The Regional Economics Applications Laboratory (REAL) is a cooperative venture between the University of Illinois and the Federal Reserve Bank of Chicago focusing on the development and use of analytical models for urban and regional economic development. The purpose of the **Discussion Papers** is to circulate intermediate and final results of this research among readers within and outside REAL. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago, Federal Reserve Board of Governors or the University of Illinois. All requests and comments should be directed to Geoffrey J. D. Hewings, Director, Regional Economics Applications Laboratory, 607 South Matthews, Urbana, IL, 61801-3671, Phone (217) 333-4740, Fax (217) 244-9339.
Web page: www.uiuc.edu/unit/real

A HIERARCHICAL BAYESIAN APPROACH FOR SPATIAL TIME SERIES MODELING

by

Yiannis Kamarianakis

REAL 03-T-15 April, 2003

A HIERARCHICAL BAYESIAN APPROACH FOR SPATIAL TIME SERIES MODELING

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

Despite the fact that the amount of datasets containing long economic time series with a spatial reference has significantly increased during the years, the presence of integrated techniques that aim to describe the temporal evolution of the series while accounting for the location of the measurements and their neighboring relations is very sparse in the econometric literature. This paper shows how the Hierarchical Bayesian Space Time model presented by Wikle, Berliner and Cressie (Environmental and Ecological Statistics, l998) for temperature modeling, can be tailored to model relationships between variables that have both a spatial and a temporal reference. The first stage of the hierarchical model includes a set of regression equations (each one corresponding to a different location) coupled with a dynamic space-time process that accounts for the unexplained variation. At the second stage, the regression parameters are endowed with priors that reflect the neighboring relations of the locations under study; moreover, the spatio-temporal dependencies in the dynamic process for the unexplained variation are being established. Putting hyperpriors on previous stages' parameters completes the Bayesian formulation, which can be implemented in a Markov Chain Monte Carlo framework. The proposed modeling strategy is useful in quantifying the temporal evolution in relations between economic variables and this quantification may serve for excess forecasting accuracy.

1. Introduction

Statistical and econometric models that aim to describe the temporal evolution and the interrelationships between variables that have a spatial reference are being referred as space-time models. Research in such modeling techniques has significantly increased during the last twenty years since it is closely related to the progress in computer technology and the existence of large databases. Despite researchers' efforts, space-time modeling techniques do not lie in an integrated theoretical framework like for example the ARIMA methodology for time series; usually, the employed techniques vary according to the kind of application that needs to be performed.

Cliff and Ord (1975) were the first to perform (in a regression framework) a model that was taking into account both spatial and temporal relationships; in the early eighties, Pfeifer and Deutsch (1980a, 1980b, 1981a, 1981b, 1981c) presented the Space-Time Autoregressive Integrated Moving Average (STARIMA) models, aiming to offer a tool for spatio-temporal modeling analogous to the ARIMA methodology for univariate time series. The STARIMA methodology has been applied in a wide variety of applications ranging from environmental (Pfeifer and Deutsch 1981a, Stoffer 1986), to epidemiological (Pfeifer and Deutsch 1980a), econometric (Pfeifer and Bodily 1990), and traffic flow (Kamarianakis and Prastacos 2002, 2003) to name just a few. Data limitations (usually in the temporal dimension) and modeling needs in regional economics' applications, forced researchers to develop space-time models different from the STARIMA ones. As significant contributions towards this direction we refer the dynamic space-time model that includes an instantaneous spatial interaction term for the response presented by Elhorst, (2001), and the Bayesian Vector Autoregressive models with spatial priors on the parameters, LeSage and Krivelyova (1999).

This paper proposes a hierarchical Bayesian method for modeling a dependent time series variable measured at different locations, relative to a set of independent time series variables that may or may not have a spatial reference. The aforementioned relationship lies in a regression framework that in the case of independent variables with spatial reference resembles the

Seemingly Unrelated Regressions (SUR) model introduced by Zellner (1962) that is often employed in the Bayesian framework (see for example Griffiths, 2001). The parameters in the regression model are endowed with priors that reflect the neighboring relations between the locations of the study; moreover, a spatio-temporal process is included to account for the unexplained variation. The followed approach is influenced by the one adopted by Wikle *et al.* (1998)[a] for modeling environmental processes as far as the design of the hierarchical Bayesian methodology and the presence of the dynamic spatio-temporal term are concerned; the main differences lie in the formulation of the spatial dependencies and the regression part in the first stage of our model.

The hierarchical steps of the Bayesian methodology are presented in the section that follows; the third section contains the distributional assumptions that characterize each of the aforementioned steps and the formulation of the spatial and spatio-temporal relations. In the sequel, the full conditional posterior distributions of the model parameters are derived. The fifth and last section contains a discussion on the Markov Chain Monte Carlo (MCMC) implementation of the model via a Gibbs sampler.

2. An overview of the Hierarchical Bayesian Methodology

The hierarchical Bayesian methodology and the MCMC estimation approach, decompose complicated estimation problems into simpler ones that rely on the conditional distributions for each parameter in the model. This innovation makes application of the Bayesian methodology far easier than past approaches that relied on analytical solution of the posterior distribution. As LeSage (2002) indicates, a result of this is that extensible toolkits for solving large classes of estimation problems can be developed at both a theoretical and applied level.

The proposed methodology models the relationship between a response variable Y measured at locations indexed by $s,\ s \in \{1,..,S\}$ which may be states, regions, prefectures, cities etc. and times

---

[a] This paper presents a model for the spatio-temporal evolution of a single environmental process measured at sites located at a grid. Modeling on the various stages' priors is based in the notion of spatial Markov fields.

t, where $t \in \{1,...,T\}$, and p explanatory variables which can be measured either at the same spatio-temporal domain or they may have no spatial reference at all. For the sake of simplicity we present the second case that can be generalized in a straightforward way. Proceeding in a similar way as in Wikle *et al.* (1998), at a first stage the model for the response variable Y is conditional on 2 processes, β, $K = \{K(s,t) : (s,t) \in D\}$, and a collection of parameters θ_1. At each location and time point the general model is of the form

$$(2.1) \quad Y(s,t) = M(X, \beta(s)) + K(s,t) + \gamma(s,t)$$

where $M(X, \beta(s))$ represents a regression model with site dependent coefficients; X is the set of p explanatory variables that may or may not have a spatial reference, and $\beta(s) = (\beta_0(s),..., \beta_p(s))$ represents the set of spatially referenced regression coefficients for each location s. $K(s,t)$ stands for a process that accounts for spatio-temporal dependencies. The $\gamma(s,t)$'s represent the unexplained variation at the first stage of the modeling process which in principle should be modeled as a $ST \times ST$ covariance matrix. However, taking into account that the K process explains much of the space-time structure of Y, one might assume that the $Y(s,t)$'s are conditionally independent random variables. That is model (2.1), can be formally written as

$$(2.2) \quad Y(s,t) \sim N\left(M(X, \beta(s)) + K(s,t), \sigma_Y^2(s)\right)$$

where

$$\theta_1 = \left(\sigma_Y^2(1),..., \sigma_Y^2(s)\right).$$

In the second stage of the hierarchical Bayesian method, the β and K processes are assumed to be independent, conditional on the second stage parameters θ_2 that can be partitioned as $\theta_2 = (\theta_\beta, \theta_K)$ leading to

$$[\beta, K \mid \theta_2] = [\beta \mid \theta_\beta][K \mid \theta_K].$$

$K(s,t)$ stands for a space time process which in general can be described by the model

$$[K \mid \theta_K] = \prod_t [K(.,t+1) \mid K(.,t)]$$

that includes

$$K_{t+1} = HK_t + \eta_{t+1}$$

as a special case and that form is that that we use from now on. In this case H is a $S \times S$ matrix of regression coefficients and η_t is an independent sequence of zero mean errors.

The third modeling stage is the specification of $[\theta_1, \theta_2 \mid \theta_3]$, where θ_3 is a collection of hyperparameters. We assume a partition $\theta_3 = (\theta_3(1), \theta_3(2))$ into hyperparameters associated with each stage and a conditional independence relation

$$[\theta_1, \theta_2 \mid \theta_3] = [\theta_1 \mid \theta_3(1)][\theta_2 \mid \theta_3(2)]$$

and $\theta_3(2)$ can be partitioned as $\theta_3(2) = (\theta_3(\beta), \theta_3(K))$ and coupled with further conditional independence assumptions

$$[\theta_2 \mid \theta_3(2)] = [\theta_\beta \mid \theta_3(\beta)][\theta_K \mid \theta_3(K)] \, .$$

Conditional independence is also assumed for the hyperpriors

$$[\theta_3] = [\theta_3(1)][\theta_3(\beta)][\theta_3(K)]$$

and the formulation can be simplified by taking $\theta_3(1)$ to be either empty or known so that the corresponding term in the above equation drops out.

3. Distributional assumptions and spatio-temporal dynamics

Equation (2.1) for the process of interest Y can be written as

$$Y_t = M_t + K_t + \gamma_t$$

where each term is a $S \times 1$ vector, M_t is a regression model with parameters endowed with priors that reflect spatial dependencies, K_t is a dynamical process that accounts for the unexplained space-time variability and γ_t is an error term. Y_t's are conditionally Gaussian such that

$$(3.1) \quad [Y_t | M_t, K_t, \sigma_\gamma^2] \sim N(M_t + K_t, \sigma_\gamma^2).$$

M_t and K_t are assumed to be mutually independent conditional on second stage parameters; the model for M_t is a system of regression equations, each one corresponding to a different location

$$M(X, \beta(s)) = \beta_0(s) + \beta_1(s) X_1 + ... + \beta_p(s) X_p .$$

X stands for a $T \times p+1$ matrix[b] that contains information on p explanatory variables in time series form that may or may not have a spatial reference. At this point we have to introduce a matrix that reflects the neighboring relations between the locations where the observations where taken; it is denoted by W and a nonzero w_{sl} element indicates a neighboring relation for the locations s, l. This matrix can be of the nearest-neighbor, spatial contiguity or inverse distance form. Each $\beta_i(s)$ is modeled as

$$(3.2) \quad \beta_i(s) \Big| \{\beta_i(l) : s \neq l\} \sim N\left(\alpha_i + \rho_i \sum_{l=1}^{S} w_{sl} \beta_i(l), \tau^2 \right)$$

---

[b] In the case of different sets of independent variables corresponding to each location we have a block diagonal $ST \times p+1$ matrix.

and the following expression holds for each S-vector β_i

$$(3.3) \quad \beta_i \left| \{\alpha_i, \rho_i, \tau^2\} \sim N\left((I-\rho_i W)^{-1}\alpha_i, \tau^2 \left[(I-\rho_i W)' (I-\rho_i W) \right]^{-1} \right). \right.$$

The space-time dynamic term is modeled as a vector autoregressive (VAR) process

$$(3.4) \quad K_t = HK_{t-1} + \eta_t$$

where H is an $S \times S$ matrix and η_t is the VAR noise term. Equation (3.4) is formally written as

$$(3.5) \quad K(s,t) = b(s)K(s,t-1) + c(s)\sum_{l=1}^{S} K(l,t-1)1(w_{s,l} \neq 0) + \eta(s,t)$$

that is the autoregressive parameter for each location under consideration varies spatially, as the parameter for the spatial dependencies. The distributional assumption for the K process is formalized as

$$(3.6) \quad K_t \left| K_{t-1}, H, \sigma_\eta^2 \sim N\left(HK_{t-1}, \sigma_\eta^2\right) \right.$$

and H is a matrix with nonzero elements at the diagonal and at positions where the corresponding elements of W are nonzero. Depending on the modeling strategies on this second stage, the implied models might not be identifiable because *K(s,t)* and *γ(s,t)* appear only through their sum in (2.1). As Wikle et al. point out: " In a Bayesian analysis with proper probabilities on all quantities, identifiability issues do not prohibit us from proceeding, though we should be careful in interpreting results from unidentified parameters." For a general discussion on this issue the interested reader is referred to Besag *et al*. (1995).

As indicated in the second section, we partition the third stage priors and assume conditional independence. The S-vector α_i in relation (3.2) is specified to be a Gaussian random variable

(3.7) $\alpha_i \sim N\left(\widetilde{\alpha}_i, \widetilde{\sigma}_a^2\right)$

and the $\widetilde{\alpha}_i, \widetilde{\sigma}_\alpha^2$ can be given values that reflect the lack of information about the α_i. For the ρ_i's we follow LeSage (2002) and put a uniform prior over the interval $\left[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right]$, where λ_{min}, λ_{max} represent the minimum and the maximum eigenvalues of the spatial weight matrix. That is we restrict the parameter ρ_i to its feasible range for row standardized W

(3.8) $\rho_i \sim U\left[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right]$.

Similar specifications hold for the parameters that correspond to the spatio-temporal dynamics

(3.9) $b(s) \sim N\left(\widetilde{b}(s), \widetilde{\sigma}_{b(s)}^2\right)$

(3.10) $c(s) \sim N\left(\widetilde{c}(s), \widetilde{\sigma}_{c(s)}^2\right)$

For the variances specified in the first two stages, we assume independence and use the conjugate priors.

(3.11) $\sigma_\gamma^2 \sim IG\left(\widetilde{q}_\gamma, \widetilde{r}_\gamma\right)$

(3.12) $\sigma_\eta^2 \sim IG\left(\widetilde{q}_\eta, \widetilde{r}_\eta\right)$

(3.13) $\tau^2 \sim IG\left(\widetilde{q}\ , \widetilde{r}\ \right)$

where *IG* refers to the inverse Gamma distribution.

4. Derivation of the full conditional distributions

This section outlines the derivation of the full conditional distributions that can be used in the Gibbs sampling framework. In general, full conditional distributions are determined by writing the joint distributions of all random quantities divided by the appropriate normalizing constant.

In hierarchical models this process is simplified due to the various conditional independence assumptions. In particular, all components of the full joint distribution that do not functionally depend on the quantity 'cancel' from the numerator and denominator of the full conditional distribution. The following derivations begin after these simplifications have been considered. The generic notation $[A|\cdot]$ and $A|\cdot$ is used to represent the conditional distribution for *A* given all other random quantities. It should be noted that the majority of the posteriors presented here are modified versions of the ones presented at Wikle *et al.* (1998).

$[K_t|\cdot]$

From (3.1), (3.6), for $t = 1,...,T-1$ the following relationship holds

$$[K_t|\cdot] \propto [Y_t|M_t, K_t, \sigma_\gamma^2][K_{t+1}|K_t, H, \sigma_\eta^2][K_t|K_{t-1}, H, \sigma_\eta^2]$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_\gamma^2}(Y_t - [M_t + K_t])'(Y_t - [M_t + K_t]) + \frac{1}{\sigma_\eta^2}(K_t - HK_{t-1})'(K_t - HK_{t-1}) + \frac{1}{\sigma_\eta^2}(K_{t+1} - HK_t)'(K_{t+1} - HK_t)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[K_t'\left(\frac{1}{\sigma_\gamma^2}I + \frac{1}{\sigma_\eta^2}H'H + \frac{1}{\sigma_\eta^2}I\right)K_t - 2\left(\frac{1}{\sigma_\gamma^2}(Y_t - M_t)' + \frac{1}{\sigma_\eta^2}K_{t+1}'H + \frac{1}{\sigma_\eta^2}K_{t-1}H'\right)K_t\right]\right\}$$

where we need the initial condition K_0. Thus

$$K_t|\cdot \sim N\left[\left(\frac{1}{\sigma_\gamma^2}I + \frac{1}{\sigma_\eta^2}H'H + \frac{1}{\sigma_\eta^2}I\right)^{-1}\left(\frac{1}{\sigma_\gamma^2}(Y_t - M_t)' + \frac{1}{\sigma_\eta^2}K_{t+1}'H + \frac{1}{\sigma_\eta^2}K_{t-1}'H'\right)', \left(\frac{1}{\sigma_\gamma^2}I + \frac{1}{\sigma_\eta^2}H'H + \frac{1}{\sigma_\eta^2}I\right)^{-1}\right].$$

Similarly, for $t=T$,

$$[K_T|\cdot] \propto [Y_T|M_T, X_T, \sigma_\gamma^2][X_T|X_{T-1}, H, \sigma_\eta^2]$$

which, as above leads to

$$K_T \left| \cdot \sim N \left[\left(\frac{1}{\sigma_\gamma^2} I + \frac{1}{\sigma_\eta^2} I \right)^{-1} \left(\frac{1}{\sigma_\gamma^2} \left(Y_T - M_T \right)' + \frac{1}{\sigma_\eta^2} K'_{T-1} H' \right)', \left(\frac{1}{\sigma_\gamma^2} I + \frac{1}{\sigma_\eta^2} I \right)^{-1} \right] \right. .$$

If $K_0 \left| \mu_{K_0}, \Sigma_{K_0} \sim N\left(\mu_{K_0}, \Sigma_{K_0}\right) \right.$ the full conditional distribution for K_0 is given by

$$[K_0|\cdot] \propto \left[K_1 \middle| K_0, H, \sigma_\eta^2 \right] \left[K_0 \middle| \mu_{K_0}, \Sigma_{K_0}\right]$$

$$\propto \exp\left\{ -\frac{1}{2\sigma_\eta^2} \left(K'_0 H'H K_0 - 2K_1 H K_0 \right) - \frac{1}{2} \left(K_0 - \mu_{K_0} \right)' \Sigma_{K_0}^{-1} \left(K_0 - \mu_{K_0} \right) \right\}$$

which leads to

$$K_0 \left| \cdot \sim N \left[\left(\frac{1}{\sigma_\eta^2} H'H + \Sigma_{K_0}^{-1} \right)^{-1} \left(\frac{1}{\sigma_\eta^2} K'_1 H + \mu'_{K_0} \Sigma_{K_0}^{-1} \right)', \left(\frac{1}{\sigma_\eta^2} H'H + \Sigma_{K_0}^{-1} \right)^{-1} \right] \right. .$$

$[\beta_i|\cdot]$

Using the distributions in (3.1) and (3.3), we can derive

$$[\beta_i|\cdot] \propto \left[\beta_i \middle| \alpha_i, \rho_i, \tau^2\right] \prod_{t=1}^{T} \left[Y_t \middle| M_t, K_t, \sigma_\gamma^2\right]$$

$$\propto \exp\left\{ -\frac{1}{2\tau^2} \left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right)' (I-\rho_i W)' (I-\rho_i W) \left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right) \right\} \times \exp\left\{ -\frac{1}{2\sigma_\gamma^2} \sum_{t=1}^{T} \left(Y_t - \sum_{i=0}^{p} \beta_i X_i - K_t \right)' \left(Y_t - \sum_{i=0}^{p} \beta_i X_i - K_t \right) \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left[\beta'_i \left(\frac{1}{\tau^2} (I-\rho_i W)' (I-\rho_i W) + \frac{T}{\sigma_\gamma^2} X_i X_i' \right) \beta_i - 2 \left(\frac{1}{\tau^2} (I-\rho_i W)^{-1} \alpha_i (I-\rho_i W)' (I-\rho_i W) + \frac{1}{\sigma_\gamma^2} \sum_{t=1}^{T} \left(Y_t - \sum_{i=0}^{p} \beta_i X_i - K_t \right)' \right) \beta_i \right] \right\} .$$

Thus

$$\beta_i \left| \cdot \sim N \left[\left(\frac{1}{\tau^2} (I-\rho_i W)' (I-\rho_i W) + \frac{T}{\sigma_\gamma^2} X_i X_i' \right)^{-1} \left(\frac{1}{\tau^2} (I-\rho_i W)^{-1} \alpha_i (I-\rho_i W)' (I-\rho_i W) + \frac{1}{\sigma_\gamma^2} \sum_{t=1}^{T} \left(Y_t - \sum_{i=0}^{p} \beta_i X_i - K_t \right)' \right)', \left(\frac{1}{\tau^2} (I-\rho_i W)' (I-\rho_i W) + \frac{T}{\sigma_\gamma^2} X_i X_i' \right)^{-1} \right] \right. .$$

$\left[\alpha_i \middle| \cdot\right]$

In this case we use (3.3) and (3.7) and derive

$$\left[\alpha_i \middle| \cdot\right] \propto \left(\left[\beta_i \middle| \alpha_i, \rho_i, \tau^2\right]\right)\left(\alpha_i \middle| \tilde{\alpha}_i, \sigma_{\alpha_i}^2\right)$$

$$\propto \exp\left\{-\frac{1}{2\tau^2}\left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right)'(I-\rho_i W)'(I-\rho_i W)\left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right)\right\} \times \exp\left\{-\frac{1}{2\sigma_{\alpha_i}^2}(\alpha_i - \tilde{\alpha}_i)'(\alpha_i - \tilde{\alpha}_i)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\alpha_i'\left(\frac{1}{\tau^2} - \frac{1}{\sigma_{\alpha_i}^2}\right)\alpha_i\right) - 2\left((I-\rho_i W)\beta_i + \frac{1}{\sigma_{\alpha_i}^2}\tilde{\alpha}_i\right)\alpha_i\right\}.$$

Thus the posterior takes the form

$$\alpha_i \middle| \cdot \sim N\left(\left(\frac{1}{\tau^2} - \frac{1}{\sigma_{\alpha_i}^2}\right)^{-1}\left((I-\rho_i W)\beta_i + \frac{1}{\sigma_{\alpha_i}^2}\tilde{\alpha}_i\right), \left(\frac{1}{\tau^2} - \frac{1}{\sigma_{\alpha_i}^2}\right)^{-1} I_s\right)$$

$\left[\rho_i \middle| \cdot\right]$

From (3.3) and (3.8) we have

$$\left[\alpha_i \middle| \cdot\right] \propto \left(\left[\beta_i \middle| \alpha_i, \rho_i, \tau^2\right]\right)\left(\rho_i \middle| \lambda_{\min}, \lambda_{\max}\right)$$

$$\propto \exp\left\{-\frac{1}{2\tau^2}\left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right)'(I-\rho_i W)'(I-\rho_i W)\left(\beta_i - (I-\rho_i W)^{-1}\alpha_i\right)\right\} \times \frac{1}{\lambda_{\max}^{-1} - \lambda_{\min}^{-1}}$$

$$\propto \exp\left\{\left(-\frac{1}{2}\left(\rho_i'\left(\frac{1}{\tau^2}\beta_i' W' W \beta_i\right)\rho_i\right)\right) - 2\left(\frac{1}{\tau^2} W \beta_i (1+\alpha_i)\right)\rho_i\right\}$$

Thus,

$$\rho_i \middle| \cdot \sim N\left(\left(\frac{1}{\tau^2}\beta_i' W' W \beta_i\right)^{-1}\left(\frac{1}{\tau^2} W \beta_i (1+\alpha_i)\right), \left(\frac{1}{\tau^2}\beta_i' W' W \beta_i\right)^{-1}\right).$$

$\left[b \middle| \cdot\right]$

From the VAR structure (3.4), (3.5), we can write the following decomposition

$$HK_{t-1} = diag(b)K_{t-1} + H_b K_{t-1} = diag(K_{t-1})b + H_b K_{t-1}$$

where H_b is the H matrix with the main diagonal replaced by zeros. Then using (3.6), (3.9)

$$[b|\cdot] \propto \prod_{t=1}^{T} \left[K_t \middle| K_{t-1}, H, \sigma_\eta^2\right] \left[b \middle| \tilde{b}, \tilde{\sigma}_b^2\right]$$

$$\propto \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - diag(K_{t-1})b - H_b K_{t-1}\right)'\left(K_t - diag(K_{t-1})b - H_b K_{t-1}\right)\right\} \times \exp\left\{-\frac{1}{2\tilde{\sigma}_b^2}\left(b-\tilde{b}\right)'\left(b-\tilde{b}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[b'\left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(diag(K_{t-1})\right)^2 + \frac{1}{\tilde{\sigma}_b^2}I\right)b - 2\left[\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - H_b K_{t-1}\right)' diag(K_{t-1}) + \frac{1}{\tilde{\sigma}_b^2}\tilde{b}\right]b\right]\right\}$$

and

$$b|\cdot \sim N\left[\left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(diag(K_{t-1})\right)^2 + \frac{1}{\tilde{\sigma}_b^2}I\right)^{-1}\left[\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - H_b K_{t-1}\right)' diag(K_{t-1}) + \frac{1}{\tilde{\sigma}_b^2}\tilde{b}\cdot I\right], \left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(diag(K_{t-1})\right)^2 + \frac{1}{\tilde{\sigma}_b^2}I\right)^{-1}\right].$$

$[c|.]$

From (3.5) we can write the decomposition

$$HK_{t-1} = c(s)K_{t-1}^{c(s)} + H_{c(s)}K_{t-1}$$

where $H_{c(s)}$ is the is the H matrix with the elements corresponding to nonzero elements of the s^{th} row of the W matrix replaced by zeros, and $K_{t-1}^{c(s)} \equiv J_{c(s)}K_{t-1}$, where $J_{c(s)}$ is an $S \times S$ matrix with ones for elements of the s^{th} row that correspond to nonzero elements of the s^{th} row of the W matrix. Then, using (3.6), (3.10)

$$[c|\cdot] \propto \prod_{t=1}^{T} \left[K_t \middle| K_{t-1}, H, \sigma_\eta^2\right] \left[c(s) \middle| \tilde{c}(s), \tilde{\sigma}_{c(s)}^2\right]$$

$$\propto \exp\left\{-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - c(s)K_{t-1}^{c(s)} - H_{c(s)}K_{t-1}\right)'\left(K_t - c(s)K_{t-1}^{c(s)} - H_{c(s)}K_{t-1}\right)\right\} \times \exp\left\{-\frac{1}{2\tilde{\sigma}_{c(s)}^2}\left(c(s) - \tilde{c}(s)\right)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[c(s)\left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T} K_{t-1}^{c(s)\prime} K_{t-1} + \frac{1}{\widetilde{\sigma}_{c(s)}^2}\right)c(s) - 2c(s)\left(\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - H_{c(s)}K_{t-1}\right)' K_{t-1}^{c(s)} + \frac{\widetilde{c}(s)}{\widetilde{\sigma}_{c(s)}^2}\right)\right]\right\}$$

$$c(s)\big|\cdot \sim N\left[\left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T} K_{t-1}^{c(s)\prime} K_{t-1} + \frac{1}{\widetilde{\sigma}_{c(s)}^2}\right)^{-1}\left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T}\left(K_t - H_{c(s)}K_{t-1}\right)' K_{t-1}^{c(s)} + \frac{\widetilde{c}(s)}{\widetilde{\sigma}_{c(s)}^2}\right), \left(\frac{1}{\sigma_\eta^2}\sum_{t=1}^{T} K_{t-1}^{c(s)\prime} K_{t-1} + \frac{1}{\widetilde{\sigma}_{c(s)}^2}\right)^{-1}\right].$$

$\left[\sigma_\gamma^2\big|\cdot\right]$

From (3.1), (3.11)

$$\left[\sigma_\gamma^2\big|\cdot\right] \propto \prod_{t=1}^{T}\left[Y_t\big|M_t, K_t, \sigma_\gamma^2\right]\left[\sigma_\gamma^2\big|\widetilde{q}_\gamma, \widetilde{r}_\gamma\right]$$

$$\propto \frac{1}{\left(\sigma_\gamma^2\right)^{\frac{ST}{2}+\widetilde{q}_\gamma+1}} \exp\left\{-\frac{1}{2\sigma_\gamma^2}\left[\frac{2}{\widetilde{r}_\gamma} + \sum_{t=1}^{T}\left(Y_t - M_t - K_t\right)' \times \left(Y_t - M_t - K_t\right)\right]\right\}$$

and

$$\sigma_\gamma^2\big|\cdot \sim IG\left(\frac{ST}{2} + \widetilde{q}_\gamma, \left[\frac{2}{\widetilde{r}_\gamma} + \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - M_t - K_t\right)' \times \left(Y_t - M_t - K_t\right)\right]^{-1}\right).$$

$\left[\sigma_\eta^2\big|\cdot\right]$

From (3.6) and (3.12)

$$\left[\sigma_\eta^2\big|\cdot\right] \propto \prod_{t=1}^{T}\left[K_t\big|H, K_{t-1}, \sigma_\eta^2\right]\left[\sigma_\eta^2\big|\widetilde{q}_\eta, \widetilde{r}_\eta\right]$$

$$\propto \frac{1}{\left(\sigma_\eta^2\right)^{\frac{ST}{2}+\widetilde{q}_\eta+1}} \exp\left\{-\frac{1}{2\sigma_\eta^2}\left[\frac{2}{\widetilde{r}_\eta} + \sum_{t=1}^{T}\left(K_t - HK_{t-1}\right)'\left(K_t - HK_{t-1}\right)\right]\right\}$$

and

$$\sigma_\eta^2\big|\cdot \sim IG\left(\frac{ST}{2} + \widetilde{q}_\eta, \left[\frac{2}{\widetilde{r}_\eta} + \frac{1}{2}\sum_{t=1}^{T}\left(K_t - HK_{t-1}\right)'\left(K_t - HK_{t-1}\right)\right]^{-1}\right)$$

$\left[\tau^2\big|\cdot\right]$

From (3.4), (3.14) we obtain

$$\left[\tau^2\middle|\cdot\right] \propto \left[\beta_i\middle|\alpha_i,\rho_i,\tau^2\right]\left[\tau^2\middle|\tilde{q},\tilde{r}\right]$$

$$\propto \frac{1}{\left(\tau^2\right)^{\frac{S}{2}}}\exp\left\{-\frac{1}{2\tau^2}\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)'\left(I-\rho_iW\right)'\left(I-\rho_iW\right)\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)\right\}\times\frac{1}{\Gamma(\tilde{q})}\tilde{r}^{-\tilde{q}}\frac{1}{\left(\tau^2\right)^{\tilde{q}+1}}\exp\left\{-\frac{1}{\tilde{r}\tau^2}\right\}$$

$$\propto \frac{1}{\left(\tau^2\right)^{\frac{S}{2}+\tilde{q}+1}}\exp\left\{-\frac{1}{2\tau^2}\left[\frac{2}{\tilde{r}}+\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)'\left(I-\rho_iW\right)'\left(I-\rho_iW\right)\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)\right]\right\}$$

and

$$\tau^2\middle|\cdot \sim IG\left(\frac{S}{2}+\tilde{q},\left[\frac{2}{\tilde{r}}+\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)'\left(I-\rho_iW\right)'\left(I-\rho_iW\right)\left(\beta_i-\left(I-\rho_iW\right)^{-1}\alpha_i\right)\right]\right).$$

5. Bayesian Estimation: Gibbs Sampler

The role of MCMC estimation is, based on conditional posteriors, to produce conclusions that are unconditional. This is accomplished by sampling over values of the conditioning parameters, rather than integrating, which is the formal procedure for inverting conditional distributions to unconditional. In our case, since the form of the conditional distributions is known we can use the "Gibbs", or "alternating conditional" sampling approach.

Given initial values for the parameters of our problem, we can draw one observation from each K_t from $\left[K_t\middle|\cdot\right]$, use these K_t's when sampling from $\left[\beta_i\middle|\cdot\right]$ to produce a first draw from the p S-vectors β_i, take draws for the S-vectors α_i, ρ_i using in their posteriors the β_i draws taken in the previous step and so on. At a second step, we update K_t by sampling from its posterior that now uses information from the first draws we took for $\sigma_\eta^2,\sigma_\gamma^2$ and each β_i, then we update similarly the β_i's and so on. This process of alternating sampling from the conditional distributions is continued until a large sample of draws has been collected. This is not an ad-hoc procedure, as formal mathematical demonstrations provided by Geman and Geman (1984), as well as Gelfand and Smith (1990), show that the stochastic process representing our parameters is a Markov chain with the correct equilibrium distribution.

While theory implies that the Markov chain is guaranteed to converge to the appropriate stationary distribution, implementation issues arise in practice. One must make choices related to the influence of starting values, how long to run the chain before convergence and how best to monitor the chain and perform the desired estimation. A common procedure is to delete the observations taken for the model parameters that correspond to the initial iterations of the Markov chain when convergence is not yet reached. For convergence diagnostics we can use a criterion like the one provided by Gelman and Rubin (1992). Finally, due to correlations of MCMC samples the Monte Carlo standard errors should be estimated by the 'batch means' approach described in Roberts (1996) with the batch size determined from examination of the lag autocorrelation plots of several parameters as obtained from pilot samples.

6. References

BESAG, J., GREEN, P., HIGDON, D. and MENGERSEN, K. (1995) Bayesian computation and stochastic systems. Statistical Science, 10, 3-66

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

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

GELMAN, A. and RUBIN, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science,* 7, 457-472.

GEMAN, S. and GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 6, 721-741.

GRIFFITHS, W.E. (2001) Bayesian inference in the seemingly unrelated regressions model. *Computer-Aided Econometrics* (editor D. Giles), Marcel Dekker.

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. *Journal of the Transportation Research Board.*

LESAGE, J.P. (2002) Application of bayesian methods to spatial econometrics, working paper, Department of Economics, University of Toledo.

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).

ROBERTS, G.O. (1996) Markov chain concepts related to sampling algorithms. *Markov Chain Monte Carlo in Practice,* Chapman and Hall, London, 45-57.

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.

WIKLE, C.K., BERLINER, M.L. and CRESSIE, N. (1998) Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5, 117-154

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