ISSN 1415-4765

TEXTO PARA DISCUSSÃO N° 905

MULTIVARIATE SPATIAL REGRESSION MODELS

Dani Gamerman*
Ajax R. B. Moreira**

Rio de Janeiro, setembro de 2002

* Do Instituto de Matemática da UFRJ.
dani@im.ufrj.br

** Da Diretoria de Estudos Macroeconômicos do IPEA.
ajax@ipea.gov.br

Governo Federal

Ministério do Planejamento, Orçamento e Gestão

Ministro – Guilherme Gomes Dias
Secretário Executivo – Simão Cirineu Dias

Fundação pública vinculada ao Ministério do Planejamento, Orçamento e Gestão, o IPEA fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

TEXTO PARA DISCUSSÃO

Uma publicação que tem o objetivo de divulgar resultados de estudos desenvolvidos, direta ou indiretamente, pelo IPEA e trabalhos que, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

SUMMARY

SINOPSE

Este artigo descreve os procedimentos para a inferência bayesiana de modelos multivariados. Esses modelos incluem uma componente espacial que é comumente utilizada em modelos econômicos. Em particular, os modelos de regressões aparentemente não-relacionadas Sure e vetores auto-regressivos são estendidos para acomodar a dependência espacial. Os procedimentos de inferência são baseados em uma variedade de esquemas de simulação desenhados para obter amostras da distribuição *a posteriori* dos parâmetros do modelo. Estes podem ser utilizados para prover também estimativas da previsão de novas observações.

ABSTRACT

This paper describes the inference procedures required to perform Bayesian inference to some multivariate econometric models. These models have a spatial component built into commonly used multivariate models. In particular, the seemingly unrelated regression and vector autoregressive models are addressed and extended to accommodate for spatial dependence. Inference procedures are based on a variety of simulation-based schemes designed to obtain samples from the posterior distribution of model parameters. They are also used to provide a basis to forecast new observations.

Multivariate spatial regression models

Dani Gamerman and Ajax R. B. Moreira
Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil
Instituto de Pesquisa Econômica Aplicada, Brazil
dani@im.ufrj.br, ajax@ipea.gov.br

Key words: Bayesian; Gibbs sampling; Hyperparameters; Markov chain Monte Carlo; Metropolis-Hastings algorithm; Vector autoregressive models.

1 Introduction

Econometric systems typically involve analysis of a collection of economic variables generated from within the system. These are called endogenous variables and will be denoted here by $y = (y_1, ..., y_q)'$. In some cases, these are affected by another collection of variables, generated from outside the system. These are called exogenous variables and will be denoted here by $x = (x_1, ..., x_p)'$. Basic models try to relate these two sets of variables by a linear form. One can obtain estimates of the strength in this relationship by estimating model parameters based on a collection of measurements repeated over an appropriate scale.

The most common scale over which measurements are made is time. In this case, one collect values of y and x over a number of time periods, for example years or quarters. In this case, one is analysing only a single system or economy as it evolves through time. An example is given by the macroeconomic study of interest rate, exchange rate and price index of the Brazilian economy after the Real Plan in 1994 (Gamerman and Moreira, 2002). In this case, the number of endogenous variables is $q = 3$ and data

was collected monthly.

Albeit the most common scale, this is by no means the only scale available for econometric data collection. Another possible scale is given by the spatial location of the system under analysis. In this case, interest lies in analysing data over a collection of spatially different economic systems. An example is given in the study of land deforestation in the Amazon region of Brazil (Andersen, Granger and Reis, 1997). In this case, the endogenous variables are given by different forms of land use and the data was collected over the municipalities of the region.

Another possiblity would be to accommodate both spatial and temporal domains in the scale. This was done in the econometric analysis of Gamerman, Moreira and Rue (2002). They built on the experience of Andersen, Granger and Reis (1997) with deforestation and included a temporal scale associated with the evolution of land uses through time.

In this paper, only the spatial scale will be considered. Multivariate regression models that accommodate the temporal scale have been extensively studies by Harvey (1989) from a frequentist perspective and West and Harrison (1997) from a Bayesian perspective. In the next section, basic results about the static multivariate regression models are presented from a frequentist and a Bayesian perspective. Section 3 introduces the spatial models considered in this work. The form of prior distribution used in the Bayesian analysis of these models is discussed in Section 4. Section 5 presents the results regarding the posterior distribution and describes the simulation schemes required to draw samples from the distribution. The paper is concluded with a few final remarks on Section 5.

2 Multivariate regression models

Let us assume there are n measurements available on both y and x. These usually are related in the econometric literaure (see Zellner, 1971; Hamilton, 1994) through the structural from

$$Ay_t = \Gamma x_t + u_t \text{ where } u_t \sim N(0, \Lambda)\,,\ t = 1, ..., n \tag{1}$$

with $y_t = (y_{t1}, ..., y_{tq})'$ and $x_t = (x_{t1}, ..., x_{tp})'$ are the vectors of endogenous and exogenous variables at unit t, A is a $q \times q$ matrix with contemporaneous (linear) relations between the endogenous variables, Γ is a $q \times p$ matrix of regression coefficients and Λ is a diagonal matrix with entries $\lambda_1^2, ..., \lambda_q^2$. The j-th line of Γ contains the vector of regression coefficients associated with the regression for y_j, $j = 1, ..., q$. Note that the only link between the q univariate regressions is through A. The model can also be written in the more compact form

$$YA = X\Gamma' + U \text{ where } U \sim N(0, I_n, \Lambda) \tag{2}$$

where $Y = (y_1, ..., y_n)'$ and $X = (x_1, ..., x_n)'$.

In many cases, it is better to work with the solution for y_t in (1). This can be obtained if A is full rank. In this case,

$$\begin{aligned} y_t &= A^{-1}\Gamma x_t + A^{-1}u_t \\ &= Bx_t + e_t \text{ where } e_t \sim N(0, \Sigma),\,,\ t = 1, ..., n \end{aligned} \tag{3}$$

or, in matrix form,

$$Y = XB' + E \text{ where } E \sim N(0, I_n, \Sigma)$$

where $B = A^{-1}\Gamma$ and $\Sigma = A^{-1}\Lambda A^{-1}$'. Equation (3) is known as the reduced form of the model.

This model could also be written in the misleading univariate regression fashion

$$y^j = X\beta_j + e^j \ , \ j = 1, ..., q, \tag{4}$$

where $z^j = (z_{1j}, ..., z_{nj})$ and β_j is the vector of regression coefficients for the j-th regression, for $j = 1, ..., q$. The β_j's may have some 0 valued components to accommodate for different exogenous variables for each regression. From expression (4), it would *seem* that the model is a collection of q unrelated univariate regressions for the q endogenous variables. Hence, the set of equations in (4) is known as seeemingly unrelated regressions (SUR). In fact, one can perform least squares separately for each of the q equations and this operation will lead to consistent estimators for the β_j's and, consequently for $B = (\beta_1, ..., \beta_q)'$. This formulation, however, forgets the dependence between the variables through the error component caused by the non-diagonal form of Σ. When this is considered, more efficient estimators can be obtained. Nevertheless, if the exogenous variables are the same for all regressions, then the ordinary least squares estimator of the β_j's obtained is the same irrespective of whether the error correlation is ignored or not.

The route pursued here considers a Bayesian paradigm for inference. Under this paradigm, a prior distribution must be specified to the model parameters. This distribution should reflect the previous knowledge of the econometrician about the process under study. This knowledge could stem from theoretical considerations, from empirical experience in similar analyses or a combination of both. The reader is referred to Migon and Gamerman (1999) for a fuller account of the Bayesian paradigm and a critical comparison with the frequentist approach.

From now on, adherence to the Bayesian perspective is assumed. In addition to the theoretical advantages of this approach, it will be seen that this approach will prove very useful from a practical point of view. It provides

a natural basis for proposing the model elaboration required to adequately describe the observation variability.

Therefore, a prior distribution for the model parameters (B, Σ) is required. A useful and flexible class of priors is given by the matricvariate Normal inverse Wishart distribution, denoted NIW. Assume the following prior distribution

$$(B, \Sigma) \sim NIW(c_0, C_0, \nu_0, S_0), \tag{5}$$

with density $p(B, \Sigma) = p(B|\Sigma)p(\Sigma)$ with $p(B|\Sigma) \propto |C_0|^{-q/2}|\Sigma|^{-p/2} \exp\{-\frac{1}{2}\text{tr}(\Sigma^{-1}(B - c_0)'C_0^{-1}(B - c_0)\}$, $p(\Sigma) \propto |\nu_0 S_0|^{\nu_0/2}|\Sigma|^{-(\nu_0+q+1)/2} \exp\{-\frac{1}{2}\text{tr}(\Sigma^{-1}\nu_0 S_0\}$ and proportionality constants do not depend on model parameters. Note also that the marginal distribution for Σ is in inverse Wishart form, denoted $IW(\nu_0, S_0)$, or $\Sigma^1 \sim W(\nu_0, S_0)$.

Then Bayes theorem gives the posterior distribution as

$$p(B, \Sigma|Y, X) \propto p(Y|X, B, \Sigma)\, p(B, \Sigma)$$

Simple matrix operations are required to show that combination of (3) with (5) gives the posterior distribution

$$(B, \Sigma|Y, X) \sim NIW(c_1, C_1, \nu_1, S_1). \tag{6}$$

where $c_1 = C_1[C_0^{-1}c_0 + X'Y]$, $C_1 = [C_0^{-1} + X'X]^{-1}$, $\nu_1 = \nu_0 + n$ and $\nu_1 S_1 = \nu_0 S_0 + (Y - Xc_1)'(Y - Xc_1) + (c_1 - c_0)'C_0^{-1}(c_1 - c_0)$. The marginal density for Y can be obtained after integrating out (B, Σ) or by noting that $p(Y) = p(Y|X, B, \Sigma)p(B, \Sigma)/p(B, \Sigma|Y)$. It is given by

$$p(Y) = k\, \frac{|C_1|^{q/2}}{|\nu_1 S_1|^{\nu_1/2}} \tag{7}$$

where the normalizing constant k does not depend on any parameter of the posterior distribution (see Press (1989), for more details).

If the data variation is entirely explained by the covariates, then all required calculations have been done and inference is performed based on this distribution. Posterior means and medians can be analytically obtained and used as point estimators of components of B and Σ. Interval estimation can also be obtained by appropriate probability manipulations with the NIW distribution. Bayesian tests to assess the relevance of any particular model component can also be performed by many different forms. For example, one can test whether a subset of B is null by evaluation of the posterior probability of parameter values more likely than 0. If this probability is small (large), the null value is very (un)likely and the component should be retained in (discarded from) the model.

Inference about non-linear parameter transformations is slightly more difficult. These are required to perform inference about components of Γ, A or about impulse response functions. Typically, analytic expressions for the desired posterior distributions will not be available in closed form. A simple and effective alternative is to draw samples from the relevant distribution. This is easily achieved because samples from (B, Σ) are easy to obtain and samples from transformations are obtain by applying the transformation to the sampled values of (B, Σ).

3 Spatial regression models

There are situations where extra variation is present due to unaccounted sources. In this case, one should try to look for better models to more appropriately describe the data. Ideally this should lead to inclusion of covariates that are responsible for this variation. Failing that, one should at least accommodate this variation in the analysis. When this variation is spatially driven, then special models should be constructed to acknowledge and assess the magnitude of this variation.

There are may possibilities available here. They can be broadly divided into two groups: a latent approach and a direct approach. In the first group, the spatial variation is incorporated into the model via its latent or unobserved components. In doing so, this specification is deferred to a latter stage or a higher level in the model hierarchy. In the second group, the spatial variation is incorporated at the first level of the hierarchy, directly to the observation equation (3).

3.1 The latent approach

The latent process formulation extends the model by allowing the parameter matrix to be region-specific, or spatially indexed. Mathematically, this means that (1) becomes

$$Ay_r = \Gamma_r x_r + u_r \text{ where } u_r \sim N(0, \Lambda)\,,\ r = 1, ..., n \tag{8}$$

or, equivalently, (3) becomes

$$y_r = B_r x_r + e_r \text{ where } e_r \sim N(0, \Sigma)\,,\ r = 1, ..., n \tag{9}$$

where $B_r = A^{-1}\Gamma_r$ and e_r and Σ are as before, $r = 1, ..., n$. Once again, B_r is composed of the q p-dimensional vectors of regression coefficients associated with the q dependent variables and can be written as $B_r = (\beta_{r1}, ..., \beta_{rq})'$. One simplification in this model is to allow only some of the effects of the exogenous variables to vary spatially. In this case, (9) simplifies to

$$y_r = B_r x_{1r} + C x_{2r} + e_r \text{ where } e_r \sim N(0, \Sigma)\,,\ r = 1, ..., n. \tag{10}$$

where the vector of explanatory variables $x_r = (x'_{1r}, x'_{2r})'$ is split into two subvector of dimensions p_1 and $p_2 = p - p_1$, for $r = 1, ..., n$. A special case of interest concerns the situation where only the vector of intercepts varies in space. In this case, $x_{1r} = 1$.

The variation of the regression coefficients B along the space enables greater adaptation by the model to changes unacconted by the covariates used in the model. It also means that there are many more parameters in the model and some identifiability restrictions over the B_r's are required. The discussion deferred to the next section. This idea has been successfully used in the univariate context. Discussion of many aspects of this class of models is provided by Gamerman, Moreira and Rue (2002).

3.2 The direct approach

The other class of models imposes the spatial constraint directly through the observation equation. One simple possibility is to include in the mean process a spatial autoregressive component that takes the spatial locations into account. This can be effectively achieved by the use of a neighbourhood matrix W that informs the proximity between any pair of sites. So, $W = (w_{ij})$ with w_{ij} representing the neighborhood between sites i and j. If they are neighbors then $w_{ij} \neq 0$ and $w_{ij} = 0$ if sites i and j are not neighbors or $i = j$. The standard value for w_{ij} is $1/m_i$ where m_i is the number of neighbors of site i but other choices can be made.

This spatial autoregressive component is frequently used in econometric applications. It can be placed in the fixed part of the model (as in Anselin, 1988), in the random part of the model (as in Pace et al., 2000) or both as in Anselin (1988) and as described below. The multivariate regression model (2) can be generalized by incorporating spatial dependence via

$$\begin{aligned} YA &= W_1 Y\Phi + X\Gamma' + U \\ U &= W_2 U\Psi + E \text{ where } E \sim N(0, I_n, \Lambda) \end{aligned} \tag{11}$$

where Λ, $\Phi = (\phi_{ij})$ and $\Psi = (\psi_{ij})$ are $q \times q$ matrices but Λ is diagonal. The matrices W_1 and W_2 provide the neighbouring structures and may be

equal to W. A special case of Φ and Ψ is given by diagonal forms with respective entries $\phi_1, ..., \phi_q$ and $\psi_1, ..., \psi_q$. The ϕ_j's and ψ_j's consist on the autoregressive coefficients of the observations and of the errors, very much like the multivariate ARMA(1,1) formulation in time series context. This is more clarly seen in the SUR framework. Assuming $A = I_q$ and diagonal forms for Φ and Ψ, equation (11) can be written in the form

$$\begin{aligned} y^j &= \phi_{jj} W_1 y^j + X\beta_j + u^j \\ u^j &= \psi_{jj} W_2 u^j + e^j \text{ where } e^j \sim N(0, \sigma_{jj}^2 I_n)\,,\ j = 1, ..., q. \end{aligned}$$

The above equations seem to indicate independence between the components of the response, as in SUR models. Nevertheless, the response is correlated because the e^j's are not independent. Spatial stationarity constraints may be placed over these coefficients. Dependence between components is also imposed on the mean structure by the contemporaneous relations in A, even if Φ and Ψ are diagonal.

The likelihood for model (11) can be obtained after (column) vectorization of Y, U and E. Letting $\tilde{u} = vec(UA^{-1})$, $\tilde{e} = vec(EA^{-1})$ gives

$$\begin{aligned} \tilde{u} &= (\Psi \otimes W_2)\tilde{u} + \tilde{e} \\ &= \tilde{\Psi}^{-1}\tilde{e} \end{aligned} \tag{12}$$

where $\tilde{\Psi} = I_{nq} - \Psi \otimes W_2$ and $\tilde{e} \sim N(0, I_n \otimes \Sigma)$, with $\Sigma = A^{-1}\Lambda A^{-1}$'. Once again, the observation equation in (11) can be rewritten as $Y = W_1 Y \Phi A^{-1} + X\Gamma' A^{-1} + UA^{-1} = W_1 Y \Phi^* + XB' + U^*$, where $\Phi^* = \Phi A^{-1}$, $B = A^{-1'}\Gamma$ and $U^* = UA^{-1}$. Letting $\tilde{y} = vec(Y)$ gives

$$\begin{aligned} \tilde{y} &= (\Phi^* \otimes W_1)\tilde{y} + \tilde{X}\beta + \tilde{u} \\ &= \tilde{\Phi}^{-1}[\tilde{X}\beta + \tilde{\Psi}^{-1}\tilde{e}], \end{aligned}$$

where $\tilde{X} = I_q \otimes X$, $\tilde{\Phi} = I_{nq} - \Phi^* \otimes W_1$ and $\beta = vec(B')$ and the last equation above follows from (12). It follows that

$$\tilde{\Psi}\tilde{\Phi}\tilde{y} = \tilde{\Psi}\tilde{X}\beta + \tilde{e}. \tag{13}$$

Equation (13) provides the likelihood for the model parameters B, Σ, Φ and Ψ as

$$l(B, \Sigma, \Phi, \Psi) \propto |\tilde{\Psi}||\tilde{\Phi}||\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}(\tilde{\Psi}\tilde{\Phi}\tilde{y} - \tilde{\Psi}\tilde{X}\beta)'(I_n \otimes \Sigma^{-1})(\tilde{\Psi}\tilde{\Phi}\tilde{y} - \tilde{\Psi}\tilde{X}\beta)\right\}.$$

This is the same likelihood as that from models in section 2 but for the presence of the term $|\tilde{\Psi}||\tilde{\Phi}|$ and the replacement of Y and X. This remark will be useful to perform Bayesian inference for these models in section 5.

Dimensionally large matrices are present but the only inversion required for the the computation of the likelihood is for the q-dimensional matrix Σ. So, likelihood-based inference becomes a computationally attractive procedure under this formulation.

Therefore, the latent and the direct approach tackle the issue of spatial dependence in very different ways. It is not clear which situations call for one approach or the other. We are not aware of any comparison between these approaches in the literature even in the simpler, univariate case. The choice between these different approaches can be made on the basis of model performance both in terms of fit and prediction.

4 Prior distributions

This section describes the form of the prior distribution specified for the spatial models described in the previous section. Again, the latent process route is tackled before.

4.1 The latent approach

In the latent case, spatial dependence is introduced in the model via its latent components, the regression coefficients. Once again, the idea of neighbourhood is used. There are many ways to achieve it through the prior, including the use of the autoregression forms (11).

In this paper, we will make use of the multivariate version of the pairwise difference prior as advocated in Besag, York and Mollié (1991). This prior was proposed by Moreira and Migon (1999) and Assunção, Gamerman, and Assunção (1999). It is a special case of the class of distributions discussed in Mardia (1988). Its density is given by

$$f_{pd}(B|\Psi) \propto |\Psi|^{n/2} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{n} k_{ij} (\beta_i - \beta_j)' \Psi (\beta_i - \beta_j) \right\} \qquad (14)$$

where $B = (B_1, ..., B_n)$, $\beta_r = vec(B_r')$, for $r = 1, ..., n$ and the k_{ij}'s provide information of the neighbourhood of the sites. A simple choice is $k_{ij} = 1$, if sites i and j are neighbours and 0, otherwise. The components of B can be placed into a vector in many different forms. An obvious choice is stacking them all in a vector of regression coefficients $\tilde{\beta} = (\beta_1', ..., \beta_n')'$. Another choice is $\beta = (\beta^{1'}, ..., \beta^{q'})'$ where $\beta^{j'} = (\beta_{1j}', ..., \beta_{nj}')'$. Note that $\tilde{\beta} = (P \otimes I_p)\beta$ for some $nq \times nq$ permutation matrix P. Simple algebra shows that the prior precision of $\tilde{\beta}$ implied by (14) is given by $-K \otimes \Psi$, where $K = (k_{ij})$ is completed with diagonal entries $k_{ii} = -\sum_j k_{ij}$, $i = 1, ..., n$. Hence, the prior precision for β is $Q = -(P \otimes I_p)(K \otimes \Psi)(P' \otimes I_p)$. Note that K and W are related via $K = (I_n - MW)$ where M is the $n \times n$ diagonal matrix with diagonal entries $m_1, ..., m_n$.

It may be reasonable to assume a Kronecker form for $\Psi = \Psi_d \otimes \Psi_e$ since each β_r is a concatenation of q p-dimensional vectors of regression coefficients, associated with the same explanatory variables. The precision Ψ_d is a $p \times p$ matrix informing the variation across the dependent variables and the precision Ψ_e is a $q \times q$ matrix informing the variation across the explanatory variables. In this case, the prior for $f_{pd}(B|\Psi)$can be written as

$$\propto |\Psi_d|^{\frac{np}{2}} |\Psi_e|^{\frac{nq}{2}} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{n} k_{ij} \text{ tr } [(B_i - B_j)\Psi_e (B_i - B_j)' \Psi_d] \right\} \qquad (15)$$

$$\propto |\Psi_d|^{\frac{np}{2}} \ |\Psi_e|^{\frac{nq}{2}} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{n} k_{ij} \text{ tr } [(B_i - B_j)' \Psi_d (B_i - B_j)\Psi_e] \right\}. \qquad (16)$$

This decomposition implies that the prior precison for $\tilde{\beta}$ is given by $-K \otimes \Psi_d \otimes \Psi_e$ and the prior precison of β is $-\Psi_d \otimes K \otimes \Psi_e$. This decomposition of Ψ also allows for specification of $\Psi_d = k_d \Sigma^{-1}$ for some suitably defined constant k_d. In this case, the prior precison is measured relative to the observational precision. This operation is called conjugate decomposition for reasons that will be clear in the next section. Also the constant k_d can be absorded without loss of generality into Ψ_e and will herafter be omitted. These results will be used to obtain the posterior distribution of β.

The prior is completed with a prior distribution for Ψ. In this paper, it is assumed in Wishart form $\Psi \sim W(n_\psi, S_\psi)$. If the Kronecker product is used, independent priors $\Psi_d \sim W(n_d, S_d)$ and $\Psi_e \sim W(n_e, S_e)$ are assumed.

4.2 The direct approach

In this case, the model parameters are given by B, Σ, Φ and Ψ. The prior for (B, Σ) will usually be in matricvariate normal inverse Wishart form. There is no obvious suggestion for the prior for Φ and Ψ. When they are diagonal matrices, it may be reasonable to assume the i.i.d. forms for the q dimensional vectors ϕ and ψ containing the diagonal elements of Φ and Ψ

$$p(\phi, \psi) = \prod_{j=1}^{q} f_\phi(\phi_j) f_\psi(\psi_j)$$

with reasonably vague specifications for densities f_ϕ and f_ψ. I.i.d. forms may also be used for all elements of the non-diagonal matrices Φ and Ψ. Another possibility is to allow partial exchangeability by separating the diagonal from the off-diagonal elements.

Note that in the latent approach, most of the efforts to model the spatial structure are concentrated in the prior with very little work required for the likelihood specification. The direct approach is the entire opposite situation, with an elaborate likelihood being constructed to account for the spatial

interactions, leaving a simple strucutre in terms of prior relations between model paramters.

Another comment relevant to both approaches regards the specification of the neighbourhood structure. So far, it has been assumed that the neighbourhood matrices W are completely known. It may be possible to specify functional forms for the w_{ij}'s depending on further unknown parameters denoted by θ. In this case, the vector of hyperparameters must be enlarged to include θ. An example is given by $w_{ij} \propto d_{ij}^{-\theta}$, with d_{ij} measuring the distance between regions i and j and $\theta > 0$.

5 Posterior inference

5.1 The latent approach

Posterior distribution is obtained via Bayes theorem by combination of lilkelihood and prior. For the latent process approach, the posterior density is given by

$$p(\beta, \Sigma, \Psi|Y) \propto l(\beta, \Sigma)p(\beta|\Psi)p(\Sigma)p(\Psi)$$

where $l(\beta, \Sigma)$ is obtained from (9) and $p(\beta|\Psi)$ is given in (14). In the case of the Kronecker decomposition of the spatial prior precision Ψ, the posterior density is

$$p(\beta, \Sigma, \Psi_d, \Psi_e|Y) \propto l(\beta, \Sigma)p(\beta|\Psi_d, \Psi_e)p(\Sigma)p(\Psi_d)p(\Psi_e).$$

where $l(\beta, \Sigma)$ is obtained from (9) and $p(\beta|\Psi)$ is given in (15) or (16). The remaining terms in both forms are Wishart for Σ or inverse Wishart for Ψ.

In any event, the joint posterior will not have closed-forms integrals and inference will have to be performed approximately. Given the high dimensionality of the quantities involved, it is usually recommended to apply MCMC methods (Gamerman, 1997). Working in componentwise fashion requires

sampling from the conditional distributions of $(B|\Sigma, \Psi, Y)$, $(\Sigma|\Psi, B, Y)$ and $(\Psi|B, \Sigma, Y)$. Vectorization of model (9) gives $y \sim N[(I_q \otimes \check{X})\beta, \Sigma \otimes I_n)$, where $\check{X} = diag(x_1', ..., x_n')$. Thus, the first of these distributions is easily obtained using (6). Combining this likelihood with prior (14) gives the first conditional

$$\beta|\Sigma, \Psi, Y \sim N(a, R) \tag{17}$$

where $a = R(\Sigma^{-1} \otimes \check{X}')y$ and $R = [(\Sigma^{-1} \otimes \check{X}'\check{X}) + Q]^{-1}$. The presence of the prior variance prevents cancelation of Σ from the expression of the posterior mean of β, as is usual with the least squares (LS) and maximum likelihood (ML) estimators. If the prior precision allows the decomposition $\Psi = \Psi_d \otimes \Psi_e$ no further simplifcation is obtained. However if a conjugate decomposition is used then full conditional moments simplify to $a = [I_q \otimes (\check{X}'\check{X} - K \otimes \Psi_e)^{-1}\check{X}']y$ and $R = \Sigma \otimes (\check{X}'\check{X} - K \otimes \Psi_e)^{-1}$. In this case, a decomposition is obtained for the moments and the posterior mean can be obtained as in the univariate response case. Further computational advantages can obtained when this decomposition is applied to sampling from (17). The nq-dimensional vector β can be broken into the q n-dimensional vectors β^j's. These can be sampled in succession from β^1 then $\beta^2|\beta^1$ up until $\beta^q|\beta^1, ..., \beta^{q-1}$ where all distributions are easily obtained from standard normal theory applied to (17).

If the observation model contains spatially varying coefficients for some covariates and static (spatially invariant) for the other ones then vectorization of (10) gives $y \sim N[(I_q \otimes \check{X}^*)\beta^*, \Sigma \otimes I_n]$, where $\check{X}^* = (\check{X}_1, X_2)$ and $\beta^* = (\beta', \gamma')'$, with $\gamma = vec(C')$, $\check{X}_1 = diag(x_{11}', ..., x_{1n}')$ and X_2 is the (usual) $n \times p_2$ design matrix having r-th row given by x_{2r}, $r = 1, ..., n$. Assuming an independent prior $\gamma \sim N(c_0, C_0)$ generalizes (17) to the full conditional of the augmented vector of regression coefficients

$$\beta^* \sim N(a^*, R^*)$$

where

$$a^*=R^*\begin{pmatrix}(\Sigma^{-1}\otimes\check{X}_1')y\\ C_0^{-1}c_0+\Sigma^{-1}\otimes X_2'y\end{pmatrix},R^*=\left[\Sigma^{-1}\otimes\begin{pmatrix}\check{X}_1'\check{X}_1 & \check{X}_1'X_2\\ X_2'\check{X}_1 & X_2'X_2\end{pmatrix}+\begin{pmatrix}Q & 0\\ 0 & C_0^{-1}\end{pmatrix}\right]^{-1}.$$

Unlike LS and ML estimators, Σ does not cancel out from the expression of the posterior mean. Once again a simplification is possible if a conjugate decomposition is used for β and the prior variance for γ is in conjugate form $\Sigma\otimes\tilde{C}_0$. Note also that the full conditional for β^* only simplifies as a product for B (or β) and C (or γ) if $\check{X}_1$ and X_2 are orthogonal. Therefore, B and C should be treated jointly.

The second and third full conditionals are

$$\Psi|B,\Sigma,Y\sim IW(n_\psi^*,S_\psi^*)\text{ and }\Sigma|B,\Psi,Y\sim IW(n_\sigma^*,S_\sigma^*)$$

where $n_\psi^*=n_\psi+n$, $S_\psi^*=S_\psi+Q(B)$, $n_\sigma^*=n_\sigma+n$ and $n_\sigma^*S_\sigma^*=n_\sigma S_\sigma+(Y-XB')'(Y-XB')$, with $Q(B)=\sum_{i,j=1}^n k_{ij}(\beta_i-\beta_j)(\beta_i-\beta_j)'$. If $\Psi=\Psi_d\otimes\Psi_e$ then the full conditional of Ψ splits into two full conditionals. These can be obtained by combination of their prior densities with (15) and (16) respectively to give

$$\Psi_d|B,\Sigma,\Psi_e,Y\sim IW(n_d^*,S_d^*)\text{ and }\Psi_e|B,\Sigma,\Psi_d,Y\sim IW(n_e^*,S_e^*)$$

where $n_d^*=n_d+np$, $n_d^*S_d^*=n_dS_d+\sum_{i,j=1}^n k_{ij}(B_i-B_j)\Psi_e(B_i-B_j)'$, $n_e^*=n_e+nq$ and $n_e^*S_e^*=n_eS_e+\sum_{i,j=1}^n k_{ij}(B_i-B_j)'\Psi_d(B_i-B_j)$. If, in addition, a conjugate decomposition is used then there is no need for the sampling Ψ_d step and the full conditional for Σ is still with the same IW form but with parameters given by $n_\sigma^*=n_\sigma+np+n$ and $n_\sigma^*S_\sigma^*=n_\sigma S_\sigma+(Y-XB')'(Y-XB')+\sum_{i,j=1}^n k_{ij}(B_i-B_j)\Psi_e(B_i-B_j)'$. Note that posterior full conditionals are all conditionally conjugate given the choices of prior that were made and easy sampling is achieved.

In a similar but univariate context, Gamerman, Moreira and Rue (2002) discuss the option of integrating out the regression coefficients (B and C,

here) analytically and work with the marginal posterior for the variance parameters only. The number of parameters for sampling reduces substantially from $O(n)$ to $O(1)$ but computations are just as demanding. Their result presents a marginally better performance of this marginal sampler but they require Metropolis-Hastings steps with tuning for appropriate chain mixing. Calculations can easily be adapted to our setting. We prefer to work with the full set of parameters given the difficulty associated with tuning the proposals for the large number of variance parameters in the multivariate setting. Therefore, the joint (or marginal) sampling scheme can be summarized as follows:

1. Sample (Σ, Ψ) from their full conditional (or marginal) posterior distribution;

2. Sample β^* from its full conditional posterior distribution;

3. Reconstruct A and Λ from the sampled value of Σ.

For the joint sampling scheme, step 1 is performed by Gibbs sampling from the relevant inverse Wishart distributions and step 2 is performed at every interation of the sampler. Step 1 of the marginal sampler can be peformed by any Metropolis chain, including slice sampling. This technique was succesfully used in the spatial context by Agarwal and Gelfand (2001). A more general alternative is the use of auxiliary variables (Damien, Wakefield and Walker, 1999). More standard techniques based on sampling directly from the marginal posterior were used in the spatial context by Gamerman, Moreira and Rue (2002). Steps 2 and 3 are only performed after convergence has been achieved in step 1.

5.2 The direct approach

The posterior here has no simple form. It is obtained from combination of the likelihood (13) with the prior for model parameters A, B, Ψ, Φ and Λ or B, Φ, Ψ and Σ. Of these, the full conditionals of Φ and Ψ are untractable but the full conditional for (B, Σ) is recognizable. It is given by $(B, \Sigma)|\Phi, \Psi, Y \sim NIW(\tilde{b}_1, \tilde{C}_1, \nu_1, \tilde{S}_1)$, where $\tilde{b}_1 = \tilde{C}_1(C_0^{-1}b_0 + Z'z)$, $\tilde{C}_1 = (C_0^{-1} + Z'Z)^{-1}$ and $\nu_1\tilde{S}_1 = \nu_0 S_0 + (z - Z\tilde{c}_1)'(z - Z\tilde{c}_1) + (\tilde{c}_1 - c_0)'C_0^{-1}(\tilde{c}_1 - c_0)$, with $Z = \Phi^* \tilde{X}$ and $z = \tilde{\Psi}\tilde{\Phi}y$. Once again, (B, Σ) can be integrated out here and sampling becomes restricted to the hyperparameters Φ and Ψ. Their marginal likelihood is $l(\Phi, \Psi) \equiv p(Y|\Phi, \Psi) = p(\Phi, \Psi|Y)p(Y)/p(\Phi, \Psi)$. But

$$p(\Phi, \Psi|Y) = \frac{p(B, \Sigma, \Phi, \Psi|Y)}{p(B, \Sigma|Y, \Phi, \Psi)} = \left\{ \frac{p(Y|B, \Sigma, \Phi, \Psi)p(B, \Sigma)}{p(B, \Sigma|Y, \Phi, \Psi)} \right\} p(\Phi, \Psi).$$

The term in brackets above is proportional to the marginal likelihood $l(\Phi, \Psi)$ and, given (Φ, Ψ), is in the same form as the marginal density of Y in section 2. Therefore, it can obtained from (7) and the remark made after equation (13) as

$$l(\Phi, \Psi) \propto |\tilde{\Psi}||\tilde{\Phi}| \, \frac{|\tilde{C}_1|^{q/2}}{|\nu_1\tilde{S}_1|^{\nu_1/2}}.$$

This can be combined with the prior for Φ and Ψ to obtain their marginal posterior. It will be untractable and will typically require Metropolis-Hastings steps. Their full conditional when (B, Σ) are not integrated out will also require proposals. Therefore, it seems reasonable to suggest the marginal apporach in this setting. However, the vast reduction in parameter dimension for the latent approach is not obtained here as B and Σ are $q \times p$ and $q \times q$ matrices.

The joint (or marginal) sampling scheme can be summarized as follows:

1. Sample (Φ, Ψ) from the full conditional (or marginal) posterior distribution;

2. Sample (B, Σ) from their posterior full conditional $(B, \Sigma)|(\Phi, \Psi)$;

3. Reconstruct A and Λ from the sampled value of Σ.

Same comments made at end of the previous subsection are also relevant here. The exception being that now no tractable distributions are available for the hyperparameters. This makes it more attractive to use the marginal sampler.

6 Applications

This section shows results obtained from data analyses using the models described in this paper. The analysis are part of a study designed to understand agricultural activity in the context of the Brazilian economy. Brazil can be divided in two distinct groups: the occupied land and the land that is still in the process of occupation, basically given by the Amazon region. The first model considers the first group while the second considers the second group.

The first analysis describes the production output in terms of the inputs and uses the latent approach. The second analysis tries to establish relationships between different economic activities and their variation through space and uses the direct approach. Both studies use data collected in the Brazilian censuses of 1970, 1975, 1980, 1985 and 1995, performed by the official government institution IBGE. The data for the first analysis is based on microregions of the occupied Brazilian territory while the second analysis uses data on municipalities in the Amazon region.

The analyses are in the form of panel data but minor adaptations are required to accommodate for the presence of repeated time measurements. They are presented here to show some of the possibilities rendered by the models introduced here. A fuller account of these studies is reported in Moreira and Paez (2002).

6.1 Agricultural productivity

This application study the factors that impact on the agricultural output (y_1), suitably aggregated over a single index (Caves, Christensen and Diewert, 1982). It considers data from the large part of Brazil with extensive agricultural activity, thus removing the Amazon region. The factors impacting production are labour (z_1), usable land (z_2), capital (z_3) and an economically impacted factor (y_2). The model assumes a production function with a variable input and the 3 fixed factors, without substitution between input and factors and with substitution between factors having unit elasticity. The model is in the form (8) with variables measured in the log scale, with only the bivariate intercept varying in space and with an additional term to account for the time variation present in the data. The matrix A has only one non-zero off-diagonal element to express the impact of the economic factor y_2 on the output. Summarizing, the model is in the form $Ay_{rt} = \alpha_r + \gamma_t + \beta z_{rt} + e_{rt}$ with $y_{rt} = (y_{1,rt}, y_{2,rt})'$, $\alpha_r = (\alpha_{1r}, \alpha_{2r})'$,

$$A = \begin{pmatrix} 1 & -\beta_{10} \\ 0 & 1 \end{pmatrix}, \beta = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix} \text{ and } z_{rt} = \begin{pmatrix} z_{1,rt} \\ z_{2,rt} \\ z_{3,rt} \end{pmatrix}.$$

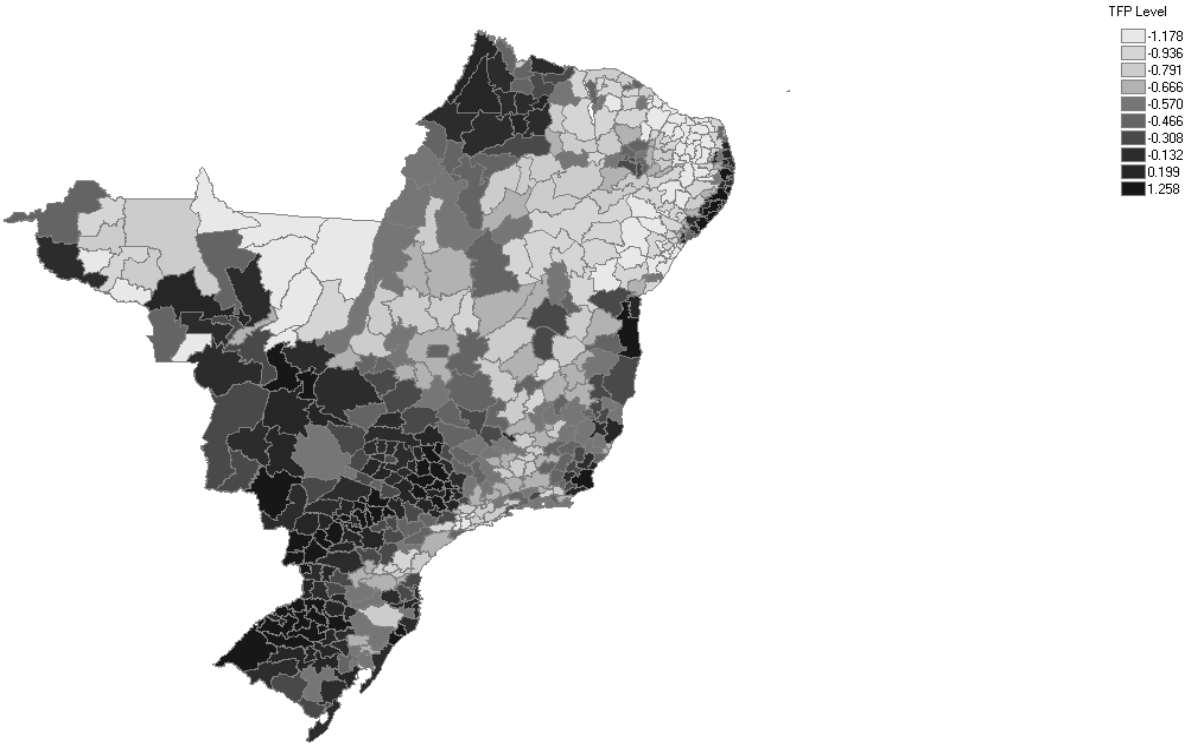

Figure 1. Geographical display of the estimated mean values of α_{2r} for each region r.

The main results obtained are listed in Table 1 below. Two models were considered: Model 1 with a space-varying intercept and Model 2 with a constant intercept. All regression coefficients are significantly distant from 0 for both models, showing the relevance of all factors in explaining the output. Also, the production scale economy given by the sum of the regression coefficients is estimated near 1 in model 1. This indicates neutrality to the size of the region, which is a sensible conclusion. The spatially-varying intercept for the output equation (α_{1r}'s) are depicted in Figure 1 geographically. It shows larger values for the microregions located in the more economically

developed regions of Brazil. Figure 2 shows that the estimated intercept experience spatial variation that is statistically relevant and should be retained in the model.

Table 1. Posterior mean elasticities

Output (y_1) equation

	β_{11}	β_{12}	β_{13}	β_{10}
Model 1	0.14	0.14	0.38	0.37
	(0.03)	(0.02)	(0.04)	(0.02)
Model 2	0.07	0.25	0.45	0.45
	(0.02)	(0.01)	(0.02)	(0.03)

Posterior s.d. provided in brackets

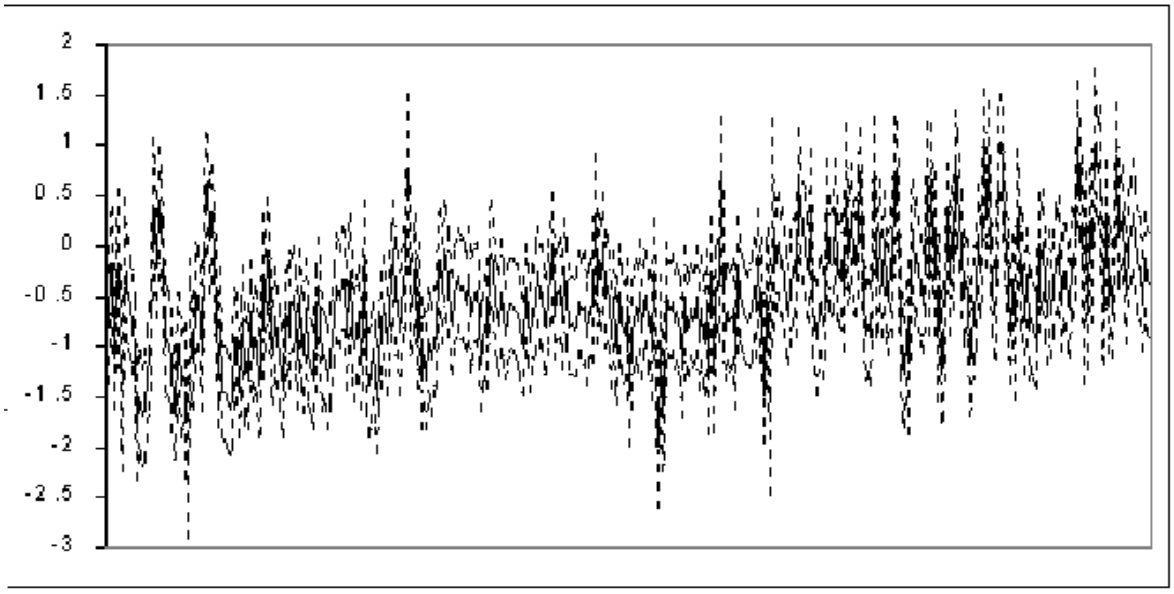

Figure 2. Linearized display of the estimated mean values of α_{1r} (full line) and the corresponding credibility limits (broken lines) for each region r.

6.2 Economic occupation

In this study, the aim is to verify evidence of spatial contagion between the intensity of economic activities and deforestation in the Amazon region of Brazil. The intensity of activities is measured in two alternative forms. Both measure number of cattle and value of crop output. They differ on the third variable which is volume of wood extracted or logit of proportion of used land. All variables are standardized with division by the area of the region.

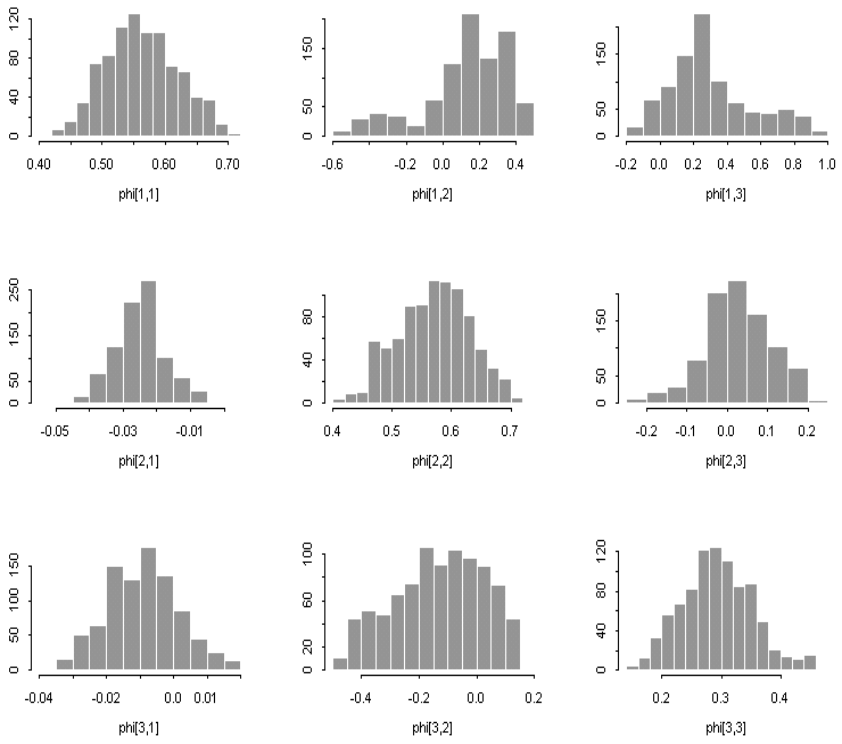

Figure 3. Posterior histograms for the components of Φ in the first analysis.

Both analyses considered $q = 3$ variables and the (11) form with $A = I_3$, $\Psi = 0$ and covariates given by 1-lag cost of transportation and by the geo-climatic characteristics of the region (climate, topography, vegetation, cost

of access to the region). Results reported in figures 3 and 4 pay special attention to the estimation of the elements of the matrix Φ. They show that each activity is affected by the neighbourhood structure due to the non-zero elements in the diagonal of Φ. In the first analysis, wood extraction does not seem to be affected by any of the other activities since the posterior distribution of ϕ_{13} and ϕ_{23} are concentrated around zero. A similar result is obtained in the second analysis that shows that used land, directly related to deforestation, is not affected by alteration of activities in the neighboring regions. In this case, used land has a positive effect over pasture activities.

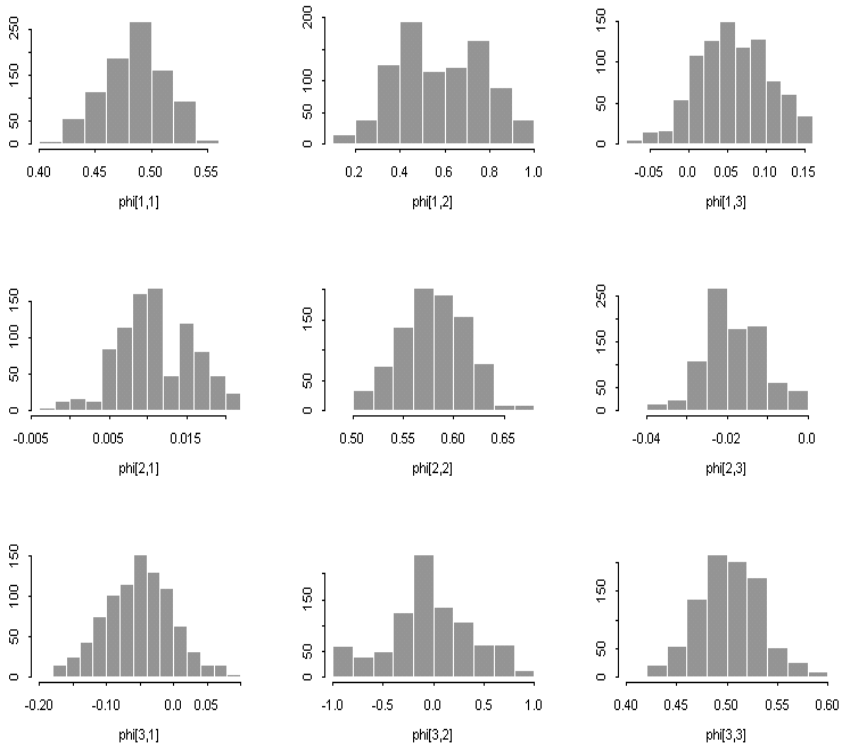

Figure 4. Posterior histograms for the components of Φ in the second analysis.

The analyses were also extended to accommodate for unknowns in the neighbourhood specification. The dependence on θ alluded to in the end of section 4 was used. The posterior distribution for θ is presented in figure 5 for both analyses. Inference is quite informative for both cases showing preference for larger spatial correlation in the first analysis. The estimates of the hyperparameters remain the same and were not reported.

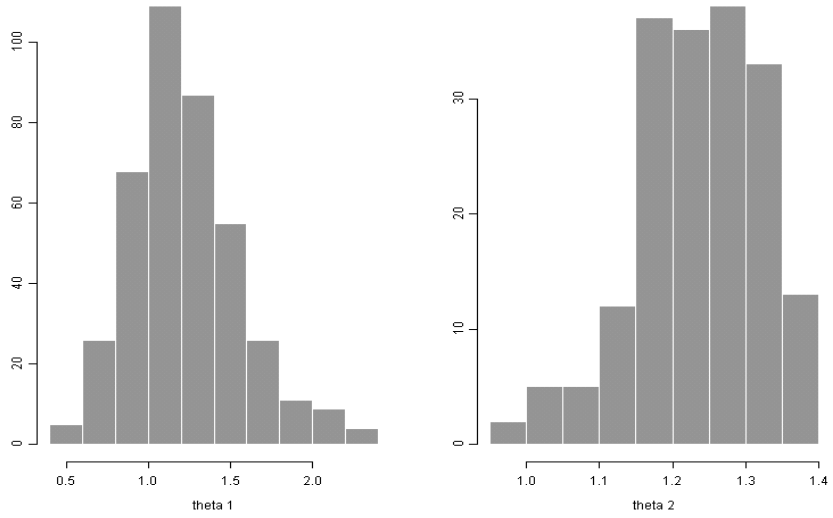

Figure 5. Posterior histograms for θ in the direct apprach: left panel - first analysis; right panel - second analysis.

7 Conclusion

This paper discusses a class of potentially useful extensions of multivariate regression models incorporating spatial variation. It deals with the spatial variation through two of the most commonly used approaches in the univariate context. New structures are thus proposed and the inferential procedures required to tackle their inference are derived. The approaches are illustrated

in the context of an economic application to real data where relevant issues are only adequately addressed through the use of the generalizations.

The Bayesian approach is preferred here but frequentist inference (e.g. likelihood-based methods) could just as well be used. In particular, the direct approach has a large body of literature using this route that could be adapted to meet this end. The main advantage of the Bayesian approach is the ability to allow full exploration of the posterior distribution of all model parameters (see figures 3 and 4). Instead, other approaches only allow for mean and variance calculations, with the latter typically obtained after asymptotic results are called for.

Attention was restricted to the posterior distribution of model parameters. However, complete inference is now available and many usual elements of standard multivariate analysis can also be obtained here. In particular, tests for the relevance of a group of components or for independence between blocks of endogenous variables can easily be obtained.

Acknowledgements

The research of the author was supported by a grant from IPEA (RedeIpea). Large part of the paper was written while the first author was visiting the Department of Statistics, University of Connecticut, whose hospitality is greatly appreciated. He would also like to thank Alan Gelfand for many comments and discussions.

REFERENCES

AGARWAL, D. K. GELFAND, A. E. *Slice Gibbs sampling for simulations based fitting of spatial data models*. University of Connecticut/Department of Statistic, 2001 (Technical Report).

ANDERSEN, L. GRANGER, C., REIS, E. A random coefficient VAR transition model of the changes in land use in the Brazilian Amazon. *Brazilian Journal of Econometrics*, v. 17, p. 1-13, 1997

ANSELIN, L. *spatial econometric: methods and models.* Kluwer, Dordrecht, 1988.

ASSUNÇÃO, J. J., GAMERMAN, D., ASSUNÇÃO, R. M. Regional differences in factor productivities of Brazilian agriculture: a space varying parameter approach. *Proceedings of the XV Latin American Meeting of the Econometric Society*, 1999.

BESAG, J., YORK, J., MOLLIÉ, A. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, v. 43, p. 1-59, 1991.

CAVES, D., CHRISTENSEN, L., DIEWERT, E. Multilateral comparison of output, input and productivity using superlative index numbers. *The Economic Journal*, v. 92. p. 73-86, 1982.

DAMIEN, P., WAKEFIELD, J., WALKER, S. Gibbs sampling for Bayesian non conjugate and hierarchical models using auxiliary variables. *Journal of the Royal Statistical Society, Series B*, v. 61, p. 331-334, 1999.

GAMERMAN, D. *Markov Chain Monte Carlo: Stochastic Simulations for Bayesian Inference*. London: Champman & Hall, 1997.

GAMERMAN, D., MOREIRA, A. R. B. Bayseian analysis of econometric times series models using hybrid integration rules. *Communication in Statistics*, v. 31, p. 49-72, 2002.

GAMERMAN, D., MOREIRA, A. R. B., RUE, H. Space-varying regression models: specifications and simulation. To appear in the special issue on *Computational Econometrics of Computational Statistics and Data Analysis*, 2002.

HAMILTON, J. *Time series analysis.* New Jersey: Princeton University Press, 1994.

HARVEY, A. C. *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press, 1989.

MARDIA, K. V. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, v. 24, p. 265-284, 1988.

MIGON, H. S., GAMERMAN, D. Statistical inference: an integrated approach. London: Arnold, 1999.

MOREIRA, A. R. B., MIGON, H. S. Spatial heterogeneity of the agricultural productivity: Brazil — 1970 to 1996. *Proceedings of the XXI Brazilian Meeting on Econometrics*, 1999.

MOREIRA, A. R. B., PAEZ, M. S. *Neighbourhood effect in the activity: measures of productivity and economic occupation of the Amazon region.* IPEA, 2002 (Technical Report).

PACE, R. K., BARRY, R., GILLEY, O. W., SIRMANS, C. F. A method for spatial-temporal forecasting with an application to real estate prices. *Internation Journal of Forecasting*, v. 16, p. 229-246, 2000.

PRESS, S. J. *Bayesian statistics: principles, models and applications*. New York: Wiley, 1989.

WEST, M., HARRISON, P. J. Bayesian forecasting and dynamics models. Second ed. New York: Springer-Verlag, 1997.

ZELLNER, A. *An introduction to Bayesian inference in econometrics*. New york: Wiley, 1971.

EDITORIAL

Coordenação
Luiz Cezar Loureiro de Azeredo

Supervisão
Helena Rodarte Costa Valente

Revisão
André Pinheiro
Elisabete de Carvalho Soares
Lucia Duarte Moreira
Luiz Carlos Palhares
Miriam Nunes da Fonseca
Tatiana da Costa (estagiária)

Editoração
Carlos Henrique Santos Vianna
Joanna Silvestre Friques de Sousa (estagiária)
Rafael Luzente de Lima
Roberto das Chagas Campos

Divulgação
Raul José Cordeiro Lemos

Reprodução Gráfica
Edson Soares

Brasília
SBS – Quadra 1 – Bloco J – Ed. BNDES,
10º andar – 70076-900 – Brasília – DF
Fone: (61) 315-5336
Fax: (61) 315-5314
Correio eletrônico: editbsb@ipea.gov.br

Home page: http://www.ipea.gov.br

Rio de Janeiro
Av. Presidente Antônio Carlos, 51, 14º andar
20020-010 – Rio de Janeiro – RJ
Fone: (21) 3804-8118
Fax: (21) 2220-5533
Correio eletrônico: editrj@ipea.gov.br

Tiragem: 130 exemplares

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo