boldsymbol{\beta}$ and the error variance is zero, as in the standard regression model, this is not the case for ρ and the error variance.

The SEM model is given by

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u} \\
\mathbf{u} &= \lambda_0 \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)
\end{aligned}
\tag{25}
$$

where

- λ_0 is the spatial autoregressive coefficient for the error lag $\mathbf{W}\mathbf{u}$,
- $\mathbf{W}$ is the spatial weight matrix,
- $\boldsymbol{\varepsilon}$ is an error $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$

Log-likelihood function

The reduced form is given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{I}_n - \lambda_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}.$$

Since $\mathbf{u} = (\mathbf{I}_n - \lambda_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}$:

- $\mathbb{E}(\mathbf{u}) = \mathbf{0}$,
- The variance-covariance matrix of $\mathbf{u}$: is

$$\mathrm{Var}(\mathbf{u}) = \mathbb{E}(\mathbf{u}\mathbf{u}^\top) = \sigma_\epsilon^2 (\mathbf{I}_n - \lambda_0 \mathbf{W})^{-1} (\mathbf{I} - \lambda_0 \mathbf{W}^\top)^{-1} = \sigma_\epsilon^2 \boldsymbol{\Omega}_u^{-1}, \quad (26)$$

  where $\boldsymbol{\Omega}_u = (\mathbf{I}_n - \lambda_0 \mathbf{W})(\mathbf{I}_n - \lambda_0 \mathbf{W})^\top$.
- So, we have heteroskedasticity in $\mathbf{u}$.
- Finally, $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{\Omega}_u^{-1})$.

Log-likelihood function

The reduced form is :

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \lambda\mathbf{W})\,\mathbf{y} - (\mathbf{I}_n - \lambda\mathbf{W})\,\mathbf{X}\boldsymbol{\beta} = \mathbf{B}\mathbf{y} - \mathbf{B}\mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{B} = (\mathbf{I}_n - \lambda\mathbf{W})$. The new error term indicates that $\boldsymbol{\varepsilon}(\mathbf{y})$. Using the **Transformation Theorem** we are able to find the joint conditional function:

$$f(y_1, ..., y_n | \mathbf{X}; \boldsymbol{\theta}) = f(\boldsymbol{\varepsilon}(\mathbf{y}) | \mathbf{X}; \boldsymbol{\theta}) \cdot \mathbf{J}.$$

Again, the Jacobian term is not equal to one, but instead is

$$\mathbf{J} = \left| \frac{\partial \boldsymbol{\varepsilon}}{\partial \mathbf{y}} \right| = |\mathbf{B}|\,.$$

Thus, the joint density function of $\boldsymbol{\varepsilon}$—which is a function of $\mathbf{y}$— equals

$$f(\boldsymbol{\varepsilon}(\mathbf{y}) | \mathbf{X}; \boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{[(\mathbf{I}_n - \lambda\mathbf{W})\,(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]^{\top}\,[(\mathbf{I}_n - \lambda\mathbf{W})\,(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]}{2\sigma^2} \right],$$

The joint density function of $\mathbf{y}$, $f(y_1, ..., y_n | \mathbf{X}; \boldsymbol{\theta})$ equals

$$f(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{B}^\top \mathbf{B}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}\right] \cdot |\mathbf{B}|$$

Finally, the log-likelihood can be expressed as

$$\log L(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} + \log|\mathbf{I}_n - \lambda\mathbf{W}| \tag{27}$$

where

$$\boldsymbol{\Omega}(\lambda) = \mathbf{B}^\top \mathbf{B} = (\mathbf{I}_n - \lambda\mathbf{W})^\top (\mathbf{I}_n - \lambda\mathbf{W})$$

Again, we run into complications over the log of the determinant $|\mathbf{I}_n - \lambda\mathbf{W}|$, which is an nth-order polynomial that is cumbersome to evaluate.

ML Estimates

Taking the derivative respect to $\boldsymbol{\beta}$ yields

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{ML}(\lambda) &= \left[\mathbf{X}^{\top}\boldsymbol{\Omega}(\lambda)\mathbf{X}\right]^{-1}\mathbf{X}^{\top}\boldsymbol{\Omega}(\lambda)\mathbf{y} \\
&= \left[(\mathbf{B}\mathbf{X})^{\top}(\mathbf{B}\mathbf{X})\right]^{-1}(\mathbf{B}\mathbf{X})^{\top}\mathbf{B}\mathbf{y} \\
&= \left[\mathbf{X}(\lambda)^{\top}\mathbf{X}(\lambda)\right]^{-1}\mathbf{X}(\lambda)^{\top}\mathbf{y}(\lambda)
\end{aligned}
\tag{28}
$$

where

$$
\begin{aligned}
\mathbf{X}(\lambda) &= \mathbf{B}\mathbf{X} = (\mathbf{I} - \lambda\mathbf{W})\mathbf{X} = (\mathbf{X} - \lambda\mathbf{W}\mathbf{X}) \\
\mathbf{y}(\lambda) &= (\mathbf{y} - \lambda\mathbf{W}\mathbf{y}).
\end{aligned}
\tag{29}
$$

If λ is known, this estimator is equal to the **GLS** estimator—$\widehat{\beta}_{ML} = \widehat{\beta}_{GLS}$—and it can be thought as the OLS estimator resulting from a regression of $\mathbf{y}(\lambda)$ on $\mathbf{X}(\lambda)$. In other words, for a known value of the spatial autoregressive coefficient, λ, this is equivalent to OLS on the transformed variables.

Remark: In the literature, the transformations:

$$\mathbf{X}(\lambda) = (\mathbf{X} - \lambda \mathbf{W}\mathbf{X})$$
$$\mathbf{y}(\lambda) = (\mathbf{y} - \lambda \mathbf{W}\mathbf{y})$$

are known as the *Cochrane-Orcutt transformation* or *spatially filtered variables*.

The ML estimator for the error variance is given by:

$$\widehat{\sigma}_{ML}^2(\lambda) = \frac{1}{n}\left(\widehat{\varepsilon}^\top \mathbf{B}^\top \mathbf{B} \varepsilon\right) = \frac{1}{n}\widehat{\varepsilon}^\top(\lambda)\widehat{\varepsilon}(\lambda) \tag{30}$$

where $\widehat{\varepsilon} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{ML}$ and $\widehat{\varepsilon}(\lambda) = \mathbf{B}(\lambda)(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{ML}) = \mathbf{B}(\lambda)\mathbf{y} - \mathbf{B}(\lambda)\mathbf{X}\widehat{\boldsymbol{\beta}}_{ML}$.

Concentrated Log-Likelihood

A concentrated log-likelihood can then be obtained as:

$$\log L(\lambda)_c = \text{const} + \frac{n}{2} \log \left[\frac{1}{n} \widehat{\varepsilon}^{\top} \mathbf{B}^{\top} \mathbf{B} \widehat{\varepsilon} \right] + \log |\mathbf{B}| \tag{31}$$

- The residual vector of the concentrated likelihood is also, indirectly, a function of the spatial autoregressive parameter.
- A one-time optimization will in general not be sufficient to obtain maximum likelihood estimates for all the parameters. Therefore an interactive procedure will be needed.
- Alternate back and forth between the estimation of the spatial autoregressive coefficient conditional upon residuals (for a value of $\boldsymbol{\beta}$), and a estimation of the parameter vector

The procedure can be summarize in the following steps:

1. Carry out an OLS of $\mathbf{BX}$ on $\mathbf{By}$; get $\widehat{\boldsymbol{\beta}}_{OLS}$
2. Compute initial set of residuals $\widehat{\epsilon}_{OLS} = \mathbf{By} - \mathbf{BX}\widehat{\boldsymbol{\beta}}_{OLS}$
3. Given $\widehat{\epsilon}_{OLS}$, find $\widehat{\lambda}$ that maximizes the concentrated likelihood.
4. If the convergence criterion is met, proceed, otherwise repeat steps 1, 2 and 3.
5. Given $\widehat{\lambda}$, estimate $\widehat{\boldsymbol{\beta}}(\lambda)$ by GLS and obtain a new vector of residuals, $\widehat{\boldsymbol{\varepsilon}}(\lambda)$
6. Given $\widehat{\boldsymbol{\varepsilon}}(\lambda)$ and $\widehat{\lambda}$, estimate $\widehat{\sigma}(\lambda)$.

Variance-Covariance Matrix

Finally, the asymptotic variance-covariance matrix is:

$$\text{AsyVar}(\boldsymbol{\beta}, \sigma^2, \lambda) = \begin{pmatrix} \frac{\mathbf{X}(\lambda)^\top \mathbf{X}(\lambda)}{\sigma^2} & 0 & 0 \\ k \times k & & \\ 0 & \frac{n}{2\sigma^4} & \frac{\text{tr}(\mathbf{W}_B)}{\sigma^2} \\ 0 & \frac{\text{tr}(\mathbf{W}_B)}{\sigma^2} & \text{tr}(\mathbf{W}_B)^2 + \text{tr}(\mathbf{W}_B^\top \mathbf{W}_B) \end{pmatrix}^{-1} \quad (32)$$

where $\mathbf{W}_B = \mathbf{W}(\mathbf{I} - \lambda \mathbf{W})^{-1}$.

The SEM model can be expanded and rewritten as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$
$$(\mathbf{I}_n - \lambda\mathbf{W})\mathbf{y} = (\mathbf{I}_n - \lambda\mathbf{W})\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\mathbf{y} - \lambda\mathbf{W}\mathbf{y} = (\mathbf{X} - \lambda\mathbf{W}\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{33}$$
$$\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\mathbf{X}(\lambda\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

Under some nonlinear restrictions we can see that (33) is equivalent to the SDM.

The unconstrained form of the model—or the SDM model—is

$$\mathbf{y} = \gamma_1 \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\gamma}_2 + \mathbf{W} \mathbf{X} \boldsymbol{\gamma}_3 + \boldsymbol{\varepsilon}, \tag{34}$$

where γ_1 is a scalar, $\boldsymbol{\gamma}_2$ is a $K \times 1$ vector (where K is the number of explanatory variables, including the constant), and $\boldsymbol{\gamma}_3$ is also a $K \times 1$ vector. Note that if $\boldsymbol{\gamma}_3 = -\gamma_1 \boldsymbol{\gamma}_2$, then the SDM is equivalent to the SEM model. Note also that $\boldsymbol{\gamma}_3 = -\gamma_1 \boldsymbol{\gamma}_2$ is a vector of $K \times 1$ nonlinear constraints of the form:

$$\gamma_{3,k} = -\gamma_1 \gamma_{2,k}, \quad \text{for} \quad k = 1, ..., K. \tag{35}$$

These conditions are usually formulated as a null hypothesis, designated as the **Common Factor Hypothesis**, and written as:

$$H_0 : \boldsymbol{\gamma}_3 + \gamma_1 \boldsymbol{\gamma}_2 = \mathbf{0}. \tag{36}$$

If the constraints hold it follows that the SDM is equivalent to the SEM model.

Estimation in R

In this example we use Anselin (1988)'s dataset. This sample corresponds to a cross-sectional dataset of 49 Columbus, Ohio neighborhoods, which is used to explain the crime rate as a function of household income and housing values. In particular, the variables are the following:

- `CRIME`: residential burglaries and vehicle thefts per thousand household in the neighborhood.
- `HOVAL`: housing value in US$1,000.
- `INC`: household income in US$1,000.

We start our analysis by loading the required packages into **R** workspace.

```
# Load packages
library("spdep")
library("memisc")                 # Package for tables
library("maptools")
library("RColorBrewer")
library("classInt")
source("getSummary.sarlm.R") # Function for spdep models
```

Estimation in R

Dataset is currently part of the **spdep** package. We load the data using the following commands.

```
# Load data
columbus    <- readShapePoly(system.file("etc/shapes/columbus.shp",
                                    package = "spdep")[1])

## Warning:   use rgdal::readOGR or sf::st_read

col.gal.nb <- read.gal(system.file("etc/weights/columbus.gal",
                                 package = "spdep")[1])
```

Estimation in R

To get some insights about the spatial distribution of `CRIME` we use the following quantile clorophet graph:

```
# Spatial distribution of crime
spplot(columbus, "CRIME",
       at = quantile(columbus$CRIME, p = c(0, .25, .5, .75, 1),
                     na.rm = TRUE),
       col.regions = brewer.pal(5, "Blues"),
       main = "")
```

Estimation in R

Figure: Spatial Distribution of Crime in Columbus, Ohio Neighborhoods
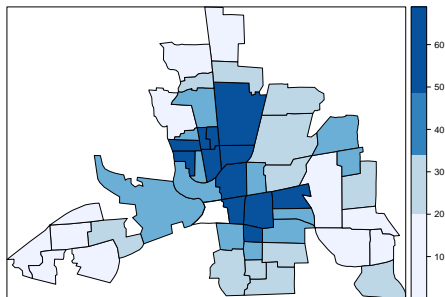
Notes: This graph shows the spatial distribution of crime on the 49 Columbus, Ohio neighborhoods. Darker color indicates greater rate of crime.

Estimation in R

Now, we compute the Moran's I for the `CRIME` variable.

```
# Moran's I test
set.seed(1234)
listw <- nb2listw(col.gal.nb, style = "W")
moran.mc(columbus$CRIME, listw = listw,
         nsim = 99, alternative = 'greater')

##
##  Monte-Carlo simulation of Moran I
##
## data:  columbus$CRIME
## weights: listw
## number of simulations + 1: 100
##
## statistic = 0.48577, observed rank = 100, p-value = 0.01
## alternative hypothesis: greater
```

Estimation in R

The functions used for each model are the following:

- OLS: `lm` function.
- SLX: `lm` function, where **WX** is constructed using the function `lag.listw` from **spdep** package. This model can also be estimated using the function `lmSLX` from **spdep** package as shown below.
- SLM: `lagsarlm` from **spdep** package.
- SDM: `lagsarlm` from **spdep** package, using the argument `type = "mixed"`. Note that `type = "Durbin"` may be used instead of `type = "mixed"`.
- SEM: `errorsarlm` from **spdep** package. Note that the Spatial Durbin Error Model (SDEM)—not shown here— can be estimated by using `type = "emixed"`.
- SAC: `sacsarlm` from **spdep** package.

All models are estimated using ML procedure outline in the previous section. In order to compute the determinant of the Jacobian we use the Ord (1975)'s procedure by explicitly using the argument `method = "eigen"` in each spatial model. That is, the Jacobian is computed as in (20).

```
# Models
columbus$lag.INC    <- lag.listw(listw,
                        columbus$INC)   # Create spatial lag of INC
columbus$lag.HOVAL <- lag.listw(listw,
                        columbus$HOVAL) # Create spatial lag of HOVAL
ols <- lm(CRIME ~ INC + HOVAL,
          data = columbus)
slx <- lm(CRIME ~ INC + HOVAL + lag.INC + lag.HOVAL,
          data = columbus)
slm <- lagsarlm(CRIME ~ INC + HOVAL,
                data = columbus,
                listw,
                method = "eigen")
sdm <- lagsarlm(CRIME ~ INC + HOVAL,
                data = columbus,
                listw,
                method = "eigen",
                type = "mixed")
sem <- errorsarlm(CRIME ~ INC + HOVAL,
                data = columbus,
                listw,
                method = "eigen")
sac <- sacsarlm(CRIME ~ INC + HOVAL,
                data = columbus,
                listw,
                method = "eigen")
```

Estimation in R

Table: Spatial Models for Crime in Columbus, Ohio Neighborhoods.

	OLS	SLX	SLM	SDM	SEM	SAC
Constant	68.619***	74.029***	46.851***	45.593***	61.054***	49.051***
	(4.735)	(6.722)	(7.315)	(13.129)	(5.315)	(10.055)
INC	−1.597***	−1.108**	−1.074***	−0.939**	−0.995**	−1.069**
	(0.334)	(0.375)	(0.311)	(0.338)	(0.337)	(0.333)
HOVAL	−0.274*	−0.295**	−0.270**	−0.300***	−0.308***	−0.283**
	(0.103)	(0.101)	(0.090)	(0.091)	(0.093)	(0.092)
W.INC		−1.383*		−0.618		
		(0.559)		(0.577)		
W.HOVAL		0.226		0.267		
		(0.203)		(0.184)		
ρ			0.404***	0.383*		0.353
			(0.121)	(0.162)		(0.197)
λ					0.521***	0.132
					(0.141)	(0.299)
AIC	382.8	380.2	376.3	378.0	378.3	378.1
N	49	49	49	49	49	49

Question:

We begin our analysis with the following question: what would happen to crime in all regions if income rose from 13.906 to 14.906 in the 30th region ($\Delta\texttt{INC} = 1$)?

We can use the reduced-form predictor given by the following formula:

$$\widehat{\mathbf{y}} = \mathbb{E}(\mathbf{y}|\mathbf{X}, \mathbf{W}) = (\mathbf{I}_n - \widehat{\rho}\mathbf{W})^{-1}\mathbf{X}\widehat{\boldsymbol{\beta}},$$

and estimate the predicted values pre- and post- the change in the income variable.

Estimation in R

```r
# The pre-predicted values
rho <- slm$rho                              # Estimated rho from SLM model
beta_hat <- coef(slm)[-1]                   # Estimated parameters
A   <- invIrW(listw, rho = rho)             # (I - rho*W)^{-1}
X   <- cbind(1, columbus$INC, columbus$HOVAL) # Matrix of observed variables
y_hat_pre <- A %*% crossprod(t(X), beta_hat) # y hat
```

Next we increase INC by 1 in spatial unit 30, and calculate the reduced-form predictions, $\widehat{\mathbf{y}}^2$.

```r
# The post-predicted values
col_new <- columbus # copy the data frame

# change the income value
col_new@data[col_new@data$POLYID == 30, "INC"] <- 14.906

# The pre-predicted values
X_d <- cbind(1, col_new$INC, col_new$HOVAL)
y_hat_post <- A %*% crossprod(t(X_d), beta_hat)
```

Estimation in R

Finally, we compute the difference between pre- and post-predictions: $\widehat{\mathbf{y}}^2 - \widehat{\mathbf{y}}^1$:

```r
# The difference
delta_y          <- y_hat_post - y_hat_pre
col_new$delta_y <- delta_y

# Show the effects
summary(delta_y)

##          V1
##  Min.   :-1.1141241
##  1st Qu.:-0.0074114
##  Median :-0.0012172
##  Mean   :-0.0336341
##  3rd Qu.:-0.0002604
##  Max.   :-0.0000081

sum(delta_y)

## [1] -1.648071
```

The predicted effect of the change would be a decrease of 1.65 in the crime rate, considering both direct and indirect effects. That is, increasing the income in US$1,000 in region 30th might generate effects that will transmit through the whole system of region resulting in a new equilibrium where the the total crime will reduce in 1.7 crimes per thousand households.

Sometimes we would like to plot these effects. Suppose we wanted to show those regions that had low and high impact due to the increase in INC. Let's define "high impacted regions" those regions whose crime rate decrease more than 0.05. The following code produces the Figure.
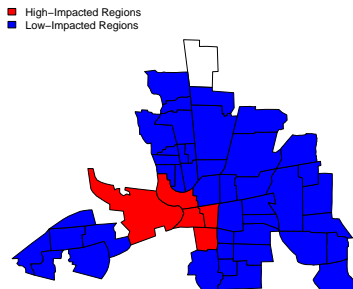
```r
# Breaks
breaks <- c(min(col_new$delta_y), -0.05, max(col_new$delta_y))
labels <- c("High-Impacted Regions", "Low-Impacted Regions")
np <- findInterval(col_new$delta_y, breaks)
colors <- c("red", "blue")

# Draw Map
plot(col_new, col = colors[np])
legend("topleft", legend = labels, fill = colors, bty = "n")
points(38.29, 30.35, pch = 19, col = "black", cex = 0.5)
```

**Figure:** Effects of a Change in Region 30: Categorization

Notes: This graph shows those regions that had low and high impact due to increase in INC in 30th. Red-colored regions are those regions with a decrease of crime rate larger than 0.05, whereas blue-colored regions are those regions with lower decrease of crime rate.

Now we map the magnitude of the changes caused by altering `INC` in region 30. The code is the following:.

```
pal5    <- brewer.pal(6, "Spectral")
cats5   <- classIntervals(col_new$delta_y, n = 5, style = "jenks")
colors5 <- findColours(cats5, pal5)
plot(col_new, col = colors5)
legend("topleft", legend = round(cats5$brks, 2), fill = pal5, bty =
```

Spillovers

Figure: Effects of a Change in Region 30: Magnitude
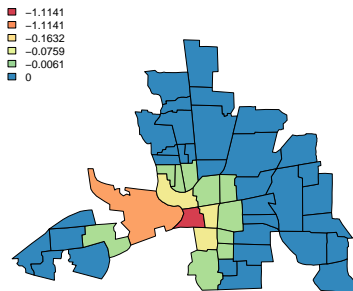
Notes: This graph shows the spatial distribution of the changes caused by altering INC in region 30.

- `impacts`: This function returns the direct, indirect and total impacts for the variables in the model.
- The spatial lag impact measures are computed using the reduced form:

$$\mathbf{y} = \sum_{r=1}^{K} \mathbf{A}(\mathbf{W})^{-1}(\mathbf{I}_n \beta_r) + \mathbf{A}(\mathbf{W})^{-1}\boldsymbol{\varepsilon} \tag{37}$$

$$\mathbf{A}(\mathbf{W})^{-1} = \mathbf{I}_n + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + ...$$

The exact $\mathbf{A}(\mathbf{W})^{-1}$ is computed when `listw` is given.

- When the traces are created by powering sparse matrices the approximation $\mathbf{I}_n + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + ...$ is used.
- When the traces are created by powering sparse matrices, the exact and the trace methods should give very similar results, unless the number of powers used is very small, or the spatial coefficien is close to its bounds.

```
impacts(slm, listw = listw)

## Impact measures (lag, exact):
##            Direct   Indirect      Total
## INC    -1.1225156 -0.6783818 -1.8008973
## HOVAL  -0.2823163 -0.1706152 -0.4529315
```

- An increase of US$1,000 in income leads to a decrease of 1.8 crimes per thousand households.
- The direct effect of the income variable in the SLM model amounts to -1.123, while the coefficient estimate of this variable is -1.074. This implies that the feedback effect is -1.123 - (-1.074) = -0.049. This feedback effect corresponds to 4.5% of the coefficient estimate.

Impacts

Let's corroborate these results by computing the impacts using matrix operations:

```
## Construct S_r(W) = A(W)^-1 (I * beta_r + W * theta_r)
Ibeta <- diag(length(listw$neighbours)) *  coef(slm)["INC"]
S <- A %*% Ibeta

ADI <- sum(diag(S)) / nrow(A)
ADI

## [1] -1.122516

n <- length(listw$neighbours)
Total <- crossprod(rep(1, n), S) %*% rep(1, n) / n
Total

##           [,1]
## [1,] -1.800897

Indirect <- Total - ADI
Indirect

##            [,1]
## [1,] -0.6783818
```

- We can also obtain the p-values of the impacts by using the argument `R`.
- This argument indicates the number of simulations use to create distributions for the impact meassures, provided that the fitted model object contains a coefficient covariance matrix.

Impacts

```
im_obj <- impacts(slm, listw = listw, R = 200)
summary(im_obj, zstats = TRUE, short = TRUE)

## Impact measures (lag, exact):
##            Direct    Indirect       Total
## INC   -1.1225156  -0.6783818  -1.8008973
## HOVAL -0.2823163  -0.1706152  -0.4529315
## ========================================================
## Simulation results (asymptotic variance matrix):
## ========================================================
## Simulated z-values:
##            Direct   Indirect       Total
## INC   -3.663948  -2.066029  -3.485143
## HOVAL -3.028626  -1.625834  -2.499986
##
## Simulated p-values:
##        Direct      Indirect Total
## INC    0.00024836  0.038826 0.00049187
## HOVAL  0.00245668  0.103985 0.01241982
```

Impacts

Now we follow the example that converts the spatial weight matrix into "sparse" matrix, and power it up using the `trW` function.

```
W <- as(nb2listw(col.gal.nb, style = "W"), "CsparseMatrix")
trMC <- trW(W, type = "MC")
im <- impacts(slm, tr = trMC, R = 100)
summary(im, zstats =  TRUE, short = TRUE)


## Impact measures (lag, trace):
##           Direct    Indirect      Total
## INC   -1.1198013 -0.6810960 -1.8008973
## HOVAL -0.2816336 -0.1712978 -0.4529315
## =========================================================
## Simulation results (asymptotic variance matrix):
## =========================================================
## Simulated z-values:
##           Direct   Indirect      Total
## INC    -3.380162 -1.763959 -3.067776
## HOVAL  -3.454579 -1.762853 -2.911180
##
## Simulated p-values:
##        Direct     Indirect Total
## INC    0.00072443 0.077739 0.0021566
## HOVAL  0.00055115 0.077925 0.0036007
```

We can also observe the cummulative impacts using the argument `Q`. When `Q` and `tr` are given in the `impacts` function the output will present the impact components for each step in the traces of powers of the weight matrix up to and including the Qth power.

Impacts

```
im2    <- impacts(slm, tr = trMC, R = 100, Q = 5)
sums2 <- summary(im2, zstats = TRUE, reportQ = TRUE, short =  TRUE)
sums2

## Impact measures (lag, trace):
##             Direct    Indirect      Total
## INC    -1.1198013 -0.6810960 -1.8008973
## HOVAL -0.2816336 -0.1712978 -0.4529315
## =================================
## Impact components
## $direct
##               INC          HOVAL
## Q1 -1.073533465 -0.2699971236
## Q2  0.000000000  0.0000000000
## Q3 -0.038985415 -0.0098049573
## Q4 -0.003424845 -0.0008613596
## Q5 -0.002722272 -0.0006846602
##
## $indirect
##             INC          HOVAL
## Q1  0.00000000  0.000000000
## Q2 -0.43358910 -0.109049054
## Q3 -0.13613675 -0.034238831
## Q4 -0.06730519 -0.016927472
## Q5 -0.02584486 -0.006500066
##
## $total
##             INC          HOVAL
## Q1 -1.07353347 -0.269997124
```