About predictions in spatial SAR models: optimal and almost optimal strategies

M. Goulard *
INRA, DYNAFOR
T. Laurent†
Toulouse School of Economics (GREMAQ), 21 allée de Brienne 31042 Toulouse, FRANCE
C.Thomas-Agnan‡
Toulouse School of Economics (GREMAQ), 21 allée de Brienne 31042 Toulouse, FRANCE

June 1, 2009

Abstract

We address the problem of prediction in a classical spatial simultaneous autoregressive model. The optimality of prediction formulas in non-spatial regression models is not immediately transposable to the framework of all spatial models. In the geostatistical literature, much attention has been devoted to this topic, with the development of the Best Linear Unbiased Prediction formulas. In contrast, in the spatial econometric literature, the classically used formulas are not always supported by theoretical optimality results. From the methodological point of view, we explore the limits of the extension of BLUP formulas in the context of the SAR models. From an equivalence between SAR and CAR we develop the best prediction formula and propose a more tractable "almost best" alternative. We consider the case of in-sample prediction as well as out-of-sample prediction. From an empirical perspective, we present data-based simulations to compare the efficiency of the most frequent formulas with the best and almost best predictions.

Very preliminary version

1 Introduction

Whereas prediction is a basic concern in geostatistics (Cressie, 1993), it has not been paid as much attention in the econometrics literature. Bivand (2002) recognizes the importance of the question: "Prediction for new data ... is a challenge

*e-mail:goulard@toulouse.inra.fr
†e-mail:
‡e-mail: cthomas@cict.fr

for legacy spatial econometric models, raising the question of what a BLUP (best linear prediction) would look like". Even though the components of the following arguments can mostly be found in the literature, nobody seem to have put them together in this fashion and we will demonstrate that the formulas used for prediction in the main softwares can thus be improved upon substantially. The reason is the aparent computational burden of these formulae. An example of application of prediction for new data is the following. Until 1999, the French population census was exhaustive and realized by the French statistical institute (INSEE) approximately every ten years. Since 2004, this exhaustive census has been replaced by a census survey which consists in annual samples and which delivers an up-to-date information. In particular, the communes with less than 10000 inhabitants at the 1999 census (called *small communes*) are sampled exhaustively every five year at the rate of one fifth per year. The sampling design of these small communes is stratified by region and inside each region, the small communes are partitioned into five rotational groups by using a balanced sample design taking into account some auxiliary socio-economics variables given by the 1999 census. Between 2004 and 2009, polling organizations needed an estimate of the population for all the small communes and of its evolution since the previous complete census of 1999. The population of all the small communes would not be delivered by the INSEE before 2009 but data sets containing the population of the two first rotational groups, corresponding to 2004 and 2005, were already known and could be used to predict the population of the other three rotational groups. In that case, out-of-sample prediction formulae were necessary for spatial models (see Lesne et al., 2008).

2 A bridge between CAR and SAR

2.1 The CAR-SAR debate

The geostatistics literature favors conditional autoregressive models (CAR) over simultaneous autoregressive models (SAR) whereas the reverse is true in the econometrics literature. Let us briefly review the arguments on both sides. The stumbling blocks of this debate are the identifiability problems, the inconsistency of OLS estimation and the modeling of spillovers.

The geostatistics literature raises identifiability issues but we would like to point out that they are only relevant to models on infinite networks: in the case of a finite network there is no problem for identification of the coefficients of a SAR model. On an infinite network equivalent to $\mathbb{Z}^d$ identification of a CAR model is always possible but identification of a SAR cannot be done without some constraints if $d > 2$. Stochastic models on an infinite network have been studied almost exhaustively by Guyon (1995). He points out the differences between $\mathbb{Z}$ and $\mathbb{Z}^d$ with $d > 2$ for the simultaneous and conditional formulation. In one dimension, any AR model admits a finite causal equivalent representation and the set of Conditional AR models and the set of AR models are equal. These two properties are not true when $d > 2$ and in particular the set of Conditional AR models is strictly larger than the set of AR

models. If a Conditional AR model can be identified by properties of its covariance, this is not the case for the AR and ARMA models if $d > 2$ because the set of models with rational spectra is strictly larger than the set of ARMA models in that case.

Concerning the inconsistency of OLS, the method of Ordinary Least Squares (OLS) is consistent for a CAR model but it is not in the case of a SAR model. For an infinite network, the consequence of the non-equivalence between AR causal representation and AR representation also results in the non-consistency of solvin he solution of the Yule-Walker equations to estimate a non-causal SAR model when $d > 2$. In the case of a finite network the inconsistency of OLS remains for a SAR when it is consistent for a CAR model. This property comes from the fact that in the case of a SAR (centered and without covariate here) :

$$Y_u = \sum_v a_v Y_v + \varepsilon_u$$

ε_u is independent of ε_v when $v \neq u$ but it is correlated with Y_v $v \neq u$. In the case of a CAR, the residual ε_u is correlated with ε_v but uncorrelated with Y_v $v \neq u$. In fact the OLS procedure is just a pseudo-likelihood procedure in the Gaussian case.

The spillover effect that economists would like to model is the fact that the characteristics of a spatial unit may be influenced by random shocks on neighboring spatial units. The CAR model does not model this effect since in this model the partial derivatives of the expected value of the dependent variable on a given spatial unit is independent of the values of the exogenous variables measured on neighboring (or any other) spatial units. For SAR models, Pace and LeSage (2006) define measures of this spillover effect through the direct, indirect and total impact.

2.2 CAR representation of a SAR model

On a finite network, since the set of CAR model and SAR model are equal, we can always transform a SAR model into a CAR model. As we know the expression of the best predictor in the CAR case, expressing the SAR model into its CAR representation yields a very simple and efficient way to define a prediction. In the general gaussian linear model the dependent variable Y

$$Y \sim \mathcal{N}(\mu, \Sigma), \tag{1}$$

has a multivariate Gaussian distribution with mean $\mu \equiv (\mu_1, \ldots, \mu_n)'$ and $(n \times n)$ variance-covariance matrix Σ. In the case of the CAR model, an additional assumption is introduced in (1) namely that Σ takes on a special form:

$$\Sigma = (I - C)^{-1} M, \tag{2}$$

where $C = (c_{ij})$, $c_{ii} = 0$; $i = 1, \ldots, n$, and $M = \text{diag}(\tau_1^2, \ldots, \tau_n^2)$ are parameters in the conditional distributions,

$$Y_i | (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n) \sim \mathcal{N}(\mu_i + \sum_{j=1}^{n} c_{ij}(Y_j - \mu_j), \tau_i^2), \tag{3}$$

for $i = 1, \ldots, n$. The $\{c_{ij}\}$ are spatial-dependence parameters and the $\{\tau_i^2\}$ are heteroskedasticity parameters that together satisfy (Besag, 1974):

$$\begin{aligned} &M^{-1}C \ \text{ is symmetric (symm.)}\,; \\ &M^{-1}(I-C) \ \text{ is positive-definite (p.d.)}\,. \end{aligned} \tag{4}$$

The covariates are included in the model linearly through $\mu = X\beta$, where β is a $(p \times 1)$ vector of regression parameters; $p < n$. It is usual to assume that the conditional variances $\tau_1^2, \ldots, \tau_n^2$ are known up to a normalizing constant ; that is, $M = \Phi\tau^2$, where $\Phi \equiv \text{diag}(\phi_1, \ldots, \phi_n)$ is a known $(n \times n)$ diagonal matrix. Usually C is a function $C(\gamma)$ of a $(q \times 1)$ vector of spatial-dependence parameters γ.
In conclusion, the CAR model is

$$Y \ \sim \ \mathcal{N}(X\beta, (I - C(\gamma))^{-1}\Phi\tau^2). \tag{5}$$

From (3), it is easy to compute the conditional expectation of Y at a given site given its values at neighboring sites by

$$\mathbb{E}(Y_i|(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)) = \mu_i + \sum_{j=1}^{n} c_{ij}(Y_j - \mu_j), \tag{6}$$

or equivalently in terms of the elements of the precision matrix $Q = \Sigma^{-1}$

$$\mathbb{E}(Y_i|(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)) = \mu_i - \sum_{j=1}^{n} \frac{q_{ij}}{q_{ii}}(Y_j - \mu_j). \tag{7}$$

Given a spatial weight matrix W, the gaussian SAR model can be written

$$Y = \rho WY + X\beta + \epsilon \tag{8}$$

or equivalently

$$Y = (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}\epsilon \tag{9}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. The ensuing covariance structure is then given by

$$\Sigma = ((I - \rho W')(I - \rho W))^{-1}\sigma^2 \tag{10}$$

Note that the corresponding precision matrix is then given by

$$Q = \Sigma^{-1} = \frac{1}{\sigma^2}((I - \rho W')(I - \rho W)) \tag{11}$$

Table 2.2 summarizes the correspondence between these two models and shows how to make the correspondence.

	Mean	Variance matrix	Precision matrix
CAR	$X\beta$	$(I - C(\gamma))^{-1}\Phi\tau^2$	$\frac{1}{\tau^2}(I - C(\gamma))\Phi^{-1}$
SAR	$(I - \rho W)^{-1}X\beta$	$((I - \rho W')(I - \rho W))^{-1}\sigma^2$	$\frac{1}{\sigma^2}(I - \rho W')(I - \rho W)$

3 Best prediction and competitors

In the general linear gaussian model when $\mu = X\beta$, the Gauss Markov theorem establishes an optimality of the predictor $\hat{Y}^T = X\hat{\beta}$ where $\hat{\beta}$ is the ordinary least squares estimator.
In the CAR model framework, it is usual to derive from (6) the following best linear unbiased predictor (see for example Guyon and Gaetan(2008))

$$\hat{Y} = \hat{\mu} - \text{Diag}(Q)^{-1}[Q](Y - \hat{\mu}) \tag{12}$$

where $\text{Diag}(Q)$ denotes the diagonal matrix containing the diagonal of Q and $[Q] = Q - \text{Diag}(Q)$.
Using the correspondence summarized in Table 2.2, it is then easy to write the best predictor in a SAR model as

$$\hat{Y}_{BP} = (I - \rho W)^{-1}X\hat{\beta} - \text{Diag}(Q)^{-1}[Q](Y - (I - \rho W)^{-1}X\hat{\beta}), \tag{13}$$

where $Q = \frac{1}{\sigma^2}(I - \rho W')(I - \rho W)$.
Similar arguments were used in LeSage and Pace (2004) but in a different perspective.

We are mainly concerned by two types of prediction problems: the in-sample and out-of-sample cases. In the in-sample prediction problem, we have n spatial units for which we observe the dependent variable Y as well as the independent variables X and we want to predict the value of Y at the observed sites after fitting the model (even though we know their values)which is the same as computing the fitted value of Y. These predicted values can be used to compute a goodness of fit crietrion. In the out-of-sample case, we have two types of spatial units: the in-sample units for which we observe the dependent variable Y_O as well as the independent variable X_0 and the out-of-sample units for which we only observe the independent variable X_0 and we want to predict the variable Y_O from the knowledge of Y_S, X_S and X_O. These two situations are illustrated in Figures 1 and 2.
In the case of in-sample prediction, we derive from (13) the predictor $\hat{Y}^{BP}$

$$\hat{Y}^{BP} = \hat{Y}^{TC} - Diag(Q)^{-1} \times [Q] \times (y - \hat{Y}^{TC}) \tag{14}$$

with $Q = I - \hat{\rho}(W' + W) + \hat{\rho}^2 W'W$.
We compare this predictor to the one $\hat{Y}^{TN}$ derived from the trend-signal-noise representation of the data introduced in Haining (1990) and detailed in Bivand (2004). Note that this predictor does not have any optimality property.

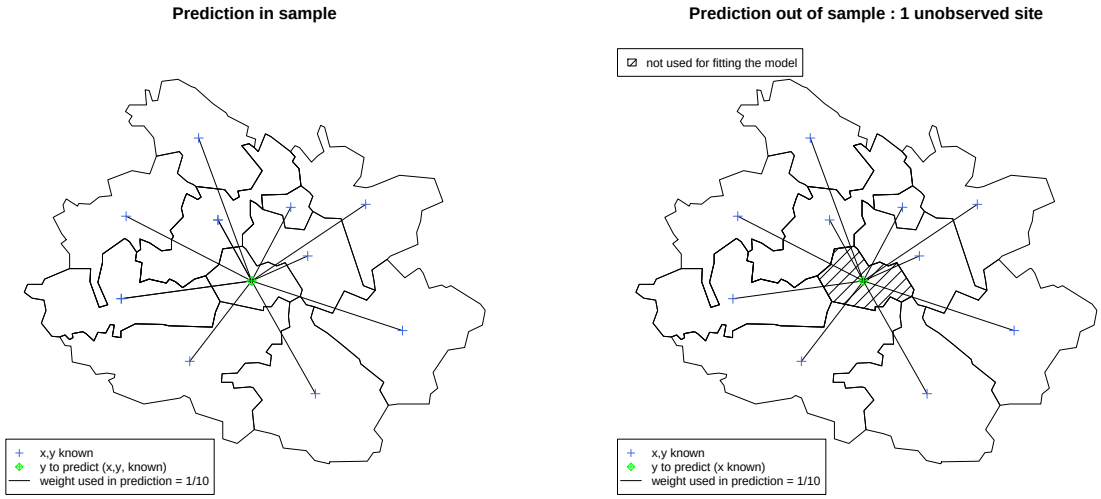

Figure 1: In and out of sample single prediction problem

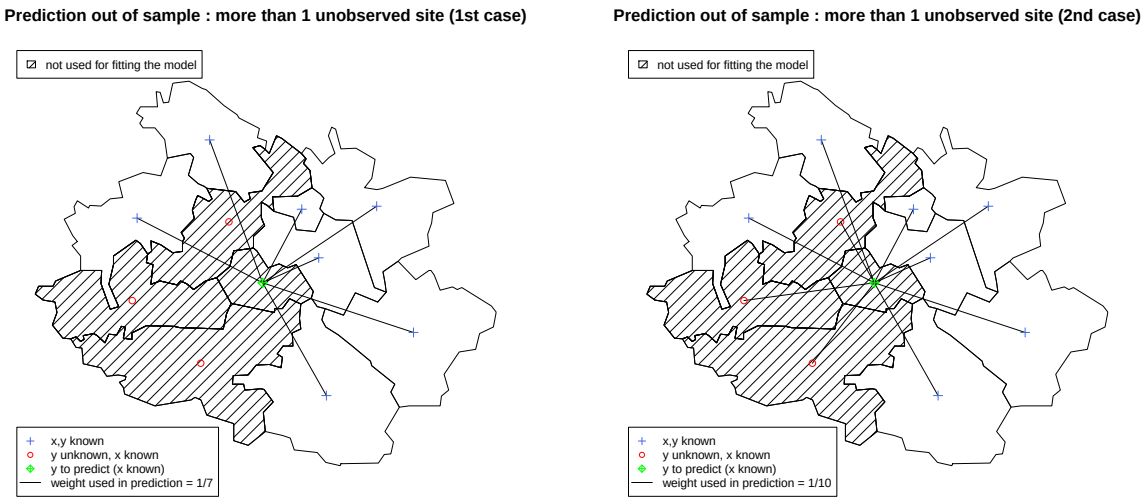

Figure 2: Out of sample multiple predictions

$$\hat{Y}^{TN} = X\hat{\beta} + \hat{\rho}WY, \tag{15}$$

where y is the vector of observed dependent variables which is available in the in-sample case.

We also compare them with the "trend-corrected" predictor $\hat{Y}^{TC}$ obtained by just using the estimated mean, which is the one most often used by practitioners.

$$\hat{Y}^{TC} = (I - \hat{\rho}W)^{-1}X\hat{\beta} \tag{16}$$

In the case of out-of-sample prediction, we first need to partition the weight matrix W into

$$W = \begin{pmatrix} W_S & W_{SO} \\ W_{OS} & W_O \end{pmatrix} \tag{17}$$

where W_S is the $(n-p)\times(n-p)$ submatrix corresponding to the neighborhood structure of the $(n-p)$ in-sample sites, W_O the $p\times p$ submatrix corresponding to the neighborhood structure of the out-of-sample sites, W_{OS} the $p\times(n-p)$ submatrix indicating the neighbors of the out-of-sample units among the in-sample units and W_{SO} the $(n-p)\times p$ submatrix indicating the neighbors of the in-sample units among the out-of-sample units.

We assume that there is a unique model corresponding to (8) with a normalized matrix W for the n observations and that the model driving the subset X_S, Y_S has the same expression but using submatrix W_S renormalized.

For the best prediction, it clear that if $(I-\hat{\rho}W) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_{n-p} - \hat{\rho}W_S & -\hat{\rho}W_{SO} \\ -\hat{\rho}W_{OS} & I_p - \hat{\rho}W_O \end{pmatrix}$ the best predictor $\hat{Y}_O^{BP}$ of the out-of-sample units is given by

$$\hat{Y}_O^{BP} = \hat{Y}_O^{TC} - Q_O^{-1}Q_{OS} \times (Y_S - \hat{Y}_S^{TC}) \tag{18}$$

$$\text{with } Q = I - \hat{\rho}(W' + W) + \hat{\rho}^2 W'W = \begin{pmatrix} Q_S & Q_{SO} \\ Q_{OS} & Q_O \end{pmatrix} \tag{19}$$

The computational difficulty of this formula lies in the inversion of the matrix Q_O and hence increases with the number of out-of-sample units.

It is clear that the trend-signal-noise predictor $\hat{Y}^{TN}$ cannot be defined in that case since it requires the value of Y_0 which is unobserved. However one can see that the trend-corrected strategy can be applied here because the diagonal of the weight matrix contains zeros and we have

$$\hat{Y}^{TC} = (I - \hat{\rho}W)^{-1}\hat{\beta}X = \begin{pmatrix} \hat{Y}_S^{TC} \\ \hat{Y}_O^{TC} \end{pmatrix} \tag{20}$$

and that

$$\hat{Y}_O^{TC} = -(D - CA^{-1}B)^{-1}CA^{-1}X_S\hat{\beta} + (D - CA^{-1}B)^{-1}X_O\hat{\beta} \tag{21}$$

4 Almost best prediction

In the out-of-sample prediction problem, the complexity of the computations is very different depending upon the number of out-of-sample units to predict. In the case of a single out-of-sample unit to predict, the formulas are quite easy to implement, even for the best strategy. When formulas will be applied simultaneously to the p out-of-sample units we will call the strategy M1. Taking advantage of this fact, if p out-of-sample units have to be predicted, we introduce a new strategy M2 consisting in applying the single "out-of-sample unit to predict" formula to each of the out-of-sample unit separately, ignoring at this stage the remaining $p-1$ units. Note that strategy M2 can be applied to the trend corrected $\hat{Y}^{TC}$ and best $\hat{Y}^{BP}$ predictors but not to the trend-signal-noise one $\hat{Y}^{TS}$.

5 Comparing the predictors

5.1 Simulation framework

In order to compare the different predictors, we design a simulation study. We use the Midi-Pyrénées region divided into $n = 283$ cantons for our study region. We construct a weight matrix W using the 10 nearest neighbors scheme (distance is based on the distance between centroids of the cantons).

We simulate three explanatory variables as follows:

- $X_1 \sim \mathcal{N}(15, 3)$
- $X_2 \sim \mathcal{B}(100, 0.45)/100$
- $X_3 \sim \log(\mathcal{U}_{[0,283]})$

We use the following spatial autoregressive regression model to generate the dependent variable

$$Y = (I - \rho W)^{-1}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon) \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{22}$$

The vector of parameters β is fixed to the value $\beta = (0, 1/4, 6, 1)$ whereas ρ and σ take a range of values. Some draws are mapped in Figures 3 and 4.

5.2 In sample comparison

For each choice of ρ and σ, we draw 500 samples of the model and we compute the maximum likelihood estimates of the parameters and the corresponding predictions. We use the total mean square error of prediction $MSE_k = \frac{1}{n}\sum_i^n (y_i - Y_i^k)^2$ for each method $k = TS, TC, BP$ to compare the quality of the predictors. Note that this criterion includes the statistical error due to parameter estimation.

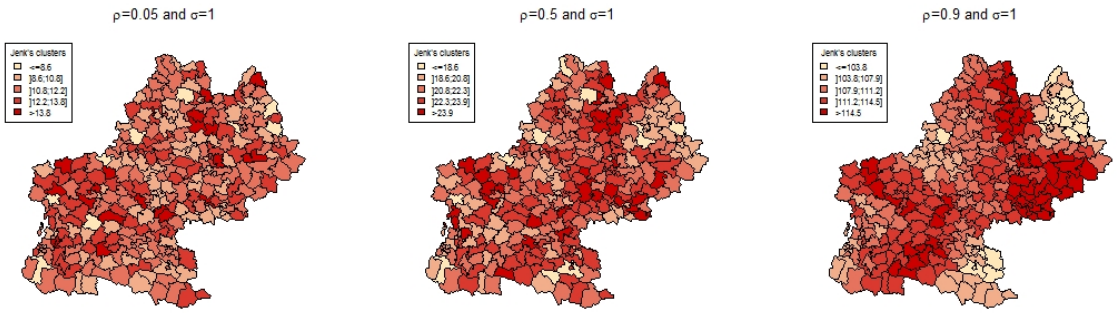

Figure 3: Map of simulated Y for different ρ and $\sigma = 1$

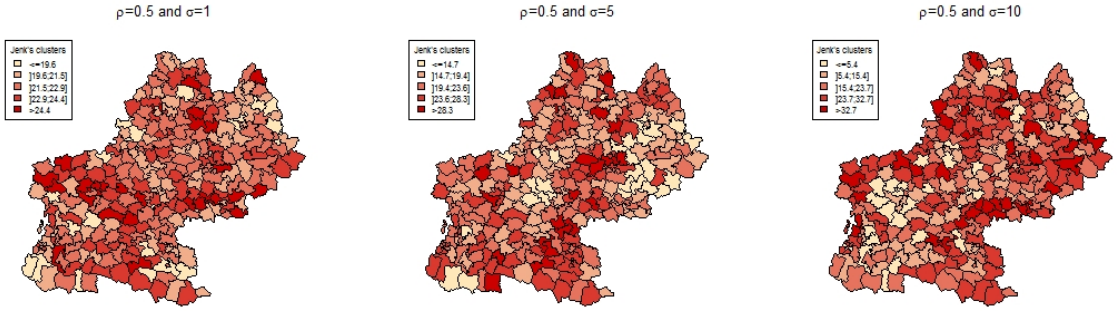

Figure 4: Map of simulated Y for different σ and $\rho = 0.5$

5.2.1 Results as ρ varies for fixed σ

Table 5.2.1 reports the mean and standard error of the mean square error of prediction for $\sigma = 1$ and $\rho = (0.05, 0.2, 0.35, 0.65, 0.8, 0.9)$. We see that the variances are stable across the different ρ. The means tend to increase as ρ increases for formula TC, to be stable for formula TS and to decrease for formula BP. Figure 5 displays

	$\bar{MSE}_{TS}$	$se(MSE_{TS})$	$\bar{MSE}_{TC}$	$se(MSE_{TC})$	$\bar{MSE}_{BP}$	$se(MSE_{BP})$
$\rho = 0.05$	0.9720	0.0833	0.9754	0.0838	0.9707	0.0832
$\rho = 0.2$	0.9884	0.0835	1.0006	0.0852	0.9850	0.0832
$\rho = 0.35$	0.9756	0.0841	1.0192	0.0896	0.9646	0.0847
$\rho = 0.5$	0.9814	0.0803	1.0890	0.1039	0.9597	0.0799
$\rho = 0.65$	0.9883	0.0799	1.2531	0.1450	0.9494	0.0790
$\rho = 0.8$	0.9871	0.0848	1.6571	0.2738	0.9308	0.0844
$\rho = 0.9$	0.9878	0.0812	2.8981	0.9635	0.9152	0.0784

the whole distribution of these mean square errors and we see that formula TC performs very poorly whereas formula TS is almost as good as formula BP.

5.2.2 Results as σ varies for fixed ρ

Table 5.2.2 reports the mean and standard error of the mean square error of prediction for $\rho = 0.5$ and $\sigma = (1, 5, 10, 15)$ and shows a stability of the results across the variance levels. The same hierarchy between the three estimators is observed.

	$\bar{MEQ}_{TS}$	$se(MSE_{TS})$	$\bar{MSE}_{TC}$	$se(MSE_{TC})$	$\bar{MSE}_{BP}$	$se(MEQ_{BP})$
$\sigma = 1$	0.9814	0.0803	1.0890	0.1039	0.9597	0.0799
$\sigma = 5$	0.9840	0.0854	1.0948	0.1083	0.9612	0.0855
$\sigma = 10$	0.9863	0.0864	1.0988	0.1092	0.9632	0.0853
$\sigma = 15$	0.9847	0.0844	1.0946	0.1045	0.9616	0.0837

5.3 Out of sample comparison

We now consider the out-of-sample prediction problem. To evaluate the performance of the different predictors, we use the same model as before to generate the samples and compute the predictive mean-square error score as follows. The number of replications is smaller than before (100 instead of 500) due to the length of the ensuing computations. We distinguish between the case when there is only one point to predict and the case of several points.

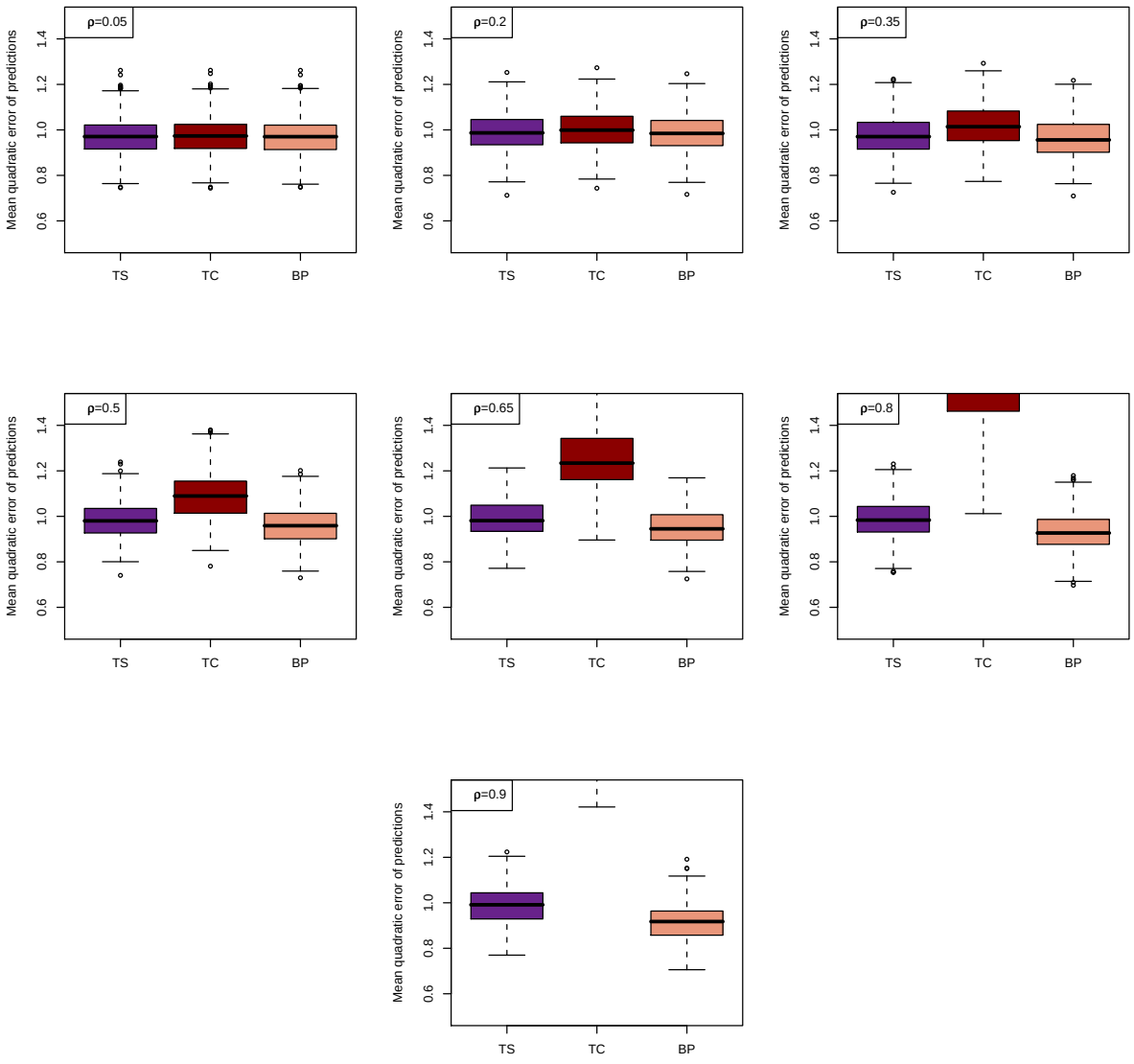

Figure 5: Boxplots of the mean square error of prediction for different values of ρ and $\sigma = 1$

5.4 Case of a single prediction

For each spatial unit i, we fit a sub-model of model (22) on the sample deprived from sample unit i. Note that there is a unique weight matrix W for the n spatial unit and we use the submatrix of dimension $n-1 \times n-1$ obtained by deleting row and column i from W for fitting the sub-model. We then predict the value of Y_i from the knowledge of the corresponding explanatory variables using the formulas of section 3. The predictive mean-square error score is then $PMSE_k = \frac{1}{n}\sum_i^n (y_i - Y_i^k)^2$ for each method $k = TS, TC, BP$ Note that this score is comparable to a leave-one-out cross-validation score in a model selection procedure. Here we use it in a different perspective to evaluate the average ability of the model to predict a unique point.

5.4.1 Results as ρ varies for fixed σ

Table 5.4.1 reports the empirical mean and standard error of the PMSE criterion for $\sigma = 1$ and $\rho = (0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.9)$. Figure **??** summarizes the boxplots of the corresponding criterion across replications for each set of parameters.

The mean and standard deviations of the PMSE criterion are stable across the values of the parameter ρ, for formula TS whereas they increase for formula TC and decrease for formula BP. Formula TC becomes very bad for large values of ρ and the gap between TS and BP gets larger as ρ increases. The efficiency of formula TS with respect to formula BP decreases to 0.953 for $\rho = 0.9$ and that of formula TC goes down as low as 0.2.

	$\bar{PMSE}_{TS}$	$se(PMSE_{TS})$	$\bar{PMSE}_{TC}$	$se(PMSE_{TC})$	$\bar{PMSE}_{BP}$	$se(PMSE_{BP})$
$\rho = 0.05$	1.0152	0.0862	1.0148	0.0857	1.0177	0.0871
$\rho = 0.2$	1.0061	0.0838	1.0182	0.0859	1.0046	0.0856
$\rho = 0.35$	1.0067	0.0798	1.0529	0.0891	0.9970	0.0806
$\rho = 0.5$	1.0251	0.0878	1.1445	0.1152	1.0029	0.0901
$\rho = 0.65$	1.0132	0.0842	1.3096	0.1576	0.9764	0.0855
$\rho = 0.8$	1.0011	0.0767	1.9147	0.3801	0.9514	0.0803
$\rho = 0.9$	1.0066	0.0872	5.1272	1.4629	0.9594	0.0971

5.4.2 Results as σ varies for fixed ρ

Table 5.4.2 reports the empirical mean and standard error of the PMSE criterion for $\rho = 0.5$ and $\sigma = (1, 5, 10, 15)$. Figure 7 summarizes the boxplots of the corresponding criterion across replications for each set of parameters. The efficiency of formula TS with respect to formula BP is stable no matter σ and is around 0.98.

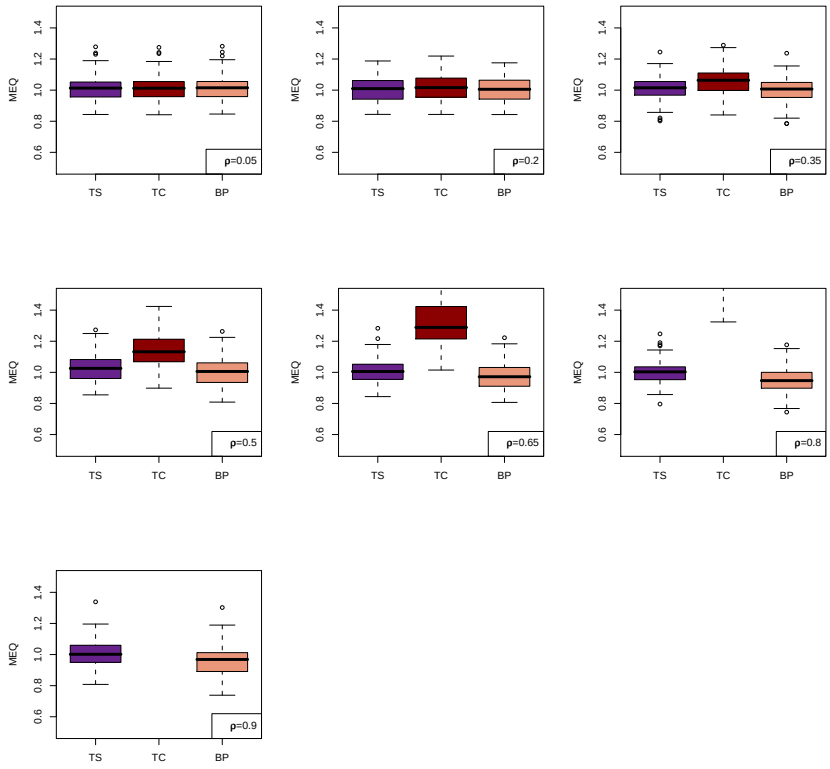

Figure 6: Parallel boxplots of PMSE criterion across replications for $\sigma = 1$ and different values of ρ.

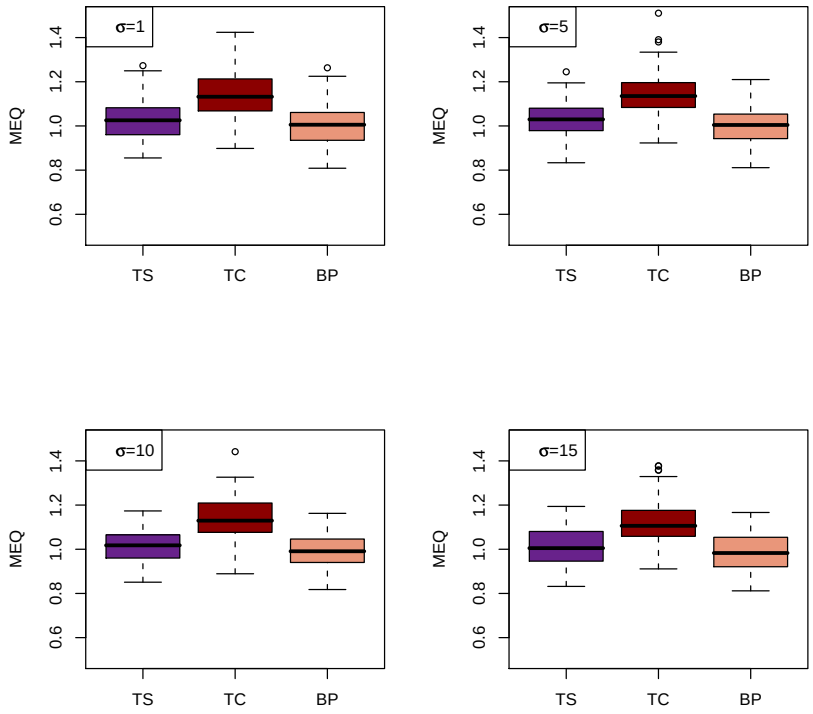

Figure 7: Parallel boxplots of PMSE criterion across replications for $\rho = 0.51$ and different values of σ

	$\bar{PMSE}_{TS}$	$se(PMSE_{TS})$	$\bar{PMSE}_{TC}$	$se(PMSE_{TC})$	$\bar{PMSE}_{BP}$	$se(PMSE_{BP})$
$\sigma = 1$	1.0251	0.0878	1.1445	0.1152	1.0029	0.0901
$\sigma = 5$	1.0276	0.0815	1.1401	0.1035	1.0029	0.0814
$\sigma = 10$	1.0156	0.0737	1.1353	0.0986	0.9907	0.0757
$\sigma = 15$	1.0113	0.0880	1.1149	0.1077	0.9889	0.0876

5.5 Case of simultaneous predictions

We consider two situations: in the first situation, the out-of-sample units are scattered in the area of interest whereas in the second case, the out-of-sample units are aggregated in some areas. The corresponding configurations of out-of-sample units is shown in Figures 5.5 and 5.5. In this last section, we apologize for the missing standard errors in the tables: there was an error in their computation that we did not have time to fix before the deadline but that will be fixed during the month of june for the next version of the preprint.

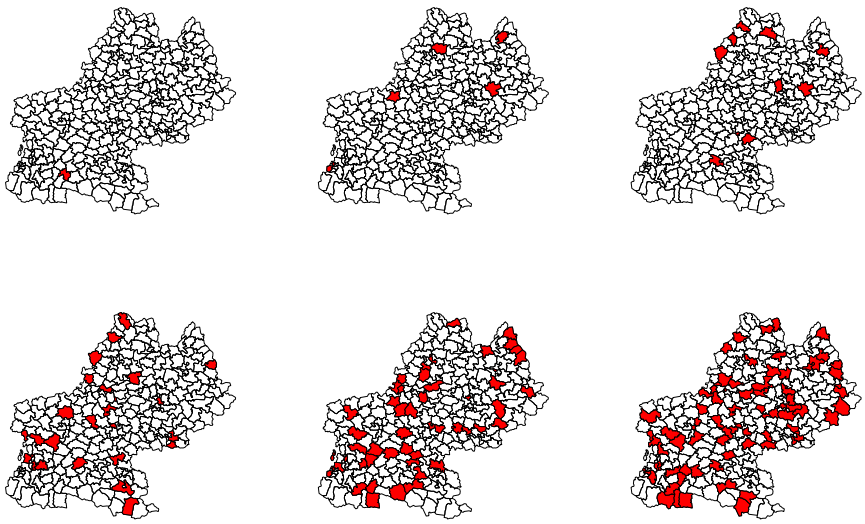

Figure 8: Configuration of out-of-sample units in the scattered case

5.5.1 Case of scattered out-of-sample units

In the case of scattered out-of-sample units, we compare formulas TS, TC and BP in their site-by-site version. We see that the trend-signal-noise predictor is around 98 per cent efficient, that strategies M1 and M2 have the same efficiency for the best predictor. The trend-corrected predictor is less efficient (from 80 to 90 pre cent efficency).

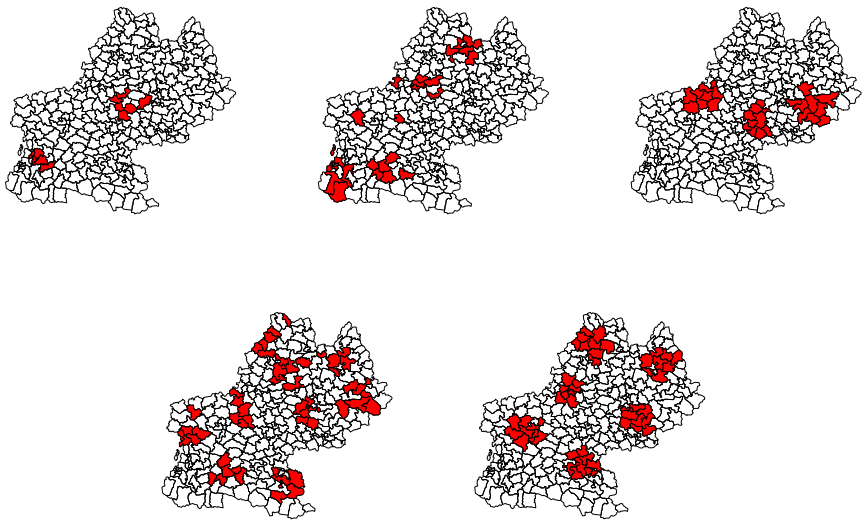

Figure 9: Configuration of out-of-sample units in the aggregated case

We suspect that the number of out-of-sample units in the neighborhood of a prediction-point should influence the quality of prediction. Figure 5.5.1 shows the distribution of the number of out-of-sample units in the neighborhood of prediction-points in the above scattered configurations.

5.5.2 Case of aggregated out-of-sample units

For the five configurations with aggregated out-of-sample units, Figure 5.5.2 presents the distribution of the number of out-of-sample units in the neighborhood of prediction-points. The following five tables display the results for configurations 1 through 5. As expected, the discrepancy between best and almost best strategy is larger in the aggregated case, decreasing as low as 60 per cent efficiency. The higher the number of out-of-sample neighbors for aa given out-of-sample points, the lower the efficiency but this effect is not as strong as the differences between formulae.

6 Conclusion

From these simulations, we conclude that, at least in the case of this particular model, the performance of the almost best predictor is almost as good as the one of the best predictor and it is much simpler to compute. We also see that the performance of the trend-signal-noise predictor is not so bad and this one is also very easy to compute. Of course, these results should also be confronted to a larger variety of parameter values and the comparison with bayesian predictors should be included. A mathematical investigation od the relative efficiency of the almost best

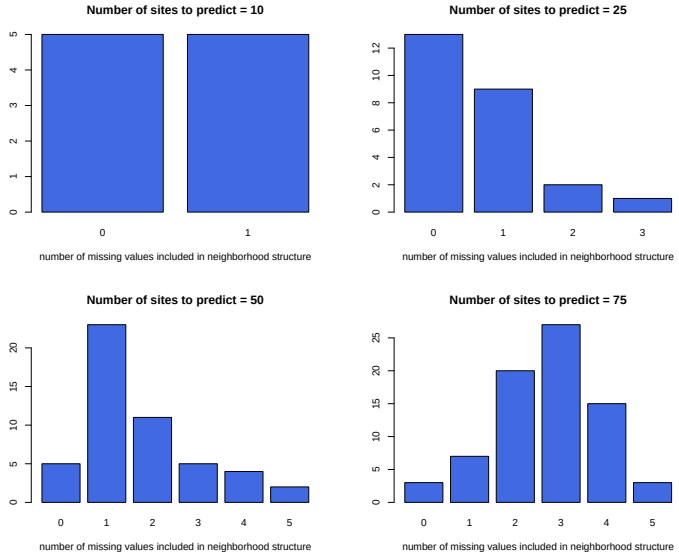

Figure 10: Distribution of the number of out-of-sample units in the neighborhood of prediction-points for the scattered configurations

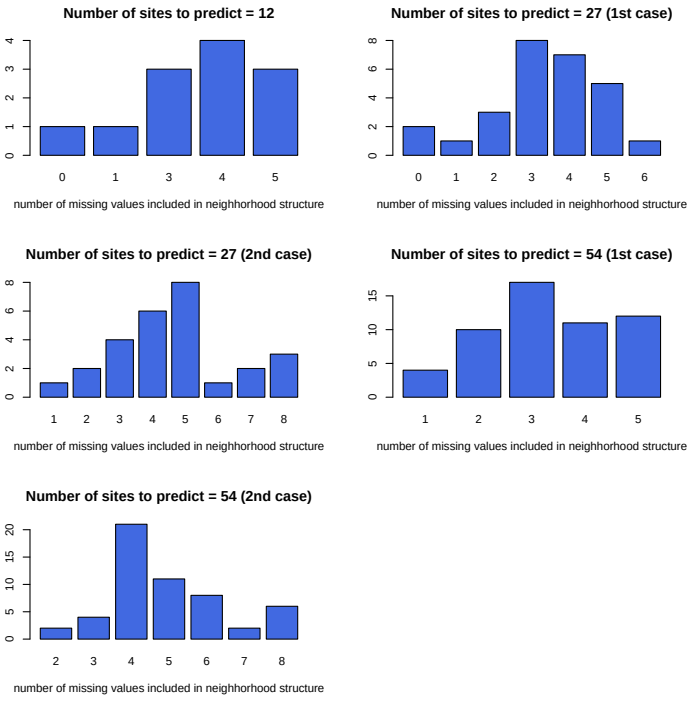

Figure 11: Distribution of the number of out-of-sample units in the neighborhood of prediction-points for the aggregated configurations

Scattered sample units case

	$\widehat{PMSE}_{TS}$	$se(PMSE_{TS})$	$\widehat{PMSE}_{TC}$		$se(PMSE_{TC})$		$\widehat{PMSE}_{BP}$		$se(PMSE_{BP})$	
	M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
$p = 1$	1.19		1.38	1.38			1.10	1.10		
$p = 5$	1.02		1.12	1.12			1.00	1.00		
$p = 10$	1.07		1.19	1.18			1.05	1.05		
$p = 25$	1.04		1.15	1.15			1.03	1.03		
$p = 50$	1.03		1.12	1.12			1.01	1.01		
$p = 75$	1.03		1.13	1.12			1.01	1.01		

$\rho = 0.5$, Aggregated case - Configuration 1

missing	number	MEQ_{TS}	$se(MEQ_{TS})$	MEQ_{TC}		$se(MEQ_{TC})$		MEQ_{BP}		$se(MEQ_{BP})$	
		M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
0	1	0.98		1.01	1.01			0.99	1		
1	1	1.09		1.23	1.22			1.55	1.03		
2	0	.	.	.	.	.	.	.	.	.	
3	3	1.02		1.10	1.09			1.00	1.00		
4	4	0.97		1.07	1.08			0.97	0.97		
5	3	1.28		1.34	1.28			1.31	1.23		
Total	12	1.08		1.15	1.14			1.07	1.05		

predictor with respect to the best would also be of interest although it seem difficult at first sight. The same problems arise in other spatial models and since prediction is ubiquitous, it is an important problem for spatial econometrics.

References

[1] R. Bivand (2002) Spatial econometrics functions in R: Classes and methods. Journal of Geographical Systems, 4, 405-421.

[2] X. Guyon (1995) Random field on a network: modeling, statistics and applications. Probability and its applications series. Springer-Verlag, 255p.

[3] C. Gaetan and X. Guyon (2008) *Modélisation et statistique spatiales*, Springer.

[4] R.P. Haining (1990) C. Gaetan and X. Guyon (2008) *Spatial data analysis in the social and environmental sciences*, Cambridge university Press.

[5] J.L. Jensen, H. Kunsch (1992) On asymptotic normality of pseudo likelihood estimate for pairwise interaction processes. Annals of the Institute of Statistical Mathematics, 46(3), 475-486.

[6] J.P. LeSage and R.K. Pace (2004) Models for spatially dependent missing data, Journal of Real Estate Finance and Economics, 29:2, 233-254.

$\rho = 0.5$, $\sigma = 1$, Aggregated case - Configuration 2

missing	number	$\bar{MEQ}_{TS}$	$se(MEQ_{TS})$	$\bar{MEQ}_{TC}$		$se(MEQ_{TC})$		$\bar{MEQ}_{BP}$		$se(MEQ_{BP})$	
		M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
0	2	1.02		1.21	1.20			0.95	0.96		
1	1	1.1		1.18	1.19			1.11	1.1		
2	3	0.80		0.91	0.91			0.78	0.77		
3	8	0.99		1.09	1.07			0.99	0.98		
4	7	1.06		1.16	1.13			1.03	1.02		
5	5	1.01		1.10	1.07			1.00	0.97		
6	1	1.54		1.63	1.38			1.53	1.35		
Total	27	1.02		1.12	1.09			1.00	0.98		

$\rho = 0.5$, $\sigma = 1$, Aggregated case - Configuration 3

missing	number	$\bar{MEQ}_{TS}$	$se(MEQ_{TS})$	$\bar{MEQ}_{TC}$		$se(MEQ_{TC})$		$\bar{MEQ}_{BP}$		$se(MEQ_{BP})$	
		M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
1	1	0.96		1.04	1.05			0.96	0.95		
2	2	0.77		0.89	0.88			0.77	0.76		
3	4	1.14		1.30	1.28			1.09	1.09		
4	6	1.11		1.19	1.18			1.11	1.08		
5	8	1.02		1.10	1.01			1.00	0.90		
6	1	1.32		1.51	1.37			1.3	1.22		
7	2	1.05		1.09	1.08			1.07	1.04		
8	3	1.12		1.15	1.03			1.15	0.96		
Total	27	1.06		1.15	1.10			1.05	0.99		

[7] J.P. Lesne, H. Tranger, A. Ruiz-Gazen and C. Thomas-Agnan (2008) Predicting population annual growth rates with spatial models, Preprint.

[8] R.K. Pace and J.P. LeSage (2006) Interpreting spatial econometrics models. Preprint.

$\rho = 0.5$, $\sigma = 1$, Aggregated case - Configuration 4

missing	number	$M\bar{E}Q_{TS}$	$se(MEQ_{TS})$	$M\bar{E}Q_{TC}$		$se(MEQ_{TC})$		$M\bar{E}Q_{BP}$		$se(MEQ_{BP})$	
		M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
1	4	0.90		1.03	1.03			0.88	0.88		
2	10	0.97		1.06	1.05			0.96	0.95		
3	17	1.06		1.17	1.14			1.05	1.03		
4	11	1.10		1.20	1.14			1.08	1.05		
5	12	1.08		1.22	1.14			1.06	1.03		
Total	54	1.05		1.16	1.12			1.03	1.01		

$\rho = 0.5$, $\sigma = 1$, Aggregated case - Configuration 5

missing	number	$M\bar{E}Q_{TS}$	$se(MEQ_{TS})$	$M\bar{E}Q_{TC}$		$se(MEQ_{TC})$		$M\bar{E}Q_{BP}$		$se(MEQ_{BP})$	
		M_1	M_1	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
2	2	1.25		1.33	1.32			1.23	1.23		
3	4	1.02		1.16	1.11			0.99	0.97		
4	21	1.09		1.20	1.18			1.06	1.06		
5	11	1.11		1.23	1.20			1.10	1.08		
6	8	1.11		1.22	1.15			1.10	1.06		
7	2	1.04		1.16	1.07			1.02	0.96		
8	6	1.35		1.39	1.11			1.34	1.04		
Total	54	1.12		1.23	1.17			1.11	1.06		