Lecture 6: Hypothesis Testing

Mauricio Sarrias

Universidad Católica del Norte

November 6, 2017

Mandatory Reading

- Anselin, L. (1990). Some Robust Approaches to Testing and Estimation in Spatial Econometrics. Regional Science and Urban Economics, 20(2), 141-163.
- Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical analysis, 20(1), 1-17.

Which model (SLM, SEM, SDM) is the more convenient for a given empiricial analysis?

- **First approach**: use a spatial specification (SLM, SEM , SDM, etc) according to some theoretical considerations.
- **Second approach**: apply a series of statistical tests.

In this lecture we present some approaches to test whether the true spatial parameters are zero or not. In other words, we would like to assess the null $H_0 : \lambda = 0$ or $H_0 : \rho = 0$, under the alternative $H_1 : \lambda \neq 0$ or $H_1 : \rho \neq 0$.

- Recall that Moran's I test allows us assess whether some observed value of a variable at one location is independent of values of that variable at neighboring locations.
- If the model has some kind of spatial dependency, but we ignore it, then the errors should be spatially correlated...
- ... then the Moran's I on the OLS residuals should detect this misspecification.

Definition (Morans's I)

A Moran I statistic for spatial autocorrelation can be applied to regression residuals in a straightforward way. Formally, this I statistic is:

$$I = \left(\frac{n}{S_0} \right) \frac{\widehat{\boldsymbol{\varepsilon}}^{\top} \mathbf{W} \widehat{\boldsymbol{\varepsilon}}}{\widehat{\boldsymbol{\varepsilon}}^{\top} \widehat{\boldsymbol{\varepsilon}}}$$

where $\widehat{\boldsymbol{\varepsilon}}$ is a vector of OLS residuals, $\mathbf{W}$ is a spatial weight matrix, n is the number of observations and S_0 is a standardization factor, equal to the sum of all elements in the weight matrix. For a weight matrix that is normalized such that the row elements sum to one, the expression simplifies to:

$$I = \frac{\widehat{\boldsymbol{\varepsilon}}^{\top} \mathbf{W} \widehat{\boldsymbol{\varepsilon}}}{\widehat{\boldsymbol{\varepsilon}}^{\top} \widehat{\boldsymbol{\varepsilon}}} \tag{1}$$

The asymptotic distribution for the Moran statistic with regression residuals was developed by Cliff and Ord (1972, 1973). In particular, the following Theorem give us the moment of the Moran's I statistic and its distribution.

Moran's I

Theorem (Moran's I)

Consider H_0 : no spatial autocorrelation, and assume that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Let the Moran's I statistic be:

$$I = \left(\frac{n}{S_0}\right) \frac{\widehat{\boldsymbol{\varepsilon}}^\top \mathbf{W} \widehat{\boldsymbol{\varepsilon}}}{\widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}} \tag{2}$$

where $\widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ is a vector of OLS residuals, $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$, $\mathbf{W}$ is a spatial weight matrix, n is the number of observations and S_0 is a standardization factor, equal to the sum of all elements in the weight matrix. Then, the moments under the null are:

$$\mathbb{E}(I) = \frac{n}{S_0} \frac{\mathrm{tr}(\mathbf{MW})}{n - K}$$

$$\mathbb{E}(I^2) = \frac{\left(\frac{n}{S_0}\right)^2 \mathrm{tr}\left(\mathbf{MWMW}^\top\right) + \mathrm{tr}\left(\mathbf{MW}\right)^2 + \left[\mathrm{tr}(\mathbf{MW})\right]^2}{(n - K)(n - K + 2)} \tag{3}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top$. Then:

$$z_I = \frac{I - \mathbb{E}(I)}{\mathrm{Var}(I)^{1/2}} \sim N(0, 1) \tag{4}$$

where $\mathrm{Var}(I) = \mathbb{E}(I^2) - \mathbb{E}(I)^2$.

Moran's I

- The interpretation of the Moran's I is not always straightforward.
- While the H_0 is the absence of spatial dependence, a precise expresssion for the H_1 does not exists.
    - So, if you reject H_0, which model is the more appropiate? SLM, SEM?
- SEM? Note that the Moran's I is mathematically equivalent to and OLS regression of egression of $\mathbf{W}\widehat{\varepsilon}$ on $\widehat{\varepsilon}$, rather than for $\widehat{\varepsilon}$ on $\mathbf{W}\widehat{\varepsilon}$, which would correspond to an autoregressive process as in SEM model.
- Moran's I is a misspecification test that has power against a host of alternatives.
- Kelejian and Prucha (2001) have criticisez Moran's I measure, arguing that the normalizing factor used by Cliff and Ord (1972) under the H_0 is not theoretically justified.

They proposed a different normalizing factor that removes this inconsistency and achieves the aim of normalizing the variance to unity. The Moran's I they proposed is the following:

$$\bar{I} = \frac{\widehat{\varepsilon}^{\top} \mathbf{W} \widehat{\varepsilon}}{\widetilde{\sigma}^2}, \tag{5}$$

with $\widetilde{\sigma}^2$ being normalizing factor that depends on the particular model chosen as an alternative hypothesis. In particular, if the alternative hypothesis is constituted by a **SEM**, the normalizing factor assumes the expression:

$$\widetilde{\sigma}^2 = \frac{\widehat{\varepsilon}^{\top} \widehat{\varepsilon} \left\{ \operatorname{tr} \left[\left(\mathbf{W}^{\top} + \mathbf{W} \right) \mathbf{W} \right] \right\}^{-1/2}}{n}. \tag{6}$$

As a consequence the test statistic can be defined as:

$$\bar{I} = \frac{n\widehat{\varepsilon}^{\top}\mathbf{W}\widehat{\varepsilon}}{\widehat{\varepsilon}^{\top}\widehat{\varepsilon}\{\operatorname{tr}\left[(\mathbf{W}^{\top} + \mathbf{W})\,\mathbf{W}\right]\}^{-1/2}}. \tag{7}$$

The two expressions reported in Equations (2) and (7) coincide if the weight matrix has **dichotomous entries** in which case $w_{ij} = w_{ij}^2$ and, therefore,

$$\sum_i \sum_j w_{ij} = \left\{\operatorname{tr}\left[(\mathbf{W}^{\top} + \mathbf{W})\,\mathbf{W}\right]\right\}^{-1/2}.$$

In their paper, Kelijan and Prucha (2001) prove that the modified Moran test $\bar{I}$ converges in distribution to a standardized normal distribution even when the priori assumption of the normality of the error is not satisfied. Even if in large samples $\bar{I} \sim \mathrm{N}(0, 1)$, in small samples its expected value and variance may be different.

Example

To carry out the Moran's *I* test on the residuals in **R** we need to pass the regression object and spatial weight object (`listw`) to the `lm.morantest` function.

```r
library("spdep")
library("maptools")
# Load data
columbus <- readShapePoly(system.file("etc/shapes/columbus.shp",
                                       package = "spdep")[1])
col.gal.nb <- read.gal(system.file("etc/weights/columbus.gal",
                                    package = "spdep")[1])
listw <- nb2listw(col.gal.nb, style = "W")
ols <- lm(CRIME ~ INC + HOVAL,
          data =  columbus)
```

Example

```
lm.morantest(ols, listw = listw, alternative = "two.sided")

##
##   Global Moran I for regression residuals
##
## data:
## model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
## weights: listw
##
## Moran I statistic standard deviate = 2.681, p-value = 0.00734
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I       Expectation            Variance
##      0.212374153      -0.033268284         0.008394853
```

- The default setting in this function is to compute the p-value for one sided test. To get a two-sided test, the `alternative` argument must be specified explicitly.

- The results show a Moran's I statistic of `0.212`, which is highly significant and reject the null hypothesis of uncorrelated error terms.

- Recall that the Moran's I statistic has high power against a range of alternatives. However, it does not provide much help in terms of which alternative model would be most appropriate.

- Morans's I does not allow to test directly $H_0 : \rho = 0$ or $H_0 : \lambda = 0$.
- We begin our discussion of the hypothesis tests by describing the ML trinity: the Wald, Likelihood Ratio (LR), and Lagrange Multiplier (LM) test.
- These tests can be thought of as a comparison between the estimates obtained after the constraints implied by the hypothesis have been imposed to the estimates obtained without the constraints.

The Trinity

- When H_0 is true, the Wald, LR, and LM tests are asymptotically equivalent.

- As $n \to \infty$, the sampling distributions of the three test $\xrightarrow{d} \chi_r^2$, where r is the number of constraints being tested.
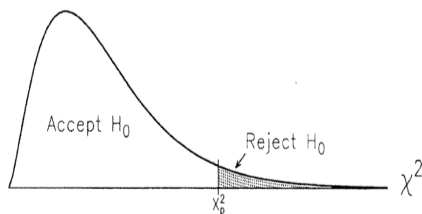
They are similar only when $n \to \infty$. In small samples this is not necessary true.

The Trinity: LR Test

- Log-likelihood function is the solid line;
- $\widehat{\beta}_U$: unconstrained estimator.
- The $H_0 : \beta = \beta^*$ imposes the constraint $\beta = \beta^*$, so that the constrained estimate $\widehat{\beta}_C = \beta^*$.
- Unless $\widehat{\beta}_U = \beta^*$, $\ln L(\widehat{\beta}_C) \leq \ln L(\widehat{\beta}_U)$.
- If the constraint **significantly** reduces the likelihood, then the null hypothesis is rejected
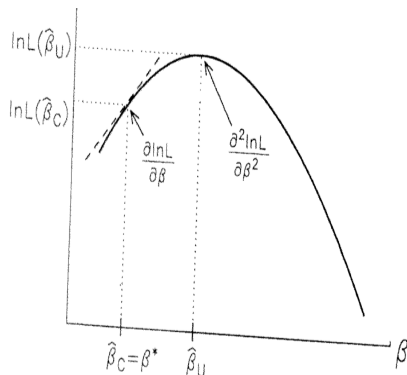
The LR test is used to compare the difference between the value of the log-likelihood of a specification considered to be unconstrained and the value of log-likelihood obtained for a constrained model specification.

We define the constrained estimate as:

$$\widetilde{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log L(\boldsymbol{\theta}) \right\} \quad \text{s.t} \quad \rho = 0 \tag{8}$$

or

$$\widetilde{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log L(\boldsymbol{\theta}) \right\} \quad \text{s.t} \quad \lambda = 0 \tag{9}$$

and the unconstrained estimate as:

$$\widehat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log L(\boldsymbol{\theta}) \right\} \tag{10}$$

Definition (Likelihood Ratio Test)

The Likelihood Ratio (LR) Test is formally defined as:

$$LR = 2 \cdot n \left(\frac{1}{n} \sum_{i=1}^{n} \log L(\widehat{\boldsymbol{\theta}}) - \frac{1}{n} \sum_{i=1}^{n} \log L(\widetilde{\boldsymbol{\theta}}) \right) \xrightarrow{d} \chi^2(r) \tag{11}$$

where r is the number of constraints.

The number of constraints imposed may vary depending on the specifications. In spatial models, the number of constrains is generally one or two, since we have the restrictions $\rho = 0$, $\lambda = 0$, or $\lambda = \rho = 0$.

LR test for the SLM

Note that the log-likelihood for the unconstrained model—that is the model for which $\rho \neq 0$—is:

$$\log L(\widehat{\boldsymbol{\theta}}) = \log |\mathbf{A}| - \frac{n \log(2\pi)}{2} - \frac{n \log(\sigma^2)}{2} - \frac{1}{2\sigma^2}(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}) \qquad (12)$$

The log-likelihood for the constrained model is found by setting $\rho = 0$ in Equation (12). Recall that if $\rho = 0$, then $\mathbf{A} = \mathbf{I} - \rho\mathbf{W} = \mathbf{I}$, then:

$$\log L(\widetilde{\boldsymbol{\theta}}) = -\frac{n \log(2\pi)}{2} - \frac{n \log(\sigma^2)}{2} - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \qquad (13)$$

Therefore, following our definition in Equation (11):

$$\begin{aligned} LR &= 2(\log L(\widehat{\boldsymbol{\theta}}) - \log(\widetilde{\boldsymbol{\theta}})) \\ &= 2\left[\log |\mathbf{A}| + \frac{1}{2\sigma^2}\left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})\right)\right] \qquad (14) \\ &= 2\log |\mathbf{A}| + \frac{1}{\sigma^2}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})\right] \end{aligned}$$

The resulting test statistic is asymptotically distributed as χ^2 with 1 degree of freedom, or, alternatively, its square root is distributed as a standard normal variate.

Note that the log-likelihood for the unconstrained model—that is the model for which $\lambda \neq 0$—is:

$$\log L(\widehat{\boldsymbol{\theta}}) = \log|\mathbf{B}| - \frac{n\log(2\pi)}{2} - \frac{n\log(\sigma^2)}{2} - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (15)$$

Then the LR for the SEM model is:

$$LR = 2\log|\mathbf{B}| + \frac{1}{\sigma^2}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}(\lambda)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \quad (16)$$

which is also distributed as $\chi^2(1)$.

To compute the test statistic LR,

1. compute the restricted MLE $\widetilde{\boldsymbol{\theta}}$ and record the value of the log-likelihood function at convergence $\log L(\widetilde{\boldsymbol{\theta}})$,

2. compute the unrestricted MLE $\widehat{\boldsymbol{\theta}}$ and record the value of the log-likelihood function at convergence $\log L(\widehat{\boldsymbol{\theta}})$,

3. and compute,

$$LR = 2\left[\log L(\widehat{\boldsymbol{\theta}}) - \log L(\widetilde{\boldsymbol{\theta}})\right]$$

This statistic is always positive because the unrestricted maximum value always exceeds the restricted one.

4. Compare LR with the critical value of chi-square distribution with 1 degrees of freedom.

Wald Test

- It compares the distance between the estimated parameters in constrained and unconstrained form.
- If the distance between the parameter estimates $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ is too high, the data fail to support the null hypothesis. In such circumstances, the null hypothesis cannot be accepted.
  - This distance can be expressed by $(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})^2$ and is influenced by the shape of the likelihood curve.
- The Wald test commonly uses unconstrained model estimates for evaluating the statistical value of W: one needs to estimate only the unconstrained model for hypothesis testing.

The Trinity: Wald Test

- The Wald test estimate the model without constraints, and assesses the constraint by considering 2 things:
- 1. It measures the distance $\widehat{\beta}_U - \widehat{\beta}_C = \widehat{\beta}_U - \beta^*$.
- The larger the distance, the less likely it is that the constraint is true. .
- 2. The distance is weighted by the curvature of the log likelihood function $\partial^2 \ln L / \partial \beta^2$
- The larger the second derivative, the faster the curve is changing.
- The LL function (dashed line) in nearly flat, so the second derivative evaluated at $\widehat{\beta}_U$ is relatively small.
- When the second derivative is small, the distance $\widehat{\beta}_U$ and $\widehat{\beta}_C$ is minor relative to the sampling variation.
- The second LL function has a larger second derivative.
- With a larger second derivative, the same distance between $\widehat{\beta}_U$ and $\widehat{\beta}_C$ might be significant.
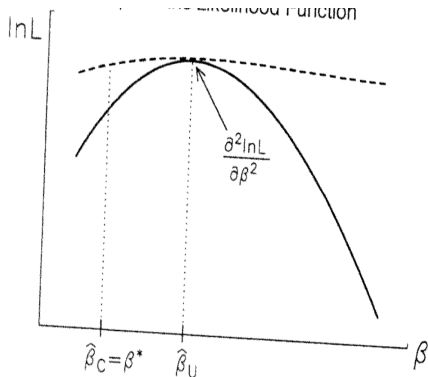
Definition (The Wald Test)

Assume that we have r nonlinear restrictions (which includes linear restriction as special case):

$$\mathbf{r}(\boldsymbol{\theta}_0) = \mathbf{0}$$

Let also

$$\mathbf{R}(\boldsymbol{\theta}) = \frac{\partial \mathbf{r}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}$$

The Wald test is given by:

$$W = n \cdot \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \left[\mathbf{R}(\widehat{\boldsymbol{\theta}}) \widehat{\mathbf{V}} \mathbf{R}(\widehat{\boldsymbol{\theta}})^\top \right] \mathbf{r}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2(r) \tag{17}$$

where r is the number of constraints.

The W statistic is:

$$W_\rho = \frac{\widehat{\rho}^2}{\widehat{\text{Var}}(\rho)} \tag{18}$$

where $\widehat{\text{Var}}(\rho)$ can be obtained from the asymptotic variance-covariane matrix as:

$$\widehat{\text{Var}}(\rho) = \left[\text{tr}(\mathbf{C}^s\mathbf{C}) + \frac{1}{\sigma^2}(\mathbf{C}\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{C}\mathbf{X}\boldsymbol{\beta})\right]^{-1} \tag{19}$$

Clearly,

$$\frac{\rho}{se(\rho)} \overset{a}{\sim} \text{N}(0, 1) \tag{20}$$

with $se(\rho)$ as the estimated standard deviation.

Computationally, the W—and LR— is more demanding since they require ML estimation under the alternative, and the explicit forms of the tests are more complicated.

The W statistic is:

$$W_\lambda = \frac{\widehat{\lambda}^2}{\widehat{\text{Var}}(\lambda)} \qquad (21)$$

where $\widehat{\text{Var}}(\lambda)$:

$$\widehat{\text{Var}}(\lambda) = \left[-\frac{\text{tr}(\mathbf{W}_B)}{\sigma^2} + \text{tr}(\mathbf{W}_B)^2 + \text{tr}(\mathbf{W}_B^\top \mathbf{W}_B) \right]^{-1} \qquad (22)$$

Wald Test: Algorithm

Let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top \right)^\top$. In general, to compute the Wald test statistic for $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$,

1. compute the unrestricted MLE $\widehat{\boldsymbol{\theta}}$,
2. compute an estimator of the variance matrix of the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, for example, the information $\mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1}$,
3. and finally compute the quadratic form:

$$W = n \cdot \widetilde{\boldsymbol{\theta}}^\top \widehat{\mathbf{V}}_w^{-1} \widehat{\boldsymbol{\theta}} \tag{23}$$

   where $\widehat{\mathbf{V}}_w$ is the $(2,2)$ block of $\mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1}$ partitioned conformably with $\boldsymbol{\theta}$: that is

$$\widehat{\mathbf{V}}_w = \left\{ \mathbf{I}_{22}(\widehat{\boldsymbol{\theta}}) - \mathbf{I}_{21}(\widehat{\boldsymbol{\theta}}) \left[\mathbf{I}_{11}(\widehat{\boldsymbol{\theta}}) \right]^{-1} \mathbf{I}_{12}(\widehat{\boldsymbol{\theta}}) \right\}^{-1} \tag{24}$$

   which is a **consistent estimator** of the asymptotic variance of $\widehat{\boldsymbol{\theta}}_2$.
4. Compare W with the critical value of chi-square distribution with $K - r$ degrees of freedom.
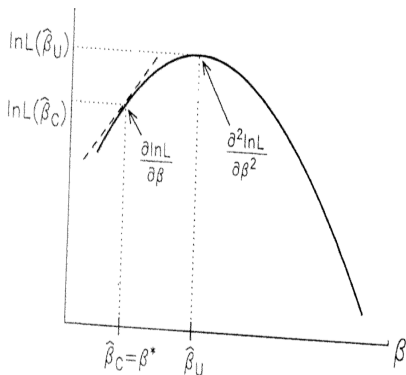
The Trinity: LM (Score) Test

- It only estimate the constrained model.

- It assesses the slope of the log likelihood function at the constraint.

- If H_0 is true, the slope (score) at the constraint should be close to 0.

- As with the Wald test, the curvature of the log likelihood function at the constraint is used to assess the significance of a nonzero slope.

- This approach is also based on the log-likelihood function curve, with the slope of the likelihood function being evaluated by the constraint type.
- The idea is that when the constraints are verified, the value of the estimated parameters $\boldsymbol{\theta}_0$ is such that the likelihood function slope at this point is zero.
- The goal is to compare, whether the slope evaluated using the constrained model is zero or strays too far from 0. In the last case, the null hypothesis must be rejected.
- The Lagrange Multiplier test (or just score test) is based on the restricted model instead of the unrestricted model.

Theorem (Lagrange Multiplier Test)

The Lagrange multiplier test statistic is:

$$LM = \left(\frac{\partial \log L(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}} \right)^{\top} \left[\mathbf{I}(\widetilde{\boldsymbol{\theta}}) \right]^{-1} \left(\frac{\partial \log L(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}} \right) \xrightarrow{d} \chi(r) \tag{25}$$

Under the null hypothesis, LM has a limiting chi-square distribution with degrees of freedom equal to the number of restrictions. All terms are computed at the restricted estimator.

- Advantage: it only requires the constrained model to be estimated.
- Less complex $\implies$ widespread use of this approach.
- Usual practice: initially use a general test for detecting residual spatial autocorrelation (Moran's *I* test for example) in order to then be able to carry out the statistical LM test to identify the specific type of the autoregressive process.

- Based on Burridge: Assume the omission of a spatial autoregressive process of the error term u_i, where $u_i = \lambda \sum_j w_{ij} u_j + \epsilon_i$.
- The null hypothesis is $H_0 : \lambda = 0$. The constrained version of the SEM model can be reduced to a standard linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

For the SEM model we need to find the score function of the log-likelihood for the constrained model:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)^\top \mathbf{B}(\lambda)^\top \mathbf{B}(\lambda)\mathbf{X}$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)^\top \mathbf{B}(\lambda)^\top \mathbf{B}(\lambda) \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) \quad (26)$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \lambda} = -\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{W}\right) + \frac{1}{\sigma^2} \left[\boldsymbol{\varepsilon}^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

LM for SEM

Under the null hypothesis $H_0 : \lambda = 0$, we get:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\bigg|_{\lambda=0} = \frac{1}{\sigma^2}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)^\top \mathbf{I}_n^\top \mathbf{I}_n \mathbf{X} = \frac{1}{\sigma^2}\widehat{\boldsymbol{\varepsilon}}_{OLS}^\top \mathbf{X}$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \sigma^2}\bigg|_{\lambda=0} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\widehat{\boldsymbol{\varepsilon}}_{OLS}^\top \widehat{\boldsymbol{\varepsilon}}_{OLS} \tag{27}$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \lambda}\bigg|_{\lambda=0} = \frac{\boldsymbol{\varepsilon}^\top \mathbf{W} \boldsymbol{\varepsilon}}{\sigma^2}$$

The test is essentially based on the score with respect to λ, i.e., on

$$s_\lambda = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \lambda}\bigg|_{\lambda=0} = \frac{\boldsymbol{\varepsilon}^\top \mathbf{W} \boldsymbol{\varepsilon}}{\sigma^2} \tag{28}$$

Recall that:

$$\text{AsyVar}(\boldsymbol{\beta}, \sigma^2, \lambda) = \begin{pmatrix} \frac{\mathbf{X}(\lambda)^\top \mathbf{X}(\lambda)}{\sigma^2} & 0 & 0 \\ k \times k & & \\ 0 & \frac{n}{2\sigma^4} & \frac{\text{tr}(\mathbf{W}_B)}{\sigma^2} \\ 0 & \frac{\text{tr}(\mathbf{W}_B)}{\sigma^2} & \text{tr}(\mathbf{W}_B)^2 + \text{tr}(\mathbf{W}_B^\top \mathbf{W}_B) \end{pmatrix}^{-1} \tag{29}$$

where $\mathbf{W}_B = \mathbf{W}(\mathbf{I} - \lambda\mathbf{W})^{-1}$. Under the null, $\mathbb{E}_{H_0}\left(\partial^2 \ln L/\partial\boldsymbol{\beta}\partial\lambda\right) = \mathbf{0}$, and $\mathbb{E}_{H_0}\left(\partial^2 \ln L/\partial\sigma\partial\lambda\right) = \mathbf{0}$ because $\mathbb{E}\left(\boldsymbol{\varepsilon}^\top\mathbf{W}\boldsymbol{\varepsilon}\right) = \sigma^2\operatorname{tr}(\mathbf{W}) = \mathbf{0}$ as $\mathbf{W}$ has a zero diagonal. Furthermore,

$$\mathbb{E}_{H_0}\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial\lambda^2}\right) = -\operatorname{tr}\left(\mathbf{W}^2 + \mathbf{W}^\top\mathbf{W}\right) \tag{30}$$

Then the expression for the LM test for a SEM specification is:

$$LM_{ERR} = \frac{1}{C}\left(\frac{\widehat{\boldsymbol{\varepsilon}}^\top\mathbf{W}\widehat{\boldsymbol{\varepsilon}}}{\widehat{\sigma}^2}\right)^2 \tag{31}$$

where $C = \operatorname{tr}\left[\left(\mathbf{W} + \mathbf{W}^\top\right)\mathbf{W}\right]$.

- Therefore, the test requires only OLS estimates.
- Under the null hypothesis, this statistic converges asymptotically to a $\chi^2(1)$. For example, if we use a significance level of 95%, the critical value is 3.84.
- Thus, we reject the null hypothesis, if the value of the statistical test LM_{ERR} is greater than 3.84.
- We can conclude in this case that spatial autocorrelation is present in the standard linear model residuals and we must proceed to estimate the SEM specification.
- Note also that it is similar in expression to Moran's I: except for the scaling factor T, this statistic is essentially the square of Moran's I.

The null hypothesis of this test is based on the significance of the
autoregressive parameter, $H_0 : \rho = 0$.
In this case:

$$s_{\rho=0} = \left. \frac{\partial \log L(\boldsymbol{\theta})}{\partial \rho} \right|_{\rho=0} = \frac{1}{\sigma^2} \boldsymbol{\varepsilon}^\top \mathbf{W} \mathbf{y} \tag{32}$$

The inverse of the information matrix was given previously. The complicating
feature of this matrix is that even under $\rho = 0$, it is not block diagonal; the
$(\rho, \boldsymbol{\beta})$ term is equal to $(\mathbf{X}^\top \mathbf{W} \mathbf{X} \boldsymbol{\beta})/\sigma^2$, obtained by inserting $\rho = 0$; i.e.,
$\mathbf{C} = \mathbf{W}$. The main problem of this is that even under $\rho = 0$, we cannot ignore
one of the off-diagonal terms. This is not the case for $s_{\lambda=0}$.

LM for SLM

Asymptotic variance of $s_{\lambda=0}$ was obtained just using the $(2,2)$ element. For the spatial lag model, asymptotic variance of $s_{\rho=0}$ is obtained from the reciprocal of the last element of: [1]

$$\text{Var}(\boldsymbol{\beta}, \sigma^2, \rho)\big|_{\rho=0} = \begin{pmatrix} \frac{1}{\sigma^2}(\mathbf{X}^\top\mathbf{X}) & \mathbf{0}' & \frac{1}{\sigma^2}\mathbf{X}^\top\mathbf{W}\mathbf{X}\boldsymbol{\beta} \\ . & \frac{n}{2\sigma^4} & \mathbf{0} \\ . & . & \text{tr}(\mathbf{W}^2 + \mathbf{W}^\top\mathbf{W}) + \frac{1}{\sigma^2}(\mathbf{W}\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{W}\mathbf{X}\boldsymbol{\beta}) \end{pmatrix}$$

Since under $\rho = 0$, $\mathbf{C} = \mathbf{W}$ and $\text{tr}(\mathbf{W}) = 0$. Recall that $T = \text{tr}\left[\left(\mathbf{W}^\top + \mathbf{W}\right)\mathbf{W}\right]$, then we can write:

$$LM_{SAR} = \frac{1}{T_1}\left(\frac{\widehat{\boldsymbol{\varepsilon}}\mathbf{W}\mathbf{y}}{\widehat{\sigma}^2}\right)^2 \tag{33}$$

where $T_1 = \left[\left(\mathbf{W}\mathbf{X}\widehat{\boldsymbol{\varepsilon}}\right)^\top\mathbf{M}\left(\mathbf{W}\mathbf{X}\widehat{\boldsymbol{\varepsilon}}\right) + T\widehat{\sigma}^2\right]/\widehat{\sigma}^2$ with $\mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$. Under the null hypothesis, the test asymptotically converges according to the χ^2 distribution to 1 degree of freedom.

---

[1] This is obtained using partitioned Inversion.

Lagrange Multiplier tests, as well as their robust forms are included in the `lm.LMtests` function. An OLS regression object and a spatial `listw` object must be passed as arguments. In addition, the tests must be specified as a character vector as illustrated below.

Example

```r
lm.LMtests(ols, listw,
           test = c("LMerr", "RLMerr", "LMlag", "RLMlag"))
```

```
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
## weights: listw
##
## LMerr = 4.6111, df = 1, p-value = 0.03177
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
## weights: listw
##
## RLMerr = 0.033514, df = 1, p-value = 0.8547
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
## weights: listw
##
## LMlag = 7.8557, df = 1, p-value = 0.005066
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
## weights: listw
##
## RLMlag = 3.2781, df = 1, p-value = 0.07021
```

Note that both `LMerr` and `LMlag` are significant. However, the robust statistics point to the lag model as the proper alternative. With this information in hand, we can select the spatial lag model as the proper model.