SPATIAL-TIME SERIES MODELING: A REVIEW OF THE PROPOSED METHODOLOGIES

by

Yiannis Kamarianakis

REAL 03-T-19 May, 2003

SPATIAL-TIME SERIES MODELING: A REVIEW OF THE PROPOSED METHODOLOGIES

Yiannis Kamarianakis

Regional Analysis Division, Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas, Vasilika Vouton, P.O. Box 1527, GR 711 10, Heraklion Crete, Greece

and Regional Economics Applications Laboratory, University of Illinois, 607 S. Mathews, #318, Urbana, Illinois, 61801-3671.

Abstract

This paper discusses three modeling techniques, which apply to multiple time series data that correspond to different spatial locations (spatial time series). The first two methods, namely the Space-Time ARIMA (STARIMA) and the Bayesian Vector Autoregressive (BVAR) model with spatial priors apply when interest lies on the spatio-temporal evolution of a single variable. The former is better suited for applications of large spatial and temporal dimension whereas the latter can be realistically performed when the number of locations of the study is rather small. Next, we consider models that aim to describe relationships between variables with a spatio-temporal reference and discuss the general class of dynamic space-time models in the framework presented by Elhorst (2001). Each model class is introduced through a motivating application.

1. Introduction

Research in statistical/econometric models that describe the spatio-temporal evolution of a single variable or multi-variable relationships in space and time started in the mid-seventies and has significantly increased during the last twenty years since it's closely related to the progress in computer technology and the existence of large databases. Cliff and Ord (1975) were the first to perform a model for the relationship between two variables in space and time[a]; since then several techniques have been developed corresponding to different inferential needs and data types. The present paper aims to summarize the proposed methodologies discussing each one through a motivating example that points out the cases where each model class is best suited.

The STARIMA model class developed at the early eighties by Pfeifer and Deutsch (1980a, 1980b, 1981a, 1981b, 1981c) is presented in section 2.1. Similar to ARIMA model building (see Box et al. 1994) for univariate time series, STARIMA model building is a three-stage procedure (identification–estimation–diagnostic checking). Although tedious in its implementation it has been applied to numerous applications ranging from environmental (Pfeifer and Deutsch 1981a, Stoffer 1986), to epidemiological (Pfeifer and Deutsch 1980a), and econometric (Pfeifer and Bodily 1990). The motivating example in this case comes from traffic flow modeling where, based on measurements taken from a set of loop detectors in a very frequent basis, a single statistical model describes the evolution of traffic conditions in an urban network. Kamarianakis and Prastacos (2002, 2003) used the hierarchical neighbor specification of the STARIMA methodology to capture the causality relations due to road network topology; moreover they performed a forecasting experiment where despite their very parsimonious formulation STARIMA models performed very well.

STARIMA models, although well suited for applications of large spatial scale they appear to be too parsimonious when the spatial time series study involves only a few measurement locations. As Giacomini and Granger (2001) point out the STARIMA class can be derived through a (nontrivial) transformation of the Vector Autoregressive

[a] Cliff and Ord's approach falls in the general model class presented at section 3.1.

Moving Average (VARMA) model; the transformation is in fact a restriction related to the neighborhood structure as revealed by a set of weight matrices. When the number of locations involved in the time series study is very small the researcher may proceed through VARMA model specification that pertains to the estimation of $\left(p \times S^2\right) + S$ parameters (*p* time lags, *S* locations). As the number of locations increases, the over-parameterized VARMA formulation leads to a large number of statistically non-significant parameters. LeSage and Krivelyova (1999) proposed a class of prior distributions for the Bayesian implementation of the VAR (BVAR) model that loosely constrains to zero the parameters that correspond to non-neighboring locations and large temporal lags. Section 2.2 presents BVAR models with spatial priors. The example application in this case is a model for employment time series that correspond to eight different American states.

The third part of the paper discusses models for multi-variable spatial time series. We focus on the general class of dynamic space-time models as formulated by Elhorst (2001). Even in the case this model class includes only temporal and spatial lags of the response as explanatory variables it differs from the models presented at the second part since it involves *instantaneous* spatial terms. A significant feature of this approach is that it can be transformed to take the form of an equilibrium correction model that permits the quantification of both long-term equilibrium relationships and short-term dynamics. Model order selection via classical procedures appears to be problematic since non-nested models may have to be compared. The example application in this case is a model for the space-time relation between employment and labor force participation.

Another class of models that apply to spatial time series are the Seemingly Unrelated Regressions (SUR) presented first by Zellner(1962); in this case each regression equation corresponds to a different location and the geographical relations are modeled implicitly in the covariance matrix of the system of equations. Anselin (1988) presented an alternative SUR formulation, the spatial SUR. In spatial SUR each equation corresponds to a different time period; in contrast with the simple SUR, to perform spatial SUR the investigator must have a dataset of a larger spatial rather than temporal dimension. When the number of cross-sections is larger than the number of time periods involved in a study, we enter the field of panel data models. Panel data and SUR models are not discussed in this paper. For a thorough discussion on the former see chapter 12 of Johnston and DiNardo (1997).

2. SPATIAL-TIME SERIES MODELS OF A SINGLE VARIABLE

2.1 The STARIMA model class

Motivating Application: Traffic Flow Modeling

Traffic flow data in metropolitan areas are collected by loop detectors located at major arterials of the road network. The detectors provide traffic volumes (number of cars that passed over the detector in a specific time interval, usually one minute), densities (proportion of time over a specific time interval that cars were over the detector) and speeds. Figure 1 depicts a set of loop detectors at the road network of Athens, Greece.

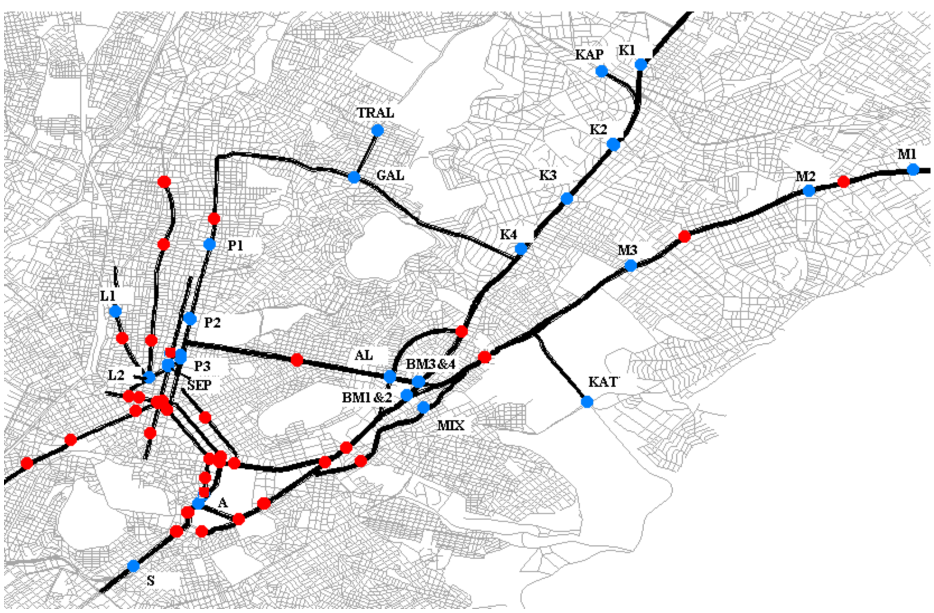

FIGURE 1: Loop detectors in a road network

In traffic flow systems tree structures are the most common method for network representation. The direction of the vectors of the tree follows the permitted traffic direction, whereas traffic flow measurements are taken at specific points of the network (Figure 2). If we assume that the traffic flow process forms a “black-box” network, i.e.

one that does not have access to any information other than past or present flows, then from Figure 2 it is clear that some measurement locations may not be connected through a path and therefore may act independently. If we also ignore any external effects and consider the distance between the measurement locations to be sufficiently long so as no congestion effects are introduced to disturb the flow pattern, no measurement location will be influenced by actions occurring downstream from it. Thus, downstream locations only depend on upstream locations but not vice versa. The question that has to be answered is how to exploit this structure in model identification and yet retain the statistical properties of the traffic flow process. The spatial topological relationships of a network as the one presented in Figure 2 can be introduced through a hierarchical ordering for the neighbors of each measurement site. This is the basis for system structuring using STARIMA model building. We shall call W_l a square $N \times N$ l^{th} order weight matrix with elements $w_{ij}^{(l)}$ that are nonzero only in the case that the measurement locations i and j are "l^{th} order neighbors". First order neighbors are understood to be closer than second order ones, which are closer than third order neighbors and so on.

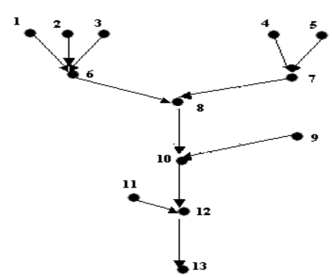

FIGURE 2. The typical road network tree structure for traffic flow. The dots represent measurement locations and the arrows the direction of flow.

The weights $w_{ij}^{(l)}$ are taken so that $\sum_{j=1}^{N} w_{ij}^{(l)} = 1$ and W_0 is the identity matrix since each site is its own zeroth order neighbour. Applying this rule to the network of Figure 2 and assigning equal weights to the l^{th} order neighbours of each site yields the following weight matrices for spatial lags 1 and 2:

$$
W_1=\begin{array}{c}
\\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\\13
\end{array}
\begin{array}{ccccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13\\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0.33 & 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}
\qquad
W_2=\begin{array}{c}
\\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\\13
\end{array}
\begin{array}{ccccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13\\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0
\end{array}
$$

Additional features such as the distances of each neighboring pair of sites are usually incorporated into the weighting matrices through an appropriate selection of weights**.**

Model Formulation

In the early eighties Pfeifer and Deutsch (1980a, 1980b, 1981a, 1981b, 1981c) introduced the STARIMA methodology. Here is a characterization of this model class by its creators:

"…*Processes amenable to modeling via this class are characterized by a single random variable observed at N fixed sites in space wherein the dependencies between the N time series are systematically related to the location of the sites. A hierarchical series of* $N \times N$ *weighting matrices specified by the model builder prior to analyzing the data is the basic mechanism for incorporating the relevant physical characteristics of the system into the model form. Each of the N time series is simultaneously modeled as a linear combination of past observations and disturbances at neighboring sites. Just as univariate ARIMA models reflect the basic idea that the recent past exerts more influence than the distant past, so STARIMA models reflect (through the specification of the weighting matrices) the idea that near sites exert more influence in each other than distant ones*."

Thus the STARIMA model class expresses each observation at time t and location i as a weighted linear combination of previous observations and innovations lagged both in space and time. The basic mechanism for this representation is the hierarchical ordering of the neighbors of each site and a corresponding sequence of weighting matrices as presented in the previous paragraph. The specification of the weighting matrices is a matter left to the model builder to capture the physical properties that are being considered endogenous to the particular spatial system being analyzed.

If Z_t is the $N \times 1$ vector of observations at time *t* at the *N* locations within the road network then the seasonal STARIMA model family is expressed as,

$$\Phi_{P,\Lambda}\left(B^S\right)\phi_{p,\lambda}(B)\nabla_S^D\nabla^d Z_t = \Theta_{Q,M}\left(B^S\right)\theta_{q,m}(B)a_t \tag{1}$$

where

$$\Phi_{P,\Lambda}\left(B^S\right) = I - \sum_{k=1}^{p}\sum_{l=0}^{\Lambda_k}\Phi_{kl}W_l B^{kS} \tag{1a}$$

$$\phi_{P,\lambda}(B) = I - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W_l B^{k} \tag{1b}$$

$$\Theta_{Q,M}\left(B^S\right) = I - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\Theta_{k,l}W_l B^{kS} \tag{1c}$$

$$\theta_{q,m}(B) = I - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W_l B^{k}\ . \tag{1d}$$

Φ_{kl} and ϕ_{kl} are respectively the seasonal and nonseasonal autoregressive parameters at temporal lag *k* and spatial lag *l*; similarly Θ_{kl} and θ_{kl} are the seasonal and nonseasonal moving average parameters at temporal lag *k* and spatial lag *l*; *P* and *p* are the seasonal and nonseasonal autoregressive orders; *Q* and *q* are the seasonal and nonseasonal moving average orders. Λ_k and λ_k are the seasonal and nonseasonal spatial orders for the k^{th} autoregressive term; M_k and m_k are the seasonal and nonseasonal spatial orders for the k^{th} moving average term; and *D* and *d* are, respectively, the number of seasonal and nonseasonal differences required, where ∇_S^D and ∇^d are the seasonal and nonseasonal difference operators, such that i.e., $\nabla_S^D = \left(I - B^S\right)^D$ and $\nabla^d = (I - B)^d$ with seasonal lag *S*. Finally, a_t is the random, normally distributed, error vector at time *t* with statistics:

$$E\{a_t\} = 0 \tag{2a}$$

$$E\{a_t a'_{t+s}\} = \begin{cases} G & if \quad s = 0 \\ 0 & if \quad s \neq 0 \end{cases} \tag{2b}$$

$$E\{Z_t a'_{t+s}\} = 0 \text{ for } s > 0\,. \tag{2c}$$

Equation (1) is referred to as a seasonal multiplicative STARIMA model of order $(p_\lambda, d, q_m)\times(P_\Lambda, D, Q_\mathrm{M})_S$.

When there is no seasonal component (quite unlikely in traffic flow) and *d*=0 the model collapses to the easier to interpret STARMA model which is of the form

$$Z_t = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W_l Z_{t-k} - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W_l a_{t-k} + a_t \tag{3}$$

where *p* is the autoregressive order, *q* is the moving average order, λ_k is the spatial order of the k^{th} autoregressive term, m_k is the spatial order of the k^{th} moving average term, ϕ_{kl} and θ_{kl} are parameters to be estimated and W_l is the $N\times N$ matrix for spatial order *l* and a_t is the random normally distributed innovation or disturbance vector at time *t*.

STARMA models can be viewed as special cases of the Vector Autoregressive Moving Average (VARMA) models (Lutkerpohl 1987, 1993). The VARMA models use general $N\times N$ autoregressive and moving-average parameter matrices to represent all autocorrelations and cross-correlations within and among the *N* time series. If the diagonal elements in these matrices are assumed to be equal (as in the case where the *N* series represent a single random process operating at different sites) and the off-diagonal elements are assumed to be a linear combination of the W_l weight matrices then the general VARMA family collapses to the STARMA model class. The VARMA model class on the other hand, can be viewed as a special case of the state-space model, which is the only multivariate technique presented in the literature of traffic-flow modelling so far. It's obvious from (1) and (3) that the STARIMA methodology provides a great reduction in the number of parameters that have to be estimated compared to the VARMA or the state-space model classes and thus facilitates the performance of applications of large spatial scale (large number of measurement locations).

Identification stage

Model identification is the first of the three stages of the iterative procedure commonly attributed to Box et al. (1994). The model form of the STARIMA class is tentatively chosen after an examination of the space-time autocorrelation and space-time partial autocorrelation functions that can be viewed as the 2-dimensional analogues of the usual autocorrelations and partials used to identify univariate ARMA models The sample space-time autocorrelation at spatial lag *l* and temporal lag *s* is calculated via

$$\rho_l(s) = \frac{T}{T-S} \frac{\sum_{t=1}^{T-s} \left[W_l Z_t\right]' Z_{t+s}}{\left(\sum_t \left[W_l Z_t\right]' \left[W_l Z_t\right] \sum_t Z_t' Z_t\right)^{1/2}} \tag{4}$$

For the space-time analogue of the Yule-Walker equations the space-time covariance function is needed

$$\gamma_{lk}(s) = E\left\{\frac{\left[W^{(l)} Z_t\right]' \left[W^{(k)} Z_{t+s}\right]}{N}\right\} \tag{5a}$$

which can be seen to be equivalent to

$$\gamma_{lk}(s) = tr\left\{\frac{W^{(k)\prime} W^{(l)} \Gamma(s)}{N}\right\} \tag{5b}$$

where $\Gamma(s) = E[Z_t Z'_{t+s}]$ and $tr[A]$ is the trace of A defined on square matrices as the sum of the diagonal elements. $\Gamma(s)$ is estimated by

$$\hat{\Gamma}(s) = \sum_{t=1}^{T-s} \frac{Z_t Z'_{t+s}}{T-s}. \tag{5c}$$

Premultiplying both sides of the general STAR model

$$Z_t = \sum_{j=1}^{k} \sum_{l=0}^{\lambda} \phi_{jl} W^{(l)} Z_{t-j} + a_t \tag{6}$$

by $\left[W^{(h)} Z_{t-s}\right]'$ gives

$$Z'_{t-s} W^{(h)\prime} Z_t = \sum_{j=1}^{k} \sum_{l=0}^{\lambda} \phi_{jl} Z'_{t-s} W^{(h)\prime} W^{(l)} Z_{t-j} + Z'_{t-s} W^{(h)\prime} a_t \tag{7}$$

Taking expected values and dividing both sides by *N* yields

$$\gamma_{h0}(s) = \sum_{j=1}^{k} \sum_{l=0}^{\lambda} \phi_{jl} \gamma_{hl}(s-j) \tag{8}$$

since $E[Z'_{t-s} a_t] = 0$ for *s*>0. This system is the space-time analogue of the Yule-Walker equations for univariate time series. The set of last coefficients ϕ'_{kl} obtained from solving the system of equations as l=0,1,..,λ for k=1,2,… forms the space-time partial correlation function of spatial order λ. Analogously to univariate time series STARMA processes are characterized by a distinct space-time partial and autocorrelation function. Purely autoregressive $\text{STAR}(p_\lambda)$ processes exhibit space-time autocorrelations that tail off both in space and time and partial autocorrelations that cut off after p lags in time and λ lags in space whereas $\text{STMA}(q_m)$ processes exhibit autocorrelations that cut off after q lags and partials that decay over time and space. Mixed models exhibit partials and autocorrelations that tail off with both time and space. For a thorough discussion on these matters the reader should consult Pfeifer and Deutsch (1980a, 1980b).

Estimation

STARIMA (p,d,q) models with $q \neq 0$ are non-linear in form so parameter estimation is performed using any of a variety of non-linear optimization techniques. As discussed in Pfeifer and Deutsch (1980a), gradient methods have found use, as has linearization, an iterative technique that at each stage "linearizes" the non-linear model using Taylor's expansion and solves approximate normal equations for the next guess at the optimum parameters. Normally, one has to minimize the expression

$$a_t = Z_t - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}Z_{t-k} + \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}a_{t-k} \tag{9}$$

where the first few alphas are functions of observations and errors at times before the initial epoch observed; this difficulty is sidestepped by substituting zero, the unconditional mean for all values of Z_t and α_t with t<1.

Diagnostic Checking

The first phase of diagnostic checking is the examination of the residuals from the fitted model; these should be distributed normally with zero mean, have a spherical variance–covariance matrix and autocovariances at nonzero lags equal to zero. Usually the sample space-time autocorrelations and partials of the residuals are computed and compared to their theoretically derived variance. If the residuals are approximately white noise, the sample space-time autocorrelation functions should all be perfectly zero; otherwise they may follow a pattern that can be represented by a STARMA model, which may be coupled with the one initially proposed and lead to a better updated model. The second phase of the diagnostic checking involves checking the statistical significance of the estimated parameters based on the approximate confidence intervals proposed by Pfeifer and Deutsch (1980a). The insignificant parameters should be removed and the resulting simpler models should be again estimated and passed through the diagnostic checking stage until all parameters are statistically significant and the residuals meet the required constraints.

Final Remarks

It appears that STARIMA modeling can be a useful tool in cases where the researcher faces datasets of large spatial and temporal dimension. Kamarianakis and Prastakos (2002, 2003) used this technique for modeling the traffic conditions of a large part of the road network depicted in figure 1 and they compared its forecasting accuracy to the one obtained by ARIMA models (one model for each detector). Although the number of parameters in the STARIMA model is about one tenth of the total number of parameters of the univariate models, they perform surprisingly well. The major gain in this case is that the researcher has a single model to explain the dynamics of traffic flow of the whole network, which can be used not only for forecasting but also for impulse control (i.e. quantification of the effect of a traffic shock to downstream locations).

2.2 Bayesian Vector Autoregressive Models with Spatial Priors

Example Application: Forecasting Regional Employment

In macroeconomic modeling the available data are much less compared to the traffic flow application presented in the previous paragraph. Consider for example the monthly employment time series from 1982 to 1995 that correspond to eight neighboring American states (Illinois, Indiana, Kentucky, Michigan, Ohio, Pennsylvania, Tennessee, and West Virginia) analyzed by LeSage and Krivelyova (1999). In this case the researcher may proceed via using the STARIMA approach taking neighboring structures into account, or he may choose a VARMA model. The former strategy appears to be too parsimonious whereas the latter over-parameterized. A STARIMA model of (both AR and MA) spatial order two and temporal order four (a possible outcome of the identification stage for the dataset we consider) would be represented by 24 parameters. On the other hand, a VARMA model of AR and MA order four would pertain to the estimation of 520 parameters and a large proportion of them is expected to be statistically insignificant. LeSage and Krivelyova (1999) circumvented this problem by implementing a Vector Autoregressive model (no-moving average terms were included) by imposing priors that loosely constrained the parameters that correspond to large temporal lags and non-neighboring locations to zero. In a detailed forecasting experiment their approach based on spatial priors provided more accurate out of sample forecasts than the conventional Bayesian VAR approach based on the so-called "Minnesota prior" (Doan, Litterman and Sims 1984).

Model Specification

A principle behind much of the modeling in regional science is that location in space is important. LeSage and Krivelyova (1999) incorporated that principle in the VAR model in form of prior information. This prior is applied to the coefficients of a VAR model shown (in compact form) in equation (10) that involves n variables, where ε_{it} denotes independent disturbances, C_i represents constants, and y_{it} for $i=1,\dots,n$ denotes the n variables in the model at time t. Model parameters $A_{ij}(l)$ take the form, $\sum_{k=1}^{m} a_{ijk} l^k$ where l is the lag operator defined by $l^k y_t = y_{t-k}$ and m is the autoregressive order of the (VAR(m)) model.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} A_{11}(l) & \cdots & A_{1n}(l) \\ \vdots & \cdots & \vdots \\ A_{n1}(l) & \cdots & A_{nn}(l) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \tag{10}$$

The n variables in our case reflect time series from n areas and the VAR structure posits a set of relationships between past lagged values of all locations in the model and the current value of each location. For example if the y_{it} represent employment in state i at time t, the VAR structure allows employment variation in each state to be explained by past employment variation in the state itself, y_{it-k}, $k=1,\ldots m$, as well as past employment variation in other states, y_{it-k}, $k=1,\ldots m$, $j \neq i$.

The Bayesian implementation of the VAR model is based in prior specification for each unknown parameter in the model; the combination of prior distributions with the likelihood obtained by the data leads to the derivation of the posterior distributions in which the researcher can base his inference. The set of prior means developed for the BVAR model in this case were motivated by first-order spatial contiguity relations of the type employed in spatial autoregressive models for cross-sectional data. Hence the prior mean for the coefficients on variables associated with first own-lag spatially contiguous variables is equal to 1/c, where *c* is the number of spatial entities contiguous to each variable in the model. In other words the spatial prior is centered on a random-walk model that averages over contiguous entities and allows for drift

$$y_{it} = a + \sum_{j=1}^{c_i} \left(\frac{1}{c_i} \right) y_{jt-1} \quad j \in C_i \tag{11}$$

where C_i is the set of c_i entities contiguous to entity *i*. Consistent with traditional approaches to BVAR modeling the prior means are set to zero for coefficients on all lags other than first lags. Bayesian approaches that specify prior means of zero for all coefficients in a model have often been successful in dealing with collinearity problems in regression models. This approach in specifying prior means requires that the time series data on the various spatial entities need to be scaled or transformed to have similar magnitudes. If this is not the case, it would make little sense to indicate that the value of a time series observation at time t was equal to the average of values from time series observations taken from spatially contiguous entities. This should be no problem as time series data can always be expressed in percentage change form or annualized growth rates.

The prior variances for the parameters in the model differ according to whether the coefficients are associated with variables from contiguous or noncontiguous entities and with the lag length. The intuitive motivation for this is the twofold belief that: 1) noncontiguous variables are less important than contiguous because there is a decay of influence with increasing distance between spatial entities; and (2) longer lags are less important than shorter lags because there is a decline of influence over time. Time-series observations from the more distant past exert a smaller influence than recent observations on the current value of the spatial time series we are modeling. These two beliefs are reflected in the prior variance specification by:

- Parameters associated with noncontiguous time series variables are assigned a smaller prior variance, so the zero prior means are imposed with more certainty.
- First own-lags of contiguous time-series variables are given a smaller prior variance, so the prior means forcing the time series to equal the average of neighboring time series are imposed tightly. Tight imposition of these prior means reflects the belief that contiguous spatial series should exhibit co-movement over time.
- Parameters associated with noncontiguous variables at lags greater than one will be given a prior variance that becomes smaller as the lag length increases, imposing the prior means of zero more tightly for longer lags. This reflects the belief that influence decays with time and noncontiguous entities are unimportant.
- Parameters associated with lags other than first own-lag of the contiguous time-series variables will have a larger prior variance, so the prior means of zero are imposed "loosely". This is motivated by the fact that there is not a great deal of confidence in the zero prior mean specification for lagged values of contiguous spatial time-series variables.

A flexible form with which to state the spatial prior means and standard deviations for variable *j* in equation *i* at length *k* is shown right below in (12)-(15),

$$\begin{aligned}
\pi(a_{ijk}) &= N\left(\frac{1}{c_i}, \sigma_c\right) \quad j \in C, \qquad k=1; \qquad i,j=1,..,n \\
\pi(a_{ijk}) &= N\left(0, \frac{\tau\sigma_c}{k}\right) \quad j \in C \qquad k=2,\ldots,m; \qquad i,j=1,..,n \qquad (12) \\
\pi(a_{ijk}) &= N\left(0, \frac{\theta\sigma_c}{k}\right) \quad j \notin C \qquad k=1,\ldots,m; \qquad i,j=1,..,n
\end{aligned}$$

where

$$0 < \sigma_c < 1, \quad (13)$$

$$\tau > 1, \quad (14)$$

$$0 < \theta < 1. \quad (15)$$

For variables j=1,..,m in equation i that are contiguous to variable i, ($j \in C$), the prior mean for lag length k=1 is set to the average of the number of entities c contiguous to variable i, and to zero for noncontiguous variables ($j \notin C$). The prior standard deviation is set to σ_c for the fist lag and obeys the restriction set forth in (13), reflecting a tight imposition of the prior mean, motivated by spatial contiguity. $\tau\sigma_c/k$ is used for lags greater than one and imposes a linear decrease in this variance as the lag length increases. Equation (14) states the restriction necessary to ensure that the prior mean of zero is imposed on the parameters associated with lags greater than one on contiguous time series loosely, relative to a tight imposition of the prior mean of $1/c_i$ on the first own lags of contiguous time series variables. $\tau\sigma_c/k$ is used for lags on noncontiguous variables whose prior means are zero, imposing a linear decrease in the variance as the lag length increases. The restriction in (15) would impose the zero means for noncontiguous states with more confidence than the zero prior means for contiguous states.

3. MODELS FOR MULTIPLE SPATIAL TIME SERIES' RELATIONS

3.1 The General First-order Serial and Spatial Autoregressive Distributed Lag Model

Example Applications and Model Illustration

Following the lines of section 2.1, in transportation literature there is extensive interest in the functional relation between traffic flows and densities. Instead of modeling the spatio-temporal evolution of traffic conditions, this time the researcher is interested in the relation between volumes and densities that are observed in space and time. The temporal dimension of the dataset in this case is much larger than the spatial one. In regional macroeconomic modeling on the other hand (section 2.2), researchers are usually confronted with the estimation of relationships between variables like GDP, employment, labor force participation, productivity, etc. that correspond to different regions (or states or prefectures) and are in the form of (usually short) time series. As an example the reader may consider twenty annual observations for two variables, employment and labor force participation, that correspond to ninety-five French (NUTS 3) regions. This dataset is part of the REGIO database provided by EUROSTAT and is similar to the one used by Elhorst (2001) for the illustration of the model that is presented in this section. In this case it is the spatial dimension that is significantly larger than the temporal one.

The general first-order serial and spatial autoregressive distributed lag model in vector form for a cross-section of observations at time *t* is represented by

$$Y_t = \tau Y_{t-1} + \delta W Y_t + \eta W Y_{t-1} + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 W X_t + \beta_4 W X_{t-1} + u_t \tag{16}$$

where Y_t denotes a $n \times 1$ vector consisting of one observation for every spatial unit ($i=1,\ldots,n$) of the dependent variable in the t^{th} time period($t=1,\ldots T$). X_t denotes a $n \times 1$ vector of the independent variable; the generalization of the model for multiple independent variables and second third etc. order is straightforward. τ, δ, η, β_1, β_2, β_3, β_4, are the response parameters, u_t is a $n \times 1$ vector containing the error terms and is normally distributed with $E(u_t) = 0$ and $E(u_t, u_t') = \sigma^2 I_n$ and W denotes an $n \times n$ weight matrix describing the geographical arrangement of the spatial units. Subscript *t-1* denotes a serially lagged variable and a variable premultiplied by W denotes its spatially lagged value. We assume that the characteristic roots of the weight matrix are known and the following relationship holds between δ the minimum and maximum characteristic roots

$$\frac{1}{\omega_{\min}} < \delta < \frac{1}{\omega_{\max}}. \tag{17}$$

The former assumption is needed to ensure that the log-likelihood function of the model can be computed whereas the latter facilitates the maximum likelihood estimation of δ and ensures invertibility of the matrix $(I-\delta W)$.

Equation (16) involves instantaneous relations between *Y*, *WY*, *X* and *WX* so it's not well suited for forecasting purposes; even without the presence of the *X* regressor in (16) this model class is different from the ones presented in the pervious sections. Its formulation is useful for empirical inference concerning long run equilibrium relationships between economic variables short run dynamics (how fast the equilibrium is approached). Reformulating (16) we obtain an equilibrium correction model

$$\left(I - \tau L - \delta W - \eta W\right)Y_t = -\left(\tau + \eta W\right)\Delta Y_t + \left(\beta_1 + \beta_2\right)X_t + \left(\beta_3 + \beta_4\right)WX_t - \beta_2 \Delta X_t + u_t \qquad (18)$$

which implies the following static long-run equilibrium relationship between *Y* and *X*

$$Y_t = \left[\left(\beta_1 + \beta_2\right)\left(I_n - \tau I_n - \delta W - \eta W\right)^{-1} + \left(\beta_3 + \beta_4\right)\left(I_n - \tau I_n - \delta W - \eta W\right)^{-1} W\right]X_t . \qquad (19)$$

A spatial unit in an equilibrium correction model is not only influenced by its local conditions but also by those of its neighbors dependent on the structure of the weight matrix. For *n* locations and *k* regressors there are $n \times k$ different "long-run" parameter estimates.

Technical Details

Let

$$B = I - \delta W , \quad A = \tau I + \eta W \qquad (20)$$

When $\left|AB^{-1}\right| < 1$ the process generating the data is stationary in time. Stationarity in space is more difficult to impose. Kelejian and Prucha (1999) formulated one necessary condition that must be satisfied: the row and the column sums of the spatial weight matrix must be bounded uniformly in absolute value as $n \to \infty$. For inverse distance matrices this condition is not automatically satisfied.

Regarding model class (16) and its generalization to higher temporal orders and multiple regressors there are still issues that need to be investigated. Estimation by maximum likelihood appears to be cumbersome; the model is also highly susceptible to multicollinearities, an issue that was not touched by Elhorst (2001). Finally model order selection tests based on Wald or Lagrange multiplier statistics do not lead to clear conclusions since model selection involves comparisons between non-nested models. Bayesian methods have been used successfully to tackle some of the above issues and it appears that this model class can be a new field for their application.

REFERENCES

BOX, G.E.P., JENKINS G.M. and REINSEL G.C. (1994) *Time Series Analysis / Forecasting and Control* (third edition). Prentice Hall, New Jersey.

CLIFF, A.D. and ORD, J.K. (1975) Space-Time modeling with an application to regional forecasting. *Transactions of the Institute of British Geographers,* 64, 119-128.

DOAN, T., LITTERMAN, R.B. and SIMS C. A. (1984) Forecasting and Conditional Projections using Realistic Prior Distributions, *Econometric Reviews*, 3, 1-100.

ELHORST, J.P. (2001) Dynamic models in space and time. *Geographical Analysis*, 33, 119-140.

JOHNSTON, J. and DiNARDO, J. (1997) Econometric Methods (Fourth Edition). Mc Graw Hill.

KAMARIANAKIS, I., and PRASTACOS, P. (2002) Space-time modeling of traffic flow. *Methods of spatial analysis – spatial time series analysis.* ERSA Proceedings.

KAMARIANAKIS, I., and PRASTACOS, P. (2003) Forecasting Traffic flow conditions in an urban network: A comparison of univariate and multivariate procedures. *Transportation Research Record.*

KELEJIAN, H.H., and PRUCHA, I.R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*, 509-522.

LUTKERPOHL, H. 1987. *Forecasting Aggregated Vector ARMA Processes.* Springer-Verlag, Berlin.

LUTKERPOHL, H. 1993. *Introduction to Multiple Time Series Analysis.* . Springer-Verlag, Berlin.

LESAGE, J.P. and KRIVELYOVA A. (1999) A spatial prior for Bayesian vector autoregressive models. *Journal of Regional Science,* 39, 297-317

PFEIFER, P.E., and BODILY, S.E. (1990) A test of space-time ARMA modeling and forecasting with an application to real estate prices, *International Journal of Forecasting,* 16, 255-272.

PFEIFER, P.E., and DEUTSCH, S.J. (1980a) A three-stage iterative procedure for space-time modeling. *Technometrics* 22(1).

__________________________. (1980b) Identification and Interpretation of First-Order Space-Time ARMA Models. *Technometrics 22 (3).*

__________________________. 1981a. Variance of the Sample-Time Autocorrelation Function of Contemporaneously Correlated Variables. *SIAM Journal of Applied Mathematics, Series A,* 40(1).

________________________. 1981b. Seasonal Space-Time ARIMA modeling. *Geographical Analysis* 13 (2).

________________________. 1981c. Space-Time ARMA Modeling with contemporaneously correlated innovations. *Technometrics* 23 (4).

STOFFER, D.S. (1986) Estimation and interpretation of Space-Time ARMAX models in the presence of missing data. Journal of the American Statistical Association, 81, 762-772.

ZELLNER, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association* 57, 500-509