Extending Predictions from Spatial Econometric Models on R

Jean-Sauveur AY <jsay.site@gmail.com>
Raja CHAKIR <chakir@grignon.inra.fr>
Julie LE GALLO <jlegallo@univ-fcomte.fr>

April 18, 2014

Abstract

This document presents an integrative framework to make predictions from spatial autoregressive models Cliff and Ord (1973, 1981); Anselin (1988). It also contains the corresponding **R** code to implement the predictors presented. The code is tangled into the `sppred` function that implements in particular the predictors from LeSage and Pace (2004, 2008) and Kelejian and Prucha (2004) for a large number of specifications of spatial autocorrelation from the `spdep` package (Bivand 2014). To use this code, save the file `sppred` in your working directory then submit `source("sppred.R")` to **R**. Some examples on the `boston` data are also available at section XX. The status of this work is actually under construction, comments are welcome.

`STILL TODO`
- Code the variances and confidence intervals of predictors
- Code the weight matrix for ex-sample predictions
- Code the predictors for `sphet` and `splm` objects

Contents

1 Theoretical Framework

1.1 Spatial econometric models

The particularity of spatial econometric models is the *ex ante* specification of interdependence between statistical units, typically from their spatial proximity. Consequently, resulting predictions not only depend on the variables of the target units (i.e., the units for which we want to predict the outcome) but potentially involve the entire set of both explanatory as well as dependent variable sample data observations. This implies that the sample pattern of the outcomes or the residuals contains additional information which may be used to modify the regression function so as to reduce the prediction variance (Goldberger 1962). Cliff and Ord (1973)

The literature about predicting from spatial econometric models is not actually unified, due to different modeling frameworks and notations: see LeSage and Pace (2004, 2008) and Kelejian and Prucha (2004). A first contribution of this document is to unify the different predictors proposed in the literature into an integrative spatial econometric framework. A second contribution is to present the **R** function `sppred` that implements all the spatial predictors and presents some empirical evidence on their relative performance on a well-known dataset (`boston`, Bivand 2014).

We present analytically the available predictors from spatial econometric models estimated on a sample of N spatial units. We begin with the more general spatial autoregressive mixed conditional (SMC) specification of the Cliff-Ord (1973, 1981) class of homoscedastic models with exogenous covariates,[1]

$$
\begin{aligned}
y &= \alpha + \rho W y + X\beta + WX\theta + u \\
u &= \lambda W u + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathbf{N}(0, \sigma^2 I_N)
\end{aligned}
$$

The term y is a $N \times 1$ vector of the continuous outcome of interest, X is a $N \times K$ matrix of the K non-constant covariates, and W is a $N \times N$ full-rank spatial weight matrix (see Anselin 1988). We limit ourselves to a same weight matrix in the outcome and error equations, but formally nothing requires this restriction. The unknown parameters (or vectors of parameters) α, ρ, β, θ, λ and σ grouped in Θ have to be estimated, as the vector ε of innovations. Classically, we assume that $\text{diag}(W) = 0$ and $|\rho|, |\lambda| < 1$. The spatial weight matrix W does not need to be symmetric.

This form is sufficiently general that the spatial autoregressive conditional (SAC) model can be recovered with $\theta = 0$, the spatial Durbin model (SDM) model with $\lambda = 0$, the spatial autoregressive (SAR) model with $\theta = \lambda = 0$, the spatial X model (SXM, also called Spatial Durban Error Model by LeSage and Pace, 2009) with $\rho = 0$ and the spatial error model (SEM) can be recovered with $\rho = \theta = 0$. All these models can be estimated by maximum likelihood through the functions `errorsarlm`, `lagsarlm`, and `sacsarlm` of the **R** package `spdep` (Bivand, 2014).

For future reference, note that the general SMC model can be reduced in $y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}(I - \lambda W)^{-1}\varepsilon$, with a distribution that involves all of the model parameters:

$$
y \mid X, W, \Theta \sim \mathbf{N}(\mu, \sigma_\varepsilon^2 \Sigma), \tag{1}
$$

[1] This terminology comes from LeSage and Pace (2009) and Bivand (2014). This model is also called spatial autoregressive model with autoregressive disturbances, SARAR(1,1), by Kelejian and Prucha (1998). The notations of spatial lag coefficients also depends on authors, we adopt here those of the **R** package `spdep` (Bivand, 2014).

with $\mu = (I - \rho W)^{-1} X\beta$, and $\Sigma = \left[(I - \rho W)(I - \lambda W)(I - \lambda W^\top)(I - \rho W^\top)\right]^{-1}$. According to the literature, we consider hereafter that the set of parameters Θ is known with certainty. This has the advantage of simplifying notations by neglecting a source of uncertainty that depends on the sample size. Moreover, all the parameters are estimated simultaneously and it makes no sens to consider Θ as partially known. Hence, the predictors presented are all conditioned by Θ, which is dropped from conditioning sets for the sake of clarity. So its not the true BLUP but an approximation.

Finally, note that geo-statistical models (typically from the kriging methods, Cressie and Cassie (1993)) are aware about the need for using BLUP predictors. Relatively to spatial econometric approaches, they involve specifying spatial dependence through the error process (like in spatial error models, see below) as opposed to the spatial lag of the outcome vector. Predicting from these error models take a simpler form than autoregressive models and can be easily recovered from the theoretical framework presented here. The reverse is not true.[2]

1.2 Implementing predictions

Basically, we consider M target spatial units for which we want to predict the values of the outcome y. These units may or may not be included in the dataset used for the estimation of the parameters. At this moment, it makes no differences and we distinguish them only in the research interest of obtaining predictions of y.[3] We discuss this point in more details at the next subsection XX.

The available information set is crucial here, as each predictor is formulated as a conditional expectation relative to it.

. in terms of conditional expectation put the focus of the available information used to construct the predictors which is important for the transparency of the code.

$$\mathbf{E}(y \mid X) = X\widehat{\beta} \tag{PRD.X}$$

$$\mathbf{E}(y \mid X, W) = X\widehat{\beta} + WX\widehat{\theta} \tag{PRD.WX}$$

$$\mathbf{E}(y \mid X, W) = (I_n - \rho W)^{-1}(X\widehat{\beta} + WX\widehat{\theta}) \tag{PRD.KP1}$$

This correspond to KP1 but also the exogenous predictor of LSP. It is the default predictor of the function `sppred`.

$$\mathbf{E}(y \mid X, W, Wy) = \rho Wy + X\widehat{\beta} + WX\widehat{\theta} + \lambda(Wy - X\beta - WX\theta) \tag{PRD4}$$

It is a bit strange but y cannot be recovered from W and Wy. KP analyses this predictor separately in KP4 and KP5 but such a distinction is not necessary. Since $\text{diag}(W) = 0$, Wy does not use individually the y_i to predict itself but needs them all to predict the entire vector.

The bias of actual predictors come from the correlation between the spatially lagged dependent variable and the error term. It is why the use of best linear unbiased predictor (Golberger) is of

[2] Does the geo-statistical literature is concerned with marginal effects of covariates? Somme papers?

[3] In particular, it could be of interest to predict outcome values for units with known y for diagnostic, goodness-of-fit purposes or computations of marginal effects of covariates.

particular importance.

$$\mathbf{E}(y \mid X, W, y_{IS}) = (I_n - \rho W)^{-1} X\widehat{\beta} + WX\widehat{\theta} + \Omega_{21}\Omega_{11}^{-1}(y_{IS} - \mathbf{E}(y_{IS} \mid X, W)) \tag{PRD5}$$

The covariance can be simplified with precision matrix (Harville) which do not necessitate the inversion of matrix: $\Omega_{21}\Omega_{11}^{-1} = \Psi_{21}\Psi_{11}^{-1}$. LSP also present a leave-one-out (loo) specification of this predictor, that is based on the same conditioning set, except that we consider the predictions individually of each OS units indexed i.

$$\mathbf{E}(y_i \mid X, W, y_{IS}) = (I_n - \rho W)^{-1} X\widehat{\beta} + WX\widehat{\theta} + \Omega_{21}\Omega_{11}^{-1}(y_{IS} - \mathbf{E}(y_{IS} \mid X, W)) \tag{prdLSP}$$

This simplification comes from...

Note that this predictor is not consistent in terms of expectation conditioned by a unique information set. In effect, the non-stochastic part is conditioned without the knowledge of the outcome variable, which is nevertheless necessary to compute the BLUP correction of the stochastic part. This misleading appears because of the writing of the reduced form before taking the conditional expectation.

$$\mathbf{E}(y_i \mid X, W, y_{IS}) = \rho W y + X\widehat{\beta} + WX\widehat{\theta} + \Omega_{21}\Omega_{11}^{-1}(y_{IS} - \mathbf{E}(y_{IS} \mid X, W)) \tag{prdKP2}$$

yopla

$$\mathbf{E}(y_i \mid X, W, y_{IS}) = \rho W y + X\widehat{\beta} + WX\widehat{\theta} + \Omega_{21}\Omega_{11}^{-1}(y_{IS} - \mathbf{E}(y_{IS} \mid X, W)) \tag{prdKP2}$$

If we put aside the differences in the conditional mean of the exogenous part, this predictor `predKP2` is identical to `prdLSP`.

Can we still maintain the signal trend distinction? Does it the same as direct and indirect effects of covariates?

We develop a framework of prediction from models with interdependent observations.

We implement the KP1 predictors, also called exogenous by LeSage and Pace.

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.

1.3 A taxonomy of spatial sample schemes

In spatial econometrics, the usual distinction between in-sample (IS) and out-of-sample (OS) predictions has firstly to be refined. It is not simply inside our outside the calibration sample but it has also some relations (or not) with it.

2 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.

3 The `sppred` extension

3.1 General Structure

Here is the general structure of the functions that call sub-functions that are defined below.

This function contents the usual verifications, with 2 more arguments: `cond.set` for the conditional set (see XX) and `mean` for the specification of the structural mean.

It is important that the same predictor is implemented when newdata are NULL or not, as when spatial matrix set.

The scan for the lagged WX is by the presence of "lag." at their name, it has to be changed.

```
sppred <- function(object, newdata = NULL, listw = NULL, yobs= object$y,
                   condset= "DEF", blup = NULL, loo = FALSE, power = NULL,
                   zero.policy = NULL, legacy = TRUE, order = 250,
                   tol= .Machine$double.eps^(3/5), ...) {
    require(spdep)
    ## USUAL VERIFICATIONS
    if (is.null(zero.policy))
        zero.policy <- get("zeroPolicy", envir = spdep:::.spdepOptions)
    stopifnot(is.logical(zero.policy))
    if (is.null(power)) power <- object$method != "eigen"
    stopifnot(is.logical(legacy)) ; stopifnot(is.logical(power))
```

```
    ## DETERMINING THE MODEL
    if (object$type== "error"){
        mod <- ifelse(object$etype== "error", "sem", "sxm")
    } else {
        mod <- switch(object$type, "lag"= "sar", "mixed"= "sdm",
                                   "sac"= "sac", "sacmixed"= "smc")
    }
    ## DATA SHAPING
    if (mod %in% c("sem", "sxm")) {lab= object$lambda ; rho= 0         }
    if (mod %in% c("sar", "sdm")) {lab= 0             ; rho= object$rho}
    if (mod %in% c("sac", "smc")) {lab= object$lambda ; rho= object$rho}
    Wlg <- substr(names(object$coefficients), 1, 4)== "lag."
    B <- object$coefficients[ !Wlg] ; Bl <- object$coefficients[ Wlg]
    if (is.null(newdata)){
        X   <- object$X[, !Wlg]
    } else {
        frm <- formula(object$call)
        mt  <- delete.response(terms(frm, data = newdata))
        mf  <- model.frame(mt, newdata)
        X   <- model.matrix(mt, mf)
        if (any(object$aliased)) X <- X[, -which(object$aliased)]
    }
    ## WEIGHT MATRIX, add an error message
    if (is.null(listw)) lsw <- eval(object$call$listw) else lsw <- listw
    ## PREDICTORS
    if (is.null(blup)){
        pt <- switch(condset, "X"= 1, "XW"= 2, "DEF"= 3, "XWy"= 4)
    } else {
        pt <- switch(blup, "LSP"= 5, "KP2"= 6, "KP3"= 7, "KPG"= 8)
    }
    prdX <- as.vector(X %*% B)
    if (pt> 1) prdWX   <- prdWX(prdX, X, Bl, mod, lsw)
    if (pt> 2 && pt!= 4) prdKP1  <- prdKP1(prdWX, rho, lsw, power, order, tol)
    if (pt> 3){
        prdWXy <- prdWX+
            rho* lag.listw(lsw, yobs)+ lab* lag.listw(lsw, yobs- prdWX)}
    if (pt==5) prdLSP <- prdLSP(prdKP1, rho, lab, lsw, yobs, loo)
    if (pt> 5 && !loo) stop("Set loo= TRUE for this blup predictor")
    if (pt==6){
        prdKP2 <- prdKP2(prdKP1, prdWXy,
                         rho, lab, lsw, yobs, power, order, tol)}
    if (pt==7){
        prdKP3 <- prdKP3(prdKP1, prdWXy,
                         rho, lab, lsw, yobs, power, order, tol)}
    if (pt==8) stop("not implemented")
    prd <- switch(pt, "1"= prdX  , "2"= prdWX , "3"= prdKP1, "4"= prdWXy,
                      "5"= prdLSP, "6"= prdKP2, "7"= prdKP3, "8"= prdKPG)
    class(prd) <- "sppred" ; as.vector(prd)
}
```

we choose to not use `object$tarX` and `object$tarY` for more transparencies. It is clear that we lost from that in terms of computation time. It is easy to predict by conditioning only on "X" because it is the same form for all the spatial models (see equation XX).

3.2 Predictors conditioned on X, W

3.2.1 exogenous predictor

```
prdWX <- function(prdX, X= X, Bl= Bl, mod= mod, lsw= lsw){
    if (!mod %in% c("sxm", "sdm", "smc")){
        prdWX <- prdX } else {
            K <- ifelse(colnames(X)[ 1] == "(Intercept)", 2, 1)
            m <- ncol(X) ; WX <- matrix(nrow= length(prdX), ncol= m+ 1- K)
            for (k in K: m){
                WX[, k+ 1- K] <- lag.listw(lsw, X[, k])
            }
            prdWX <- prdX+ (WX %*% Bl)
        }
    prdWX
}
```

3.2.2 endogenous predictor

```
prdKP1 <- function(prdWX, rho= rho, lsw= lsw,
                   power= power, order= order, tol= tol){
    if (power){
        W <- as(as_dgRMatrix_listw(lsw), "CsparseMatrix")
        prdKP1 <- c(as(powerWeights(W, rho= rho, X= as.matrix(prdWX),
                                    order= order, tol= tol), "matrix"))
    } else {
        prdKP1 <- c(invIrW(lsw, rho) %*% prdWX)
    }
    prdKP1
}
```

3.3 Predictors conditioned on X, W, y

3.3.1 biased predictors

The predictors equivalent to KP4 and KP5, we do not let the choice (because the omitted combination can be recovered from previous predictors) and we can eventually add a KP6 for SAC and SMC models. The computations are in the general.

3.3.2 BLUP LSP

It can make sens to distinguish one shot to one leave one.

```
prdLSP <- function(prdKP1, rho= rho, lab= lab,
                   lsw= lsw, yobs= yobs, loo= loo){
```

```
    ZL <- diag(length(prdKP1))- (lab* listw2mat(lsw))
    ZR <- diag(length(prdKP1))- (rho* listw2mat(lsw))
    Z  <- ZL %*% ZR ; P22 <- t(Z) %*% Z
    if (loo){
        prdLSP <- matrix(NA, ncol= 1, nrow= length(prdKP1))
        for (i in 1: length(prdKP1)){
            prdLSP[ i] <- prdKP1[ i]-
                (P22[i, -i] %*% (yobs[ -i]- prdKP1[ -i])/ P22[i, i])
        }
    } else {
        P11 <- P22
        prdLSP <- prdKP1+ ((solve(P22) %*% P11 %*% (yobs- prdKP1)))
    }
    prdLSP
}
```

3.3.3 BLUP KP2

```
prdKP2 <- function(prdKP1, prdWXy= prdWXy, rho= rho, lab= lab, lsw= lsw,
                   yobs= yobs, power= power, order= order, tol= tol){
    if (power){
        W <- as(as_dgRMatrix_listw(lsw), "CsparseMatrix")
        GL <- as(powerWeights(W, rho= lab, order= order, tol= tol,
                              X= diag(length(prdWXy))), "matrix")
        GR <- as(powerWeights(W, rho= rho, order= order, tol= tol,
                              X= diag(length(prdWXy))), "matrix")
    } else {
        GL <- invIrW(lsw, rho) ; GR <- invIrW(lsw, lab)
    }
    sum.u <- GL %*% t(GL) ; sum.y <- GR %*% sum.u %*% t(GR)
    prdKP2 <- matrix(NA, ncol= 1, nrow= length(prdWXy))
    for (i in 1: length(prdKP2)){
        WM <- listw2mat(lsw)[i, ]
        rg <- (sum.u[i, ] %*% GR %*% WM)/ (WM %*% sum.y %*% WM)
        prdKP2[ i] <- prdWXy[ i]+ (rg %*% WM %*% (yobs- prdKP1))
    }
    prdKP2
}
```

3.3.4 BLUP KP3

```
prdKP3 <- function(prdKP1, prdWXy= prdWXy, rho= rho, lab= lab, lsw= lsw,
                   yobs= yobs, power= power, order= order, tol= tol){
    if (power){
        W <- as(as_dgRMatrix_listw(lsw), "CsparseMatrix")
        GL <- as(powerWeights(W, rho= lab, order= order, tol= tol,
                              X= diag(length(prdWXy))), "matrix")
        GR <- as(powerWeights(W, rho= rho, order= order, tol= tol,
                              X= diag(length(prdWXy))), "matrix")
```

```
    } else {
        GL <- invIrW(lsw, lab) ; GR <- invIrW(lsw, rho)
    }
    sum.u <- GL %*% t(GL) ; sum.y <- GR %*% sum.u %*% t(GR)
    prdKP3 <- matrix(NA, ncol= 1, nrow= length(prdWXy))
    for (i in 1: length(prdKP3)){
        rg <- sum.u[i, ] %*% GR[, -i] %*% solve(sum.y[-i, -i])
        prdKP3[ i] <- prdWXy[ i]+ (rg %*% (yobs[i ]- prdKP1[-i ]))
    }
    prdKP3
}
```

3.4 Predictors conditioned by hand

Try with

> rg ¡- sum.u[i,] %*% GR %*% solve(sum.y) prdKP3[i] ¡- prdWXy[i]+ (rg %*% (yobs- prdKP1))

4 How it works

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, for the inverse integrating directly the code from powerWeigths?

4.4 The predictors 1 for OS predictions

5 The `boston` example

5.1 The spatial samples

We use here the `boston` dataset, available from the `spdep` package (Bivand 2014). It consists in XX. From the whole sample, we draw randomly 211 observations from 250 drawn of a uniform distribution and dropping duplicated units. This allows to distinguish two sample: IS and OS.

```
library(spdep) ; data(boston) ; library(rgdal)
bst.sp <- readOGR(system.file("etc/shapes", package= "spdep"), "boston_tracts")
set.seed(84) ; cal <- unique(round(runif(250)* 500))
bst.sp$SMP <- "OS" ; bst.sp$SMP[cal] <- "IS"
ins.sp <- subset(bst.sp, bst.sp$SMP== "IS")
ots.sp <- subset(bst.sp, bst.sp$SMP== "OS")
cbind("ALL"= dim(bst.sp), "IN SMP"= dim(ins.sp), "OUT SMP"= dim(ots.sp))
```

```
OGR data source with driver: ESRI Shapefile
Source: "/home/jsay/R/x86_64-pc-linux-gnu-library/3.0/spdep/etc/shapes", layer: "boston_tracts"
with 506 features and 21 fields
Feature type: wkbPolygon with 2 dimensions
     ALL IN-SMP OUT-SMP
[1,] 506    211     295
[2,]  22     22      22
```

Spatial units without neighbors are kept in these spatial data.frames.

5.2 Neighborhood relationships

The following code use the function `nbsubset` described in Appendix XX and tangled in the file `spsubset.R`. As it is shown by the following Figure XX, this function allows to compute the `nb` equivalent to the spatial weight matrix W_I, W_O, W_{IO} and W_{OI}. The logic behind `nbsubset` is rather different than the classical `subset.nb` function because some absent units have to be referred by present units. So we do not change the dimension of the `nb` matrix, we put some 0 instead but this leads to some problems to compute the spatial weights after (see the commentaries in Appendix A.2).

```
source("spsubset.R")
bst.nb <- poly2nb(bst.sp) ; bst.lw <- nb2listw(bst.nb)
WII <- nbsubset(bst.nb, bst.sp$SMP== "IS", target= bst.sp$SMP== "IS")
WIO <- nbsubset(bst.nb, bst.sp$SMP== "IS", target= bst.sp$SMP== "OS")
WOI <- nbsubset(bst.nb, bst.sp$SMP== "OS", target= bst.sp$SMP== "IS")
WOO <- nbsubset(bst.nb, bst.sp$SMP== "OS", target= bst.sp$SMP== "OS")
par(mar= c(0, 0, 0, 0)) ; plot(bst.sp, lwd= 5)
plot(ots.sp, col= "grey70", add= TRUE, border= FALSE)
plot(ins.sp, col= "grey30", add= TRUE, border= FALSE)
```

```
plot(bst.nb, coordinates(bst.sp), add= T, col= "black" , pch= 20, cex= .2)
plot(WII, coordinates(bst.sp), add= T, col= "green", pch= 20, cex= .2)
plot(WIO, coordinates(bst.sp), add= T, col= "blue" , pch= 20, cex= .2)
plot(WOO, coordinates(bst.sp), add= T, col= "red"  , pch= 20, cex= .2)
legend("topleft", col= c("blue", "green", "red"), lwd= 1.5,
       c("PRD NBH", "IS NBH", "OS NBH"), bty= "n", cex= 1.25)
legend("right", fill= c("grey30", "grey70"),
       c("IN-SAMPLE", "OUT-OF-SAMPLE"), bty= "n", cex= 1.25)
```

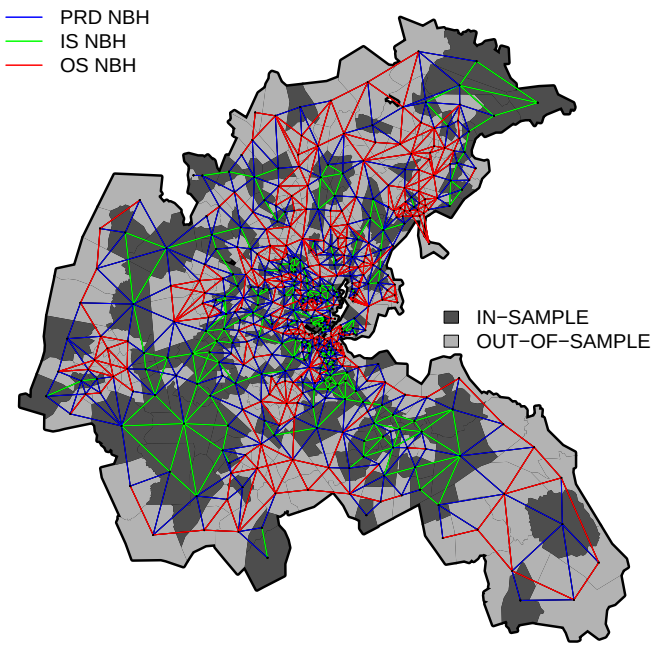

Figure 1: Neighboring relationships for the different Boston subsamples

From a first order neighboring point of view, the symmetry of the matrix produce a clear distinction between the spatial relationships. They are identified through different colors according to... The green links are of central interest for information transmission between units. Be careful about the island that present yellow links (see at the bottom right).

LES ILOTS "IS" NE PEUVENT PAS ETRE UTILISES POUR ESTIMER LE MODELE MAIS PEUVENT ETRE UTILE POUR PREDIRE

Il ne faut pas que je change les id et donc la dimension il faut simplement mettre de zéro? Et ensuite à partir d'autres subsets sur cette matrice on devrait obtenir la bonne dimension. Dans cette deuxième étape on devrait alors enlever les ilots.

Une autre possibilité serait de spécifier la matrice avec tout le monde, la traduire en W et sélectionner les lignes et les colonnes qui nous intéressent.

5.3 Estimating the models

We estimate on the IS sample the 6 spatial econometric models from the `spdep` package, with the classical specification presented in Bivand (2002). Before the estimation (and to be OK with the

computation of spatial weights) we have to drop some units without neighbors with the help of the `card` function.

```
## DROP OUT THE UNITS WITHOUT NEIGHBORS
ins.lw <- subset(bst.lw, card(WII)!= 0)
inz.sp <- subset(bst.sp, card(WII)!= 0)
## DEFINE THE MODEL SPECIFICATION
eqn.bst <- log(CMEDV)~ CRIM    + ZN + INDUS   + CHAS    + I(NOX^2)+
                       I(RM^2) + AGE+ log(DIS)+ log(RAD)+ TAX     +
                       PTRATIO + B  + log(LSTAT)
## ESTIMATE THE MODELS AND REPORT THE AICs
sem <- errorsarlm(eqn.bst, data= inz.sp, ins.lw, method= "Matrix")
sxm <- errorsarlm(eqn.bst, data= inz.sp,
                  ins.lw, etype= "emixed"   , method= "Matrix")
sar <- lagsarlm(  eqn.bst, data= inz.sp, ins.lw, method= "Matrix")
sdm <- lagsarlm(  eqn.bst, data= inz.sp,
                  ins.lw,  type= "mixed"    , method= "Matrix")
sac <- sacsarlm(  eqn.bst, data= inz.sp, ins.lw, method= "Matrix")
smc <- sacsarlm(  eqn.bst, data= inz.sp,
                  ins.lw,  type= "sacmixed" , method= "Matrix")
spmod <- list(sem= sem, sxm= sxm, sar= sar, sdm= sdm, sac= sac, smc= smc)
data.frame(lapply(spmod, AIC))
```

```
        sem       sxm       sar       sdm       sac       smc
1 -132.4525 -164.1509 -139.7872 -163.6526 -142.8058 -163.1881
```

The SXM is the more parsimonious in terms of AC

5.4 Predicting for in-sample

We next compute the RMSE for each model $\times$ predictor combination.

```
source("sppred.R") ; library(plyr)
rmse <- function(prd, mod) sqrt(mean(I(prd- mod$y)^2))
RMSE <- data.frame("X"=   ldply(spmod, function(x)
                            rmse(sppred(x, condset= "X"), x))$V1,
                   "XW"=  ldply(spmod, function(x)
                            rmse(sppred(x, condset= "XW"), x))$V1,
                   "KP1"= ldply(spmod, function(x) rmse(sppred(x), x))$V1,
                   "XWy"= ldply(spmod, function(x)
                            rmse(sppred(x, condset= "XWy"), x))$V1,
                   "LSP"= ldply(spmod, function(x)
                            rmse(sppred(x, blup= "LSP", loo= TRUE), x))$V1,
                   "KP2"= ldply(spmod, function(x)
                            rmse(sppred(x, blup= "KP2", loo= TRUE), x))$V1,
                   "KP3"= ldply(spmod, function(x)
                            rmse(sppred(x, blup= "KP3", loo= TRUE), x))$V1)
row.names(RMSE) <- names(spmod) ; round(RMSE, 4)
```

The results are intriguing, the fact that KP3 does not always perform best is fuzzy.

Table 2: The RMSE from IS sample for the 6 spatial models and 7 predictors

	X	XW	KP1	XWy	LSP	KP2	KP3
sem	0.2124	0.2124	0.2124	0.1449	0.137	0.185	0.2015
sxm	0.9832	0.1661	0.1661	0.1292	0.1221	0.1464	0.1539
sar	1.3541	1.3541	0.1846	0.1504	0.1409	0.1423	0.1119
sdm	1.5871	1.5536	0.1627	0.1306	0.1231	0.128	0.1127
sac	0.941	0.941	0.1899	0.3152	0.1373	0.327	0.302
smc	1.9819	0.692	0.17	0.4797	0.1208	0.4632	0.4714

5.5 Recovering impact measures

5.6 Recovering trend and signal terms

6 Summary

6.1 Changes relative to `predict.sarlm`

- Implement predictions for SARAR and Mixed SARAR models from respectively `sac` and `sacmixed` classes.
- Compute BLUP and almost BLUP spatial predictors
- About the in-sample / out of sample structure (`newdata`)
- About the distinction between trend and signal
- The simplification of the in-sample predictions

6.2 About the intercept

We change the scan of the intercept, in particular in presence of WX in the regression. If W is row standardized, we have to drop the intercept to avoid collinearity. The initial function add the constant at the end of the computations, we only drop the intercept in the presence of WX.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, *4*. Springer, Boston.

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*, *5*. Pion, London.

Cliff, A. D. and J. K. Ord (1981). *Spatial Processes, Models & Applications*. Pion, London.

Cressie, N. A. and N. A. Cassie (1993). *Statistics for Spatial Data*, *900*. Wiley New York.

A Appendix

A.1 Subset function for `nb` objects

The following function `nbsubset` is a first step to decompose the spatial weight matrix, in order to compute ex-sample predictions. Based on the function `subset.nb` from the package `spdep`, the `subset` argument is replaced by two new arguments: `focal` and `target`. In raw matrix terms, `focal` corresponds to rows and `target` corresponds to columns. As `subset`, these arguments have to be logical vector of the same size that the `nb` object. Note that we can recover the same `nb` relations from the `subset.nb` by putting the same logical vector both for `focal` and `target`. See XX for an example of use.

```
nbsubset <- function (x, focal= rep(TRUE, length(x)),
                      target= rep(TRUE, length(x)), ...){
    if (!inherits(x, "nb"))
        stop("not a neighbours list")
    if (!is.logical(focal))
        stop("subset not a logical vector")
    if (!is.logical(target))
        stop("target not a logical vector")
    n <- length(x)
    if (n != length(focal))
        stop("neighours list and subset focal different lengths")
    if (n != length(target))
        stop("neighours list and subset target different lengths")
    tgt.ids <- seq(1, n)[ target]
    x <- sym.attr.nb(x)
    xattrs <- names(attributes(x))
    z <- ifelse(focal, x, 0L)
    for (i in seq(along= 1: n)[ focal]){
        zi <- z[[ i]]
        zn <- zi[zi %in% tgt.ids]
        if (length(zn)== 0) z[[ i]] <- 0L
        else z[[ i]] <- sort(unique(zn))
    }
    attr(z, "region.id") <-attr(x, "region.id")
    for (i in 1:length(xattrs)) {
        if (xattrs[i] != "region.id")
            attr(z, xattrs[i]) <- attr(x, xattrs[i])
    }
    z <- sym.attr.nb(z)
    z
}
```

A.2 Subset function for `listw` objects

Below the `lwsubset` function, that is the symmetric of previous `nbsubset` function for the `listw` objects. The code is not finished because of the function `nb2listw` that cannot admits units without neighbors and the presence of rectangular `nb` relationships from `nbsubset`.

```
lwsubset <- function (x, focal= rep(TRUE, length(x)),
                      target= rep(TRUE, length(x)), zero.policy = NULL, ...){
    if (!inherits(x, "listw"))
        stop("not a weights list")
    if (is.null(zero.policy))
        zero.policy <- get("zeroPolicy", envir = .spdepOptions)
    stopifnot(is.logical(zero.policy))
    if (!is.logical(focal))
        stop("focal not a logical vector")
    if (!is.logical(target))
        stop("target not a logical vector")
    nb <- x$neighbours
    vlist <- x$weights
    if (attr(vlist, "mode") != "binary")
        stop("Not yet able to subset general weights lists")
    style <- x$style
    n <- length(nb)
    if (n != length(focal))
        stop("neighbours list and focal vector different lengths")
    if (n != length(target))
        stop("neighbours list and target vector different lengths")
    subnb <- nbsubset(nb, focal, target)
    sublistw <- nb2listw(neighbours= subnb, glist= NULL,
                         style= style, zero.policy= zero.policy)
    sublistw
}
```