Regressions in Spatially Dynamic Factor Models

Regressione e modelli fattoriali spazio-temporali

Lara Fontanella, Luigi Ippoliti, Pasquale Valentini

Abstract This paper discusses a number of conceptual issues pertaining to the study of the relationships existing between two groups of variables which are supposed to be spatially and temporally correlated. Since it is assumed that this relationships can be studied in a reduced latent space, we provide an overview of the motivations for including spatial effects in a dynamic factor model, both from a theory-driven as well as from a data-driven perspective. Considerable attention is paid to the inferential framework necessary to carry out estimation and to the different assumptions, constraints and implications embedded in the various model specifications.

Abstract *Il presente articolo riguarda lo studio delle relazioni esistenti tra due gruppi di variabili che assumiamo siano spazialmente e temporalmente correlate. Si presume che lo studio di queste relazioni possa essere effettuato in uno spazio ridotto generato attraverso la stima Bayesiana di un modello fattoriale gerarchico per processi spazio-temporali. Commentando alcune possibili parametrizzazioni, si dimostra che il modello é sufficientemente flessibile ed adatto a risolvere molti dei problemi statistici comunemente presenti nelle analisi di dati spazio-temporali.*

Key words: Dynamic factor models, spatio-temporal processes, Bayesian inference, spatial econometrics

Lara Fontanella
University G. d'Annunzio, Chieti-Pescara, Italy, e-mail: lfontan@unich.it

Luigi Ippoliti
University G. d'Annunzio, Chieti-Pescara, Italy, e-mail: ippoliti@unich.it

Pasquale Valentini
University G. d'Annunzio, Chieti-Pescara, Italy e-mail: pvalent@unich.it

1 Introduction

The idea of borrowing information from different but related sources can be very powerful for statistical analysis. It proved to be very useful in the last decades where complex data structures began to be tackled, as they required sophisticated modeling strategies. In this paper we consider the problem of modeling high-dimensional multivariate, spatially and temporally referenced data. This problem has enjoyed widespread popularity in the last years and requires the definition of general and flexible statistical models where the temporal and cross-sectional dependencies must be accommodated.

Spurred by recent advances in Geo-spatial data acquisition technologies, it is often desirable for these data to examine the relationships existing between one or more dependent variables and some other linked covariates. This can be achieved in a number of ways, though there might be no single approach which can be considered uniformly as being the most appropriate.

For example, general vector autoregressive (VAR) models are recognized as allowing for simultaneous modeling of variables in a multivariate context. Traditionally, VAR models use a small number of variables to avoid inflating the number of parameters to be estimated and, in general, they do not allow for a direct modeling of locational spillovers. A spatial adaptation of VARs, denoted as SpVAR models, explicitly considers the potential impacts of specific events in neighboring sites (regions) and has been discussed in Kuethe and Pede (2011). The SpVAR is a specific version of the Spatio-Temporal Auto-Regressive Moving Average - STARMA - model where the linear dependencies are lagged in both space and time. Since STARMAs are an extension of the ARMA class of models they are particularly useful to produce temporal forecasts of the variables of interest but they are not suitable to provide spatial predictions if these are required. As discussed in Valentini et al. (2013), the STARMA specification also suffers from some other disadvantages. Seemingly Unrelated Regression (SUR) and error correction panel data models have also been largely used with spatial and time effects. Apart from their rather complex structure, as STARMAs, these models are not suitable when the number of regions is relatively large. In fact, the application of an unrestricted SURE-GLS approach to large N (cross section dimension) and T (time series dimension) panels involves nuisance parameters that increase at a quadratic rate as the cross section dimension of the panel is allowed to rise (Pesaran, 2006).

Among the different methodologies proposed in the literature, dynamic factor models (DFMs) have grown significantly in popularity and have been shown to be very useful for exploratory analysis, policy analysis and forecasting in a data-rich environment. DFMs have been widely developed in both methodological and practical issues, and have become a standard tool for increasingly high-dimensional modeling of time series. They have been extensively used in macroeconomics and finance with the core idea of explaining the common dynamic structure of the multivariate time series through a set of common (time series) factors. This is achieved by the introduction of flexible temporal correlation structures for the latent factors, previously assumed to be independent. This renders the DFM capable of assessing the complexity of time series data. For a recent review on DFMs the reader may refer to Gamerman and Salazar (2013) and references therein.

Special attention will be devoted here to the use of a dynamic factor analytic approach in the framework of spatial statistics. It will be shown that this is not only an important area of application but also that this area can receive several benefits from this modeling approach.

A key property of much spatio-temporal data is that observations at nearby sites and times will tend to be similar to one another. Then, factor analysis assumes that the cross dependence can be characterized by a finite number of unobserved common factors, possibly due to common shocks that affect all the spatial sites, albeit with different intensities. Thus, the strong co-movement and the high correlation among the series, amplified by the presence of spatial correlation, suggest that both observable and unobservable factors must be at place.

In this paper, we thus approach the analysis of multivariate spatio-temporal processes from the perspective of recent developments of dynamic factor models. Through a fully Bayesian approach, we contribute to the recent literature by melding together dynamic factor models, spatial regression models and geostatistical techniques, in order to explain the multifactorial nature of many spatio-temporal data. We assume that the relationships existing between the groups of dependent and regressor variables can be studied through a temporally dynamic and spatially descriptive model, hereafter referred to as *spatial dynamic structural equation* model (SD-SEM).

The proposed model has an intuitive appeal and enjoys several advantages. First, our model formulation exploits the spatio-temporal nature of the data and explicitly defines a non-separable spatio-temporal covariance structure of the multivariate process. Second, since the data have a multivariate and multidimensional structure, in that several time series can be measured at specific spatial sites, the temporal relationships between dependent and regressor variables is modeled in a latent space. The observed processes are thus described by a potentially small set of common dynamic latent factors with the advantage of overcoming the difficulties of interpreting the relationships under study due to collinearity and low signal-to-noise ratio issues. Temporal forecasts of the variables of interest can also be obtained by only modeling the dynamics of a few common factors. Third, by modeling the spatial variation via spatially structured factor loadings, we entertain the possibility of identifying clusters of spatial sites that share common time series components. Through the spatial modeling of the factor loadings, spatial interpolations of the observed variables are also straightforward. Fourth, several general structures that make use of different covariate information, can be easily accommodated in the different levels of the hierarchy. Fifth, the SD-SEM offers a unified approach suitable to deal with variables and indicators measured at different scales and coming from different spatial sources. Hence, the model provides a simple solution to the misalignment problems which, for example, normally occurs in health care research. Lastly, the model specification is not limited to normally distributed variables, but it can be extended to handle more types of variables from an exponential family.

The remainder of the paper is organized as follows. In section 2, we describe the general model and discuss how regression ideas can be incorporated into the factor model setting. In section 3 we discuss the prior specification with special attention on structures which define general forms of spatial correlation and cross-correlation between variables at different locations. In section 4 we discuss how to perform posterior inference. Finally, section 5 concludes the paper with a discussion on specific inferential problems, possible uses of the model and directions for further work.

2 The spatial dynamic structural equation model

Often observations are multivariate in nature, i.e. we obtain vector of time series at locations across space. For such data, together with the study of the temporal dynamics of the variables, we need to model both association between measurements at a location as well as association between measurements across locations. With increased collection of such multivariate spatial data, there arises the need for flexible explanatory stochastic models in order to improve estimation precision and to provide simple descriptions of the complex relationships existing among the variables.

There are cases of interest in which the association between dependent variables and the set of explanatory variables can be better investigated through the estimation of some latent factors rather than the variables individually. In the following, a model formulation which describes the structural relations among the variables in a lower dimensional space is thus presented.

Assume initially that $\mathbf{Y}$ and $\mathbf{X}$ are two multivariate Gaussian spatio-temporal processes observed at temporal instants $t \in \{1,2,\dots\}$ and generic locations, $\mathbf{s} \in \mathscr{D}_y$ and $\mathbf{u} \in \mathscr{D}_x$, respectively. For the two different processes, the spatial sites $\mathbf{s}$ and $\mathbf{u}$ can denote the same location but, in general, they need not be the same. Furthermore, both $\mathscr{D}_y$ and $\mathscr{D}_x$ may represent two different spatial domains of interest.

Let n_y be the number of observed variables for $\mathbf{Y}$ and n_x the number of observed variables for $\mathbf{X}$. The most informative case is represented by the isotopic configuration where, for each multivariate process, $\mathbf{Y}$ or $\mathbf{X}$, all variables are measured at all their respective sites. In this case, we can thus write $\mathbf{Y}(\mathbf{s},t) = [Y_1(\mathbf{s},t),\dots,Y_{n_y}(\mathbf{s},t)]'$ and $\mathbf{X}(\mathbf{u},t) = [X_1(\mathbf{u},t),\dots,X_{n_x}(\mathbf{u},t)]'$. The opposite case is the completely heterotopic case where not all the variables can be observed at the same site.

Without loss of generality, for the sake of simplicity, we describe here the isotopic case and assume that $\mathbf{Y}$ or $\mathbf{X}$ can be observed at N_y and N_x spatial sites, respectively. Let $\tilde{n}_y = n_y N_y$ and $\tilde{n}_x = n_x N_x$. Then, at a specific time t, the $(\tilde{n}_y \times 1)$ and $(\tilde{n}_x \times 1)$ dimensional spatial processes, $\mathbf{Y}$ and $\mathbf{X}$, are denoted as $\mathbf{Y}(t) = [\mathbf{Y}(\mathbf{s}_1,t)',\dots,\mathbf{Y}(\mathbf{s}_{N_y},t)']'$ and $\mathbf{X}(t) = [\mathbf{X}(\mathbf{u}_1,t)',\dots,\mathbf{X}(\mathbf{u}_{N_x},t)']'$. Since it is also assumed that $\mathbf{X}$ is a predictor of $\mathbf{Y}$, which is thus the process of interest, we work within the well known framework of transfer response models covered in many standard time series books.

Our model assumes that each multivariate spatial process, at a specific time t, has the following linear structure

$$\mathbf{X}(t) = \mathbf{m}_x(t) + \mathbf{H}_x \mathbf{f}(t) + \mathbf{u}_x(t) \tag{1}$$
$$\mathbf{Y}(t) = \mathbf{m}_y(t) + \mathbf{H}_y \mathbf{g}(t) + \mathbf{u}_y(t) \tag{2}$$

where $\mathbf{m}_y(t)$ and $\mathbf{m}_x(t)$ are $(\tilde{n}_y \times 1)$ and $(\tilde{n}_x \times 1)$ mean components modeling the smooth large-scale temporal variability, $\mathbf{H}_y$ and $\mathbf{H}_x$ are measurement (factor loadings) matrices of dimensions $(\tilde{n}_y \times m)$ and $(\tilde{n}_x \times l)$, respectively, and $\mathbf{g}(t)$ and $\mathbf{f}(t)$ are m- and l-dimensional vectors of temporal common factors. Also, $\mathbf{u}_y(t)$ and $\mathbf{u}_x(t)$ are Gaussian error terms for which we assume $\mathbf{u}_y(t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{u_y})$ and $\mathbf{u}_x(t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{u_x})$. For simplicity, throughout the paper it is assumed that $\boldsymbol{\Sigma}_{u_y}$ and $\boldsymbol{\Sigma}_{u_x}$ are both diagonal matrices and that $m \ll \tilde{n}_y$ and $l \ll \tilde{n}_x$.

In the second level of the hierarchy the temporal dynamics of the common factors are then modeled through the following equations

$$\mathbf{g}(t) = \sum_{i=1}^{p} \mathbf{C}_i \mathbf{g}(t-i) + \sum_{j=0}^{q} \mathbf{D}_j \mathbf{f}(t-j) + \boldsymbol{\xi}(t) \tag{3}$$
$$\mathbf{f}(t) = \sum_{k=1}^{s} \mathbf{R}_k \mathbf{f}(t-k) + \boldsymbol{\eta}(t) \tag{4}$$

where $\mathbf{C}_i$ $(m \times m)$, $\mathbf{D}_j$ $(m \times l)$, and $\mathbf{R}_k$ $(l \times l)$ are coefficient matrices modeling the temporal evolution of the latent vectors $\mathbf{g}(t) = [g_1(t), \ldots, g_m(t)]'$ and $\mathbf{f}(t) = [f_1(t), \ldots, f_l(t)]'$, respectively. Finally, $\boldsymbol{\xi}(t)$ and $\boldsymbol{\eta}(t)$ are independent Gaussian error terms for which we assume $\boldsymbol{\xi}(t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi})$ and $\boldsymbol{\eta}(t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\eta})$.

Equation (3) represents the structural part of the model formulation and appears as a VAR model in which the variables in $\mathbf{g}(t)$, are controlled for the effects of other variables in $\mathbf{f}(t)$. Equations (1-4) thus provide the basic formulation of the SD-SEM. One clear advantage of this model is that in the context of increasingly high-dimensional time series, the temporal relationship between dependent and regressor variables can be modeled in a reduced latent space. Also, temporal forecasts of the variables of interest, $\mathbf{Y}$, can be obtained by modeling the dynamics of a few common factors represented in $\mathbf{g}(t)$. Also, the model is spatially descriptive in that it can be used to identify possible clusters of locations whose temporal behavior is primarily described by a potentially small set of common dynamic latent factors. As it will be shown in the next section, flexible and spatially structured prior information regarding such clusters can be specified through the columns of the factor loading matrix. Model completion requires specific forms for patterns of time-variation in the elements of $\mathbf{m}_y(t)$ and $\mathbf{m}_x(t)$, and for the elements of the factor loadings matrices $\mathbf{H}_y$ and $\mathbf{H}_x$. We note that the pure factor model is the special case given by equation (2), with $\mathbf{m}_y(t) = \mathbf{0}$, and equation (3) with $\mathbf{D}_j = \mathbf{0}$ for all j, when the typical assumption that the common factors $\mathbf{g}(t)$ are zero-mean and independent over time yields a linear factor representation of the conditional variance matrix of $\mathbf{Y}(t)$. When $\mathbf{m}_y(t) = \mathbf{A}_y \mathbf{z}(t)$, where $\mathbf{A}_y$ is a $(\tilde{n}_y \times r)$ matrix of regression coefficients and

$\mathbf{z}(t)$ is a r-vector of observed variables, the model is what has become known as a factor-augmented vector autoregression (FAVAR).

3 Prior specification

Since the model is developed under the Bayesian paradigm, an important part of the model specification deals with the choice of prior distributions.

FACTOR LOADINGS SPECIFICATION
The loadings are useful quantities as they allow for the identification of common features and for the interpretation of the relationship or correlation structure between the different series. Let $\mathbf{h}_{y_j} = \left[\mathbf{h}_{y_j}(\mathbf{s}_1)', \mathbf{h}_{y_j}(\mathbf{s}_2)', \dots, \mathbf{h}_{y_j}(\mathbf{s}_{N_y})'\right]'$ denote the j-th column of $\mathbf{H}_y$ and let $\mathbf{h}_{y_j}(\mathbf{s}_i) = \left[h_{y_j,1}(\mathbf{s}_i), \dots, h_{y_j,n_y}(\mathbf{s}_i)\right]'$, denote the n_y-dimensional vector of loadings defined for the spatial site $\mathbf{s}_i$. With obvious notation, the same holds for $\mathbf{h}_{x_j}$ but, for simplicity, assume henceforth we only focus on the process of interest $\mathbf{Y}$. In general, taking care of the identifiability constraints (see Lopes and West, 2004), for the factor loadings one can take independent normal priors. However, for spatio-temporal data, one may be more interested in including conditional dependencies within the elements of $\mathbf{Y}(t)$. In fact, because of their spatial nature, the factor loadings are characterized by spatial patterns and consequently, a number of papers have discussed the possibility of inducing flexible correlation structures into the columns of $\mathbf{H}_y$. For example, in the framework of geostatistics, Wikle and Cressie (1999) assume $\mathbf{Y}(t)$ is a univariate spatially continuous process (i.e. $n_y = 1$ and $\mathscr{D}_y \subseteq \mathscr{R}^2$) and model the columns of $\mathbf{H}_y$ as orthonormal basis functions. Calder (2007), on the other hand, suggested the use of smoothed deterministic kernels to build $\mathbf{H}_y$. Alternatively, Lopes et al. (2008) introduced a spatial DFM where the columns of the factor loadings matrix follow conditionally independent Gaussian Random fields, that is $\mathbf{h}_{yj} \sim N\left(\boldsymbol{\mu}_{h_{y_j}}, \boldsymbol{\Sigma}_{h_{y_j}}\right)$, where $\boldsymbol{\mu}_{h_{y_j}}$ is a mean vector and $\boldsymbol{\Sigma}_{h_{y_j}} = \tau^2_{y_j}\mathbf{R}(\varphi_{y_j})$. For a valid correlation function (e.g. exponential, Matérn, etc), $\mathbf{R}(\varphi_{yj})$ denotes the $(N_y \times N_y)$ spatial correlation matrix where its generic entry is given by $\rho(|\mathbf{s}_i - \mathbf{s}_{i'}|; \varphi_{y_j})$, and φ_{y_j} is the ”effective” range parameter (i.e. the distance which makes the correlation negligible). To take into account the effect of some explanatory variables, it is possible to parameterize the mean vector, $\boldsymbol{\mu}_{h_{y_j}}$, through the definition of a suitable design matrix, $\boldsymbol{\Delta}^*$, such that $\boldsymbol{\mu}_{h_{y_j}} = \boldsymbol{\Delta}^*\boldsymbol{\beta}_{h_{y_j}}$, with $\boldsymbol{\beta}_{h_{y_j}}$ a vector of parameters. The presence of two sets of variables introduces further possibilities beyond standard DFM. In particular, relationships between the loading matrices $\mathbf{H}_y$ and $\mathbf{H}_x$ may be introduced. For example, Ippoliti et al. (2012) use $\mathbf{H}_x$ as a (latent) design matrix for the mean of $\mathbf{H}_x$.

If $\mathbf{Y}(t)$ is a multivariate $(n_y > 1)$ continuous spatial process observed on $\mathscr{D}_y \subseteq \mathscr{R}^2$, each column of the measurement matrix $\mathbf{H}_y$ can still be assumed to be a Gaussian spatial process and its correlation structure defined through the linear model of coregionalization (LMC, Schmidt and Gelfand, 2003). LMC assumes that a multivariate spatial process can be written as a linear combination of simpler univariate spatial process. Hence, at the spatial site $\mathbf{s}_i$, we may write $\mathbf{h}_{y_j}(\mathbf{s}_i) = \mathbf{A}_j \boldsymbol{\omega}(\mathbf{s}_i), \quad i = 1,2,\ldots,N_y,\ j = 1,2,\ldots,m$, where $\mathbf{A}_j$ is a $(n_y \times n_y)$ full-rank (lower-triangular) matrix and the components of $\boldsymbol{\omega}(\mathbf{s}_i)$, $\omega_k(\mathbf{s}_i),\ k = 1,\ldots,n_y$, are spatial processes independent across k, with mean 0, unit variances and spatial correlation matrix $\mathbf{R}(\varphi_{kj})$ function. It can thus be shown (Schmidt and Gelfand, 2003) that complex covariance structures can be obtained as

$$\boldsymbol{\Sigma}_{h_{y_j}} = \sum_{k=1}^{n_y} \mathbf{R}(\varphi_{kj}) \otimes \mathbf{a}_{kj}\mathbf{a}'_{kj}, \quad j = 1,2,\ldots,m$$

where $\mathbf{a}_{kj}$ represents the k-th column vector of $\mathbf{A}_j$. Note that a separable covariance specification represents the simplest form of LMC.

Large amounts of essentially-continuous spatial data are associated with the nodes or interiors of a regular rectangular lattice, or with irregularly-spaced sites or irregularly-shaped regions. For example, pixellated images are associated with the interiors of rectangular lattices, and epidemiological, ecological and economic data are usually associated with irregular sites or regions. The modeling of the spatial correlation for factor loadings corresponding to univariate and multivariate spatial processes observed on a lattice was discussed by Valentini et al. (2013). In this case $\mathbf{h}_{y_j}$ is assumed to be a conditional autoregressive CAR process, also known as Gaussian Markov random field, for which a simple form for the spatial dependence is based on a 0/1 neighbourhood adjacency matrix.

Common latent factors specification

By means of an appropriate concatenation of observables $\left[\mathbf{Y}(t)'\ \mathbf{X}(t)'\right]'$ and common factors $\left[\mathbf{g}(t)'\ \mathbf{f}(t)'\right]'$, equations (1-4) can be rewritten in state-space form by appropriately enlarging the state vector according to the order of lagged dependence of the latent factors (see, for example, Ippoliti et al., 2012; Valentini et al., 2013). It can be very hard to estimate this model in its full expression for typical applications. Hence, natural simplifications are usually obtained by restricting the order p, q, and s of the auto- and cross-regressions to small values, say 1 or 2. By assuming without loss of generality that $p \geq max(s,q)$, $\mathbf{D}_i = \mathbf{0}$ for $i > q$ and $\mathbf{R}_j = \mathbf{0}$ for $j > s$, it is useful to specify the joint generation process for $\mathbf{g}(t)$ and $\mathbf{f}(t)$ as a VAR(p) process of the type

$$\mathbf{d}(t) = \boldsymbol{\Phi}_1 \mathbf{d}(t-1) + \ldots + \boldsymbol{\Phi}_p \mathbf{d}(t-p) + \boldsymbol{\varepsilon}(t) \tag{5}$$

where

$$\mathbf{d}(t) = \begin{bmatrix} \mathbf{g}(t) \\ \mathbf{f}(t) \end{bmatrix}, \quad \boldsymbol{\Phi}_i = \begin{bmatrix} \mathbf{C}_i & \mathbf{D}_i \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix}, \quad \boldsymbol{\varepsilon}(t) = \begin{bmatrix} \boldsymbol{\xi}(t) \\ \boldsymbol{\eta}(t) \end{bmatrix}.$$

Let $\boldsymbol{\alpha}(t) = \left[\mathbf{d}(t)'\ \mathbf{d}(t-1)' \ldots \mathbf{d}(t-p+1)'\right]'$. Then, in the second level of the hierarchy the evolution of the joint common factors can be represented by the following transition equation $\boldsymbol{\alpha}(t) = \boldsymbol{\Gamma}\ \boldsymbol{\alpha}(t-1) + \boldsymbol{\zeta}, \quad \boldsymbol{\zeta} \sim N(0,\boldsymbol{\Lambda})$, where $\boldsymbol{\Gamma}$ is a

$(k \times k)$ block coefficient matrix, with $k = (m+l)p$, characterizing the dynamic evolution of the joint common factors and $\boldsymbol{\Lambda}$ is a covariance matrix with elements $\lambda_{ij}, \; i, j = 1, \ldots, k$. The prior for the latent process $\boldsymbol{\alpha}(t)$ is completed by $\boldsymbol{\alpha}(0) \sim N(\mathbf{a}_0, \boldsymbol{\Sigma}_{d0})$, with known hyperparameters $\mathbf{a}_0$ and $\boldsymbol{\Sigma}_{d0}$.

Assuming $\mathbf{D}_j = \mathbf{0}$, for all j, the state equation represents exactly the dynamic evolution of a standard DFM. In this case, many specifications can be envisaged for the autoregressive parameters in $\mathbf{C}_i$ and λ_{ij}. One possibility for the $\boldsymbol{\Lambda}$ matrix is a diagonal form with elements λ_i having independent Gamma distributions as priors. Otherwise, *inverted Wishart* distribution can be chosen. Similarly, independence assumptions can be made for the autoregressive matrices $\mathbf{C}_i$. One possibility is to consider $\mathbf{C}_i = diag(c_{1j}, \ldots, c_{mj})$ such that, $c_i \sim N(0, \gamma)$ independent, for some large value of γ if one want to represent vague prior information. If one is concerned with the possibility of unit roots and non-stationarity, a mixture prior may also be assumed for the autoregressive coefficients as shown in Huerta and West (1999). Considering a full model specification for the latent process $\boldsymbol{\alpha}(t)$, a general prior (which does not involve the restrictions inherent in the natural conjugate prior) can be specified through the independent Normal-Wishart prior. Using this prior, the joint posterior does not have a convenient form that would allow easy Bayesian analysis (e.g. posterior means and variances do not have analytical forms). However, the conditional posterior distributions do have convenient forms and Gibbs sampler which sequentially draws from the Normal and the Wishart distributions can be programmed up in a straightforward fashion. Lopes et al. (2008) also discuss the use of alternative prior specifications and we refer to them for known results. Other priors that are enjoying increasing popularity are called Stochastic Search Variable Selection (SSVS, George, Sun and Ni, 2008). These allow for shrinkage of the auto- and cross-regression coefficients and lead to restricted formulations in an automatic fashion that require only minimal prior input from the researcher. Notice that the SSVS approach can also be thought of as automatically selecting a restricted model and the relationship between such a strategy and conventional model selection techniques using an information criteria (e.g. the Akaike or Bayesian information criteria) is discussed in Fernandez, Ley and Steel (2001). By using a vector error correction model representation of equation (5), Valentini et al. (2013) also used a SSVS prior for testing the cointegration structure *within* $\mathbf{g}(t)$ and $\mathbf{f}(t)$ and *between* the two processes.

Large scale dynamic factors, periodic or cyclical components can also be directly specified in the model either through the mean level or through the common dynamic factors. In the first case, the same pattern is assumed for all locations, while in the other case, the common seasonal factors receive different weights for different columns of the factor loading matrix, so allowing different seasonal patterns for the spatial locations. By considering trend models to be of the form $\nabla^j \gamma(t) = \omega(t)$, where ∇^j is the j-th order difference operator and $\omega(t)$ a normally distributed zero-mean sequence with unknown variance ϕ^2, locally linear trend models can also be easily included in the model formulation.

4 Posterior inference

Posterior inference for the proposed class of spatial dynamic factor models is facilitated by MCMC algorithms. Standard MCMC for dynamic linear models are adapted to our model specification such that posterior and predictive analysis are readily available. Conditional on r and m, the number of common factors, the MCMC scheme described in Lopes and West (2004) can be easily adapted where the common factors are jointly sampled via the well known forward filtering backward sampling (FFBS) scheme (Carter and Kohn, 1994). All other full conditional distributions are ”standard” multivariate Gaussian or Gamma distributions. Exceptions refer to the spatial correlation parameters (range parameters and conditional autoregressive parameters under specific priors) which are sampled using a Metropolis-Hastings step.

5 Concluding remarks

This paper was concerned with the discussion on the use of factor models for multivariate time series observed at multiple sites. Our presentation has focused exclusively on the discussion about model building and a number of other issues were not addressed. We will briefly comment upon them now.

Two important issues are model identification and model selection. Many ways to handle the problem of identification can be found in the literature. Basically, the idea is to impose constraints on $\mathbf{H}_y$ and $\mathbf{H}_x$ as, for example, the lower constraint of Geweke and Zhou (1996). As discussed by Lopes and West (2004), an advantage of using this approach is that the order of the sites in the measurement matrices has no effect on the resulting model nor on predictive inferences under the specified model. However, the spatial structure imposed into the columns can be exploited in order to avoid the problem. In some applications, identifiability of the factor loadings is not required, especially for covariance matrix estimation, variable selection and prediction - see Bhattacharya and Dunson (2011) for more details. Uncertainty about the number of latent factors has been studied in different ways. One possibility is fitting the model for different choices of m and r and then using selection criteria like AIC, BIC or the Predictive Model Choice statistic of Gelfand and Ghosh (1998) for model selection. Lopes and West (2004) proposed fully Bayesian inference on the number of factors through a reversible jump MCMC. Other important items not discussed yet regard the possible uses of the model. Since the DS-SEM is highly structured and flexible, it can be used to solve most of the statistical problems commonly encountered in the analysis of spatio-temporal data. By discussing two examples regarding the analysis of environmental and macroeconomic data, Ippoliti et al. (2012) and Valentini et al. (2013) show how to obtain both unconditional and conditional forecasts of $\mathbf{Y}$, its spatial predictions and how to apply a multiplier analysis to describe the reaction over time of the dependent variable $\mathbf{Y}$ to exogenous impulses. A final important consideration is that the SD-SEM can also be easily extended to allow for non Gaussian observations. In this case the SD-SEM appears as a hierarchical model with first level measurement equations for the conditionally independent variables, $Y_k(\mathbf{s},t)$, $k = 1,\ldots,n_y$ and $X_j(\mathbf{u},t)$, $j = 1,\ldots,n_x$, in the one-parameter natural exponential family. The natural parameters can then be related to the linear predictors which, in the same spirit of equations (1) and (2), are defined by a linear combination of spatial and temporal components.

References

1. Bhattacharya, A., Dunson, D. B.: Sparse Bayesian infinite factor models. Biometrika, **98**, 291–306 (2011)
2. Calder, C.: Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. Environmental and Ecological Statistics, **14**, 229–247 (2007)
3. Carter, C.K., Kohn, R.: On Gibbs sampling for state space models. Biometrika, **81**, 541–553 (1994)
4. Fernandez, C., Ley, E., Steel, M.: Benchmark priors for Bayesian model averaging. Journal of Econometrics, **100**, 381–427 (2001)
5. Gamerman, D., Salazar E.: Hierarchical modeling in time series: the factor analytic approach. In book *Bayesian Theory and Applications (eds. P. Damien, P. Dellaportas, N. G. Polson, and D. A. Stephens)*, pp. 167-182, Oxford University Press (2013)
6. Gelfand, A.E., Ghosh, S.K.: Model choice: a minimum posterior predictive loss approach. Biometrika, **85**, 1–11 (1998)
7. George, E., Sun, D., Ni, S.: Bayesian stochastic search for VAR model restrictions. Journal of Econometrics, **142**, 553–580 (2008)
8. Geweke, J. F., Zhou, G.: Measuring the pricing error of the arbitrage pricing theory. The Review of Financial Studies, **9**, 557–587 (1996)
9. Huerta, G., West, M.: Priors and component structures in autoregressive time series models. Journal of the Royal Statistical Society, Series B, **61**, 881–899 (1999)
10. Kuethe, T., Pede, V.: Regional Housing Price Cycles: A Spatio-temporal Analysis Using US State-level Data. Regional Studies, **45**, 563–574 (2011)
11. Ippoliti, L., Valentini, P., Gamerman, D.: Space-Time Modelling of Coupled Spatio-Temporal Environmental Variables. Journal of the Royal Statistical Society, Series C (Applied Statistics), **61**, 175-200 (2012)
12. Lopes, H. F., West, M.: Bayesian model assessment in factor analysis. Statistica Sinica, **14**, 41–67 (2004)
13. Lopes, H. F., Salazar, E., Gamerman, D.: Spatial dynamic factor analysis. Bayesian Analysis, **3**, 759–792 (2008)
14. Pesaran, M.H.: Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, **74**, 967–1012 (2006)
15. Schmidt, A., Gelfand, A. E.: A Bayesian coregionalization model for multivariate pollutant data. Journal of Geophysics Research, **108**, 8783 (2003)
16. Valentini, P., Ippoliti, L., Fontanella, L.: Modeling US housing prices by spatial dynamic structural equation models. The Annals of Applied Statistics, **7**, 763-798 (2013)
17. Wikle, C. K., Cressie, N.: A dimension-reduced approach to space-time Kalman filtering. Biometrika, **86**, 815–829 (1999)