

*Economics Department*  
*Working Papers in Economics*

---

*Boston College*

*Year 2002*

---

Aggregation of Space-Time Processes

Raffaella Giacomini\*

Clive W.J. Granger†

\*Boston College,

†University of California, San Diego,

This paper is posted at eScholarship at Boston College.

[http://escholarship.bc.edu/econ\\_papers/24](http://escholarship.bc.edu/econ_papers/24)

# Aggregation of Space-Time Processes

Raffaella Giacomini and Clive W. J. Granger\*

*University of California, San Diego*

July 2002

## Abstract

In this paper we compare the relative efficiency of different methods of forecasting the aggregate of spatially correlated variables. Small sample simulations confirm the asymptotic result that improved forecasting performance can be obtained by imposing *a priori* constraints on the amount of spatial correlation in the system. One way to do so is to aggregate forecasts from a Space-Time Autoregressive model (Cliff *et al.*, 1975), which offers a solution to the ‘curse of dimensionality’ that arises when forecasting with VARs. We also show that ignoring spatial correlation, even when it is weak, leads to highly inaccurate forecasts. Finally, if the system satisfies a ‘poolability’ condition, there is a benefit in forecasting the aggregate variable directly.

**KEYWORDS:** Spatial correlation, Aggregation, Forecast efficiency, Space-Time models, VAR.

**J.E.L. Codes:** C33, C43, C53

---

\*Contact information: Raffaella Giacomini, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92039-0508, USA. Email: rgiacomini@ucsd.edu. The authors would like to thank Andrew Patton, Marjorie Flavin, Ross Starr and two referees for helpful comments.

# 1 Introduction

Many variables of economic interest are contemporaneous aggregates of variables observed over time and across a number of different regions. When interested in forecasting the aggregate across regions, the analyst might ask whether it will be more efficient to forecast the aggregate series directly or to model the individual components separately and then aggregate the forecasts. The literature that tries to answer the question is fairly large, but does not provide clear guidelines; most papers focus on particular econometric models under restrictive assumptions and the results of simulations and empirical investigations are often contradictory. Whatever the model used for forecasting, it is plausible to think that the final verdict on whether to aggregate or not will be sensitive to the degree of interdependence existing among the variables measured in the different regions. The early literature in the field restricts attention to the standard regression framework, where temporal and cross-sectional dependence among the variables is ignored (e.g., Grunfeld and Griliches, 1960, Aigner and Goldfeld, 1974) or only generic contemporaneous correlation in the errors is allowed (e.g., Pesaran, Pierse and Kumar, 1989). The case where the correlation between variables is due to the presence of a common factor is analyzed by Granger (1980, 1987), who illustrates the implications of aggregating variables that depend on both common and idiosyncratic factors. It is shown that, when aggregation is across a large number of units, the common factors will dominate the process for the aggregate, even though they might be relatively unimportant at the individual level. It follows that there might be a benefit in forecasting the disaggregated variables, provided the common factor is appropriately taken into account (a similar conclusion is reached in an empirical application by Zellner and Tobias, 2000). A different strand of literature conducts the analysis of contemporaneous aggregation in the context of vector ARMA processes, that allow the variables to be dependent in both time and cross-sectional directions. A comprehensive treatment of this case and an extended bib-

liography can be found in Lütkepohl (1987). In this paper, we consider the situation where one is interested in forecasting the aggregate of a number of variables observed in a moderately large number of regions and over a long time period. As the cross-sectional dimension increases, forecasting with vector ARMA models becomes quickly infeasible, as the so-called ‘curse of dimensionality’ makes it difficult to estimate the model accurately and raises the issue of the effect of estimation uncertainty on the model’s forecasting performance. To restrict the amount of interdependence in our system, we make use of the concept of ‘spatial autocorrelation’, which arises when observations at one region are systematically dependent on the observations at neighboring regions, while the dependence vanishes for regions far apart. In some sense, our position is intermediate between the approaches taken in the literature, where the cross-sectional dependence among the units to aggregate is either ignored or assumed to be due to a common factor. The latter case, in particular, corresponds to an assumption of constant spatial autocorrelation across regions, whereas our approach assumes that the spatial correlation disappears for regions sufficiently distant in space. The issue of spatial autocorrelation is relevant for a wide range of economic fields, such as environmental economics, urban economics, industrial organization and international economics. Econometric models that explicitly account for spatial dependence were originally proposed by Cliff and Ord (1973) and Cliff, Haggett, Ord, Bassett and Davies (1975), but their application in economics was initially restricted to a few specific fields such as regional science and real estate economics. More recently, there has been a renewed interest for models of spatial dependence in traditional economics, perhaps due to the growing attention in economic theory for models that explicitly account for interaction among heterogeneous agents and, on the empirical side, to the increasing availability of highly disaggregated and spatially referenced data from the Geographic Information Systems. An excellent review of the current state of the field of spatial econometrics can be found in Anselin (1988). See also Haining (1990) and references therein. In this paper, we will analyze contemporaneous aggregation within

a class of stochastic models known as Space-Time Autoregressive (Space-Time *AR*) models introduced by Cliff *et al.* (1975) and generalized by Pfeifer and Deutsch (1980)<sup>1</sup>. More recent discussions and applications of the Space-Time *AR* model in econometrics are Elhorst (2000) and Szulc (2000), while a generalization of the model to continuous space is proposed by Brown, Karesen, Roberts and Tonellato (2000).

The paper is organized as follows. Section 2 briefly anticipates some of the issues involved with aggregation of spatially correlated variables in a very simplified example. Section 3 introduces the forecasting model that will be used in the paper to describe the spatial interactions among the variables of interest. Different ways to forecast the aggregate of spatially correlated variables are proposed and their efficiency compared in Section 4, both in a framework that assumes the parameters of the models to be known and in the more realistic setting where parameter estimation uncertainty is present. Section 5 explores the small sample behavior of the different forecasts of the aggregate and compares their efficiency in a Monte Carlo experiment. Section 6 concludes. All proofs are in the Appendix.

## 2 A simple model of spatial dependence

To illustrate the issues involved with the aggregation of processes that are spatially dependent, we will initially consider a very simple example in which a variable  $x_t$  is measured over time in three neighboring regions  $i - 1$ ,  $i$  and  $i + 1$ .<sup>2</sup> Because of the spatial proximity of the regions, we can assume that the value of the variable at time  $t$  in region  $i$  depends on the values of the variable at all three locations at time  $t - 1$ , a situation that can be visually

---

<sup>1</sup>Both authors propose a general Space-Time ARMA class of models, which are a generalization of univariate ARMA models that explicitly incorporate spatial dependence. For simplicity, we will only focus on the autoregressive subgroup

<sup>2</sup>Here and in the following, we adopt the convention of using lowercase letters for either random variables or their realizations.

represented as follows.

$$\begin{array}{ccc}
 x_{i-1,t-1} & x_{i,t-1} & x_{i+1,t-1} \\
 & \searrow \downarrow \swarrow & \\
 & x_{i,t} &
 \end{array}$$

Since we are not including in our analysis regions on either sides of the diagram, we have that the time- $t$  values of  $x$  in regions  $i-1$  and  $i+1$  are only functions of their own lag and of the lagged variable in region  $i$ , which introduces so-called ‘edge effects’ into our framework.<sup>3</sup>

Suppose that the dependence in the diagram can be expressed as

$$x_{i,t} = \phi x_{i,t-1} + \psi_1 x_{i-1,t-1} + \psi_2 x_{i+1,t-1} + \varepsilon_{i,t}, \quad (1)$$

where  $\varepsilon_{i,t}$  is a zero-mean white noise uncorrelated across regions. Denoting the spatial aggregate measured in region  $i$  and at time  $t$  with  $S_{i,t}(x) = \sum_{k=i-1}^{i+1} x_{k,t}$ , we have

$$S_{i,t}(x) = \phi S_{i,t-1}(x) + \psi_1 S_{i-1,t-1}(x) + \psi_2 S_{i+1,t-1}(x) + S_{i,t}(\varepsilon). \quad (2)$$

If we assume that the edge effects are negligible, it follows that the aggregate variables  $S_{i,t-1}$ ,  $S_{i-1,t-1}$  and  $S_{i+1,t-1}$  are approximately equal and therefore (2) becomes

$$S_t(x) = (\phi + \psi_1 + \psi_2) S_{t-1}(x) + S_t(\varepsilon). \quad (3)$$

In sum, the process for the aggregate is approximately an  $AR(1)$  and the coefficients of spatial dependence ( $\psi_1$  and  $\psi_2$ ) are incorporated into the autoregressive coefficient. We can thus conclude that aggregation causes in this case a ‘simplification’ of the dynamic properties of the process.

---

<sup>3</sup>‘Edge effects’ result from excluding from the data set spatial units that are related to the units in the sample. This problem can be compared to the issue of initial values in time series, but it presents further complications due to the particular nature of spatial dependence (see Anselin, 1988, pp.172-176).

### 3 The Space-time Auto Regressive model

A space-time model is a time-series model that explicitly takes into account linear dependence between the variables that is lagged both in time and in space. Suppose that the variable of interest  $x_{it}$  is observed at each of  $k$  fixed locations ( $i = 1, \dots, k$ ) and over time ( $t = 1, \dots, T$ ). In the remainder of the paper, the locations will be referred to as ‘regions’, but they can represent a variety of levels of data agglomeration, from city districts and counties to states and countries. The construction of space-time models relies on the assumption that the relationships among the variables in the various regions under examination depend in a systematic way on the regions’ relative distance. The conditional mean of the variable  $x_{it}$  is thus modeled as a linear function of past observations of the variable at region  $i$  and at neighboring regions. To be able to relate a variable at one region to the observations for the same variable at other regions, it is necessary to introduce the concept of spatial lag, which presents some complications relative to the fairly intuitive definition of time lag. While the temporal lag operator shifts the variable by one or more periods in time, in space the direction of shift is not unique and the definition could change depending on the spatial arrangement of the data. The first step thus consists in identifying the neighbors of each region according to some prespecified criterion (e.g., border-sharing), and grouping them in neighboring sets (e.g., first-order, second-order neighbors and so forth).

Once the sets of neighbors for each region have been identified, the spatial lag operator can be defined as the weighted average of all the observations in a given neighboring set (see, e.g., Anselin, 1988, pp.22-26). Formally, if  $x_i$  is the observation in region  $i$  and  $J_s$  is the set of neighbors of order  $s$ , the spatial lag of order  $s$  is given by

$$L^{(s)}x_i = \sum_{j \in J_s} w_{ij}^{(s)} x_j \quad s = 1, 2, \dots \quad (4)$$

We see that a spatial lag is in practice a distributed lag, rather than a shift in a given direction like in the time series case. The choice of the weights  $w_{ij}^{(s)}$  in (4) is a crucial issue in

spatial econometrics. The weights are usually assumed to be exogenous, nonstochastic and to satisfy the following properties:

- 1)  $w_{ij}^{(s)} \geq 0$ ,
- 2)  $w_{ii}^{(s)} = 0$ ,
- 3)  $\sum_{j \in J_s} w_{ij}^{(s)} = 1$ .<sup>4</sup>

Typically, the spatial weights are chosen *a priori* by the researcher to reflect geographical characteristics of the regions under consideration (e.g., distance, length of common borders, number of roads etc.), but alternative specifications, for example based on definitions of economic distance, have also been used in the literature (e.g., Pinkse and Slade, 1998).<sup>5</sup> The overwhelming majority of the spatial econometric literature assumes the spatial weight matrix to be known in advance. As a consequence, as pointed out by Stetzer (1982), any inference conducted with a spatial model will be sensitive to the problem of misspecification of the weight matrix, with possible resulting inconsistency of the parameter estimates and misleading assessment of forecast performance. Detailed references and a discussion of weight matrix specification can be found in Anselin (1988). In this paper, we will conform to the standard practice in the spatial econometric literature of considering the weight matrix to be known *a priori* and to be correctly specified, while fully acknowledging that the presence of estimation uncertainty in the weight's specification may affect the conclusions we draw.

The central issue of the paper, forecasting with data that exhibit spatial dependence, can

---

<sup>4</sup>The normalization of the weights (property n. 3) is usually motivated by an ease of interpretation of the model's coefficients. Anselin (1988, p. 24) points out that this assumption in some situations implies a view of spatial interaction that is not economically meaningful. For example, if the weights for the N neighbors of a given region are simply equal to  $1/N$ , it follows that a higher number of neighbors implies less individual influence, an assumption not always justifiable on economic grounds.

<sup>5</sup>When economic theory is not helpful in guiding the choice of the appropriate weights, and if the model is manageably small, one could alternatively conduct a model-selection search by fitting different weight specifications and choosing the one that maximizes the likelihood.

be illustrated using the simplest form of space-time model for the conditional mean of the zero-mean variable  $x_{it}$ , a Space-Time  $AR(1,1)$ , which ignores dependence beyond the first temporal and spatial lags

$$x_{it} = \phi x_{it-1} + \psi \sum_{j=1}^k w_{ij} x_{jt-1} + \varepsilon_{it} \quad i = 1, \dots, k \quad t = 1, \dots, T. \quad (5)$$

The weights  $w_{ij}$  sum to one for each  $i$  and are non-zero only for neighbors of region  $i$ . Collecting the weights  $w_{ij}$  in a  $k \times k$  spatial weight matrix  $W = (w_{ij})$ , we can rewrite model (5) in vector form as

$$\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \psi W \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T. \quad (6)$$

The first element on the right hand side of equation (6) represents the first temporal lag of  $\mathbf{x}_t$ , while the second element is the first spatial lag of the vector at time  $t-1$ . In the paper, we assume for simplicity that initial and edge effects are negligible, corresponding to the ideal situation of a long time series which is measured near the ‘center’ of a spatial configuration of regions. Edge effects are a known problem in spatial econometrics, and some partial - and often arbitrary - solutions have been proposed in the literature (see Anselin, 1988, for a discussion). Even though the theoretical results assume the absence of edge effects, the impact of this assumption will be briefly discussed in the simulation in section 5.

The Space-Time  $AR$  model in (6) should not be confounded with the Spatial  $AR$  model, denoted  $SAR$ , which is an  $AR$  model that contains a simultaneous spatially lagged dependent variable. The Space-Time  $AR$  model is more appropriate than the  $SAR$  to discuss the effects of spatial correlation on forecast efficiency, since the model assumes that the spatial effects take one period to become manifest, and can thus be exploited for forecasting.

It is easily seen that the Space-Time  $AR(1,1)$  model defined in (6) is a special case of a vector autoregressive model ( $VAR$ ) of order 1, where the autoregressive coefficient matrix is restricted to equal  $\phi I_k + \psi W$  (the generalization to higher spatial and temporal orders is straightforward). It is in this sense that the space-time model we consider in this paper is a

refinement of a multivariate time-series model. Whereas a *VAR* model allows all variables in a system to be related to all other variables, the Space-Time *AR* model imposes *a priori* restrictions on this interdependence due to the spatial allocation of the variables, exploiting the fact that contiguous regions could be related in a systematic way.

In the remainder of the paper, we will assume model (6) to be correctly specified for the conditional mean of  $\mathbf{x}_t$ . Clearly, the model considered is very simple, and does not claim to be an accurate description of reality. Potentially important features ignored by the specification (6) are, e.g., exogenous explanatory variables, contemporaneous spatial lags of the dependent variable and a number of cross-sectional units which increases with the sample size (whereas we consider it to be fixed as  $T \rightarrow \infty$ ). We feel however that use of a more sophisticated model would unnecessarily complicate the analysis without providing any more insight into the problem of interest.

## 4 Forecasting aggregated space-time processes

Suppose that the goal is to forecast  $y_t \equiv \sum_{i=1}^k x_{it}$ , the aggregate across regions of a set of  $k$  variables  $x_{it}$ , related by spatial dependence.

When time series of the  $k$  variables  $\{x_{it}\}_{t=1}^T$  and of the aggregate  $\{y_t\}_{t=1}^T$  are available, the researcher can obtain a forecast for  $y_t$  in several different ways:

- f1. The aggregate  $y_t$  can be forecasted directly by fitting a univariate forecasting model to the series  $\{y_t\}_{t=1}^T$ .
- f2. Univariate forecasts of each variable  $x_{it}$ ,  $i = 1, \dots, k$  can similarly be obtained and then aggregated.
- f3. The vector  $\mathbf{x}_t$  may be forecasted by fitting a multivariate forecasting model to the series  $\{\mathbf{x}_t\}_{t=1}^T$ . A forecast for  $y_t$  is obtained by aggregating the resulting forecasts for

each  $x_{it}$ .

- f4. A Space-Time *AR* model can be used to forecast each variable  $x_{it}$ ,  $i = 1, \dots, k$ . The forecasts for each component are then aggregated.

Intuitively speaking, the different scenarios correspond to forecasting using different information sets. The first scenario only utilizes information contained in the univariate aggregated series. The second scenario treats each component separately and for each component only uses the information in the relative univariate series. The third and fourth methods further incorporate information about possible or known interdependencies among the individual series.

There are various reasons why one would consider forecasting using the different scenarios. The multivariate models in f3 and f4 have the advantage of utilizing the maximum amount of information present in the data, but they can become easily unmanageable when the dimension of the system increases. Forecasting with a univariate model of the aggregate, on the other hand, offers a quick and easy alternative and has the advantage of being tightly parameterized, but it implies a loss of information. Aggregating univariate forecasts of the components can be a viable alternative, since it is based on more information than the direct forecast of the aggregate, but it entails ignoring the relationships among the component series.

In the rest of the paper, we will attempt to rank the four different approaches in terms of their relative efficiency.

To formalize, we make the following assumptions about the data-generating process (DGP):

**Assumption 1.** The  $k \times 1$  data vector  $\mathbf{x}_t$  is generated according to  $\mathbf{x}_t = \phi_0 \mathbf{x}_{t-1} + \psi_0 W \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$ ,  $t = 1, \dots, T$ ,  $\boldsymbol{\varepsilon}_t \sim i.i.d. N(0, \Sigma_\varepsilon)$ .  $W$  is the nonstochastic weight matrix defined in (??),  $W \neq 0$ . The aggregate variable is then given by  $y_t = \boldsymbol{\iota}' \mathbf{x}_t$ , where  $\boldsymbol{\iota}$  is a  $k \times 1$  vector of ones.

**Assumption 2.** The eigenvalues of the matrix  $\phi_0 I_k + \psi_0 W$  lie inside the unit circle.

The assumption of normality for the variable of interest is quite restrictive and unrealistic, but it is dictated by the desire to obtain tractable analytical solutions.

The forecasts we will be referring to are optimal (minimum mean squared error) linear predictors, based on the relevant information set. To illustrate, the optimal one-step-ahead forecast of  $y_t$  using the Space-Time  $AR(1,1)$  model in (6) (f4) is given by  $y_{t-1}(1) = \iota'(\phi \mathbf{x}_{t-1} + \psi W \mathbf{x}_{t-1})$ . For scenario f3, we have instead  $y_{t-1}(1) = \iota' \mathbf{B} \mathbf{x}_{t-1}$ , where  $\mathbf{B}$  is the  $k \times k$  coefficient matrix in a  $VAR(1)$  model of  $\mathbf{x}_t$ . For the univariate forecasting situations f1 and f2, the optimal forecasts of  $y_t$  and of  $x_{it}$ ,  $i = 1, \dots, k$  can be based on the  $ARMA$  representations of  $y_t$  and  $x_{it}$ . Such  $ARMA$  representations (with finite orders) exist due the fact that any linear transformation of an  $ARMA$  process has an  $ARMA$  representation (see, e.g., Lütkepohl, 1987, Corollary 4.2.2), and both  $y_t$  and  $x_{it}$  are linear transformations of the  $VAR(1)$  process  $\mathbf{x}_t$ . We will make the following assumption about the orders of the  $ARMA$  representations of the variables of interest in each forecasting situation.

**Assumption 3.** The orders of the  $ARMA$  representations of  $y_t$ ,  $\mathbf{x}_t$  and  $x_{it}$ ,  $i = 1, \dots, k$  are minimal and they are either known or estimated consistently.

In practice, the forecasts formulated in all scenarios depend on parameters, which will have to be estimated. From the above discussion, it is clear that a comparison of the efficiency of different forecasting methods will rely on the interplay between how well a given model exploits the information available and the effects of estimation uncertainty, as we will explore in detail in the following sections. A large part of this section is related to the results collected in Lütkepohl (1987). Our treatment differs from Lütkepohl (1987)'s in that we assume the presence of spatial correlation in the system.

## 4.1 Forecast comparisons with known parameters

To establish a benchmark, we will initially compare the relative efficiency of forecasting the aggregate in scenarios f1-f4 under the assumption of no parameter estimation uncertainty. In order to simplify the analysis, in this section and in the remainder of the paper we will restrict attention to one-step-ahead forecasts of  $y_t$ , denoted by  $y_{t-1}(1)$ , and choose as a measure of forecast accuracy the forecasts' Mean Squared Error (MSE):

$$MSE(y_{t-1}(1)) \equiv E[(y_t - y_{t-1}(1))^2]. \quad (7)$$

**Proposition 1 (Efficiency comparison with known parameters)** *Let  $MSE^{(i)}$ ,  $i = 1, \dots, 4$  denote the MSE of the one-step-ahead optimal linear forecast for the aggregate  $y_t$  obtained under each of the forecasting scenarios f1-f4, in the absence of parameter estimation uncertainty. Given assumptions 1-3, the following ranking between alternative forecasting methods can be established*

a) *The aggregate of forecasts from a VAR is more efficient than the aggregate of univariate forecasts for each component:  $MSE^{(3)}(y_{t-1}(1)) < MSE^{(2)}(y_{t-1}(1))$ .*

b) *The aggregate of forecasts from a VAR is (weakly) more efficient than the forecast based on the aggregated data:  $MSE^{(3)}(y_{t-1}(1)) \leq MSE^{(1)}(y_{t-1}(1))$ .*

c) *The aggregate of forecasts from a VAR is as efficient as the aggregate of univariate Space-Time AR forecasts for each component:  $MSE^{(3)}(y_{t-1}(1)) = MSE^{(4)}(y_{t-1}(1))$ .*

As pointed out by Wei and Abraham (1981), the relative efficiency of forecasting the aggregate variable directly or aggregating univariate forecasts for the components (i.e., of methods f1 and f2) cannot be established for a general data-generating process.

Kohn (1982) shows that the equality of MSE in Proposition 1-b) holds for a general gaussian, stationary  $VAR(p)$  process  $B(L)\mathbf{x}_t = \boldsymbol{\varepsilon}_t$  if and only if

$$\boldsymbol{\iota}'B(L) = b(L)\boldsymbol{\iota}', \quad (8)$$

where  $b(L)$  is a scalar lag polynomial of order  $p$ . In other words, condition (8) guarantees that the aggregate of a  $VAR(p)$  is an  $AR(p)$  process, in which case it will be equivalent to forecast the disaggregated system and then aggregate the forecasts or to forecast the aggregate directly (we will therefore call (8) the ‘poolability’ condition). In our case, the data-generating process is a Space-Time  $AR(1, 1)$ , and condition (8) becomes  $\mathbf{1}'(\phi_0 I_k + \psi_0 W) = b\mathbf{1}'$ , with  $b$  scalar, which corresponds to requiring the matrix  $\phi_0 I_k + \psi_0 W$  to have equal column sums. Since the typical column sum is of the form  $s_j = \phi + \psi \sum_{i=1}^k w_{ij}$ , it follows that poolability is attained when the spatial weight matrix has equal column sums. Loosely speaking, each column sum of  $W$  represents the total spatial ‘influence’ of region  $j$  over all other regions and therefore condition (8) means that we can forecast the aggregate directly if the total spatial influence is relatively uniform across regions (which will happen for a set of regions that share similar spatial and/or economic conditions).

Kohn (1982) exploits condition (8) to devise a simple test for poolability of the variables in the different regions, which in our case consists in regressing the aggregate variable  $y_t$  on  $y_{t-1}$  and on  $k - 1$  components of  $\mathbf{x}_{t-1}$ . Testing the null hypothesis of poolability is equivalent to a test of zero coefficients on all the included components of  $\mathbf{x}_{t-1}$ .

To summarize the results, in the absence of estimation uncertainty and if the poolability condition is not satisfied, it is optimal to forecast the fully disaggregated  $VAR$  system and aggregate the forecasts or, equivalently, to aggregate the forecasts from a Space-Time  $AR$  model. Under the same conditions, both the aggregation of univariate forecasts for the components and a direct forecast of the aggregate will be sub-optimal.

## 4.2 Forecast comparisons with estimated parameters

In practice, the parameters that appear in the expressions for the optimal forecast of  $y_t$  in each situation are not known, and must be estimated. We see next how the estimation variability affects the efficiency ranking results that were found in the previous section. In

the remainder of the paper, the estimation method considered will be maximum likelihood (ML), and  $\hat{\theta}_T$  will denote the ML estimator of the true parameter  $\theta_0$  for a sample of size  $T$ .

A standard result shows that ML estimators of the parameters of a stationary gaussian (vector) *ARMA* are consistent for the true parameters and asymptotically normal.

**Lemma 2** *Given assumptions 1-3, for each of the models estimated in scenarios f1-f4, the ML estimator of the parameters satisfies  $\hat{\theta}_T \xrightarrow{p} \theta_0$  and  $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \Sigma_{\hat{\theta}})$ , where  $\Sigma_{\hat{\theta}}$  is the Hessian,  $\Sigma_{\hat{\theta}} = T \left[ -E \frac{\partial^2 \ln l}{\partial \theta \partial \theta'} \right]^{-1}$ .*

We use the estimated models to build forecasts and compare their relative efficiency in terms of asymptotic MSE. Given that all models used for forecasting in situations f1-f4 are members of the (vector) *ARMA* class, we will first derive the asymptotic MSE for the one-step-ahead forecast of the generic  $k \times 1$  variable  $z_t$  based on a vector *ARMA* model for  $z_t$ . For simplicity, we make the following assumption.

**Assumption 4.** Estimation and forecasting are based on independent and identically distributed processes.

Asymptotic MSE for dynamic models that do not rely on this last assumptions have been derived in the literature (see, e.g., Brown and Mariano, 1989), but at the cost of increased complexity.

Lütkepohl (1987) proves the following result.

**Lemma 3** *Given assumptions 1-4, the forecast  $\hat{z}_{t-1}(1)$ , obtained by replacing the true parameter  $\theta_0$  of the optimal linear predictor of  $z_t$  ( $z_{t-1}(1)$ ) by its ML estimator  $\hat{\theta}_T$ , is a consistent estimator of  $z_{t-1}(1)$  and*

$$\sqrt{T}[\hat{z}_{t-1}(1) - z_{t-1}(1)] \xrightarrow{d} N(0, \Omega_z), \quad (9)$$

where

$$\Omega_z = E \left[ \frac{\partial z_{t-1}(1)}{\partial \theta'} \Sigma_{\hat{\theta}} \frac{\partial z_{t-1}(1)}{\partial \theta} \right]. \quad (10)$$

The asymptotic MSE of  $\hat{z}_{t-1}(1)$  is given by

$$MSE(\hat{z}_{t-1}(1)) = MSE(z_{t-1}(1)) + \frac{1}{T}\Omega_z. \quad (11)$$

Equation (11) shows that the presence of parameter estimation uncertainty adds a positive definite term  $\frac{1}{T}\Omega_z$  (asymptotically vanishing) to the MSE that would be obtained if the parameters of the data-generating process were known. When comparing the performance of forecasts constructed in the different scenarios, one must thus take into account both components of the asymptotic MSE. The expression for  $\Omega_z$  is rather complicated for general ARMA processes (see, e.g., Yamamoto, 1981 and Baillie, 1980), but it can be seen that its magnitude depends on the number of parameters estimated. This fact can be proven explicitly for the asymptotic MSE of the aggregate forecast obtained in situation f3.

**Lemma 4** *Let  $\mathbf{x}_t$  be a stationary,  $k$ -dimensional VAR( $p$ ) process  $\mathbf{B}(\mathbf{L})\mathbf{x}_t = \boldsymbol{\varepsilon}_t$ , with  $\boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \Sigma_\varepsilon)$ . Then we have*

$$\sqrt{T}[\hat{\mathbf{x}}_{t-1}(1) - \mathbf{x}_{t-1}(1)] \xrightarrow{d} N(0, \Omega_{\mathbf{x}}), \quad (12)$$

with  $\Omega_{\mathbf{x}} = kp\Sigma_\varepsilon$ .

**Proposition 5 (Asymptotic MSE for situation f3)** *Under assumptions 1-4, the forecast for  $y_t$  obtained by estimating a VAR(1) model on the  $k \times 1$  vector  $\mathbf{x}_t$  and aggregating the forecasts of the components has asymptotic MSE given by*

$$MSE^{(3)}(\hat{y}_{t-1}(1)) = \boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} + \frac{k}{T}\boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} \quad (13)$$

The above result indicates that, in the presence of estimation uncertainty, the forecast scenario f3, which is optimal when the data-generating process is known, may lose its optimality properties. To see why, assume for the time being a diagonal covariance matrix for the disturbances,  $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \sigma^2 I_k$ , in which case the asymptotic MSE for the estimated aggregate becomes  $MSE(\hat{y}_{t-1}(1)) = \sigma^2 k + \frac{1}{T}\sigma^2 k^2$ . We see that the component of the MSE

that is due to estimation error grows at the rate of  $k^2$ , where  $k$  is the number of units to aggregate, while the component that would be obtained in the absence of estimation uncertainty only grows at rate  $k$ . This fact introduces a trade-off between the optimality of specifying the fully disaggregated system (no information loss) and the reduction in efficiency due to estimation error and thus the relationships between the  $MSE$ 's in Proposition 1 - a), b) are no longer valid in general. Instead, for each given specification of the DGP, one could find integers  $k_1$  and  $k_2$  such that the forecast of the aggregate from f1 is more efficient than the forecast from f3 for  $k > k_1$  and the forecast from f2 is more efficient than the one from f3 for  $k > k_2$ . The only ranking we can establish is the following.

**Proposition 6 (Efficiency comparison with estimated parameters)** *Given assumptions 1-4, in the presence of estimation uncertainty, the following relative efficiency results among forecasting methods f1-f4 can be established:*

a) *The aggregate of forecasts from a Space-Time AR model is (weakly) more efficient than the aggregate of forecasts from a VAR.  $MSE^{(4)}(\hat{y}_{t-1}(1)) \leq MSE^{(3)}(\hat{y}_{t-1}(1))$ .*

b) *If the poolability condition (8) is satisfied, it is more efficient to forecast the aggregate directly rather than to aggregate forecasts from a VAR.  $MSE^{(1)}(\hat{y}_{t-1}(1)) < MSE^{(3)}(\hat{y}_{t-1}(1))$ .*

To summarize the results of this section, we found that in the presence of estimation uncertainty aggregating forecasts from the Space-Time  $AR(1, 1)$  model is weakly more efficient than aggregating forecasts from a  $VAR(1)$ . If the poolability condition (8) cannot be rejected, it is also always preferable to forecast the aggregate directly than to aggregate forecasts from the  $VAR(1)$ , and the benefits increase proportionally to the number of regions considered.

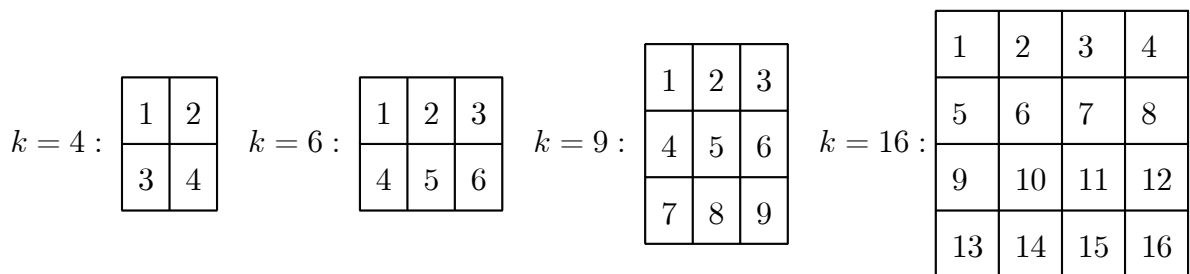
## 5 Monte Carlo experiment

### 5.1 Simulation design

In this section the small sample behavior of the forecasts of the aggregate obtained in scenarios f1-f4 are investigated by means of a Monte Carlo experiment. The process used for the simulation is the Space-Time  $AR(1, 1)$  in (6)

$$\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \psi W \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

where the disturbances are  $i.i.d.N(0, I_k)$ . We consider the case of  $k = 4, 6, 9, 16$  regions arranged on a regular grid as follows



Because of this particular spatial arrangement of the regions, the system will be affected by edge effects. For  $k = 9$ , for example, region 5 is the only one with all four first-order neighbors in the system, while all other regions are situated on the edges and therefore are plausibly affected by units not included in the system. We therefore perform the simulation with two alternative specifications of the spatial weight matrix  $W$ . The first specification is obtained by dividing the weights equally between all first order neighbors of each unit. For

example, for  $k = 4$  and  $6$ , the matrices  $W$  will be

$$W^{(4)} = \begin{bmatrix} 0 & .5 & .5 & 0 \\ .5 & 0 & 0 & .5 \\ .5 & 0 & 0 & .5 \\ 0 & .5 & .5 & 0 \end{bmatrix}, \quad W^{(6)} = \begin{bmatrix} 0 & .5 & 0 & .5 & 0 & 0 \\ .33 & 0 & .33 & 0 & .33 & 0 \\ 0 & .5 & 0 & 0 & 0 & .5 \\ .5 & 0 & 0 & 0 & .5 & 0 \\ 0 & .33 & 0 & .33 & 0 & .33 \\ 0 & 0 & .5 & 0 & .5 & 0 \end{bmatrix} \quad (14)$$

where row 1 contains the weights for region 1's first order neighbors, row 2 the weights for region 2's first order neighbors and so forth. The matrix  $W$  for  $k = 9$  and  $16$  is constructed in a similar fashion. For this particular choice of weights, the spatial weight matrix  $W$  has equal column sums only for  $k = 4$ , in which case the process satisfies the poolability condition. For all other values of  $k$  the poolability condition is not verified. The second specification of the spatial weight matrix assumes that each region has four first-order neighbors, with weights equally split among them. As a consequence, the weight matrices will be modified so that each non-zero entry in the matrices in (14) will equal  $0.25$ .<sup>6</sup>

For each  $k$ , we generate data from four different processes, reflecting different assumptions about the value of the autoregressive coefficient  $\phi$  and of the spatial coefficient  $\psi$ . We consider the four specifications 1)  $\phi = .45$ ,  $\psi = .45$ , 2)  $\phi = .45$ ,  $\psi = .1$ , 3)  $\phi = .1$ ,  $\psi = .45$ , 4)  $\phi = .1$ ,  $\psi = .1$ . These values of the parameters, together with the two alternative specifications of  $W$ , are such that Assumption 2 is satisfied.

The predictors for scenarios f1-f4 described in Section 4 are computed as follows (all parameters are estimated by ML).

---

<sup>6</sup>In some sense the second specification of the weight matrices can be thought of as being obtained from a larger weight matrix in which a 'buffer zone' around the edges has been eliminated. This is similar to the common practice of dropping a number of observations at the beginning of the series when generating time series data, so that the influence of the initial value is reduced.

- f1. A univariate *ARMA* model is fitted to the time series for the aggregate using a *BIC* selection criterion with maximum number of lags  $k_{\max} = 4$  and the model is used to forecast.
- f2. A univariate *ARMA* model is fitted to each component using a *BIC* selection criterion with  $k_{\max} = 4$ . The univariate forecasts for each component are then aggregated.
- f3. A *VAR*(1) is estimated and the forecasts for each component are aggregated.
- f4. A Space-Time *AR*(1, 1) model is estimated and the forecasts aggregated.

When estimating the *VAR*(1), especially for large  $k$ , a large number of coefficients will not be significantly different from zero, due to the fact that the matrix  $W$  contains a large number of zeros. We thus consider a further forecasting scenario f3new that tries to eliminate this possible source of inefficiency.

- f3new. A *VAR*(1) is estimated equation by equation, the non-significant coefficients dropped, and the system re-estimated only using the significant lags. Forecasts for each component are then aggregated.

For each of the above methods, we divide the sample in two parts. Estimation is performed using the first  $T = 200$  observations, the estimated model is used to produce a sequence of  $n = 100$  forecasts for the out-of-sample period and the MSE is computed. We repeat the procedure for 1000 times for all four simulated processes and for all  $k$ . The results are reported in Tables 1 and 2 and Figures 1 and 2.

## 5.2 Interpretation of the results

As the results are robust to both specifications of the weight matrix, we only discuss the case of the first weight specification in (14). Figures 1 and 2 report the average MSE over the 1000

Monte Carlo replications for each data-generating process and for the different forecasting methods considered. The figures reveal that, in general, the aggregate of forecasts from the Space-Time *AR* model has the lowest average MSE, while the aggregate of univariate forecasts has the highest, with a widening gap as the number of regions increases. The inefficiency of forming univariate forecasts of the components is not surprising when the data exhibit high spatial dependence, since this dependence is ignored by the procedure. A less expected result is the clear ranking that emerges from Figure 2, where the data are characterized by weak spatial dependence. Even in this case, we conclude that forecasting with the Space-Time *AR* outperforms all competitors, and that the aggregate of univariate forecasts leads to high MSE. This result suggests that even though the spatial dependence among regions is weak, ignoring it altogether will lead to inefficiencies, especially when the number of units in the system increases. Table 1 reports the proportion of times each one of the five methods has the lowest MSE in the Monte Carlo iterations, while Table 2 analyzes the performance of the Space-Time *AR* compared to each one of the other methods.

The results in Table 1 reveal different dynamics for the various specifications of the data-generating process. In the majority of cases, the aggregate forecast from the Space-Time *AR* model (f4) is seen to be the most efficient, and a similar conclusion emerges from Table 2. Exceptions to the optimality of method f4 that can be seen in both Table 1 and Table 2 are the data-generating process with high AR and low spatial coefficients and the case of  $k = 4$ , for which the most efficient forecast is the direct forecast of the aggregate (f1). For  $k = 4$  the process satisfies the poolability condition (8), and therefore the Monte Carlo simulation seems to confirm the asymptotic result of optimality of the aggregate forecast when the poolability condition is satisfied (Proposition 6).

Another conclusion that emerges from Table 1 is that the worst performing method is in the majority of cases the aggregation of univariate forecasts for the components (f2). This is not surprising, since forecasting each component separately ignores the interrelationships

between regions and thus causes a loss of information, as well as introducing a large amount of parameter estimation uncertainty. In confirmation of the conclusion from Figure 2, notice that even in the case of low spatial coefficient (and low AR coefficient) it is never advisable to forecast with univariate models, a fact symbolized by zero entries in the table.

The relative performance of forecasting with the Space-Time *AR* model versus the *VAR*, i.e., of *f4* versus *f3* can be analyzed by looking at Table 2. It is shown that *f4* is more efficient than *f3*, with only one exception for  $k = 4$ . We can thus conclude that both asymptotic results derived in Proposition 6 hold even in relatively small samples.

A surprising conclusion that emerges from the tables is the bad performance of method *f3new*, the re-estimated *VAR*. From Table 2, we see that the Space-Time *AR* model is always more efficient than the re-estimated *VAR* and the efficiency improvement from imposing the spatial relationship instead of reducing the dimension of the system through a general-to-specific search is in general greater for higher number of components. These results seem to shed a negative light on the procedure of model selection through a general-to-specific search. It is known that this model selection criterion is not consistent, in the sense that asymptotically it has a non-zero probability of overfitting (see, e.g., Hall, 1994), and our results might in part be a reflection of this fact.

## 6 Conclusion

When forecasting the aggregate of variables measured over time and in different regions, it is plausible to assume that the individual components will be spatially correlated. A system is characterized by spatial correlation (see, e.g., Cliff and Ord, 1973) when data collected in neighboring regions are related in a systematic way, while the correlation disappears for regions that are far apart in space. In this paper, we analyzed different methods of forecasting the aggregate of spatially correlated disaggregates and compared their relative efficiency.

We presented both asymptotic results and small sample simulation results, that suggest that imposing constraints dictated by economic theory (in this case in the form of a spatial weight matrix) leads to improved forecast performance. One way to impose such constraints is to aggregate forecasts from a Space-Time Autoregressive model (see, e.g., Cliff *et. al.*, 1975), which was shown to be a restricted VAR with constraints dictated by *a priori* considerations on the amount of spatial correlation in the system. Forecasting with a Space-Time AR model offers a solution to the ‘curse of dimensionality’ that arises when we try to model panel data with a moderately large cross-sectional dimension using a VAR. The forecasting performance of a VAR was revealed to rapidly deteriorate in our Monte Carlo simulation as the number of regions increased. We also showed that ignoring spatial dependence, and simply aggregating univariate forecasts for each region, leads to highly inaccurate forecasts. The same conclusion was also reached in the case when the spatial dependence in the system is only weak, as has been found in some empirical investigations (e.g., Bronars and Jansen, 1987). Another result that was proven in the paper is that if the variables measured in different regions satisfy the ‘poolability’ condition, and especially for a large number of regions, there is a benefit in forecasting the aggregate variable directly. The simple test for poolability suggested by Kohn (1982) under the simplifying assumption of no estimation uncertainty is applicable in our framework, and it was seen to lead to similar inferences, even once estimation uncertainty is taken into account. Finally, we analyzed aggregating forecasts from a parsimonious VAR obtained through a general-to-specific search that eliminates the insignificant variables and re-estimates the model, a procedure which yielded highly inefficient forecasts.

Our results support some of the recent findings in the literature on forecasting with heavily parameterized econometric models. An example is Clements and Hendry (1998, Ch. 12), who show that the forecast accuracy of a number of models can be improved by imposing zero restrictions on coefficients that are close to zero, a procedure that in some situations outperforms forecasting with the true data-generating process. This last finding, in particu-

lar, offers a hopeful answer to the concern of misspecification of the constraints embedded in the Space-Time AR specification, since it implies that even invalid zero restrictions on some coefficients can potentially improve forecast accuracy. The greater accuracy of forecasting with parsimonious models could also explain the success in forecasting macro-variables with Bayesian VAR, which has been documented, among others, by Doan, Litterman and Sims (1984).

The analysis in the paper relied on many simplifications of the actual complexity of data measured in space and time and therefore it does not claim to be exhaustive. Our hope is that this paper will highlight issues of importance in the modeling of aggregates of variables that are possibly spatially dependent. We believe accounting for even weak spatial effects can lead to improved forecasts, and further econometric results in this direction are awaited.

## References

- [1] Aigner, D. J. and Goldfeld, S. M. (1974): ‘Estimation and Prediction from Aggregate Data when Aggregates are Measured More Accurately than Their Components’, *Econometrica*, 42, 113-134
- [2] Akaike, H. (1969): ‘Fitting Autoregressive Models for Prediction’, *Annals of the Institute of Statistical Mathematics*, 21, 243-247
- [3] Anderson, T. W. (1980): ‘Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models’ in D.R. Brillinger and G. C. Tiao (eds.), *Directions in Time Series*, Institute of Mathematical Statistics, 49-59
- [4] Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht
- [5] Baillie, R. T. (1980): ‘Prediction from ARMAX Models’, *Journal of Econometrics*, 12, 365-374
- [6] Bronars, S. G. and Jansen, D. W. (1987): ‘The Geographic Distribution of Unemployment Rates in the U.S.’, *Journal of Econometrics* 36, 251-279
- [7] Brown, P. E., Karsen, K. F., Roberts, G. O. and Tonellato, S. (2000): ‘Blur-Generated Non-Separable Space-Time Models’, *Journal of the Royal Statistical Society*, B, 62, 847-860
- [8] Brown, B. W. and Mariano, R. S. (1989): ‘Predictors in Dynamic Nonlinear Models’, *Econometric Theory*, 5, 430-452
- [9] Clements, M. P. and Hendry, D. F. (1998): *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting

- [10] Cliff, A. D. and Ord, J. K. (1973): *Spatial Autocorrelation*, Pion, London
- [11] Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K., Davies, R. (1975): *Elements of Spatial Structure*, Cambridge University Press, Cambridge
- [12] Doan, T., Litterman, R. and Sims, C. A. (1984): ‘Forecasting and Conditional Projection using Realistic Prior Distributions’, *Econometric Reviews*, 3, 1-100
- [13] Elhorst, J. P. (2001): ‘Dynamic Models in Space and Time’, *Geographical Analysis*, 33, 119-140
- [14] Granger, C. W. J. (1980): ‘Long Memory Relationships and the Aggregation of Dynamic Models’, *Journal of Econometrics*, 14, 227-238
- [15] Granger, C. W. J. (1987): ‘Implications of Aggregation with Common Factors’, *Econometric Theory*, 3, 208-222
- [16] Grunfeld, Y. and Griliches, Z. (1960): ‘Is Aggregation Necessarily Bad?’, *Review of Economics and Statistics*, 42, 1-13
- [17] Haining R., (1990): *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge
- [18] Hall, A. (1994): ‘Testing for a Unit Root in Time Series With Pretest Data-Based Model Selection’, *Journal of Business and Economic Statistics*, 12, 461-470
- [19] Kohn, R. (1982): ‘When is an Aggregate of a Time Series Efficiently Forecast by its Past?’, *Journal of Econometrics*, 18, 337-350
- [20] Lütkepohl, H. (1987): *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin Heidelberg

- [21] Pesaran, M. H., Pierse, R. G. and Kumar, M. S. (1989): ‘Econometric Analysis of Aggregation in the Context of Linear Prediction Models’, *Econometrica*, 57, 861-888
- [22] Pfeifer, P. E. and Deutsch, S. J. (1980): ‘A Three-Stage Iterative Procedure for Space-Time Modeling’, *Technometrics*, 22, 35-47
- [23] Pinkse, J. and Slade, M. E. (1998): ‘Contracting in Space: and Application of Spatial Statistics to Discrete-Choice Models’, *Journal of Econometrics* 85, 125-154
- [24] Stetzer, F. (1982): ‘Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments’, *Environment and Planning A*, 14, 571-584
- [25] Szulc, E. (2000): ‘Modelling the Space-Time Structure of the Economic Process on the Example of Unemployment’, in Zielinski, Z. (eds.) *Dynamic Econometric Models 4*, Torún, Poland
- [26] Wei, W. W. S. and Abraham, B. (1981): ‘Forecasting Contemporaneous Time Series Aggregates’, *Communications in Statistics-Theory and Methods*, A10, 1335-1344
- [27] Yamamoto, T. (1981): ‘Prediction of Multivariate Autoregressive-Moving Average Models’, *Biometrika*, 68, 485-492
- [28] Zellner, A. and Tobias, J. (2000): ‘A Note on Aggregation, Disaggregation and Forecasting Performance’, *Journal of Forecasting*, 19, 457-469

## 7 Appendix

**Proof of Proposition 1.** The weak inequality results for part a) and b) are proven by Lütkepohl (1987), Section 4.2 and Kohn (1982). The strong inequality in part a) follows from  $W \neq 0$  and Corollary 4.1.3 of Lütkepohl (1987), which states that the inequality is strong for a non-diagonal *VAR* operator. To prove part c), notice that the forecast of the vector  $\mathbf{x}_t$  from the *VAR* and the univariate forecasts of each component  $x_{it}$  from the Space-Time *AR* are based on the same information set  $\{\mathbf{x}_{t-j}; j \geq 1\}$  and thus coincide if the parameters of the DGP are known. ■

**Proof of Lemma 2.** See, e.g., Anderson (1980). ■

**Proof of Lemma 3.** See proof of Prop. 3.1 in Lütkepohl (1987). ■

**Proof of Lemma 4.** See, e.g., Akaike (1969). ■

**Proof of Proposition 5.** From (11), we have that  $MSE(\hat{y}_{t-1}(1)) = MSE(\hat{y}_{t-1}(1)) + \frac{1}{T}\Omega_y$ . The first component of the sum is derived in the following way: in the case of known parameters, the best one-step ahead linear forecast of  $\mathbf{x}_t$  from the *VAR*(1) model  $\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$  is  $\mathbf{x}_{t-1}(1) = \mathbf{B}\mathbf{x}_{t-1}$ , and therefore the forecast for the aggregate  $y_t = \boldsymbol{\iota}'\mathbf{x}_t$  is given by  $\hat{y}_{t-1}(1) = \boldsymbol{\iota}'\mathbf{x}_{t-1}(1) = \boldsymbol{\iota}'\mathbf{B}\mathbf{x}_{t-1}$ . The forecast error for  $\hat{y}_{t-1}(1)$  is then obtained as  $e_t(1) = y_t - \hat{y}_{t-1}(1) = \boldsymbol{\iota}'\boldsymbol{\varepsilon}_t$ , which leads to

$$MSE(\hat{y}_{t-1}(1)) = E(e_t(1)e_t(1)') = \boldsymbol{\iota}'E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\boldsymbol{\iota} = \boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} \quad (15)$$

From Lemma 4, it follows that for the forecast of the aggregate the asymptotic covariance matrix will be

$$\frac{1}{T}\Omega_y = \frac{1}{T}\boldsymbol{\iota}'\Omega_{\mathbf{x}}\boldsymbol{\iota} = \frac{1}{T}\boldsymbol{\iota}'k\Sigma_\varepsilon\boldsymbol{\iota} = \frac{k}{T}\boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} \quad (16)$$

and therefore the asymptotic *MSE* of the estimated aggregate forecast is obtained by summing (15) and (16). ■

**Proof of Proposition 6.** a) From Proposition 1 - c) and (11) it follows that

$$MSE^{(3)}(\hat{y}_{t-1}(1)) - MSE^{(4)}(\hat{y}_{t-1}(1)) = \frac{1}{T}(\Omega_y^{(3)} - \Omega_y^{(4)}). \quad (17)$$

For both forecasting scenarios f3 and f4, we have that  $\hat{y}_{t-1}(1) = \boldsymbol{\iota}'\hat{x}_{t-1}(1)$ . In f3,  $\hat{x}_{t-1}(1)$  is obtained by substituting the true parameter vector  $\theta_0^{(3)}$  with its ML estimator  $\hat{\theta}_T$  into  $x_{t-1}(1)$  and, similarly,  $\hat{x}_{t-1}(1)$  in f4 is obtained by substituting the true parameter vector  $\theta_0^{(4)}$  with its ML estimator  $\theta_T^*$  into  $x_{t-1}(1)$ . We have shown that the Space-Time  $AR(1, 1)$  model can be seen as a  $VAR(1)$  model with restrictions on the parameters, which implies that  $\theta_T^*$  is a restricted ML estimator of  $\theta^{(3)}$ . If the restrictions are true, the estimator  $\theta_T^*$  is more efficient than  $\hat{\theta}_T$ , i.e.,  $\Sigma_{\hat{\theta}} - \Sigma_{\theta^*}$  is positive semidefinite, where  $\Sigma_{\theta}$  is the asymptotic matrix of the ML estimator defined in Lemma 2. This, in turn, implies that

$$\Omega_x^{(3)} - \Omega_x^{(4)} = E \left[ \frac{\partial x_{t-1}(1)}{\partial \theta'} (\Sigma_{\hat{\theta}} - \Sigma_{\theta^*}) \frac{\partial x_{t-1}(1)}{\partial \theta} \right] \quad (18)$$

is positive semidefinite, which gives

$$MSE^{(3)}(\hat{y}_{t-1}(1)) - MSE^{(4)}(\hat{y}_{t-1}(1)) = \frac{1}{T}(\Omega_y^{(3)} - \Omega_y^{(4)}) = \frac{1}{T}\boldsymbol{\iota}'(\Omega_x^{(3)} - \Omega_x^{(4)})\boldsymbol{\iota} \geq 0. \quad (19)$$

b) If condition (8) is satisfied, the aggregate  $y_t$  is an  $AR(1)$ ,

$$y_t = \boldsymbol{\iota}'\mathbf{x}_t = \boldsymbol{\iota}'\mathbf{B}\mathbf{x}_{t-1} + \boldsymbol{\iota}'\boldsymbol{\varepsilon}_t = by_{t-1} + v_t \quad (20)$$

where  $v_t \sim i.i.d.N(0, \boldsymbol{\iota}'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{\iota})$ . Applying Lemma 4, it follows that

$$MSE^{(1)}(\hat{y}_{t-1}(1)) = MSE^{(1)}(y_{t-1}(1)) + \frac{1}{T}\Omega_y = \boldsymbol{\iota}'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{\iota} + \frac{1}{T}\boldsymbol{\iota}'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{\iota}. \quad (21)$$

The asymptotic  $MSE$  for forecasting method f3 is derived in Proposition 5:  $MSE^{(3)}(\hat{y}_{t-1}(1)) = \boldsymbol{\iota}'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{\iota} + \frac{k}{T}\boldsymbol{\iota}'\Sigma_{\boldsymbol{\varepsilon}}\boldsymbol{\iota}$  and thus, for  $k > 1$ , we have that  $MSE^{(1)}(\hat{y}_{t-1}(1)) < MSE^{(3)}(\hat{y}_{t-1}(1))$ . ■

**Table 1**

Proportion of times each forecasting method has lowest MSE

		Low spatial coefficient					High spatial coefficient				
N. regions		f1	f2	f3	f3new	f4	f1	f2	f3	f3new	f4
Low AR coeff.	4	.32	<u>0</u>	.26	0	.42*	.35*	<u>.01</u>	.22	.17	.25
	6	.35	<u>0</u>	.21	.01	.44*	.27	<u>.01</u>	.15	.13	.44*
	9	.35	<u>0</u>	.15	.01	.49*	.25	<u>0</u>	.11	.09	.55*
	16	.39	<u>0</u>	.08	0	.53*	.37	<u>0</u>	.08	.06	.49*
High AR coeff.	4	.35*	.15	.18	<u>.13</u>	.19	.35*	<u>.01</u>	.24	.19	.21
	6	.34*	.17	.15	<u>.11</u>	.23	.24	<u>.01</u>	.14	.17	.44*
	9	.38*	.13	.12	<u>.12</u>	.25	.12	<u>0</u>	.11	.21	.56*
	16	.38*	.15	<u>.07</u>	.13	.27	.32	<u>.01</u>	.12	.08	.48*

This table shows the proportion of times that each forecasting method has the lowest MSE in the Monte Carlo simulation. If forecasts from the five methods were all equally efficient, the number of times one particular model has the lowest MSE would be distributed as a Binomial  $(T, 0.2)$ , where  $T$  is the sample size. For  $T=1000$ , one can use the Normal approximation to the Binomial and derive a 95% confidence interval  $\hat{p} \pm 0.025$  for each proportion  $\hat{p}$  reported in the table. A ‘\*’ indicates that the forecasting method in the corresponding column is the best the highest number of times, while a ‘\_’ denotes the worst performing method. For example, for a number of regions  $k = 4$  and low spatial and AR coefficients, the aggregate of Space-Time AR forecasts (f4) is the best, since it has the lowest MSE 42% of the time. For the same data-generating process, the aggregate of univariate forecasts (f2) fares the worst, since it has lowest MSE 0% of the time.

**Table 2**

Proportion of times the Space-Time AR has lower MSE than other methods

		Low spatial coefficient				High spatial coefficient			
N. regions		f1	f2	f3	f3new	f1	f2	f3	f3new
Low AR coeff.	4	.58*	.99*	.64*	.98*	.43	.97*	.58*	.71*
	6	.56*	.99*	.72*	.98*	.61*	.98*	.73*	.82*
	9	.59*	.99*	.78*	.98*	.66*	.98*	.73*	.82*
	16	.56*	1*	.88*	.99*	.56*	.98*	.87*	.92*
High AR coeff.	4	.32	.67*	.59*	.64*	.40	.97*	.49	.67*
	6	.45	.68*	.69*	.66*	.66*	.98*	.69*	.78*
	9	.43	.69*	.76*	.69*	.71*	.99*	.78*	.85*
	16	.43	.68*	.87*	.74*	.59*	.99*	.78*	.88*

This table shows the proportion of times that the aggregate of Space-Time AR forecasts has a lower MSE than the other four methods in the Monte Carlo simulation. If forecasts from the two methods were equally efficient, the number of times the model in the column has lower MSE would be distributed as a Binomial  $(T, 0.5)$ , where  $T$  is the sample size. For  $T=1000$ , the 95% confidence interval for each proportion  $\hat{p}$  in the table is given by  $\hat{p} \pm 0.03$ . A ‘\*’ indicates that the Space-Time AR significantly outperforms the alternative method at the 5% significance level. For example, for a number of regions  $k = 4$  and low spatial and AR coefficients, the aggregate of Space-Time AR forecasts has a lower MSE than the forecast based on the aggregate (f1) 58% of the time.

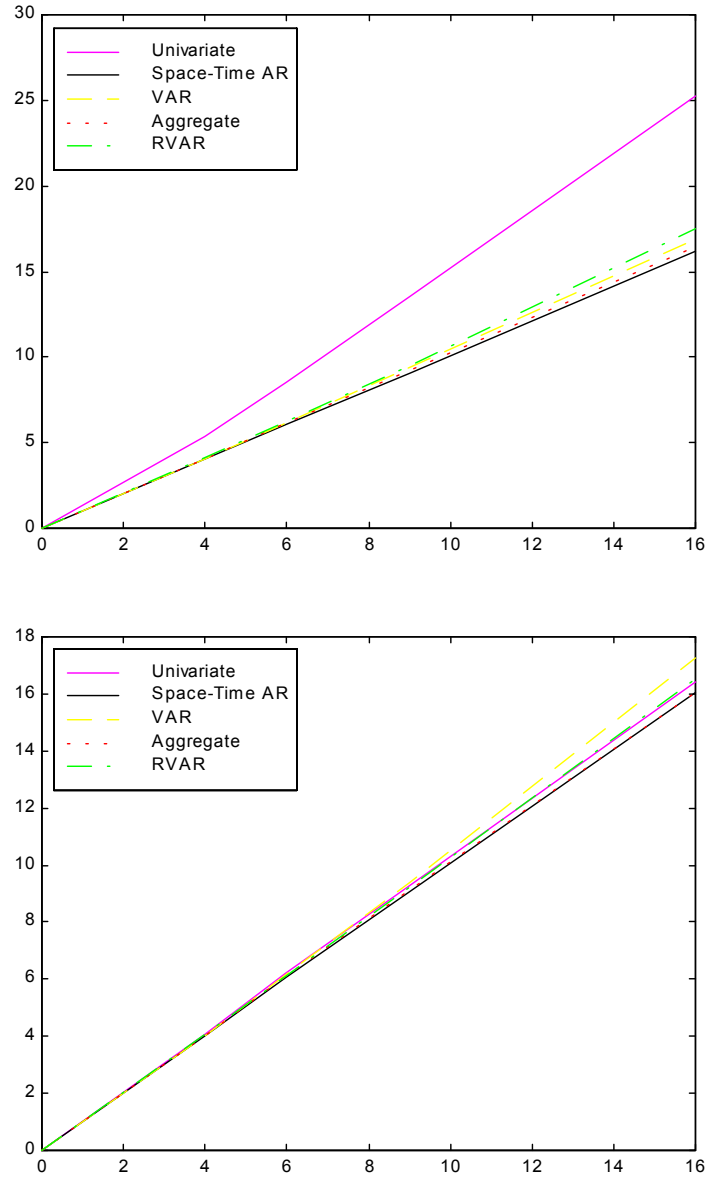


Figure 1: The Figure shows the average MSE (vertical axis) over the Monte Carlo replications for increasing number of regions (horizontal axis). Upper panel represents the DGP with  $\phi = .45$  and  $\psi = .45$ . Lower panel represents the DGP with  $\phi = .45$  and  $\psi = .1$ .

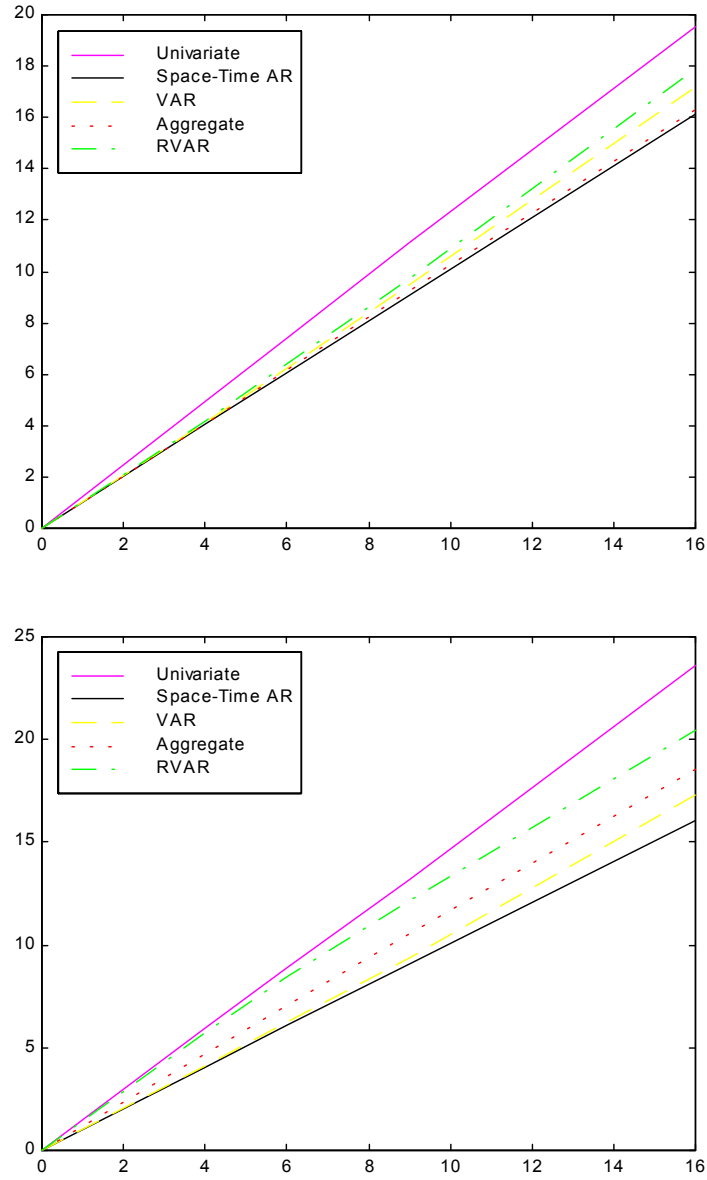


Figure 2: The Figure shows the average MSE (vertical axis) over the Monte Carlo replications for increasing number of regions (horizontal axis). Upper panel represents the DGP with  $\phi = .1$  and  $\psi = .45$ . Lower panel represents the DGP with  $\phi = .1$  and  $\psi = .1$ .