Spatial models in `gretl`: the SPM package

Chiara Casoli[*], Luca Pedini[*], and Francesco Valentini[*]

[*]Università Politecnica delle Marche

Abstract

This package allows practitioners to estimate cross-sectional spatial models in `gretl`. The package, presented in Casoli et al. (2019) can handle three types of models: Spatial Autoregressive Models (SAR), Spatial Durbin Models (SDM) and Spatial Error Models (SEM). Computation of the Hessian matrix is performed in both analytical and mixed ways. Some speed-up procedures for the computation of the log-determinant term are available.

Contents

1 The Spatial Models

SPM package allows estimation of three types of spatial models: the SAR, the SDM and the SEM. For further details see LeSage and Pace (2009).

1.1 The Spatial Autoregressive model and the Spatial Durbin Model

The SAR model includes spatial lags of the dependent variable only, whereas the SDM adds also spatial lags of the covariates. Generalising, and defining $Z = [\iota_n X \; W \; X]$ and $\delta = [\alpha \; \beta \; \theta]'$, it is possible to write:

$$y = \rho W y + Z\delta + \varepsilon \tag{1}$$
$$y = (I_n - \rho W)^{-1} Z\delta + (I_n - \rho W)^{-1}\varepsilon \tag{2}$$
$$\varepsilon \sim N(0, \sigma^2 I_n),$$

where equation (1) denotes the SAR if $Z = [\iota_n X]$ or the SDM if $Z = [\iota_n X W X]$, and equation (2) the related DGP.

We denote y as an $n \times 1$ vector of the dependent variable, W as the $n \times n$ spatial weight matrix, ι_n as the constant term, X as the $n \times k$ matrix of explanatory variables, and ε as the error component. Spatial dependence is captured by the parameter ρ.

Estimation of parameters ρ, δ and σ^2 is implemented via Maximum Likelihood. In particular, assuming ρ as known, defined as ρ^*, the model becomes $y - \rho^* W y = Z\delta + \varepsilon$, suggesting that parameters δ and σ^2 can be easily estimated as follows: $\hat{\hat{\delta}} = (Z'Z)^{-1} Z'(I_n - \rho^* W)y$ and $\sigma^2 = n^{-1}(y - \rho^* W y - Z\delta)'(y - \rho^* W y - Z\delta)$. The log-likelihood function is given by:

$$\ln L = -(n/2)\ln(\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e'e}{2\sigma^2} \tag{3}$$
$$e = y - \rho W y - X\beta$$
$$\rho \in (min(\omega)^{-1}, max(\omega)^{-1}), \tag{4}$$

in which ω contains the eigenvalues of the spatial weights matrix. If W has been scaled such to have the maximum eigenvalue equal to 1, it is possible to restrict the interval such that $\rho \in (min(\omega)^{-1}, 1)$. The choice of matrix W is up to users; the SPM automatically provides a row-standardisation.

The optimisation problem can be easily handled using the concentrated log-likelihood (equation (5)) as a function of the only parameter ρ. δ and σ^2 can be consequently derived as a function of the estimated ρ. This can be summarised in:

$$\ln L(\rho) = c + \ln|I_n - \rho W| - (n/2)\ln[(e_0 - \rho e_d)'(e_0 - \rho e_d)] \quad (5)$$
$$e_0 = y - Z\delta_0 \quad (6)$$
$$e_d = Wy - Z\delta_d \quad (7)$$
$$\delta_0 = (Z'Z)^{-1}Z'y \quad (8)$$
$$\delta_d = (Z'Z)^{-1}Z'Wy, \quad (9)$$

in which c is a constant term, δ_0, e_0, δ_d and e_d are computed *ex ante* from two auxiliary regressions of y and Wy on Z respectively. The Maximum Likelihood estimates of parameters $\hat{\delta}$, $\hat{\sigma^2}$ and the associated disturbances variance-covariance matrix $\hat{\Omega}$ are given by: $\hat{\delta} = \delta_0 - \hat{\rho}\delta_d$, $\hat{\sigma^2} = n^{-1}(e_0 - \hat{\rho}e_d)'(e_0 - \hat{\rho}e_d)$ and $\hat{\Omega} = \hat{\sigma^2}[(I_n - \hat{\rho}W)'(I_n - \hat{\rho}W)]^{-1}$. Finally, to calculate standard errors and the related t statistics, the variance-covariance matrix of the parameters is computed in two different ways: pure analytical and mixed (analytical/numerical), suggested in LeSage and Pace (2009). The choice of the technique is left to the users.

1.2 The Spatial Error Model

The SEM contains spatial dependences in the *disturbances*, as shown in equation (10), with (11) being the DGP.

$$y = X\beta + u \quad (10)$$
$$u = \lambda W u + \varepsilon$$
$$y = X\beta + (I_n - \lambda W)^{-1}\varepsilon \quad (11)$$
$$\varepsilon \sim N(0, \sigma^2 I_n).$$

Here the spatial dependence is expressed by the parameter λ; the other variables follow the notation described above.

The full log-likelihood is given by:

$$\ln L = -(n/2)\ln(\pi\sigma^2) + \ln|I_n - \lambda W| - \frac{e'e}{2\sigma^2} \quad (12)$$
$$e = (I_n - \lambda W)(y - X\beta).$$

Again, it is possible to concentrate the log-likelihood, as a function of the only parameter λ, and then recovering β and σ^2; unlike the previous case, however, $e(\lambda)'e(\lambda)$ is not a simple quadratic form of the parameters, but is derived from moment matrices as in (14)

$$\ln L(\lambda) = c + \ln|I_n - \lambda W| - (n/2)\ln(e(\lambda)'e(\lambda)) \quad (13)$$

$$
\begin{aligned}
A_{XX}(\lambda) &= X'X - \lambda X'WX - \lambda X'W'X + \lambda^2 X'W'WX \\
A_{Xy}(\lambda) &= X'y - \lambda X'Wy - \lambda X'W'y + \lambda^2 X'W'Wy \\
A_{yy}(\lambda) &= y'y - \lambda y'Wy - \lambda y'W'y + \lambda^2 y'W'Wy \\
\beta(\lambda) &= A_{XX}(\lambda)^{-1} A_{Xy}(\lambda) \\
e(\lambda)'e(\lambda) &= A_{yy}(\lambda) - \beta(\lambda)' A_{XX}(\lambda)\beta(\lambda)
\end{aligned} \tag{14}
$$

The values for $\hat{\beta}$ and $\hat{\sigma^2}$ can be recovered, again, straightforwardly (LeSage and Pace, 2009). The variance-covariance matrix is computed in analytical way.

2 The functions

The package provides 4 public functions. Via scripting, the functions are the following:

- `sr()`: this function provides estimation of a SAR model or a SDM.
- `sem()`: allows estimation of a SEM
- `printres()`: prints the results of `sr()` and `sem()`
- `spatial_GUI()`: boh.

2.1 The function `sr()`

This function requires the following inputs:

- `y`: the dependent variable series
- `X`: the list of regressors (without constant)
- `W`: the weight matrix
- `sdm`: a boolean (0 for the SAR, 1 for the SDM)
- `hess_form`: a boolean (0 for analytical Hessian, 1 for mixed)
- `lik_type`: an integer determining the log-determinant computation method. See below
- `moments`: see below
- `n_rep`: see below
- `poly_ord`: see below

The integer `lik_type` assumes value 0 for analytical computation, 1 for the Ord decomposition of the log-determinant, 2 for Monte Carlo numerical approximation and 3 for approximation with Chebychev polynomials (LeSage and Pace, 2009). The default is 0. If `lik_type` = 2, the `moments` and `n_rep` scalars define the number of moments and the number of replications, respectively, necessary for the Monte Carlo approximation. If omitted, the defaults are 50 moments and 100 replications. Instead, if `lik_type` = 3, the scalar `poly_ord` specifies the Chebychev polynomial order. The default is 10.

The output of the function is a bundle containing:

- `model`: the estimated model
- `variables`: list of regressors
- `beta`: estimated coefficients
- `betastderr`: standard errors of estimated coefficients
- `beta_t`: the t-statistic for $\hat{\beta}$
- `rho`: estimated ρ
- `rhostderr`: standard errors of estimated ρ
- `rho_t`: the t-statistic for $\hat{\rho}$
- `s2`: estimated variance of the error
- `s2stderr`: standard errors of $\hat{\sigma^2}$
- `s2_t`: the t-statistic for $\hat{\sigma^2}$
- `hess`: Hessian computation type
- `lkt`: log-determinant computation type
- `lk`: log-likelihood
- `Sigma`: covariance matrix of errors
- `CPUtime`: elapsed time in seconds

2.2 The function `sem()`

This function requires the following inputs:

- `y`: the dependent variable series
- `X`: the list of regressors (without constant)

- W: the weight matrix
- lik_type: an integer determining the log-determinant computation method. See below
- moments: see below
- n_rep: see below
- poly_ord: see below

The integer lik_type assumes value 0 for analytical computation, 1 for the Ord decomposition of the log-determinant, 2 for Monte Carlo numerical approximation and 3 for approximation with Chebychev polynomials (LeSage and Pace, 2009). The default is 0. If lik_type = 2, the moments and n_rep scalars define the number of moments and the number of replications, respectively, necessary for the Monte Carlo approximation. If omitted, the defaults are 50 moments and 100 replications. Instead, if lik_type = 3, the scalar poly_ord specifies the Chebychev polynomial order. The default is 10.

The output of the function is a bundle containing:

- model: the estimated model
- variables: list of regressors
- beta: estimated coefficients
- betastderr: standard errors of estimated coefficients
- beta_t: the t-statistic for $\hat{\beta}$
- lambda: estimated λ
- lambdastderr: standard errors of estimated λ
- lambda_t: the t-statistic for $\hat{\lambda}$
- s2: estimated variance of the error
- s2stderr: standard errors of $\hat{\sigma^2}$
- s2_t: the t-statistic for $\hat{\sigma^2}$
- lkt: log-determinant computation type
- lk: log-likelihood
- Sigma: covariance matrix of errors
- CPUtime: elapsed time in seconds

3 Example

References

Casoli, C., Pedini, L., and Valentini, F. (2019). Spatial models in gretl: the spm package. *GNU Regression, Econometrics and Time Series Library*, page 23.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.