Extending Predictions from Spatial Econometric Models on R

Jean-Sauveur AY
<jsay.site@gmail.com>

Julie LE GALLO
<jlegallo@univ-fcomte.fr>

March 31, 2014

Abstract

This repository presents some improvements of the `R` function used to make predictions from spatial models (`predict.sarlm`). The status is actually under construction: *¡2014-03-24 lun.¿*.

TODO

- Code the variances and confidence intervals of predictors

1 Getting started

1.1 Changes relative to current function

- Implement predictions for SARAR and Mixed SARAR models
- Compute BLUP and almost BLUP spatial predictors

1.2 TODO Determine output structure

1.3 Changes in computational details

- About the in-sample / out of sample structure (`newdata`)
- About the inclusion of the intercept in mixed models
- About the distinction between trend and signal
- The simplification of the in-sample predictions

2 Econometric Theory

2.1 Spatial Econometric Framework

From the more general from of the Cliff-Ord (1973, 1981) class of models,[1]

[1] This model has different names in the literature: spatial autoregressive model with autoregressive disturbances (SARAR(1,1), Kelejian and Prucha, 1998) or Spatial Autoregressive Conditional (SAC, XX). We retain XX here.

$$
\begin{aligned}
y &= \rho W y + X\beta + \gamma W X + \varepsilon \\
\varepsilon &= \lambda W \varepsilon + u
\end{aligned}
$$

with $u \sim \mathbf{N}(0, \sigma^2 \cdot I_N)$. The y is a $N \times 1$ vector continuous outcome, X is a $N \times K$ matrix of the K covariates, and W is a $N \times N$ spatial weight matrix. We limit ourselves to a same weight matrix in the outcome and error equations, but nothing precludes this restriction. The unknown parameters ρ, γ, λ and σ have to be estimated, as the vector u of residuals. Classically, we assume that $\text{diag}(W) = 0$, $\mid \rho \mid < 1$, $\mid \lambda \mid < 1$. *standardization of W?* Not the same notations than KP 2007.

This model is sufficiently general that the SARAR(1,1) model can be recovered with $\theta = 0$ (Kelejian2007) (also called SAC by Biva02, BPGR13), the spatial error model (SEM) can be recovered with $\rho = \theta = 0$, the spatial X model (SXM) with $\rho = 0$, the spatial autoregressive (SAR) model with $\theta = \lambda = 0$; and the spatial Durbin model (SDM) model can be recovered when $\lambda = 0$. Another useful non exclusive distinction is the error models (SEM, SDM and SARAR) and the lag models (SAR, SDM and SARAR)

2.2 Making Predictions

The bias of actual predictors come from the correlation between the spatially lagged dependent variable and the error term.

Can we still maintain the signal trend distinction?

We have to explain the differences between in-sample, out-of-sample and ex-sample in a spatial context. Ex-sample is not necessary linked to temporal, it is also interesting to counterfactual simulations. The prediction in out-of-sample needs a certain spatial embedding between the two spatial samples, not having sampled neighbors does not mean no neighbors. But in a spatial segregative case, this corresponds to a ex-sample case.

3 Current function from `spdep`

Our code is an extension of the function `predict.sarlm()` actually the default function from the package `spdep` (Bivand).

```
library(spdep) ; predict.sarlm
```

predict-sarlm.R

The current function, accessible through previous link, implement different predictor according to the absence of the presence of newdata. For the in-sample predictions (`if(newdata== NULL)`), the predictors are computed as Eq. XX using BLUP. For the out of sample predictions (`if(newdata!= NULL)`), the predictors are computed as Eq. XX using biased and inefficient predictors. It produces inconsistencies by not implementing the same predictions if we put the data that are used to fit the model in the `newdata` argument (cf. XX example below). Another shortcoming of the current function is the class of objects from SEM and SXM: they are not vectors. Lastly, if we put `sacmixed` objects in the current function, they are not recognized as such and produce some errors about matrix dimension.

At the center of this distinction is the observability of the outcome variable y.

Some other particularities are present in the current function. The OS predictor for error models is KP1 but not directly for lag models. For that, we have to put `legacy== FALSE`. The signal is computed by difference for the lag models in out of sample.

4 The New Functionalities

4.1 Choosing a type of predictor

Our new `R` function for spatial predictions – called `sppred` for the moment – admits a first additional argument `predictor` that specify the computed predictor. Knowing that predictors corresponding to larger information sets are more complex, flexibility is needed to let the user makes its own trade-off between simplicity and prediction efficiency. The following table define the available predictors.

Table 1: The available values for the new `predictor` argument

`predictor`	label	equation (see XX)
″1″	minimum information	(XX)
″2″	heuristic BLUP	(XX)
″3″	BLUP	(XX)
″4″	heuristic data	(XX)

The `predictor` 4 is currently the default for IS prediction in `predict.sarlm` (it corresponds to the predictor KP4 for lag models and KP5 for error models).

4.2 Specifying

4.3 General structure, usual checks, and IS predictions

Here the code

4.4 The predictors 1 for OS predictions

5 Testing