Title
Aggregationn of Space-Time Processes

Authors
Giacomini, Raffaella
Granger, Clive W.J.

Publication Date
2001-05-01

2001-07

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

AGGREGATION OF SPACE-TIME PROCESSES

BY

RAFFAELLA GIACOMINI

AND

CLIVE W.J. GRANGER

DISCUSSION PAPER 2001-07
MAY 2001

Aggregation of Space-Time Processes

Raffaella Giacomini and Clive W. J. Granger*

University of California, San Diego

May 2001

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

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 and Ross Starr for their 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 assumed to exist 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 between the variables is ignored (Grunfeld and Griliches (1960), Aigner and Goldfeld (1974)) or only generic contemporaneous correlation in the errors is allowed (Pesaran, Pierse and Kumar (1989)). The case when the correlation between variables is due to the presence of a common factor is analyzed in Granger (1980, 1987), who illustrates the implications of aggregating variables that depend on a common and an idiosyncratic factor. 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 bibliography can be found in Lütkepohl (1987). In the present paper, we consider an ideal situation where the variable of interest is observed in a moderately large

number of regions and over a long time period. As the cross-sectional dimension increases, modeling with vector ARMA processes 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, we can say that 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 Auto Regressive (Space-Time AR) models introduced by Cliff *et al.* (1975) and generalized by Pfeifer and Deutsch (1980)[1].

[1]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

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).

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 $i+1$. 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 $t-1$

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

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 the so-called 'edge effect' into our framework.[2] Suppose that the dependence in the diagram can be expressed as

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

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

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

the autoregressive subgroup

[2]The 'edge effect' results 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).

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

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

In sum, the process for the aggregate is approximately an $AR(1)$ and the coefficients of spatial dependence (ψ_1 and ψ_2) are incorporated in the autoregressive coefficient. Also, no matter what the process describing $x_{i,t}$ is, the aggregate process will be well approximated by an $AR(1)$ and we can thus conclude that aggregation causes in this case a 'simplification' of the dynamic properties of the variable.

3 The Space-time Auto Regressive model

A space-time model is a univariate time-series model that explicitly takes into account linear dependence lagged both in time and in space. Suppose that the variable of interest x_{it} is observed at each of k fixed locations ($i = 1, ..., k$) and over time ($t = 1, ..., T$). 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 relationship between the variables in the various regions under examination depends 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 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. Figure 1 shows the simplifying example of data collected over a regular grid. One of the possible criteria to define the neighboring sets for a region 0 is to consider as first-order neighbors

the cells that share a border with it (region 1), in which case the second-order neighbors for region 0 will be the first-order neighbors of the region 1 cells, and so on.

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

Figure 1. First and second order neighbors for region 0

The above definition of neighboring sets for a given region can be easily extended to the more realistic setting of observations arranged irregularly in space. 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 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, we can define the spatial lag of order s as

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

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. They are exogenous, nonstochastic and satisfy the following properties[3]:

1) $w_{ij}^{(s)} \geq 0$,

2) $w_{ii}^{(s)} = 0$,

3) $\sum_{j \in J_s} w_{ij}^{(s)} = 1$.

[3]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.

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 (Pinkse and Slade, 1998). 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. The overwhelming majority of the spatial econometric literature assumes the spatial weight matrix to be known in advance. As a consequence, 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 (Stetzer, 1982). Detailed references and a discussion of weight matrix specification can be found in Anselin (1988). In the remainder of the paper, we will conform to the standard practice in the spatial econometric literature of considering the weight matrix to be correctly specified.

The central issue of the paper, forecasting with data that exhibit spatial dependence, can be illustrated using the simplest form of space-time model, a Space-Time $AR(1,1)$ for the zero-mean variable x_{it}, which ignores dependence beyond the first temporal and spatial lags

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

The weights w_{ij} sum to one for each i and are non-zero only for first-order neighbors of region i. The disturbance ε_{it} is a white noise. The weights w_{ij} can be collected in a $k \times k$ matrix $W = (w_{ij})$, the spatial weight matrix, which allows one to define the first spatial lag for the $k-$dimensional vector $\mathbf{x}_t$ as

$$L^{(1)}\mathbf{x}_t = W\mathbf{x}_t.$$

Again, the spatial weight matrix has rows summing up to one and non-zero elements only for first-order neighbors of the region in row i.

With this notation for the spatial lag, the Space-Time $AR(1,1)$ model can be written in vector form as

$$\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \psi W \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{6}$$

where $\boldsymbol{\varepsilon}_t$ is a k-dimensional white noise with covariance matrix[4] $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Sigma_\varepsilon$. 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 $t-1$. From equation (6) it is clear that the Space-Time $AR(1,1)$ model is just a special case of a vector autoregressive model of order 1 ($VAR(1)$), where the autoregressive coefficient matrix is restricted to equal $\phi I_k + \psi W$. The generalization to higher spatial and temporal orders is straightforward. A Space-Time AR process can thus be obtained from a VAR by imposing a set of constraints on the coefficient matrices. It is in this sense that a space-time model is a refinement of a multivariate time-series model. Whereas a multivariate model allows all variables in a system to be related to all other variables, a space-time 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.

4 Forecasting aggregated space-time processes. Asymptotic results

Suppose that the goal is to forecast $y_t = \sum_{i=1}^{k} x_{it}$, the aggregate across regions of a set of k variables x_{it}, possibly related by spatial dependence. In this situation, there are several ways to form a forecast for the aggregate.

- f1. Forecast y_t directly by an $ARMA$ model fitted to the aggregate series

[4]The original formulation of the Space-Time AR model assumes uncorrelated disturbances across space (Pfeifer and Deutsch, 1980). Since our focus of interest is forecasting performance, we can relax such restrictive assumption and allow for contemporaneous correlation in the disturbances

- f2. Estimate and forecast univariate $ARMA$ models for each variable x_{it}, $i = 1, ..., k$. Aggregate the forecasts to obtain a prediction of the aggregate y_t.

- f3. Estimate a VAR on all k variables, use it to form forecasts of the individual components and aggregate.

- f4. Exploit the spatial relationships between the variables in the system and estimate a Space-Time AR model for each variable x_{it}, $i = 1, ..., k$. Construct the forecasts for each x_{it} and aggregate them.

In the following, we will attempt to determine which method of forecasting the aggregate is optimal. To simplify the analysis, we will restrict attention to the one-step-ahead optimal linear forecast of y_t, denoted by $y_{t-1}(1)$, and choose as a measure of forecast accuracy the forecasts' Mean Squared Error (MSE). In the remainder of the paper, we make use of the assumptions listed below:

Assumption 1. The data-generating process is the Space-Time $AR(1,1)$ in (6)

Assumption 2. The data-generating process is covariance stationary (i.e., the eigenvalues of the matrix $\phi I_k + \psi W$ are inside the unit circle)

Assumption 3. The orders of the estimated $ARMA$ models are either known or estimated consistently

A large part of this section is an extension to our space-time framework of the results collected in Lütkepohl (1987).

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 that the data-generating processes are known.

Proposition 1 *Let* $MSE^{(i)}(y_{t-1}(1))$, $i = 1, ..., 4$ *denote the asymptotic 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. The following ranking between alternative forecasting methods can then be established

a) The aggregate of forecasts from a VAR is more efficient than the aggregate of univariate $ARMA$ 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: $MSE^{(3)}(y_{t-1}(1)) = MSE^{(4)}(y_{t-1}(1))$.

Proof. 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 Corollary 4.1.3 in Lütkepohl (1987), which states that the inequality is strong for a non-diagonal VAR operator, as is the case when the system exhibits spatial correlation. 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 process are known. ■

As pointed out by Wei and Abraham (1981) the relative efficiency of forecasting the aggregate directly or aggregating univariate $ARMA$ processes 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 $VAR(p)$ process $B(L)\mathbf{x}_t = \boldsymbol{\varepsilon}_t$ if and only if

$$\boldsymbol{\iota}' B(L) = b(L)\boldsymbol{\iota}', \tag{7}$$

where $\boldsymbol{\iota}$ is a $k \times 1$ vector of ones and $b(L)$ is a scalar lag polynomial of order p. In other words, condition (7) guarantees that the aggregate of a $VAR(p)$ follows 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 (7) the 'poolability' condition). If the data-generating process is a Space-Time $AR(1,1)$, condition (7) becomes $\boldsymbol{\iota}'(\phi I_k + \psi W) = b\boldsymbol{\iota}'$,with b scalar, which corresponds to requiring the matrix $\phi I_k + \psi W$ to

have equal column sums. Since the typical column sum is of the form $\nu_j = \phi + \psi \sum_{i=1}^{k} w_{ij}$, we have that poolability is attained when the spatial weight matrix has equal columns sums. Loosely speaking, each column sum of W represents the total spatial 'influence' of region j over all other regions and therefore condition (7) says 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 (7) 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 data-generating process of the variable of interest is not known, and must be estimated. We see next how the estimation variability affects the asymptotic results that were discussed in the previous section. We further introduce the following assumptions.

Assumption 4. The coefficients of all the models considered are estimated by Maximum Likelihood (ML). The ML estimator of the parameter θ will be denoted as $\hat{\theta}_T$ where T is the sample size.

Assumption 5. Estimation and forecasting are based on independent and identically distributed processes.

For each model estimated in scenarios f1-f4 above, and under standard conditions, the estimated parameters of the forecasting models are asymptotically 'well - behaved'.

Proposition 2 *For each of the models estimated in scenarios f1-f4 the ML estimator of the parameters satisfies*

$$\hat{\theta}_T \xrightarrow{p} \theta \tag{8}$$

and

$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, \Sigma_{\hat{\theta}}), \tag{9}$$

where $\Sigma_{\hat{\theta}}$ *is the Hessian,*

$$\Sigma_{\hat{\theta}} = T\left[-E\frac{\partial^2 \ln l}{\partial\theta\partial\theta'}\right]^{-1}$$

Proof. See, e.g., White (1994). ■

We use the estimated models to build forecasts and compare their relative efficiency in terms of asymptotic MSE. In section 3 we showed how a Space-Time AR model can be obtained from a VAR model by imposing linear restrictions on the coefficient matrices. This implies that all models used for forecasting in f1-f4 are members of the (vector) $ARMA$ class, and thus we will first discuss the MSE measure of forecast performance for a general k-dimensional vector $ARMA$ process z_t. Lütkepohl (1987) proves the following result[5]

Proposition 3 *(Prop. 3.1, Lütkepohl (1987)). The forecast* $\hat{z}_{t-1}(1)$ *obtained by replacing the true parameter* $\theta = (\theta_1, ..., \theta_N)$ *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*

$$\sqrt{T}[\hat{z}_{t-1}(1) - z_{t-1}(1)] \xrightarrow{d} N(0, \Omega_z), \tag{10}$$

where

$$\Omega_z = E\left[\frac{\partial z_{t-1}(1)}{\partial\theta'}\Sigma_{\hat{\theta}}\frac{\partial z_{t-1}(1)}{\partial\theta}'\right], \tag{11}$$

[5]Although the result was given for a general forecast horizon h, we will restrict our attention to 1-step-ahead forecasts.

$$\frac{\partial z_{t-1}(1)}{\partial \theta'} = \left[\frac{\partial z_{t-1i}(1)}{\partial \theta_n}\right] \quad i = 1, ..., k; \quad n = 1, ..., N.$$

Also, the asymptotic MSE of $\hat{z}_{t-1}(1)$ *is given by*

$$MSE(\hat{z}_{t-1}(1)) = MSE(z_{t-1}(1)) + \frac{1}{T}\Omega_z \tag{12}$$

Equation (12) shows that the presence of parameter estimation adds a positive definite term $\frac{1}{T}\Omega_z$ (asymptotically vanishing) to the asymptotic MSE that would be obtained if 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 general expression for Ω_z is rather complicated for general stationary and invertible vector 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 the k-dimensional VAR(p) process* $\mathbf{B(L)x_t} = \boldsymbol{\varepsilon}_t$*, with* $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Sigma_\varepsilon$*. Then we have*

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

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

Proof. See, e.g., Akaike (1969, 1971). ■

Proposition 5 *Let* $\mathbf{x}_t$ *be the process in (6). When a forecast for the aggregate* y_t *is obtained by estimating a* $VAR(1)$ *model on a sample of size* T *and aggregating the forecasts of the components, the asymptotic MSE of the aggregate forecast is*

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

Proof. From (12), we have that $MSE(\hat{y}_{t-1}(1)) = MSE(y_{t-1}(1)) + \frac{1}{T}\Omega_y$. The first component of the sum is derived in the following way: in the case of known data-generating

process, 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 $y_{t-1}(1) = \boldsymbol{\iota}'\mathbf{x}_{t-1}(1) = \boldsymbol{\iota}'\mathbf{B}\mathbf{x}_{t-1}$. The forecast error for $y_{t-1}(1)$ is then obtained as $e_t(1) = y_t - y_{t-1}(1) = \boldsymbol{\iota}'\boldsymbol{\varepsilon}_t$, which leads to

$$MSE(y_{t-1}(1)) = E(e_t(1)e_t(1)') = \boldsymbol{\iota}'E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\boldsymbol{\iota} = \boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} \tag{14}$$

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

$$\frac{1}{T}\Omega_y = \frac{1}{T}\boldsymbol{\iota}'\Omega_{\mathbf{x}}\boldsymbol{\iota} = \frac{1}{T}\boldsymbol{\iota}'k\Sigma_\varepsilon\boldsymbol{\iota} = \frac{k}{T}\boldsymbol{\iota}'\Sigma_\varepsilon\boldsymbol{\iota} \tag{15}$$

and therefore the asymptotic MSE of the estimated aggregate forecast is obtained by summing (14) and (15). ■

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 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) and b) are no longer valid in general. Instead, for each given specification of the Space-Time $AR(1,1)$ model, 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 rankings we can establish are the following.

Proposition 6 *In the presence of estimation uncertainty, the following relative efficiency results among forecasting methods f1-f4 can be established:*

a) The aggregate of univariate Space-Time AR forecasts for each component 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 (7) is satisfied, it is more efficient to forecast the aggregate directly rather than aggregate forecasts from a VAR. $MSE^{(1)}(\hat{y}_{t-1}(1)) < MSE^{(3)}(\hat{y}_{t-1}(1))$.

Proof. a) From Proposition 1, c) and (12) it follows that

$$MSE^{(3)}(\hat{y}_{t-1}(1)) - MSE^{(4)}(\hat{y}_{t-1}(1)) = \frac{1}{T}(\Omega_y^{(3)} - \Omega_y^{(4)}). \tag{16}$$

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 $\theta^{(3)}$ 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 $\theta^{(4)}$ with its ML estimator θ_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 θ_T^* is a restricted ML estimator of $\theta^{(3)}$. If the restrictions are true, the estimator θ_T^* is more efficient than $\hat{\theta}_T$, i.e., $\Sigma_{\hat{\theta}} - \Sigma_{\theta^*}$ is positive semidefinite, where Σ_θ is the asymptotic matrix of the ML estimator in (9). This, in turn, implies that

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

is positive semidefinite, which gives

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

b) If condition (7) is satisfied, the aggregate y_t follows an $AR(1)$,

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

with $E(v_t^2) = \boldsymbol{\iota}'\Sigma_e\boldsymbol{\iota}$. Applying Lemma 4, it follows that

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

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

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 (7) cannot be rejected, it is also always preferable to forecast the aggregate directly than to aggregate forecasts from the $VAR(1)$, and the benefit increases 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

$k = 4$:

1	2
3	4

$k = 6$:

1	2	3
4	5	6

$k = 9$:

1	2	3
4	5	6
7	8	9

$k = 16$:

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

.

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

$$W^{(4)} = \begin{bmatrix} 0 & .5 & .5 & 0 \\ .5 & 0 & 0 & .5 \\ .5 & 0 & 0 & .5 \\ 0 & .5 & .5 & 0 \end{bmatrix}, \quad W^{(6)} = \begin{bmatrix} 0 & .5 & 0 & .5 & 0 & 0 \\ .33 & 0 & .33 & 0 & .33 & 0 \\ 0 & .5 & 0 & 0 & 0 & .5 \\ .5 & 0 & 0 & 0 & .5 & 0 \\ 0 & .33 & 0 & .33 & 0 & .33 \\ 0 & 0 & .5 & 0 & .5 & 0 \end{bmatrix} \tag{17}$$

where row 1 contains the weights for region 1's first order neighbors (2 and 3 in $W^{(4)}$, 2 and 4 in $W^{(6)}$), row 2 the weights for region 2's first order neighbors (1 and 4 in $W^{(4)}$, 1, 3 and 5 in $W^{(6)}$) 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 entry in the matrices in (17) above will equal 0.25.[6]

For each k, we generate data from four different processes, reflecting different assumptions about the value of the autoregressive coefficient ϕ and of the spatial coefficient ψ. We consider the four following specifications

- 1. $\phi = .45$, $\psi = .45$
- 2. $\phi = .45$, $\psi = .1$
- 3. $\phi = .1$, $\psi = .45$

[6] 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.

- 4. $\phi = .1$, $\psi = .1$

The predictors described above in scenarios f1-f4 are computed as follows

- 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 f3* that tries to eliminate this possible source of inefficiency;

- f3*. A $VAR(1)$ is estimated equation by equation, the non-significant coefficients dropped, and the system re-estimated only using the significant lags. Forecasts are then formed for each component and 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 observations and the relative 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, 2 and Figures 2-5.

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 (17). Figures 2-5 report the average MSE over the 1000 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 gap widening 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 5, where the data are characterized by weak spatial dependence. Even in this case, it is seen 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. For each of the four model specifications and for each of the four levels of k, both the best and the worst performing models are emphasized. 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 $T = 1000$, we can use the Normal approximation to the Binomial and derive a 95% confidence interval $\hat{p} \pm 0.025$ for each proportion reported in Table 1. Table 2 analyzes the performance of the Space-Time AR compared to each one of the other methods. The benchmark of comparison is in this case a proportion of 0.5, indicating that the two methods under examination are equally efficient (the confidence interval will be in this case $\hat{p} \pm 0.03$).

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 (7), and therefore the Monte Carlo result is a confirmation of the asymptotic optimality of forecasting the aggregate when such condition is satisfied (Prop. 6). For higher levels of k, the weight matrix has unequal column sums, which in principle would make the process fail condition (7). When the spatial coefficient is small and the AR coefficient is high, however, the stronger effect of the equal AR coefficients dominates the heterogeneity in the spatial component, in which case the process would most likely pass the test for poolability suggested by Kohn (1982).

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 5, 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 entries in the table that are virtually zero.

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 f3*, 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 (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 (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 the present 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 (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 known data-generating process 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 was seen to lead to 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 particular, 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] Akaike, H. (1971): "Autoregressive Model Fitting for Control", *Annals of the Institute of Statistical Mathematics,* 23, 163-180

[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., Karesen, 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] Clements, M. P. and Hendry, D. F. (1998): *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting

[9] Cliff, A. D. and Ord, J. K. (1973): *Spatial Autocorrelation*, Pion, London

[10] Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K., Daview, R. (1975): *Elements of Spatial Structure*, Cambridge University Press, Cambridge

[11] Doan, T., Litterman, R. and Sims, C. A. (1984): "Forecasting and Conditional Projection using Realistic Prior Distributions ", *Econometric Reviews*, 3, 1-100

[12] Elhorst, J. P. (2000): "Dynamic Models in Space and Time ", manuscript, University of Groningen, Netherland

[13] Granger, C. W. J. (1987): "Implications of Aggregation with Common Factors ", *Econometric Theory*, 3, 208-222

[14] Granger, C. W. J. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models ", *Journal of Econometrics,* 14, 2

[15] Grunfeld, Y. and Griliches, Z. (1960): "Is Aggregation Necessarily Bad? ", *Review of Economics and Statistics*, 42, 1-13

[16] Haining R., (1990): *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge

[17] 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

[18] Kohn, R. (1982): "When is an Aggregate of a Time Series Efficiently Forecast by its Past? ", *Journal of Econometrics*, 18, 337-350

[19] Lütkepohl, H. (1987): "Forecasting Aggregated Vector ARMA Processes ", Springer-Verlag, Berlin Heidelberg

[20] 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

[21] Pfeifer, P. E. and Deutsch, S. J. (1980): "A Three-Stage Iterative Procedure for Space-Time Modeling ", *Technometrics,* 22, 35-47

[22] 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

[23] Stetzer, F. (1982): "Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments ", *Environment and Planning A,* 14, 571-584

[24] 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

[25] Wei, W. W. S. and Abraham, B. (1981): "Forecasting Contemporal Time Series Aggregates ", *Communications in Statistics-Theory and Methods,* A10, 1335-1344

[26] White, H. (1994): *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge

[27] Yamamoto, T. (1981): "Prediction of Multivariate Autoregressive-Moving Average Models ", *Biometrika*, 68, 485-49227-238

[28] Zellner, A. and Tobias, J. (2000): "A Note on Aggregation, Disaggregation and Forecasting Performance ", *Journal of Forecasting*, 19, 457-469

TABLE 1: Proportion of times each method has lowest MSE.

		LOW SPATIAL COEFF.					HIGH SPATIAL COEFF.				
	Number of regions	f1	f2	f3	f3*	f4	f1	f2	f3	f3*	f4
		Aggr.	Univ.	VAR	RVAR	Space-Time AR	Aggr.	Univ.	VAR	RVAR	Space-Time AR
LOW AR COEFF.	K=4	0.317	0.003	0.258	0.005	0.417*	0.352*	0.012	0.219	0.175	0.242
	K=6	0.345	0.001	0.202	0.012	0.440*	0.266	0.012	0.149	0.129	0.444*
	K=9	0.351	0.001	0.152	0.008	0.488*	0.252	0.009	0.113	0.088	0.538*
	K=16	0.395	0.000	0.078	0.003	0.524*	0.370	0.009	0.079	0.057	0.485*
HIGH AR COEFF.	K=4	0.352*	0.151	0.178	0.129	0.190	0.349*	0.014	0.244	0.185	0.208
	K=6	0.339*	0.172	0.145	0.110	0.234	0.238	0.011	0.142	0.168	0.441*
	K=9	0.382*	0.136	0.115	0.121	0.246	0.115	0.002	0.114	0.213	0.556*
	K=16	0.377*	0.147	0.068	0.134	0.274	0.318	0.006	0.117	0.080	0.479*

This table presents the proportion of times that each forecasting method has the lowest MSE. An ' * ' indicates that the forecasting method in the corresponding column is the best the highest number of times, while an '_' denotes the worst performing method. For example, for 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 41.7% of the time. For the same data-generating process, the aggregate of univariate forecasts (f2) is the worst, since it has lowest MSE 0.3% of the time.

TABLE 2: Proportion of times Space-Time AR has lower MSE than other 4 methods.

		LOW SPATIAL COEFF.				HIGH SPATIAL COEFF.			
	Number of regions	f1	f2	f3	f3*	f1	f2	f3	f3*
		Aggr.	Univ.	VAR	RVAR	Aggr.	Univ.	VAR	RVAR
LOW AR COEFF	K=4	0.580*	0.995*	0.636*	0.986*	0.432	0.975*	0.585*	0.708*
	K=6	0.569*	0.998*	0.724*	0.984*	0.611*	0.981*	0.730*	0.818*
	K=9	0.591*	0.998*	0.787*	0.989*	0.659*	0.981*	0.800*	0.873*
	K=16	0.564*	1.000*	0.886*	0.991*	0.562*	0.982*	0.871*	0.921*
HIGH AR COEFF	K=4	0.422	0.673*	0.592*	0.644*	0.404	0.976*	0.494	0.670*
	K=6	0.458	0.682*	0.693*	0.662*	0.661*	0.981*	0.696*	0.782*
	K=9	0.431	0.690*	0.758*	0.694*	0.712*	0.996*	0.778*	0.855*
	K=16	0.437	0.689*	0.872*	0.741*	0.595*	0.991*	0.780*	0.883*

This table presents the proportion of times that the aggregate of Space-Time AR forecasts has a lower MSE than the other 4 methods. An ‘ * ’ indicates that the Space-Time AR significantly outperforms the alternative method at the 5% level. For example, for 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) 58.0% of the time.

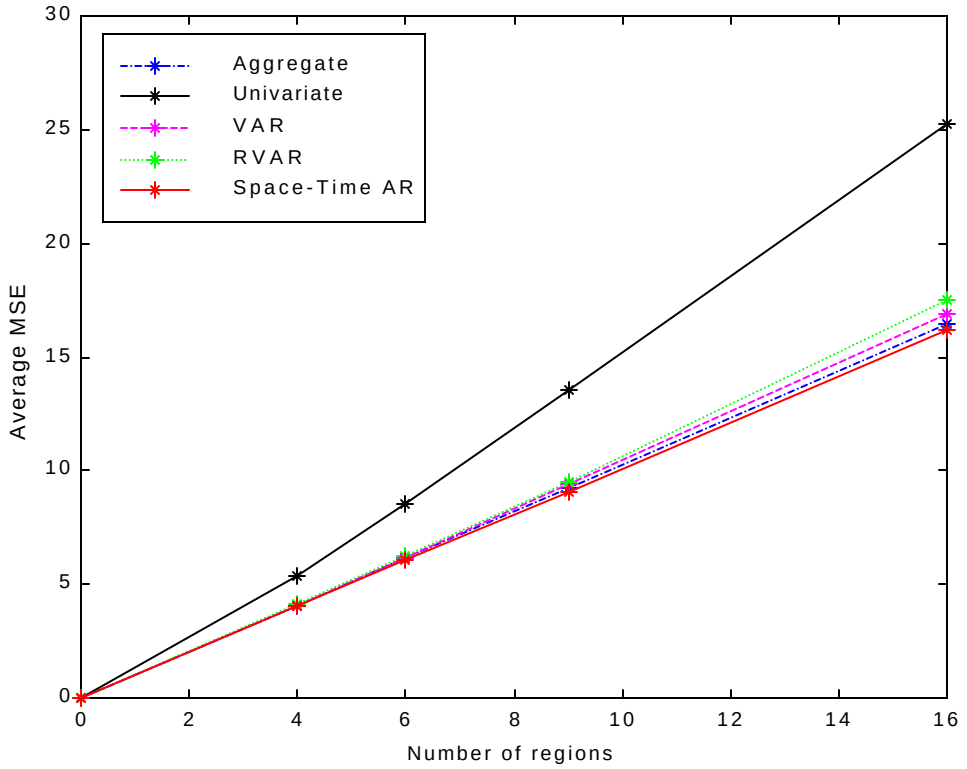

Figure 2. Average MSE. $f = 0.45$ $y = 0.45$

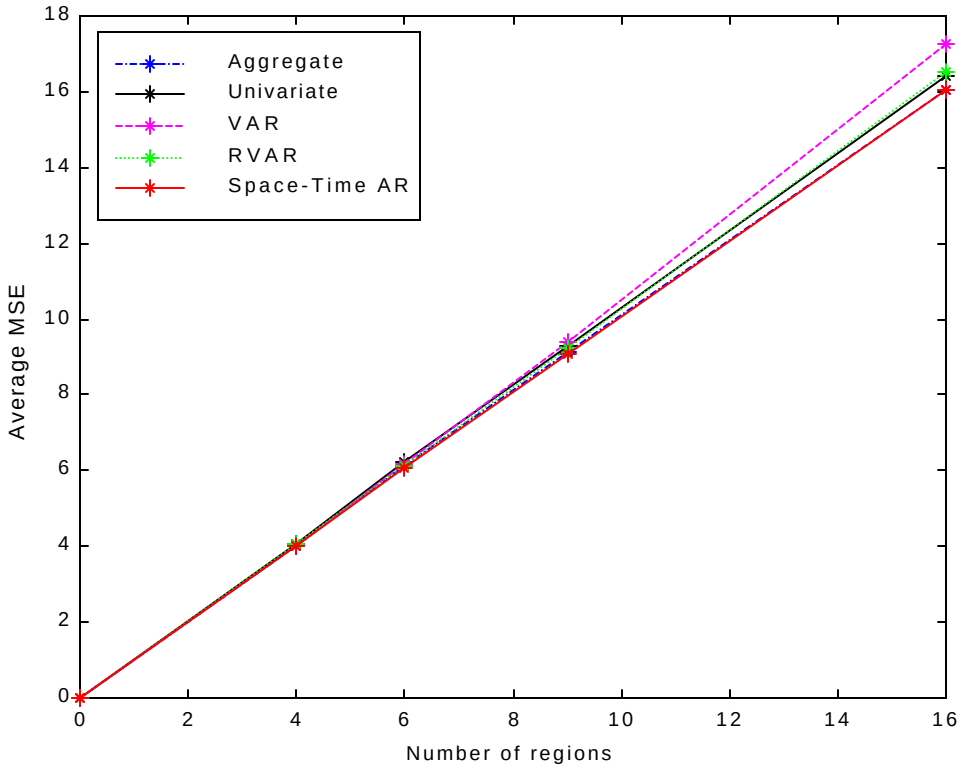

Figure 3. Average MSE. $f = 0.45$ $y = 0.1$

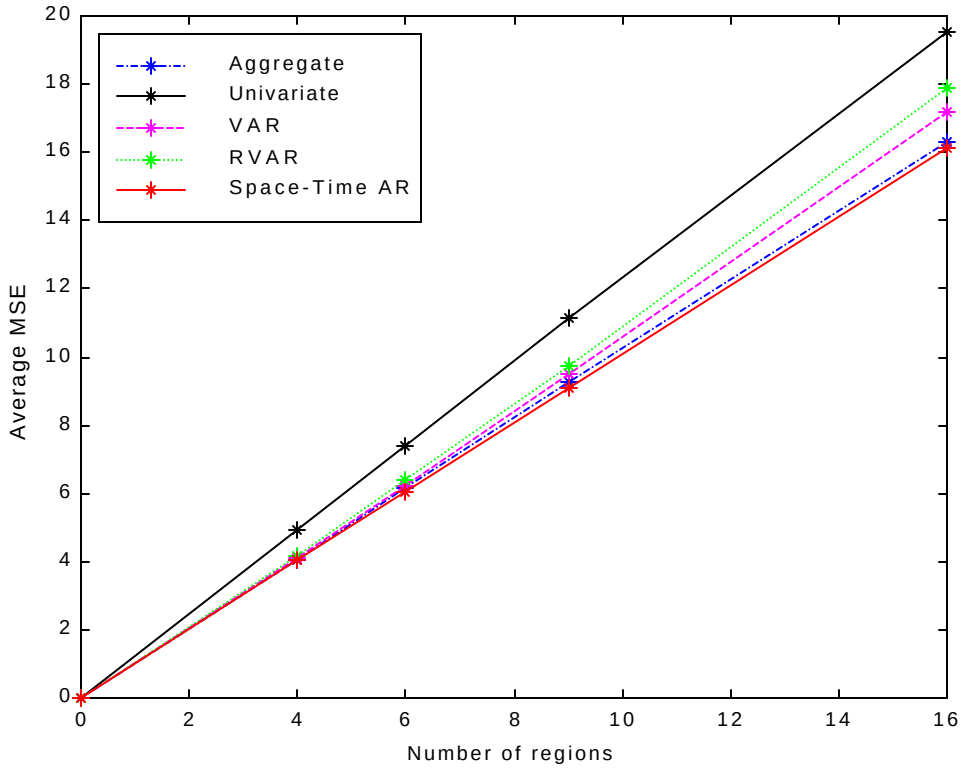

Figure 4. Average MSE. $f = 0.1$ $y = 0.45$

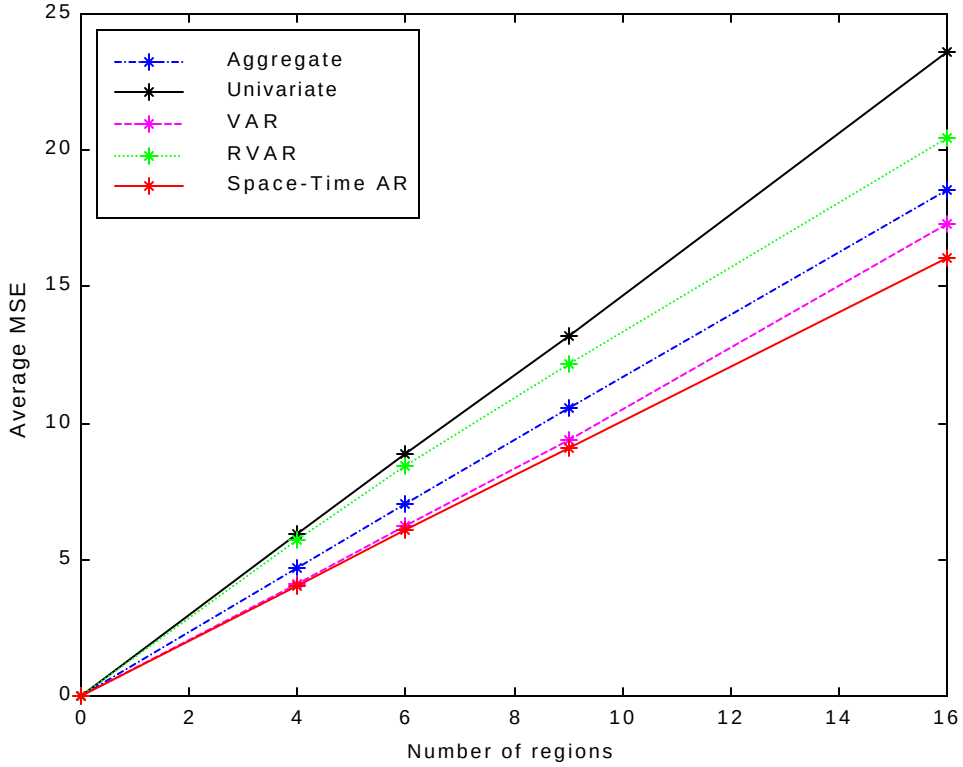

Figure 5. Average MSE. $f = 0.1$ $y = 0.1$