Gamerman, Dani; Moreira, Ajax Reynaldo Bello; Rue, Håvard

Working Paper

Space-varying regression models: Specifications and simulation

Discussion Paper, No. 102

Provided in Cooperation with:
Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Gamerman, Dani; Moreira, Ajax Reynaldo Bello; Rue, Håvard (2015) : Space-varying regression models: Specifications and simulation, Discussion Paper, No. 102, Institute for Applied Economic Research (ipea), Brasília

This Version is available at:
https://hdl.handle.net/10419/220191

102
DISCUSSION PAPER
Originally published by Ipea in July 2001 as
number 809 of the series Texto para Discussão.

ipea

SPACE-VARYING REGRESSION MODELS: SPECIFICATIONS AND SIMULATION

Dani Gamerman
Ajax R. B. Moreira
Havard Rue

ipea

102
DISCUSSION PAPER

Originally published by Ipea in July 2001 as number 809 of the series Texto para Discussão.

Brasília, January 2015

SPACE-VARYING REGRESSION MODELS: SPECIFICATIONS AND SIMULATION

Dani Gamerman[1]
Ajax R. B. Moreira[2]
Håvard Rue[3]

1. Do Instituto de Matemática da UFRJ.
2. Da Diretoria de Estudos Macroeconômicos do Ipea.
3. Da Norwegian University of Science and Technology.

Federal Government of Brazil

Secretariat of Strategic Affairs of the Presidency of the Republic
Minister Roberto Mangabeira Unger

ipea Institute for Applied Economic Research

A public foundation affiliated to the Secretariat of Strategic Affairs of the Presidency of the Republic, Ipea provides technical and institutional support to government actions – enabling the formulation of numerous public policies and programs for Brazilian development – and makes research and studies conducted by its staff available to society.

President
Sergei Suarez Dillon Soares

Director of Institutional Development
Luiz Cezar Loureiro de Azeredo

Director of Studies and Policies of the State, Institutions and Democracy
Daniel Ricardo de Castro Cerqueira

Director of Macroeconomic Studies and Policies
Cláudio Hamilton Matos dos Santos

Director of Regional, Urban and Environmental Studies and Policies
Rogério Boueri Miranda

Director of Sectoral Studies and Policies, Innovation, Regulation and Infrastructure
Fernanda De Negri

Director of Social Studies and Policies, Deputy
Carlos Henrique Leite Corseuil

Director of International Studies, Political and Economic Relations
Renato Coelho Baumann das Neves

Chief of Staff
Ruy Silva Pessoa

Chief Press and Communications Officer
João Cláudio Garcia Rodrigues Lima

URL: http://www.ipea.gov.br
Ombudsman: http://www.ipea.gov.br/ouvidoria

DISCUSSION PAPER

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

Discussion paper / Institute for Applied Economic

Research.- Brasília : Rio de Janeiro : Ipea, 1990-

ISSN 1415-4765

1. Brazil. 2. Economic Aspects. 3. Social Aspects. I. Institute for Applied Economic Research.

CDD 330.908

SUMÁRIO

RESUMO

Os modelos de regressão com parâmetros variando no espaço são uma generalização dos modelos lineares em que é permitido aos coeficientes da regressão mudarem ao longo do espaço. A estrutura espacial é especificada por uma extensão multivariada de uma distribuição *a priori* que considera as diferenças entre os coeficientes de regiões vizinhas. Isso permite a incorporação da informação da vizinhança espacial.

Para estimar o modelo utilizamos a abordagem bayesiana e o algoritmo do MCMC considerando diferentes esquemas de amostragem. Esses esquemas foram comparados em termos da autocorrelação da cadeia de Markov, e em termos dos resultados obtidos.

Foram discutidas diferentes especificações *a priori* que admitem estruturas espaciais semelhantes. Os resultados são ilustrados com dados simulados e com um conjunto real de informações.

Space-varying regression models:

specifications and simulation

Dani Gamerman, Ajax R. B. Moreira and Håvard Rue
Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil
Instituto de Pesquisa Econômica Aplicada, Brazil
Norwegian University of Science and Technology, Norway

Abstract

Space-varying regression models are generalizations of standard linear models where the regression coefficients are allowed to change in space. The spatial structure is specified by a multivariate extension of pairwise difference priors thus enabling incorporation of neighboring structures and easy sampling schemes. Different sampling schemes are available and may be used in an MCMC algorithm. These schemes are compared in terms of chain autocorrelation and resulting inference. We also discuss different prior specifications that accommodate the spatial structure. Results are illustrated with simulated data and applied to a real dataset.

Key words: Bayesian; Hyperparameters; Gibbs sampling; Markov chain Monte Carlo; Markov random fields; Metropolis-Hastings algorithm; Sampling schemes.

1 Introduction

Consider initially the multiple linear regression model where independent observations $y_1, ..., y_n$ follow the distribution $y_i \sim N(\mu_i, \sigma^2)$, for $i = 1, ..., n$. In standard regression, the means μ_i are described by a linear relation $x_i'\beta$ where x_i is the r-dimensional vector of explanatory variables for the ith observational unit ($i = 1, ..., n$) and β is the vector of regression coefficients. Inference about β informs about the strength of the x's in terms of explaining the variability of the observations y.

The extensions we consider below allow the regression coefficients to change with the observation unit. Hence, we will have $\mu_i = x_i'\beta_i$, for $i = 1, ..., n$ and unknown quantities (β, ϕ) where $\beta = \text{vec}(B)$, the column vectorization of the $n \times r$ matrix B given by

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2r} \\ & & \vdots & \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nr} \end{pmatrix} = \begin{pmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_n' \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \beta^1 & \beta^2 & \cdots & \beta^r \end{pmatrix}$$

and $\phi = 1/\sigma^2$. The likelihood is

$$l(\beta, \phi) = (2\pi)^{-n/2}\phi^{n/2} \exp\left\{-\frac{\phi}{2}\sum_{i=1}^{n}(y_i - x_i'\beta_i)^2\right\}. \quad (1)$$

Under a Bayesian formulation, the model must be completed by a prior for (β, ϕ), where assumptions about the relation between the components of β are made. Note also that this prior may also depend on other unknown quantities. They are typically variance parameters and are denoted by Ψ. Priors for these hyperparameters must also be specified. In general, these are taken as reasonably vague, reflecting the difficulty in the incorporation of substantial knowledge about them.

There are many possible routes to be taken from here. These different routes are associated with qualitatively different assumptions about the relationship between the β_i's. They are in general based on weak quantitative assumptions, generally through vague prior on hyperparameters. These models can only lead to precise posterior inference if the dimension r of the β_i's is substantially smaller than the number n of observational units.

One possibility is to allow the regression coefficients to vary across units without any particular structure. Hierarchical models introduced by Lindley and Smith (1972) suggest in their simplest forms that the β_i's should be allowed to vary as a random sample from a given distribution, indexed by Ψ. The most common example is to have β_i's as hypothetical draws from a $N(\mu, \Sigma)$ distribution and in this case $\Psi = (\mu, \Sigma)$. Another possibility popularized by West and Harrison (1997) is to use a dynamic model whereby regression coefficients form a generalized random walk $\beta_i = G_i\beta_{i-1} + w_i$ for random inputs w_i's. This formulation is particularly useful when the units are sequentially observed and a lagged effect between successive regression coefficients is expected.

For spatially distributed data, a number of possibilites are available. One can use the geostatistical approach (Cressie, 1991, ch. 2-5) where the regression errors $y_i - \mu_i$ are spatially correlated or the discrete approach of pairwise different priors (Besag et al., 1991) where unit-specific random effects have their distribution specified according to the neighbouring arrangement of the units. Both cases however can be accommodated in the above form by taking $x_i'\beta_i = z_i'\beta + v_i$ with $v = (v_1, ..., v_n)$ having a joint n-variate normal distribution, $x_i = (z_i, 1)$ and $\beta_i = (\beta, v_i)$, $i = 1, ...n$. Usually the mean of the normal distribution is 0 and the aproaches differ by directly specifying the variance or precison matrix, respectively. In this paper, we will pursue the second route through a multivariate generalization.

Even though the likelihood does not depend on Ψ, the posterior distribution for all model parameters must be evaluated. This is given by

$$\pi(\beta, \phi, \Psi|y) \propto l(\beta, \phi)f(\beta|\phi, \Psi)p(\phi)p(\Psi) \quad (2)$$

where it was assumed that the joint prior density f for β depends on Ψ and ϕ

and ϕ and Ψ are prior independent. In the sequel, it will be sometimes useful to define $\theta = (\phi, \Psi)$.

The observational precision ϕ is generally given a Gamma prior with density $\phi^{(\nu_\phi/2)-1} \exp\{-\nu_\phi S_\phi \phi/2\}$, denoted $G(\nu_\phi/2, \nu_\phi S_\phi/2)$, and, when Ψ is a precision matrix, it is generally given a Wishart prior with density $|\Psi|^{(\nu_\psi/2)-(r+1)/2}$ $\exp\{-tr(\nu_\psi S_\psi \Psi)/2\}$, denoted $W(\nu_\psi/2, \nu_\psi S_\psi/2)$.

The rest of the paper is organized as follows. The next section introduced the spatial model used in the regression context and derives the posterior distribution. Section 3 presents 4 different schemes to undertake sampling-based inference using MCMC. Section 4 presents other forms of prior specification. Section 5 provides some results from application to simulation studies and a real dataset and Section 6 draws some concluding comments.

2 Model definition

A suitable model to represent spatial situations is provided by Markov random fields (MRF). In very simple terms, a collection $X = (X_1, ..., X_n)$ of random quantities is said to form a MRF if the joint distribution of X satisfies the property that $(X_i|X_{-i}) \sim (X_i|X_{\partial i})$ where $\partial i = \{j : j \text{ is a neighbour of } i\}$, for $i = 1, ..., n$. When the regions are ordered, $\partial i = \{i-1,\ i+1\}$ and the condition reduces to $(X_i|X_{-i}) \sim (X_i|X_{i-1}, X_{i+1})$, for all $i \neq 1, n$.

There are many possible prior models for β that can follow a MRF. An interesting example is provided by the pairwise difference prior

$$f_{pd}(\beta|\phi, \Psi) \propto \phi^{nr/2} \, |\Psi|^{n/2} \, \exp\left\{ -\frac{\phi}{2} \sum_{i,j=1}^{n} w_{ij} (\beta_i - \beta_j)' \Psi (\beta_i - \beta_j) \right\} \qquad (3)$$

where w_{ij} are weights associated with the neighboring structure. For example,

$$w_{ij} = \begin{cases} 1 & , \text{ if } i \sim j \\ 0 & , \text{ otherwise} \end{cases} , \qquad (4)$$

where $i \sim j$ means that regions i and j are neighbors. Of course, the matrix Ψ need not be scaled by ϕ but this can be useful. Removal of the dependence of (3) on ϕ implies trivial changes on calculation. The form (3) was proposed independently by Moreira and Migon (1999) and Assunção et al. (1999), without scaling on ϕ. It is basically a generalization of the univariate pairwise difference prior used by Besag et al. (1991). The idea of using pairwise differences as a model for regression coefficients was first considered in a paper by Assunção et al. (1998). Models based on (1) and (3) will be called pure space-varying regression models (pure SVRM), because all r regression coefficients are subject to a spatial structure imposed by the prior.

This prior basically attributes larger probability to regions of the β space that have similar values for neighboring β_i's. It is an improper prior because the variance of β is $\phi^{-1}W^{-1} \otimes \Psi^{-1}$ where $W = (k_{ij})$ and

$$k_{ij} = \begin{cases} w_{i+} & , \text{ if } i = j \\ -w_{ij} & , \text{ if } i \sim j \\ 0 & , \text{ otherwise} \end{cases} \qquad \text{and } w_{i+} = \sum_{j \sim i} w_{ij}.$$

Consequently, the rows of the $n \times n$ matrix W add up to zero. However, this is a useful prior representation of spatial structure and leads to proper posteriors and sensible results, provided proper priors are specified for ϕ and Ψ.

Simple calculations show that (3) can also be rewritten as

$$f_{pd}(\beta|\phi, \Psi) \propto \phi^{nr/2} \, |\Psi|^{n/2} \exp\left\{ -\frac{\phi}{2} Q(\beta) \right\}$$

where $Q(\beta)$ can be written in many different ways. The most useful ones are

$$Q(\beta) = \sum_{i,j=1}^{n} k_{ij} \beta_i' \Psi \beta_j = \beta' \left(W \otimes \Psi \right) \beta = \text{tr}[Q_s(\beta)\Psi],$$

where $Q_s(\beta) = B'WB = \sum_{i,j=1}^{n} k_{ij} \beta_i \beta_j' = \sum_{i,j=1}^{n} w_{ij} (\beta_i - \beta_j)(\beta_i - \beta_j)'$.

It can also be shown that the full prior conditional distribution of β_i is given by

$$\beta_i | \beta_{-i}, \phi, \Psi \sim \beta_i | \beta_{\partial i}, \phi, \Psi \sim N\left(\bar{\beta}_{\partial i}, \frac{1}{\phi w_{i+}} \Psi^{-1} \right) \text{ where } \bar{\beta}_{\partial i} = \frac{1}{w_{i+}} \sum_{j \sim i} w_{ij} \beta_j \quad (5)$$

is the average of the w_{i+} neighboring β's, for $i = 1, ..., n$.

Combination of all model assumptions gives

$$\begin{aligned} \pi(\beta, \phi, \Psi|y) \propto & \; \phi^{\frac{n}{2}} \exp\left\{ -\frac{\phi}{2} \sum_{i=1}^{n} (y_i - x_i' \beta_i)^2 \right\} \phi^{\frac{nr}{2}} |\Psi|^{\frac{n}{2}} \exp\left\{ -\frac{\phi}{2} \beta' \left(W \otimes \Psi \right) \beta \right\} \\ & \times \; \phi^{(\nu_\phi/2)-1} \exp\left\{ -\frac{1}{2} \nu_\phi S_\phi \phi \right\} |\Psi|^{(\nu_\psi/2)-r} \exp\left\{ -\frac{1}{2} tr(\nu_\psi S_\psi \Psi) \right\} \\ & \propto \exp\left\{ -\frac{\phi}{2} \left[\nu_\phi S_\phi + \beta' \left(X'X + W \otimes \Psi \right) \beta - 2\beta' X'y + y'y \right] \right\} \\ & \times \; \phi^{[(\nu_\phi + n + nr)/2]-1} |\Psi|^{[(\nu_\psi + n)/2]-r} \exp\left\{ -\frac{1}{2} tr(\nu_\psi S_\psi \Psi) \right\} \qquad (6) \end{aligned}$$

where the design matrix $X = \text{diag}(x_1', ..., x_n')$ is in a slightly unusual block diagonal form. If Ψ is not scaled by ϕ in (3) then the posterior becomes

$$\begin{aligned} \pi(\beta, \phi, \Psi|y) \quad \propto \quad & \exp\left\{ -\frac{\phi}{2} \left[\nu_\phi S_\phi + \beta' \, X'X \, \beta - 2\beta' X'y + y'y \right] - \frac{1}{2} \beta' \left(W \otimes \Psi \right) \beta \right\} \\ \times \quad & \phi^{[(\nu_\phi + n)/2]-1} |\Psi|^{[(\nu_\psi + n)/2]-r} \exp\left\{ -\frac{1}{2} tr(\nu_\psi S_\psi \Psi) \right\} \qquad (7) \end{aligned}$$

These distributions are not easily summarized and special approximating schemes are needed. We shall concentrate here on Markov chain Monte Carlo (MCMC) methodology where samples are repeatedly taken from Markov chain kernels to reproduce a stationary trajectory towards chain equilibrium (Gamerman, 1997). A central issue is the determination of fast sampling schemes based on full conditionals. Several such schemes are detailed in the next section.

3 Sampling schemes

It should be noted that many possibilities are available for sampling from (6). We shall deal with each one of them in detail, commenting where possible on the anticipated advantages and disadvantages of each scheme. Broadly speaking, we will be considering with the following sampling scheme:

A. sampling from β_1, ... , β_n, ϕ and Ψ;

B. sampling from (β, ϕ) and Ψ;

C. sampling from β^1, ... , β^p, ϕ and Ψ.

D. sampling from (ϕ, Ψ) only;

A few comments are needed before detailing each of the sampling schemes above. First, comparison between schemes should be based on overall cost that includes cost of a single draw **and** convergence rate. Secondly, for complicated posterior forms multimodality is very likely to occur. The priors used in the models here are known to have convergence problems (Knorr-Held and Rue, 2000). It is important that sampling schemes reliably converge to the more relevant regions of the parameter space.

Note that, for scheme (D), samples must be drawn from the marginal posterior of $\Psi|y$. Samples from β and ϕ are then obtained by noting that

$$\pi(\beta, \phi, \Psi|y) = \pi(\beta, \phi|\Psi, y)\pi(\Psi|y).$$

Draws from (β, ϕ) are obtained by drawing a value Ψ^* from the marginal posterior $\pi(\Psi|y)$ and then sampling from the tractable distribution $\pi(\beta, \phi|\Psi^*, y)$. Therefore, scheme D is actually sampling all model parameters jointly.

For schemes A to C, the full posterior conditionals for ϕ and for precision matrices Ψ are trivially obtained from (6) as $G\{(\nu_\phi + n + nr)/2, [\beta'(X'X + W \otimes \Psi)\beta - 2\beta' X'y + y'y]/2\}$ and $W\{(\nu_\psi + n)/2, [\nu_\psi S_\psi + \phi Q_s(\beta)]/2\}$ distributions, respectively. We therefore shall concentrate on sampling from the regression coefficients β. The case when Ψ is not a full precision matrix is dealt with in the next section but note that it is always true that $\pi(\Psi|\beta, \phi, y) \propto p(\Psi)W[\nu_\psi/2, \phi Q_s(\beta)/2]$.

When the pairwise difference precision Ψ is not scaled by ϕ, the full conditionals for ϕ and Ψ are $G\{(\nu_\phi + n)/2, [\beta' X'X\,\beta - 2\beta' X'y + y'y]/2\}$ and $W\{(\nu_\psi + n)/2, [\nu_\psi S_\psi + Q_s(\beta)]/2\}$ distributions, respectively.

3.1 Scheme A

The full posterior conditional of β_i is easily obtained as

$$\pi(\beta_i|\beta_{-i},\phi,\Psi,y) \propto l(\beta_i) f_{pd}(\beta_i|\beta_{-i},\phi,\Psi).$$

The likelihood term is simply $\exp\{-\phi(y_i - x_i'\beta_i)^2/2\}$ and the prior term is given by (5). Combination of these results gives $(\beta_i|\beta_{-i},\phi,\Psi,y) \sim (\beta_i|\beta_{\partial i},\phi,\Psi,y) \sim N(a_i,\phi^{-1}R_i)$ where $a_i = R_i(x_i y_i + w_{i+}\Psi\bar{\beta}_{\partial i})$ and $R_i = (x_i x_i' + w_{i+}\Psi)^{-1}$, for $i = 1,...,n$. When Ψ is not scaled by ϕ, the expression of a_i and R_i change to $a_i = R_i(\phi x_i y_i + w_{i+}\Psi\bar{\beta}_{\partial i})$ and $R_i = (\phi x_i x_i' + w_{i+}\Psi)^{-1}$, for $i = 1,...,n$. These distributions are easily sampled from and the most computationally demanding task is the inversion of $r \times r$ variance matrices.

Therefore, sampling cost is not an issue here as r is usually rather small (in the order of 10^1) but convergence rate may be. Neighbouring β_i's are expected to be highly correlated due to their prior form and if this correlation is large this may cause considerable delay is reaching chain equilibrium. This problem may be severe since there may be many β_i's (typically in the order of 10^3 to 10^4).

One could also consider yet another sampling scheme where each β_{ij} is sampled univariately. Given the ease of operation of scheme (A), this scheme does not seem necessary. Note also that scheme A is sometimes preferred (with respect to the other schemes in this paper) in the literature now available when pairwise difference priors are used.

3.2 Scheme B

Let us first define $R = (X'X + W \otimes \Psi)^{-1}$ and $a = R\,X'y$. The full posterior conditional of (β,ϕ) is obtained from (6) as

$$\begin{aligned}\pi(\beta,\phi|\Psi,y) &\propto \phi^{[(\nu_\phi+n+nr)/2]-1}\exp\left\{-\frac{\phi}{2}\left[\nu_\phi S_\phi + \beta' R^{-1}\beta - 2\beta' R^{-1}a + y'y\right]\right\} \\ &\propto \phi^{[(\hat{\nu}_\phi+nr)/2]-1}\exp\left\{-\frac{\phi}{2}\left[(\beta-a)'R^{-1}(\beta-a) + \hat{S}_\phi\right]\right\} \qquad (8)\end{aligned}$$

where $\hat{\nu}_\phi = \nu_\phi + n$ and $\hat{S}_\phi = \nu_\phi S_\phi + (y - Xa)'(y - Xa)$. It is clear from (8) that $(\beta,\phi|\Psi,y) \sim NG(a,R,\hat{\nu}_\phi/2,\hat{S}_\phi/2)$, ie $(\beta|\phi,\Psi,y) \sim N(a,\phi^{-1}R)$ and $(\phi|\Psi,y) \sim G(\hat{\nu}_\phi/2,\hat{S}_\phi/2)$.

When Ψ is not scaled by ϕ, the full posterior conditional of (β,ϕ) is no longer in closed NG form. Nevertheless, it is straightforward to obtain the full conditional of β for (7) as $(\beta|\phi,\Psi,y) \sim N(a,R)$, where the expression of the moments are changed to $R = (\phi X'X + W \otimes \Psi)^{-1}$ and $a = \phi R\,X'y$.

These distributions are also simple to generate values from but now the computational demand has been substantially increased by the need to invert $nr \times nr$ dimensional matrices to obtain R. Fortunately, this is not required

here. Rue (2001) showed that substantial computational savings are obtained by appropriately exploring MRF properties of R^{-1}.

A summary of Rue's strategy follows below. The structure of R^{-1} is typically very sparse with many null elements due to the MRF form of W and the block diagonal form of $X'X$. Therefore, R^{-1} can be rearranged without any loss in a band diagonal form with smallest possible bands. There are known permutation schemes that allows this operation to be performed (Saad, 1996). Substantial gains are obtained and the Cholesky decomposition $R^{-1} = LL'$ is easily performed, where the lower triangular matrix L will also be band diagonal. Therefore, $U = L'\beta - L'a$ will have $N(0, I)$ full posterior conditional and can be quickly generated. To obtain a draw from β, let $c = L'a$. Then, $c = L'RX'y = L^{-1}X'y$ or $X'y = Lc$. This allows easy calculation of c by working downwardly along the rows of L. With the value of c and the drawn value of U, β is obtained from $L'\beta = U + c$ by working upwardly along the rows of L'. Full details and generating codes are available in Rue (2000). In our computations, the algorithm described in Golub (1996) was used.

The main advantage of this approach (over scheme A) is the ability to perform block sampling over the possibly highly correlated components β_i. The main disadvantage is the computational cost involved in the permutations and Cholesky decomposition. Note that no inversion operations are required and that the permutations do not require splitting any of the β_i's. This scheme was used by Besag and Higdon (1999).

3.3 Scheme C

Let us first define for any $m \times r$ matrix (or $1 \times r$ vector) A, the $m \times (r-1)$ matrix (or $1 \times (r-1)$ vector) A^{-j} as the matrix (vector) A without its jth column and $X^j = \text{diag}(x_{1j}, ..., x_{nj})$, $j = 1, ..., r$. Scheme C can be easily obtained by noting that given the value of B^{-j}, the model for the *observations* $\check{y}_i = y_i - x_i^{-j\prime}\beta_i^{-j}$, $i = 1, ..., n$ is $\check{y} \sim N(X^j\beta^j, \phi^{-1}I)$.

The full prior conditional for β^j induced from (3) is

$$f_{pd}(\beta^j|\beta^{-j}, \phi, \Psi) \propto \exp\left\{-\frac{\phi\psi_{jj}}{2}\beta^{j\prime}W\beta^j\right\},$$

where ψ_{jj} is the jth diagonal element of Ψ, $j = 1, ..., p$. Combining with the likelihood above gives

$$\begin{aligned}\pi(\beta^j|\beta^{-j}, \phi, \Psi, y) &\propto \exp\left\{-\frac{\phi}{2}(\check{y} - X^j\beta^j)'(\check{y} - X^j\beta^j)\right\}\exp\left\{-\frac{\phi\psi_{jj}}{2}\beta^{j\prime}W\beta^j\right\} \\ &\propto \exp\left\{-\frac{\phi}{2}\left[\beta^{j\prime}(\psi_{jj}W + X^{j\prime}X^j)\beta^j - 2\beta^j X^{j\prime}\check{y}\right]\right\}.\end{aligned}$$

Defining $R^j = (\psi_{jj}W + X^{j\prime}X^j)^{-1}$ and $a^j = R^j\, X^{j\prime}\check{y}$, gives that $(\beta^j|\beta^{-j}, \phi, \Psi, y) \sim N(a^j, R^j)$, for $j = 1, ..., r$. Once again, when Ψ is not scaled by ϕ in (3), the

expression of the moments above are changed to $R^j = (\psi_{jj} W + \phi X^{j\prime} X^j)^{-1}$ and $a^j = \phi R^j X^{j\prime} \tilde{y}$, for $j = 1, ..., r$.

Scheme C requires r generations of n-dimensional normal variates, in comparison with the nr generations required for scheme B. Note that as in scheme B, inversion of R^j is not required and that R^j also shares the nice band diagonal form due to the diagonal form of X^j and the sparse nature of W induced by the MRF prior. Therefore, drawing from the β^j's follows exactly the same strategy described for scheme B.

The main disadvantage over scheme B is the possible slow mixing due to the correlation between the β^j's. This effect may be alleviated by linear reparametrization via centering of the explanatory variables. In standard linear regression, this helps to orthogonalize the design matrix and hence leads to less correlated parameters. Sampling may be performed for these transformed parameters and the β^j's are easily recovered. The main advantage over scheme C may be speed of computations. Cholesky decompositions for scheme B requires operations of order $\mathcal{O}(n^2 r^2)$ whereas for scheme they only require order $\mathcal{O}(n^2)$ operations. Computational savings may be substantial even for very small n because n is typically large.

3.4 Scheme D

It is easy to see from (8) that the proportionality constant required to complete the expression of the full posterior conditional density of (β, ϕ) is

$$(2\pi)^{nr/2} |R|^{-1/2} \frac{(\hat{S}_\phi/2)^{\hat{\nu}_\phi/2}}{\Gamma(\hat{\nu}_\phi/2)}.$$

Note, however, that $\pi(\Psi|y) = \pi(\beta, \phi, \Psi|y)/\pi(\beta, \phi|\Psi, y)$. Combining the above result with (6) and (8) and discarding constants gives

$$\begin{aligned} \pi(\Psi|y) &\propto |R|^{1/2} (\hat{S}_\phi)^{-\hat{\nu}_\phi/2} |\Psi|^{n/2} p(\Psi) \\ &\propto \frac{|\Psi|^{n/2}}{|R^{-1}|^{1/2}} (\hat{S}_\phi)^{-\hat{\nu}_\phi/2} p(\Psi) \\ &\propto \frac{|\Psi|^{n/2}}{\prod_{i=1}^{nr} l_{ii}} (\hat{S}_\phi)^{-\hat{\nu}_\phi/2} p(\Psi) \end{aligned} \tag{9}$$

where we made use of the facts that $|R^{-1}| = |L|^2 = (\prod_{i=1}^{nr} l_{ii})^2$ and $R^{-1} a = X' y$. Note that $\hat{S}_\phi$ depends on Ψ.

When Ψ is not scaled by ϕ in (3), only β can be integrated out in (7). Using again the results obtained for scheme B, it is not difficult to obtain that

$$\begin{aligned} \pi(\phi, \Psi|y) \propto \phi^{[(\nu_\phi+n)/2]-1} \exp\left\{-\frac{\phi}{2}\hat{S}_\phi\right\} |\phi X'X + W \otimes \Psi|^{-1/2} \\ \times |\Psi|^{(\nu_\psi+n-r-1)/2} \exp\left\{-\frac{1}{2} tr(\nu_\psi S_\psi \Psi)\right\}. \end{aligned} \tag{10}$$

In this case, ϕ cannot be integrated out analytically as before.

Although it is possible to analytically obtain (9) or (10) up to a proportionality constant, it is not easy to devise direct sampling schemes. Indirect sampling schemes such as SIR (Rubin, 1988) or adaptive SIR (Schmidt et al., 1999) may be applied here if the dimensions involved are not large. As a general purpose scheme, MCMC with Metropolis-Hastings proposals will be used here. Two possible proposal forms are random walks and sampling from an approximating density such as the Wishart for Ψ and Gamma for ϕ, where applicable.

For example, if a $W(\nu_\psi/2, \nu_\psi S_\psi/2)$ prior and a $W[\hat{\nu}_\psi/2, (\nu_\psi S_\psi + y'y)/2]$ proposal are used for Ψ in (9), the acceptance probability becomes

$$\alpha(\Psi^{(c)}, \Psi^{(p)}) = 1 \wedge \prod_{i=1}^{nr} \frac{l_{ii}^{(c)}}{l_{ii}^{(p)}} \exp\left\{\frac{1}{2}\mathrm{tr}[y'y(\Psi^{(p)} - \Psi^{(c)})]\right\} \left(\frac{\nu_\phi S_\phi + y'y - a^{(c)\prime} X'y}{\nu_\phi S_\phi + y'y - a^{(p)\prime} X'y}\right)^{\hat{\nu}_\phi/2}$$

where the superscripts (c) and (p) stand for current and proposed, respectively.

Another possibility is to use Wishart (and Gamma) forms for Ψ (and ϕ) centered around the previous chain values and number of degrees of freedom (d.o.f.) ν tuned to achieve reasonable rates of chain moves. This form of chain is very similar to random walk forms. This is more easily seen in the scalar case where one can rewrite the proposal as $\log \phi^{(c)} = \log \phi^{(p)} + \eta$ where $e^\eta \sim G(\nu, \nu)$. In the applications of this paper, this form of proposal was used.

4 Extensions

There are many interesting extensions to the basic model described above. In this section, we consider some of them, namely, mixed SVRM, other forms of prior specification for β and special forms for the hyperparameter Ψ.

4.1 Mixed SVRM

In the pure form, SVRM have all regression coefficients subject to a spatial structure. This may be unrealistic and in practice there may be effects that do not vary with space. These effects cause the model to include effects of different natures. In general, these effects can have other influences form a variety of sources due to an unstructured hierarchical form or a temporal element to them. In this paper, however, we restrain ourselves to spatial effects and static effects.

Therefore, the mean responses of the observations are now given by $\mu_i = z_i'\mu + x_i'\beta_i$. The β_i's are still related by the spatial structure (3) but the static regression coefficient μ is not. There are no restrictions to variables that enter the vectors of explanatory variables z_i and x_i. In fact, they can even be the same. Identifiability conditions require that whenever a covariate enters both z and x then the associated component of the β_i's must have a fixed sum. Usually,

one imposes that $\sum_{i=1}^{n} \beta_{ij} = 0$ so that the corresponding component of μ can be interpreted as an overall basic effect.

Typically but not necessarily the model is completed with an independent prior $\mu|\phi \sim N(m, \phi^{-1}M)$, for some vector m and positive definite matrix M. This causes the parameter vector to increase but componentwise MCMC analysis remains virtually unchanged. When μ is known, new *observations* $\breve{y}_i = y_i - z_i'\mu$ can be formed, very much like in scheme C, and they restore the likelihood (1). Sampling for the other model parameters remains as before with observation vector $\breve{y}$ replacing y.

Sampling μ given the other parameters is trivial as the likelihood is based on *observations* $\breve{y}_i = y_i - x_i'\beta_i \sim N(z_i'\mu, \phi^{-1})$. Combining it with the prior gives the full posterior conditional $(\mu|\beta, \phi, \Psi, y) \sim N(\hat{m}, \phi^{-1}\hat{M})$ where $\hat{M} = (Z'Z + M^{-1})^{-1}$, $\hat{m} = M(Z'\breve{y} + M^{-1}m)$ and Z is the (static) design matrix with rows $z_1', ..., z_n'$.

4.2 Proper prior specifications

One of the (theoretical) disadvantages of (3) is that it is improper. The impropriety stems from the fact that the prior is only informing about distances between β_i's and not informing about the β_i's themselves. One simple way to correct that is to *pin* the β_i's down to some point in their space of variation. A simple extension is provided by the prior

$$f(\beta|\phi, \Psi, \lambda) \propto \phi^{nr/2} |\Psi|^{n/2} \exp\left\{ -\frac{\phi}{2} \left[Q(\beta) + \lambda \sum_{i=1}^{n} (\beta_i - b_i)'\Psi(\beta_i - b_i) \right] \right\}, \quad (11)$$

where $Q(\beta)$ is as in (3) and the additional parameter λ controls the relative weight attached to the *pinning down* part of the prior. The proportionality constant is now finite and can be analytically obtained. Useful choices for the b_i's are 0 or the MLE obtained in the static (or some local) regression. Needless to say, genuine prior mean specifications should be used whenever they are available. A special case of this form is used as a prior for spatial effects by Fernandez (1999).

This prior is equivalent to $(\beta|\phi, \Psi, \lambda) \sim N[C^{-1}(\lambda I \otimes \Psi)\mu, C]$ where $\mu = (\mu_1', ..., \mu_n')'$ and $C^{-1} = (\lambda I + W) \otimes \Psi$. As a consequence, the full prior conditional of β_i is

$$\beta_i|\beta_{-i}, \phi, \Psi, \lambda \sim \beta_i|\beta_{\partial i}, \phi, \Psi, \lambda \sim N\left(w_i\bar{\beta}_{\partial i} + (1 - w_i)\mu_i, \frac{1}{\phi(w_{i+} + \lambda)}\Psi^{-1} \right)$$

where $w_i = w_{i+}/(w_{i+} + \lambda)$, i = 1, ... , n

4.3 Special forms for the hyperparameter

One special case of Ψ considers a diagonal form with diagonal entries $\psi_1, ..., \psi_r$. Then, $Q(\beta) = \sum_{j=1}^{r} Q_j(\beta^j)$ where

$$Q_j(\beta^j) = \sum_{i,l=1}^{n} w_{ij}\psi_j(\beta_{ij} - \beta_{lj})^2 = \psi_j {\beta^j}' W \beta^j. \tag{12}$$

Sampling schemes for β do not suffer any change although scheme C may have beneficial mixing properties due to lack of prior correlation. Also, scheme D does not benefit from further computational simplifications to (9). If prior independence is assumed for the ψ_j's then (12) ensures that their joint full conditional distributions will consist on a product of independent densities. For example, if independent $G(\nu_\psi/2, \nu_\psi S_\psi/2)$ priors are specified for the ψ_j's, their full posterior conditional will be a product of $G[(\nu_\psi + n)/2, (\nu_\psi S_\psi + {\beta^j}' W \beta^j)/2]$ distributions.

Another interesting special case is when the precision matrix (for $r = 2$) is given by

$$\Psi = \psi_1 \begin{pmatrix} 1 & \psi_2 \\ \psi_2 & 1 \end{pmatrix}.$$

The hyperparameter ψ_1 informs about the magnitude of the dispersion and ψ_2 informs about the possible correlation between the components of the β_i's. In this case, a prior for the pair $\psi = (\psi_1, \psi_2)'$ must be specified. No simplification is achieved for any of the sampling schemes and now even sampling from the components of ψ is non standard and possibly indirect methods or Metropolis-Hastings steps may be applied.

Similar comments follow for the case when the weights w_{ij}'s in the expression of (3) depend on further hyperparameters. An example is given by continuous, geostatistical methods where the weights typically depend on the distance between the regions (i.e., $w_{ij} = \exp(-bd_{ij})$, where d_{ij} is the distance between sites i and j). In any case, the above calculations are easily adapted. A prior for the distance attenuation parameter b must be specified and a sampling step for b must be included in the MCMC scheme. This is an advantage of MCMC schemes: introduction of a further unknown quantity only implies an extra step in each iteration with steps associated with the other parameters remaining the same.

5 Applications

In order to study the performance of the schemes and the ability of the posterior to estimate the model parameters, a simulated dataset was generated. The first difficulty appears as the improper form of (3) prevents direct generation from the model. An approximating alternative must necessarily be used. We have

chosen an approximating device inspired by the spatial structure used by Anselin (1988). Also, we opted for not scaling Ψ by ϕ.

The spatial structure used is that given by the 558 homogeneous microregions of Brazil, as defined by the official Brazilian Statistics Institute IBGE, and illustrated in figure 1. The neighbouring structure is defined by the existence of a common border of any length and $w_{ij} = 1$ for such pairs and 0, otherwise. We used $r = 2$ explanatory variables that were uniformly generated. The algorithm was run with 2 parallel chains. Initial points were provided by values generated from the hyperparameters' prior distribution for all sampling schemes. Given that this prior was proper but had very large variances, these values provide reasonable reassurance of exploration of different regions of the parameter space. It should be unlikely for both chains of the same scheme to converge to the same local mode. Convergence was diagnosed according to Geweke's (1992) test within chains and the Gelman and Rubin (1992) shrinkage factor.

The analyses report results obtained from the last 1000 values from each chain totalling 2000 values for all model parameter. Scheme D was run with the random walk form for the proposal described before and d.o.f. tuned to acceptance rates around 40%. In general, iterations from scheme A are around 6 times faster than scheme C that is around 3 times faster than schemes B and D. In addition, the burn-in period seems to be unaffected by changes to the hyperparameters' prior and indicate faster convergence for schemes A and B (after around 400 iterations) than for schemes B and D (after around 800 iterations). We have opted to allocate the same CPU time to all schemes. Since iterations of scheme A are the fastest ones, each chain was allowed to run for a further iterations and only when time limit was reached values were stored.

Table 1 below shows the mean value of $\log \pi(\beta, \theta)$ where $\theta = (\phi, \Psi)$ for schemes B and D and a number of prior value specifications for $\nu_\phi = \nu_\psi = \nu$ and $S_\psi = S_\phi I = SI$, for the simulated dataset. The prior means are, therefore, given by $E(\phi) = E(\Psi_{ii}) = S^{-1}$, for $i = 1, ..., r$.

The results show that for a wide range of prior specifications schemes B and D lead ot the same posterior values, indicating convergence to the same parameter region. It is worth stressing here that entries on table 1 were compiled after the chains were diagnosed to be stationary according to the criteria mentioned above.

Table 1. Joint posterior density (in log)

prior		Scheme	
d.o.f. (ν)	mean ($1/S$)	B	D
10	1	-2297	-2296
	100	-2103	-2103
50	1	-2579	-2579
	100	-1912	-1912
100	1	-2746	-2746
	100	-1671	-1669

5.1 Simulated dataset

Anselin (1988) proposed the study of spatial regression model by introduction of spatial correlation into the observational error structure. His idea can be adapted to the construction of spatially correlated parameters. Specifically, the β's can be generated from the model where

$$\beta_i = \rho\bar{\beta}_{\partial i} + e_i\,, \quad e_i \sim N\left(0, \frac{1}{w_{i+}}\Psi^{-1}\right) \tag{13}$$

independently, for $i = 1, ..., n$. In matrix notation, this means that $B = \rho W^* B + e$ with $\text{vec}(e) \sim N(0, \Psi^{-1} \otimes K^{-1})$ where $\text{vec}(X)$ is the column vectorization of the matrix X, W^* given by

$$W^* = (k^*_{ij}) \text{ where } k^*_{ij} = \begin{cases} 0 & , \text{ if } i = j \\ w_{ij}/w_{i+} & , \text{ if } i \sim j \end{cases}.$$

and $K = diag(w_{1+}, ..., w_{n+})$. The spatial correlation is induced by the neighbouring matrix W^* and it is not difficult to check that W^* and W are closely related through $W = K(I - W^*)$.

The independence imposed by the right variance K means that the prior for β is given by $f(\beta) \propto \prod_{i=1}^{n} f_i(\beta_i; \beta_{-i})$ where each f_i is given by (13). When $\rho = 1$, these are the full prior conditionals in (5). This prior can therefore be seen as an approximation to the prior in (3) in the same spirit as the pseudo-likelihood suggested by Besag (1975) approximates the true likelihood arising with observations from a Markov random field (Qian and Titterington, 1991).

Parameter generation was carried out with $\rho = 0.999$. The dataset generation is completed with observations drawn from model (1) with $r = 2$, a diagonal matrix $\Psi = diag((0.03)^{-2}, (0.15)^{-2})$ and $\phi = 0.25$. Results must be interpreted with care since there is no clear correspondence between this spatial structure and (3). Note that prior (13) also becomes improper when $\rho = 1$ since $I - W^*$ is rank defficient.

Figures 1, 2 and 3 shows the results of the estimation in space (fig. 1) and over the line (figs. 2 and 3), respectively, of the spatially varying regression coefficients along with uncertainty bounds for the prior with $\nu = 10$ and mean 1. The strong spatial pattern of the process is clear from the figures. The estimates appear to reproduce well the true spatial pattern. Important differences in estimation appear when the schemes are summarized. Squared deviations can be defined as $(\beta_i - \hat{\beta}_i)^2$ where $\hat{\beta}_i$ is the posterior mean of β_i. These quantities can be summed up over all regions to provide an overall measure of fit named here SSD_i. These figures, provided in table 2, are substantially larger for scheme C whereas no apparent difference seems to exist between scheme A, B and D. This provide indication of convergence of scheme C to a different region of the parameter space. The priors used in table 2 show a variety of opinions ranging from strong to weak prior information, as measured by the number of degrees of feeedom, and from small to large expected values for the hyperparameters.

Convergence of scheme C seems to worsen with change of prior means. For the other schemes, the SSD figures change very little as the parameters of the prior distribution are changed. This indicates that there is strong information in the likelihood for spatially varying regression coefficients: they are reliably estimated and the results are not changed even after substantial changes in the prior. The same is not true for the hyperparameters.

The estimated chain autocorrelation for the hyperparameters is depicted in Figure 4. The scheme with smaller autocorrelations is scheme B followed by schemes A and D and then followed by scheme C. We also present below the results from point estimation of hyperparameters in Table 3. Posterior estimates tend to concentrate around prior means indicating scarcity of information in the likelihood. Credibility intervals are not shown for brevity but indicate similarity between schemes A, B and D and their lengths decrease with increasing prior d.o.f. as expected. None of the regions found by the schemes include the generated value but it is unwise here to require correct estimation since we are using different models for simulation and estimation and even Ψ has different meanings in these models. Diferences between the estimated posterior means with schemes A, B and D are well within their Monte Carlo uncertainty. Figure 5 confirms these findings with the scatterplot of pairs of hyperparameters sampled from the posterior showing the same pattern, except for scheme C.

Table 2. Estimation for simulated dataset: SSD

	prior		Scheme			
	d.o.f. (ν)	mean ($1/S$)	A	B	C	D
β_1	10	1	0.042	0.041	0.050	0.041
		100	0.039	0.039	0.398	0.039
	50	1	0.052	0.052	0.059	0.052
		100	0.039	0.039	0.281	0.039
	100	1	0.055	0.055	0.076	0.055
		100	0.039	0.039	0.212	0.039
β_2	10	1	0.073	0.073	0.080	0.073
		100	0.071	0.071	0.383	0.071
	50	1	0.081	0.081	0.085	0.081
		100	0.071	0.071	0.306	0.070
	100	1	0.083	0.083	0.096	0.083
		100	0.071	0.071	0.222	0.071

Table 3. Point estimation for simulated dataset: θ

	prior		Scheme			
	d.o.f. (ν)	mean ($1/S$)	A	B	C	D
ϕ (true= 0.25)	10	1	1.93	1.66	1.96	1.49
		100	160.3	120.8	150.6	121.4
	50	1	1.14	1.14	1.10	1.13
		100	103.9	104.0	101.4	104.2
	100	1	1.04	1.09	1.05	1.09
		100	100.4	104.7	101.9	104.7
Ψ_{11} (true= 1111)	10	1	21.73	22.05	24.78	23.12
		100	160.3	120.8	150.6	121.4
	50	1	5.91	5.90	6.64	6.12
		100	77.79	77.71	14.20	81.67
	100	1	3.58	3.58	3.90	3.69
		100	82.26	82.16	8.82	86.80
Ψ_{12} (true= 0)	10	1	5.54	5.48	3.19	5.31
		100	10.08	9.94	2.99	9.25
	50	1	1.82	1.81	1.33	1.82
		100	8.74	8.76	5.49	8.35
	100	1	1.06	1.06	0.80	1.05
		100	7.78	7.75	3.98	7.37
Ψ_{22} (true= 44.44)	10	1	10.05	10.01	10.90	10.51
		100	12.97	12.89	4.28	13.39
	50	1	4.71	4.69	5.21	4.93
		100	13.93	13.87	9.39	14.46
	100	1	3.13	3.14	3.43	3.29
		100	15.20	15.24	6.31	15.71

5.2 Real dataset

The Amazon region is a vast area in the North of Brazil that is recently becoming an agricultural frontier. A question of interest is to determine the pace with which land use is changing. Andersen, Granger and Reis (1997) (AGR, hereafter) proposed a VAR-like model to land usage that enables projection into the future. Denoting by y_{jkt} the proportion of land in county j used for purpose k in time t and x_{jt}, the proportion of land in county j still covered by forest ($k = 1$), the model proposed is

$$y_{jkt} = \beta_{jk0}\Delta y_{j1t} + \sum_{m=2}^{K} \beta_{jkm} y_{jm,t-1} + e_{jkt}$$

where $j = 1, ..., J$, $k = 1, ..., K$, $t = 1, ..., T$ and Δ is the time lag opetator. The authors acknowledge the presence of heterogeneity between counties by adding an extra layer to the model where, the regression coefficients β_{jkm} vary across

counties. They did not, however, consider the spatial location of the counties in this variation. Spatially determined variation seems a more reasonable assumption to account for the similarity between neighboring counties due to stage of land ocupation by settlers or proximity to large cities.

In this application, data from $J = 228$ administrative counties in the Amazon region of Brazil depicted in Figure 6, with $K = 4$ land usages: forest, crop land, pasture and fallow land, collected at years 1970, 1975, 1980 and 1985 ($T = 4$) was used. Variation of coefficients across counties was modelled with the spatial prior (3) and only data for pasture ($k = 2$) was used. A more complete analysis of this dataset will be reported elsewhere. Note, however, that the data strucutre contain many observations per county and therefore, the likelihood must be adapted accordingly.

The main results of the analysis are reported in Table 4 and Figures 6 and 7. Table 4 presents a comparative analysis of the results obtained by AGR with those from the models with fixed and spatially varying vector of regression coeficients. For the AGR and spatial models, where regression coefficients vary across counties, the mean value of the average is reported in the Table. There is reasonable agreement between these averages across models even though data used by AGR was different due to minor data adjustments. We have also caried out a comparison between the fixed and the spatial models using the predictive fit statistic D proposed by Gelfand and Ghosh (1998). Their statistic measures goodness of fit but combines it with a penalization for model complexity using decision theoretic justifications. The values obtained where $D_{static} = 0.035 > 0.016 = D_{spatial}$, showing a preference for the more complex and more realistic model with spatial variation.

Table 4. Estimated values of β

coefficient	AGR	Fixed	Spatial
β_0	0.2919	0.1384	0.1420
β_1	-0.0440	-0.0176	0.2760
β_2	1.1543	1.2771	1.0230
β_3	0.1114	0.0545	0.0730

Figure 6 shows the spatial distribution of the posterior mean of the β_{j21}'s. These coefficents measure the effect of change from agriculture in the previous time to proportion of land used for pasture in each county. The results show a clear pattern of change as one moves in the North-South direction. The effect of change from agriculture to pasture is more intense in the South of the region. This area is known to be the main entrance to the region. Human occupation in the region is typically characterized by low technology agriculture which quickly exhausts the land resources and leads to pasture use of the land. Our findings seem to confirm this prior indication.

Figure 7 shows the spatial variation arranged over a line for all regression coefficients along with corresponding credibility limits. The coefficients β_{j20}'s

are marginally significant and all the other coefficients are clearly significant with the β_{j21}'s and β_{j23}'s showing a similar spatial behaviour.

6 Concluding remarks

We have considered analyses of different sampling schemes with a spatial regression model. The results obtained with our simulations must be interpreted with care since they are all based on empirical evidence. Given that we have emulated a few different prior forms we are reasonably confident that results hold for a wider class of spatial processes. The analyses were performed on software that is freely available on `http://www.ipea.gov.br`. This software provides a useful tool for analyses of regression models with the presence of a spatial pattern infuencing the effect of the explanatory variables.

Empirical results suggest that scheme D where advantage of the analytic integrability of state parameters is taken has a good performance. This in not an entirely surprising result and confirms those obtained in state space models by Gamerman and Moreira (1998). It is also important to stress that convergence speed of scheme D may be crucially dependent on the form the hyperparameters are sampled.

Similarities between spatial and time series models are known although not heavily explored in the literature. It is well known that joint sampling of state parameters improves convergence in state space models (Shephard, 1994). The results obtained here seem to point at that direction but with less importance than in the state space context. Nevertheless, we would still recommend joint sampling of state parameters as our first choice. Given that state space models are a special, limiting case of the spatial models considered here, this assertion indicate an interesting area of investigation.

All simulations and models considered here assume normal observations. An obvious next step is to consider these models under different observational sampling schemes. There are many instances where regional observations arise in the form of counts or proportions. Unfortunately, scheme D is no longer available here and schemes A, B and C must be adapted to incorporate Metropolis acceptance steps. This poses another challenging problem of practical relevance.

Acknowledgments

The research of the first author was partially supported by grants from CNPq, FAPERJ and Ministério de Ciência e Tecnologia (PRONEX).

References

Andersen, L., Granger, C. and Reis, E. (1997). A random coefficient VAR transition model of the changes in land use in the Brazilian Amazon. *Brazilian Journal of Econometrics*, **17**, 1-13.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.

Assunção, R. M., Assunção, J. J. and Lemos, M. (1998). Induced technical change: a Bayesian spatial varying parameter model. Proceedings of the XX Meeting of the Brazilian Econometric society.

Assunção, J. J., Gamerman, D. and Assunção, R. M. (1999). Regional differences in factor productivities of Brazilian agriculture: a space varying parameter approach. Technical report, Statistical Laboratory, Universidade Federal do Rio de Janeiro.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, **24**, 179-195.

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society, Series B*, **61**, 691-746.

Besag, J, York, J. and Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.

Fernandez, C. (1999). Bayesian inference in mixture models with spatial correlation. Paper presented at the Second European conference oh Highly Structured Stochastic Systems, Pavia, Italy.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London.

Gamerman, D. and Moreira, A. R. B. (1998). Bayesian analysis of econometric time series models using hybrid integration rules. Technical report, Statistical Laboratory, Universidade Federal do Rio de Janeiro.

Gelfand, A. and Ghosh, S. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**, 1-11.

Gelman, A. and Rubin, D. R. (1992). Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, **7**, 457-511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion), in *Bayesian Statistics 4* (eds. J. M. Bernardo et al.), 169-193, University Press.

Golub, G and Van Loan, C.(1996) *Matrix Computations* The Johns Hopkins University Press

Knorr-Held, L. and Rue, H. (2000). Efficient block sampling in Markov random field models for disease mapping. Paper in preparation.

Lindley, D. V. and Smith, A. F. M. (1972) Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**,

1-41.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, **21**, 1087-91.

Moreira, A. R. B. and Migon, H. S. (1999). Spatial Heterogeneity of the Agricultural Productivity: Brazil - 1970 to 1996. Proceedings of the XXI Brazilian Meeting on Econometrics.

Qian, W. and Titterington, D. M. (1991). Estimation of parameters in hidden Markov models. *Phil. Trans. R. Soc. London A*, **337**, 407-428.

Rue, H. (2000). Fast sampling of Gaussian Markov random fields with applications. Technical Report no. 1/2000, Norwegian University for Science and Technology.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society, Series B*, **63**, XX-XX.

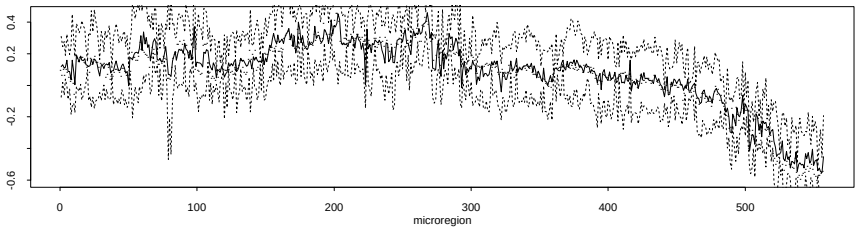
Rubin, D. B. (1988) Using the SIR algorithm to simulate posterior distributions (with discussion), in *Bayesian Statistics 3* (eds J. M. Bernardo et al.), Oxford University Press, Oxford, pp. 395-402.

Saad, Y. (1996). Iterative methods for sparse linear systems, The PWS series in computational science, PWS Publishing Company.

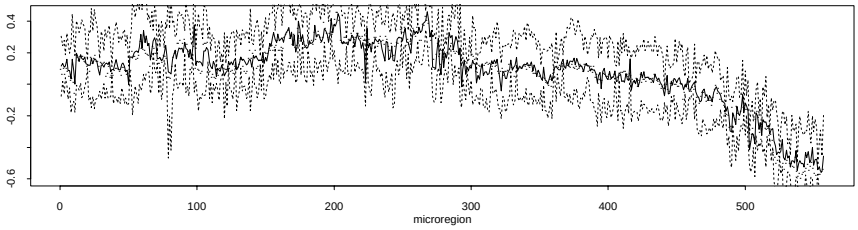
Schmidt, A. M., Gamerman, D. and Moreira, A. R. B. (1999). An adaptive resampling scheme for cycle estimation. *Journal of Applied Statistics*, **26**, 619-641.

West, M. and Harrison, P. J. (1997) *Bayesian Forecasting and Dynamic Models*, 2nd edn, Springer Verlag, New York.

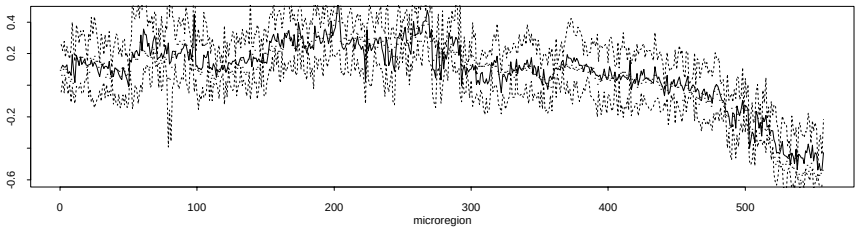
(a) Scheme A



(b) Scheme B



(c) Scheme C



(d) Scheme D

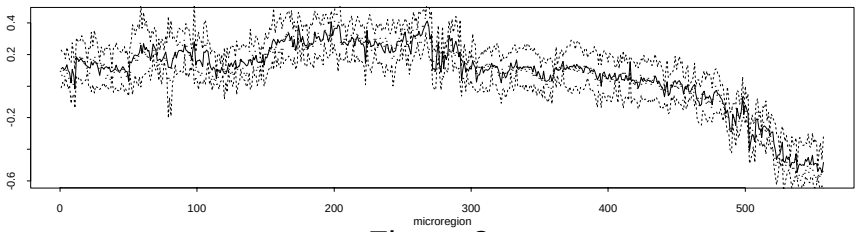
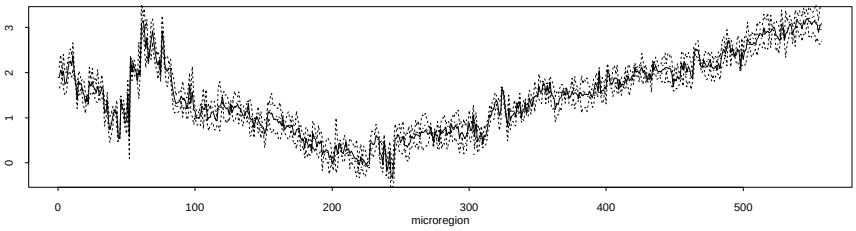
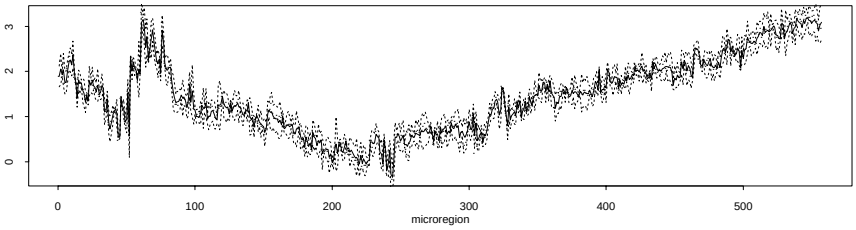


Figure 2

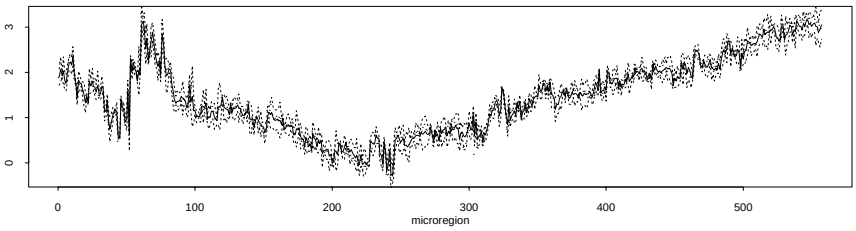
(a) Scheme A



(b) Scheme B



(c) Scheme C



(d) Scheme D

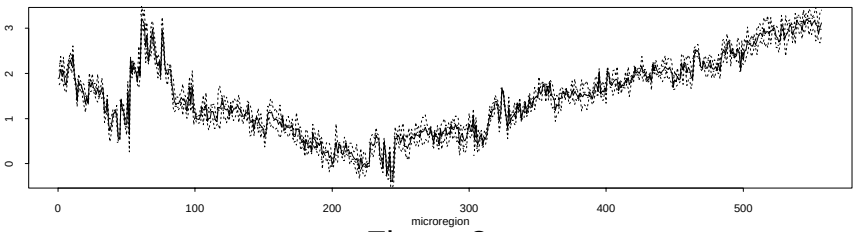
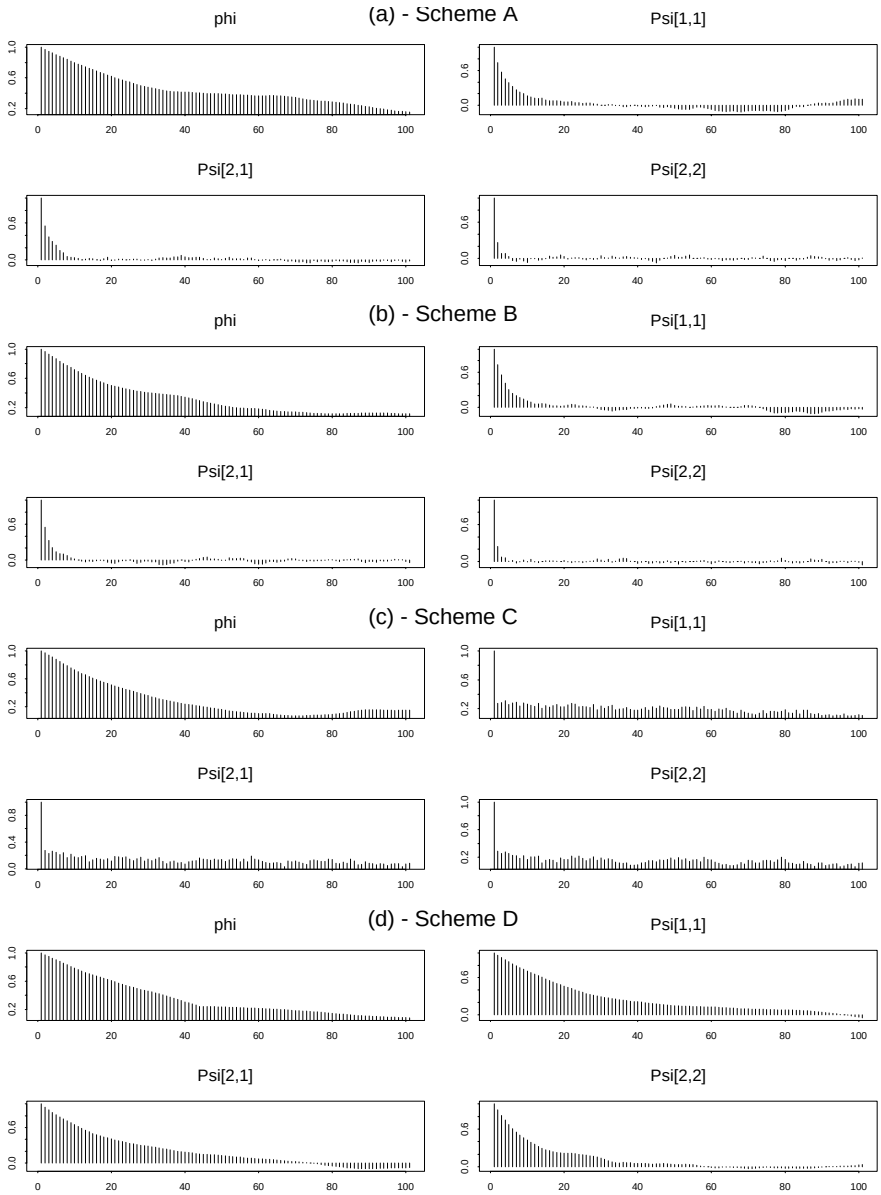
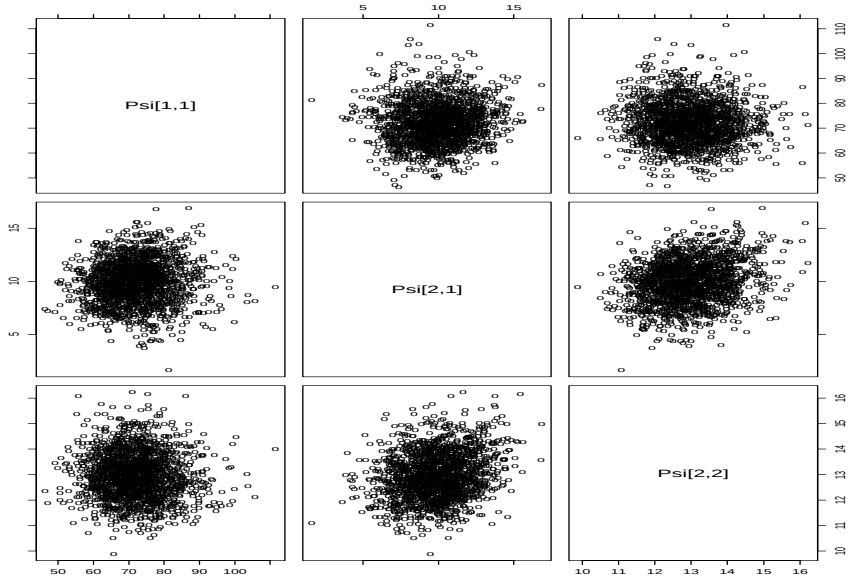


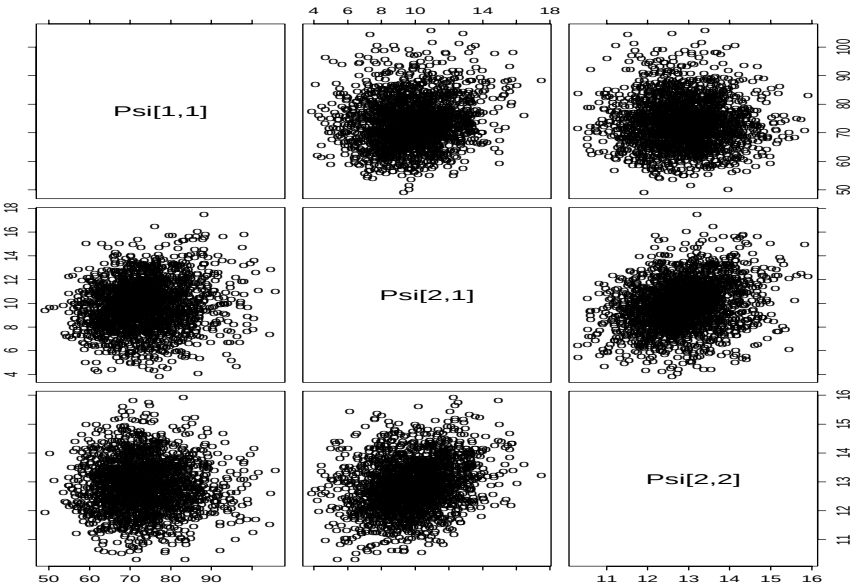
Figure 3


(a) - Scheme A
phi
Psi[1,1]
Psi[2,1]
Psi[2,2]
(b) - Scheme B
phi
Psi[1,1]
Psi[2,1]
Psi[2,2]
(c) - Scheme C
phi
Psi[1,1]
Psi[2,1]
Psi[2,2]
(d) - Scheme D
phi
Psi[1,1]
Psi[2,1]
Psi[2,2]

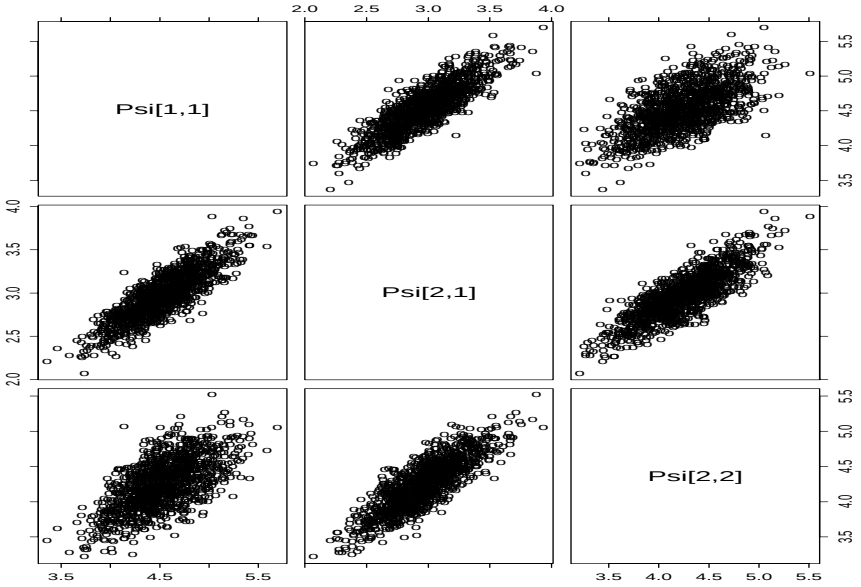
(a) - Scheme A



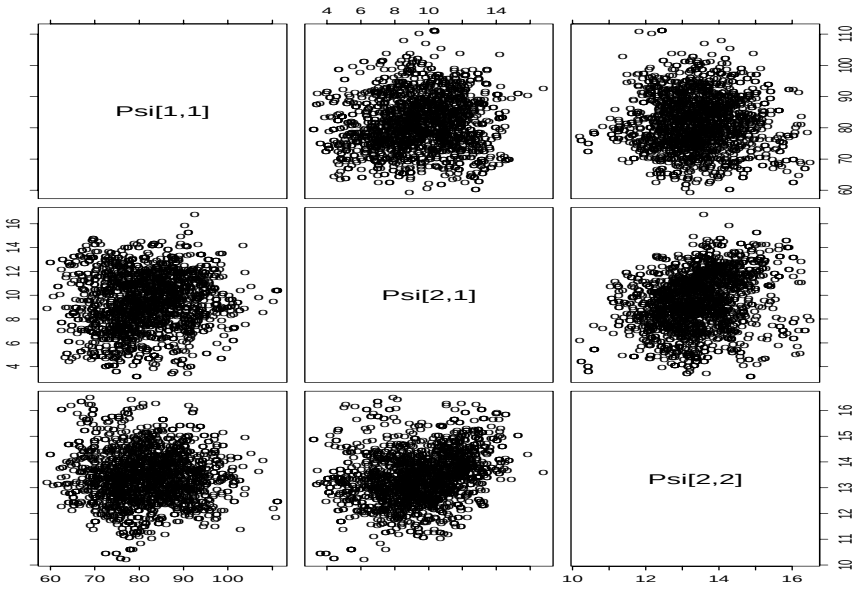
(b) - Scheme B

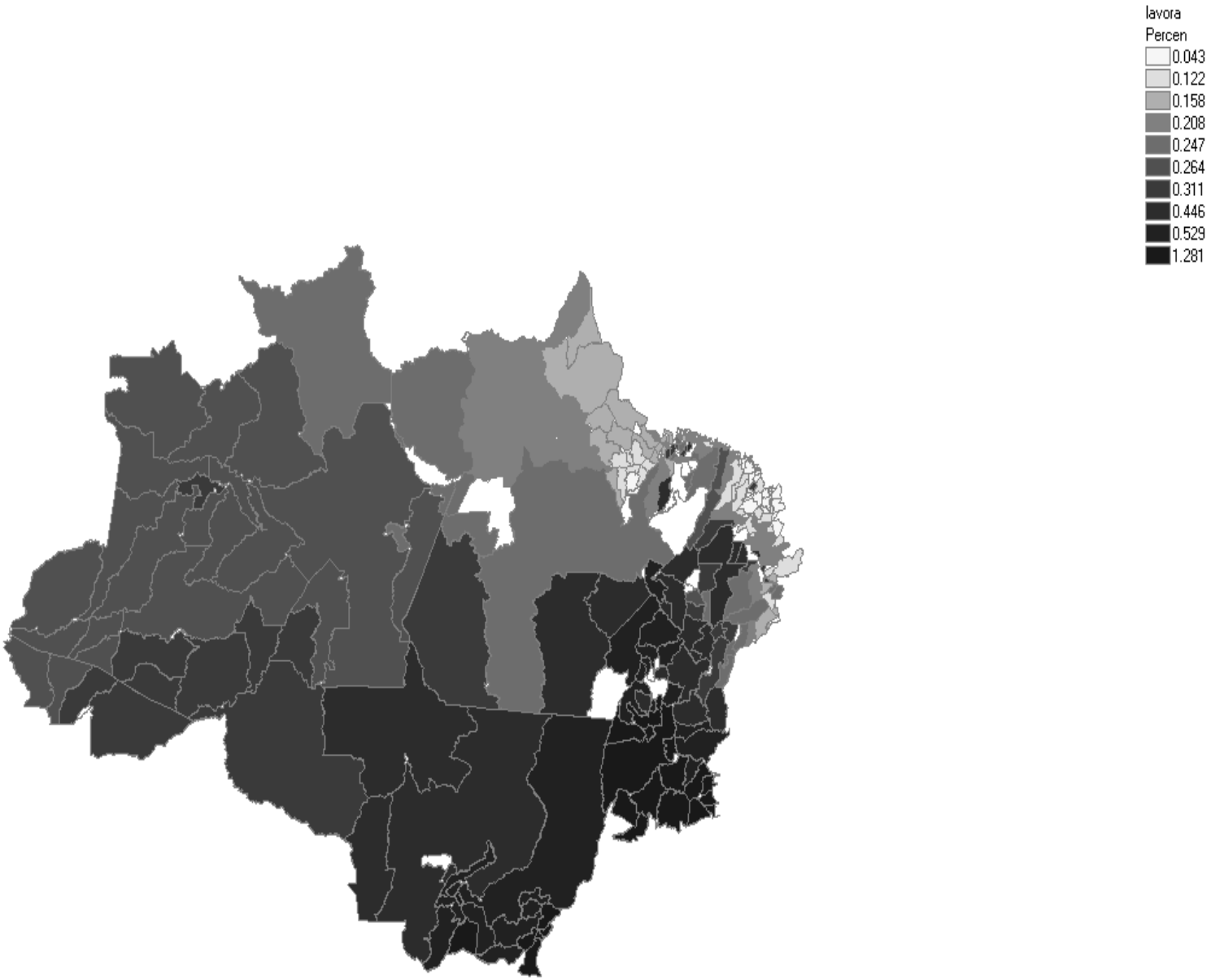


(c) - Scheme C

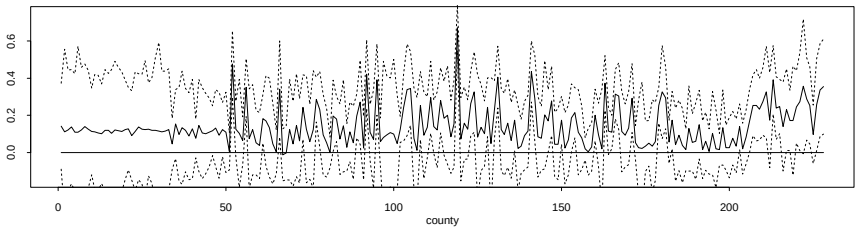


(d) - Scheme D

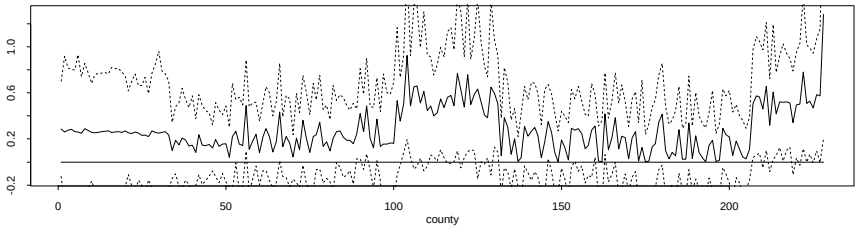



lavora
Percen
0.043
0.122
0.158
0.208
0.247
0.264
0.311
0.446
0.529
1.281

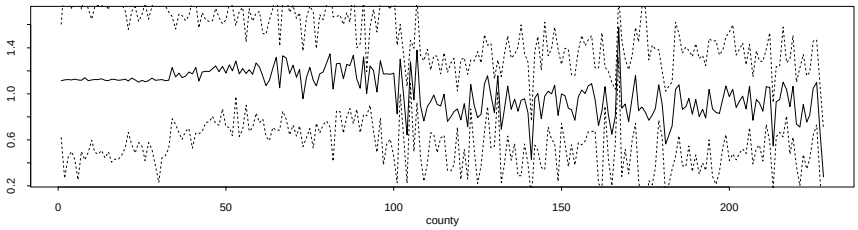
(a) beta[0]



(b) beta[1]



(c) beta[2]



(d) beta[3]

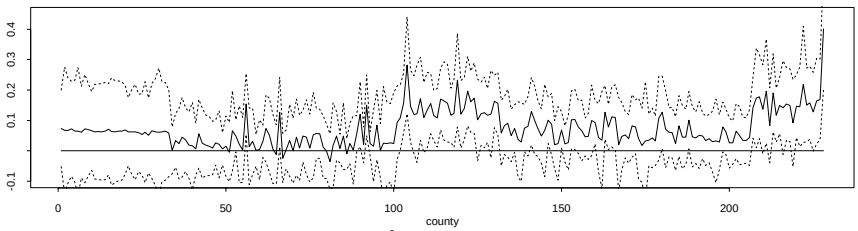


Figure 7

Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Coordination
Cláudio Passos de Oliveira

Supervision
Everson da Silva Moura
Reginaldo da Silva Domingos

Typesetting
Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Diego André Souza Santos
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa

Cover design
Luís Cláudio Cardoso da Silva

Graphic design
Renato Rodrigues Buenos

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea Bookstore

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo
70076-900 – Brasília – DF
Brazil
Tel.: + 55 (61) 3315 5336
E-mail: livraria@ipea.gov.br

Composed in Adobe Garamond 11/13.2 (text)
Frutiger 47 (headings, graphs and tables)
Brasília – DF – Brazil

Ipea's mission
Enhance public policies that are essential to Brazilian development by producing and disseminating knowledge and by advising the state in its strategic decisions.

ISSN 1415-4765

ipea Institute for Applied Economic Research

Secretariat of Strategic Affairs

BRAZILIAN GOVERNMENT BRASIL