SpatialEconometrics.jl: a Package for Estimating Spatial Econometrics Models in Julia

Alan Leal[*]

Abstract

Spatial Econometrics models and estimators are multiple. Their usage in different scientific areas is relevant. Lacking in the Julia environment is a package that present to the user the oppportunity to fit and analyze spatial econometrics models. The package SpatialEconometrics.jl aims to fill this gap. In this release (v0.1.0), our goal is to implement the SAR, SEM, and SAC (SARAR) models, using maximum likelihood estimators for these three models. When appropriate (for SAR and SARAR), we also implement function detailing direct, indirect, and total effects. Our implementation is consistent with literature as results are presented in formatted tables to allow for fast analysis by users.

1 Introduction

Spatial Econometrics is a branch of Econometrics populated with several methods aimed at analyzing the relation among unities, when space is not neutral. In that way, Spatial Econometrics deals directly to spillover and feedback effects between a spatial unity and their neighbors. Methods in this area are multiple, with models ranging from cross-section, to panel data, endogenous models, regime models, and much more.

The package SpatialEconometrics.jl aims to fill the gap in the Julia ecosystem of a package that implements Spatial Econometrics models and estimators. Initially, we implement the models SAR, SEM, and SAC, using their maximmum likelihood estimators.

The paper is divided in, besides this Introduction, a Methods sections, followed by Usage, and finally a Final Remarks.

2 Methods

We work here with three cross section spatial models, which are:

- SAR:
$$y = \rho W y + X\beta + u$$
  with $u \sim N(0, \sigma^2)$
- SEM:
$$y = \rho X\beta + u$$

[*]Ph.D. candidate in Economics at Universidade of São Paulo

with $u = \lambda W u + e$ and $e \sim N(0, \sigma^2)$

- SAC (or SARAR):

$$y = \rho W y + X\beta + u$$

with $u = \lambda W u + e$ and $e \sim N(0, \sigma^2)$

The three models are estimated by the maximization of their log likelihood. The variance-covariance matrix of the coefficients is estimated using the property that the this matrix equals the negative of the inverse of the information matrix of each estimator, that is,

$$var(\hat{\beta}) = -(Inf(f, \hat{\beta}))^{-1}$$

Where Inf is the information matrix associated to the f log likelihood function and $\hat{\beta}$ is our estimated coefficients.

2.1 Log Likelihood of the models

The models considered have by their log-likelihood the following expressions:

- SAR

$$\begin{aligned} logL(\beta, \rho, \sigma^2) = -\frac{n}{2} log(2\pi\sigma^2) + log|I_n - \rho W| - \\ \frac{1}{2\sigma^2}(y - X\beta)'((I_n - \rho W')((I_n - \rho W)(y - X\beta) \end{aligned}$$

Where y is the vector of dependent variable, X is a matrix containing the explanatory variables, including the constant, β is the coefficients vector, ρ is the spatial autocorrelation parameter, W is the spatial weight matrix, n is the number of observations in σ^2 is the variance of the error term.

- SEM

$$logL(\beta, \lambda, \sigma^2) = \frac{n}{2} log(2\pi\sigma^2) + log|I_n - \lambda W| - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

Where y is the vector of dependent variable, X is a matrix containing the explanatory variables, including the constant, β is the coefficients vector, λ is the spatial error parameter, W is the spatial weight matrix, n is the number of observations in σ^2 is the variance of the error term.

- SAC (or SARAR)

$$\begin{aligned} logL(\beta, \rho, \lambda, \sigma^2) = \frac{n}{2} log(2\pi\sigma^2) + log|I_n - \rho W_1| + log|I_n - \lambda W_2| \\ - \frac{1}{2\sigma^2}(y - \rho W_1 y - X\beta)'(y - \rho W_1 y - X\beta) \end{aligned}$$

Where y is the vector of dependent variable, X is a matrix containing the explanatory variables, including the constant, β is the coefficients vector, ρ is the spatial autocorrelation parameter,λ is the spatial error parameter, W_1 is the spatial weight matrix related to the dependent variable, W_2 is the spatial matrix related to the error term, n is the number of observations in σ^2 is the variance of the error term.

2.2 Effects

The effects of SAR and SAC models are calculated in the following manner:

- Direct Effects

$$\frac{1}{n}\sum_i \phi_{ii}\beta_r = \frac{1}{n}tr(I-\rho W)^{-1}\beta_r$$

- Total Effects

$$\frac{1}{n}\sum_i\sum_j \phi_{ij}\beta_r$$

- Indirect Effects

$$\frac{1}{n}\sum_i\sum_j \phi_{ij}\beta_r - \frac{1}{n}\sum_i \phi_{ii}\beta_r = \frac{1}{n}tr(I-\rho W)^{-1}\beta_r$$

Where n is the number of observations and $\phi_{i,j} = (I-\rho_W)^{-1}[i,j]$.

3 Usage

Initially, you can install the SpatialEconometrics.jl package by running:

```
using Pkg
Pkg.add(url="https://github.com/alanleal-econ/SpatialEconometrics.jl")
using SpatialEconometrics
```

We use the zika ceara dataset here, coming from Amaral, Resende de Carvalho, Hernandes Rocha, da Silva, and Vissoci (2019).

3.1 SAR

The function sar is called in the following manner:

```
sar_model=sar(X,y,W)
names_col=names(df)
sar_summary(sar_model,names_col)
```

Where X is the matrix of independent variables, incluiding a constant column, y is the dependent variable, and W is a matrix, built from a spatial matrix. Moreover, df is the original dataframe with data that contains the names of the variables. This command produces the following results:

```
.------------.---------.---------.----------.----------.-----------.
Maximum Likelihood Estimation of SAR Model
.------------.---------.---------.----------.----------.-----------.
Log-Likelihood: -737.902
Number of observations: 184
sigma2: 176.214
.------------.---------.---------.----------.----------.----------.
|  Variable  |   beta   | Std Dev | Lower CI | Upper CI | p-value  |
:------------+---------+---------+----------+----------+----------:
|  constant  | 16.627  | 39.880  | -62.078  |  95.332  | 0.677241 |
|   ln_gdp   | -2.600  |  6.923  | -16.263  |  11.063  | 0.707744 |
```

```
|    ln_pop    |   1.228   |  8.970   | -16.475  |  18.931  | 0.89124  |
|  mobility    | -14.590 | 35.789   | -85.221  |  56.041  | 0.684016 |
|  environ     |  5.229   | 11.160   | -16.796  |  27.254  | 0.639967 |
| sanitation   |  3.247   | 10.118   | -16.721  |  23.216  | 0.748625 |
'------------'---------'---------'----------'----------'----------'
rho: 0.238, Standard Deviation: 0.113
```

The direct, indirect, and total effects are used in the following manner:

```
beta_complet=vcat(sar_model.sigma2,sar_model.rho[1],sar_model.coefs
    [:,1])
sar_effects=effects_sar(y,X,W,beta_complet)
effects_summary(sar_effects,names_col[2:end])
```

Which produces the following results:

```
.------------.----------------.------------------.---------------.
|  Variable  | Direct Effects | Indirect Effects | Total Effects |
:------------+----------------+------------------+---------------:
|   ln_gdp   |     -2.630     |      -0.782      |    -3.411     |
|   ln_pop   |     1.242      |      0.369       |     1.612     |
|  mobility  |    -14.758     |      -4.386      |    -19.144    |
|  environ   |     5.289      |      1.572       |     6.861     |
| sanitation |     3.285      |      0.976       |     4.261     |
'------------'----------------'------------------'---------------'
```

3.2 SEM

The function sem is called in the following manner:

```
sem_model=sem(X,y,W)
names_col=names(df)
sar_summary(sem_model,names_col)
```

Which produces the following results:

```
.------------.---------.---------.----------.----------.-----------.
Maximum Likelihood Estimation of SEM Model
.------------.---------.---------.----------.----------.-----------.
Log-Likelihood: -737.902
Number of observations: 184
sigma2: 176.214
.------------.---------.---------.----------.----------.----------.
|  Variable  |   beta   | Std Dev | Lower CI | Upper CI | p-value  |
:------------+---------+---------+----------+----------+----------:
|  constant  | 16.627  | 39.880  | -62.078  |  95.332  | 0.677241 |
|   ln_gdp   | -2.600  |  6.923  | -16.263  |  11.063  | 0.707744 |
|   ln_pop   |  1.228  |  8.970  | -16.475  |  18.931  | 0.89124  |
|  mobility  | -14.590 | 35.789  | -85.221  |  56.041  | 0.684015 |
|  environ   |  5.229  | 11.160  | -16.796  |  27.254  | 0.639967 |
| sanitation |  3.247  | 10.118  | -16.721  |  23.216  | 0.748625 |
'------------'---------'---------'----------'----------'----------'
lambda: 0.238, Standard Deviation: 0.113
```

3.3 SAC

The function sarar is called in the following manner:

```
    M=W
    sarar_model=sarar(X,y,W,M)
    sarar_summary(sarar_model,names_col)
```

Which produces the follwoing results:

```
    .------------.---------.---------.----------.----------.-----------.
Maximum Likelihood Estimation of SAC Model
.------------.---------.---------.----------.----------.-----------.
Log-Likelihood: -737.764
Number of observations: 184
sigma2: 176.676
.------------.---------.---------.----------.----------.----------.
|  Variable  |    beta   | Std Dev | Lower CI | Upper CI | p-value  |
:------------+---------+---------+----------+----------+----------:
|  constant  | 15.360  | 40.525  | -64.620  |  95.341  | 0.705119 |
|   ln_gdp   | -2.890  |  7.135  | -16.973  |  11.193  | 0.685943 |
|   ln_pop   |  1.486  |  9.109  | -16.491  |  19.463  | 0.870596 |
|  mobility  | -13.968 | 37.093  | -87.175  |  59.240  | 0.706961 |
|  environ   |  5.964  | 11.523  | -16.777  |  28.706  | 0.605369 |
| sanitation |  4.277  | 10.677  | -16.796  |  25.351  | 0.689199 |
'------------'---------'---------'----------'----------'----------'
rho: 0.085, Standard Deviation: 0.322
lambda: 0.171, Standard Deviation: 0.311
```

The direct, indirect, and total effects are used in the following manner:

```
beta_complet=vcat(sarar_model.sigma2,sarar_model.rho[1],sarar_model.
   lambda[1],sarar_model.coefs[:,1])
SARAR_effects=effects_sarar(y,X,W,M,beta_complet)
effects_summary(SARAR_effects,names_col[2:end])
```

Which produces the following results:

```
.------------.----------------.-----------------.---------------.
|  Variable  | Direct Effects | Indirect Effects | Total Effects |
:------------+----------------+-----------------+---------------:
|   ln_gdp   |     -2.894     |      -0.264     |     -3.159    |
|   ln_pop   |     1.488      |      0.136      |      1.624    |
|  mobility  |    -13.987     |      -1.278     |    -15.265    |
|  environ   |     5.973      |      0.546      |      6.518    |
| sanitation |     4.283      |      0.391      |      4.675    |
'------------'----------------'-----------------'---------------'
```

4 Final Remarks

References

Amaral, P., Resende de Carvalho, L., Hernandes Rocha, T. A., da Silva, N. C., & Vissoci, J. R. N. (2019). Geospatial modeling of microcephaly and zika virus spread patterns in brazil. *PloS one*, *14*(9), e0222668.