Probit regression model for spatially dependent discrete choice data: a simulation study

Ricardo Tadashi Takeyama
tadashi@ime.unicamp.br

Emanuel Pimentel Barbosa
emanuel@ime.unicamp.br

Universidade Estadual de Campinas - UNICAMP
Instituto de Matemática, Estatística e Computação Científica - IMECC

July 2004

Abstract

This paper is about spatial econometrics in the context of discrete choice models for area data. It is presented here a simulation study of a (very useful) Bayesian econometric model for the probit regression in binary responses with spatially structured random effects in the latent variable.

Because of the large number of parameters to be estimated by stochastic simulation and the model complexity itself, in some cases it is not certain the convergence in the Gibbs sampling procedure. With this respect, it is proposed in this paper some modifications in the MCMC procedure in order to improve convergence. Also a couple of particular cases of this spatial model, which are more known in the literature, are included in the analysis as reference for comparison.

The simulation study considers the spatial structure of the geographical (administrative) region of Campinas, SP, Brazil, composed by 90 counties. The model is simulated according to different settings, with three distinct sample sizes, varying the number of agents in each region, where are generated 150 samples at each area. Also, it is considered six different values for the spatial parameter. In all settings, it is compared the estimation performance of the proposed procedure with the references in the econometric literature and the results are discussed. Also, a short illustration of the main procedures considered is presented involving real data about the 2002 Brazilian presidential election.

Keywords: Spatial econometrics, discrete choice model, random effects, probit regression, Gibbs sampling, simulation.

1 Introduction

The aim of this paper is to study econometric regression models with binary responses for cross-sectional data with spatial dependence, useful in problems of discrete choice, where the parameters estimation process may present some difficulties.

Econometric models for endogenous continuous variables with spatial dependence are largely spread out, as in (ANSELIN, 1988, 2002),(LESAGE, 1997), (KELEJIAN; PRUCHA, 1999), and others. However, discrete models, particulary probit regression models for area data and related tools, which is the subject of this article, although of recognized practical importance, it has received less attention in the literature. It happens partially because of the aditional complexity due to the spatial dependence structure incorporated in the model (FLEMING, 2002).

Some estimation techniques have been proposed for discrete models with spatial structure as (MCMILLEN, 1992) using the EM algorithm for a spatial probit model, (PINKSE; SLADE, 1998) used the Generalized Method of Moments (GMM) for a similar probit model. Other examples of discrete choice models in a spatial context may be found in the econometric literature, as in (LESAGE, 2000) or (HOLLOWAY; SHANKAR; RAHMAN, 2002), both considering the Bayesian approach.

An important reference for the present paper is (SMITH; LESAGE, 2003) where it is presented a Bayesian formulation for the spatial dependence through a random effect similar to the one introduced by (BESAG; YORK; MOLLIE, 1991) and also used in geostatistics (DIGGLE; TAWN; MOYEED, 1998). In this model it is assumed that the agents or individuals are grouped in regions where agents inside the regions are homogeneous suggesting that the spatial dependence occur only among regions.

This model and its corresponding Bayesian estimation method based on MCMC is studied in the present paper and an improvement in the estimation process is proposed in order to get better efficiency in the estimation algorithm and to avoid eventual convergence difficulties. Both implementation procedures (the original and the alternative one proposed here) are compared in a simulative study as well as some other relate models of a particular or simpler type.

It is considered here a sampling plan for the Gibbs sampler different from the one used by (SMITH; LESAGE, 2003), where we fix the spatial dependence parameter for a certain period in the Markov chain. Also in the Metropolis-in-Gibbs procedure to sample the posterior distribution of the spatial coefficient, instead of considering random walk chains as in (LESAGE, 2000; SMITH; LESAGE, 2003) and (HOLLOWAY; SHANKAR; RAHMAN, 2002), we propose the use of independent chains. With these proposed modification, it is obtained better results in the estimates of the quantities of interest such as regression and spatial coefficients.

This article is organized as follows. In Section 2, the spatially structured random effect regression model is presented as a discrete choice model. In Section 3, the Bayesian hierarchical model is defined as well as the corresponding Gibbs sampler for its estimation. In the Section 4 it is presented a simulation study where it is assessed how the estimation process performs in different settings with the advantages of the proposed procedure being stressed. Finally, it is presented in the Section 5 a small numerical example with real data about the

Brazilian presidential election of 2002 for a certain geographical region and a final discussion about the results obtained is presented in Section 6, followed by the bibliographical references.

2 Discrete choice model

It is supposed the existence of data about the choices of a set of agents or individuals distributed in m areas in a certain geographical region. In particular, it is considered that only two mutually exclusive choices are relevant and they are referred as 0 and 1. The observed choice for each individual $k = 1, 2, ..., n_i$ in the i-th area $(i = 1, 2, ..., m)$ is the considered as realization of a random variable Y_{ik} where,

$$Y_{ik} = \begin{cases} 1 & \text{if individual } k \text{ of region } i \text{ chooses } 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

It is supposed that the choices are based on a random utility function (SMITH; LESAGE, 2003; GREENE, 2003) where the k-th utility, for each one of the two alternatives, have the form,

$$U_{ik0} = \gamma^t w_{ik0} + \alpha_0^t s_{ik} + \theta_{i0} + \varepsilon_{ik0} \qquad U_{ik1} = \gamma^t w_{ik1} + \alpha_1^t s_{ik} + \theta_{i1} + \varepsilon_{ik1}$$

, where w_{ika} is a vector of dimension of observed attributes for the alternative $a = 0, 1$ and s_{ik} is a s-dimensional vector of observed attributes relative to individual k. The term $\theta_{ia} + \varepsilon_{ika}$, for $a = 0$ or 1, represents a contribution to the utility of all other attributes non-observed for region i, individual k and alternative a. The non-observable regional effect θ_{ia} represents the utility relative to the choice a common to all individuals in region i, and the individual effect, ε_{ika} represents all other non-observable components relative to individual k. Taking utility differences for the two possible choices for individual k, we have.

$$Y_{ik}^* = U_{ik1} - U_{ik0} = x_{ik}^t \beta + \theta_i + \varepsilon_{ik}, \tag{2.2}$$

where $\beta = (\gamma^t, \alpha_0^t - \alpha_1^t)^t$ is the vector of parameters, $x_{ik} = [(w_{ik1} - w_{ik0})^t, s_{ik}^t]^t$ is the vector of attributes, $\theta_i = (\theta_{i1} - \theta_{i0})$ is the regional effect, and $\varepsilon_{ik} = (\varepsilon_{ik1} - \varepsilon_{ik0})$ are the individual effects, which are independent of the regional effect. As consequence,$P[Y_{ik} = 1] = [Y_{ik}^* > 0]$.

For the non-observable quantities in the model it is assumed that all dependence among the utility difference for the agents in different regions are captured by the dependence among the regional effects $(\theta_i, i = 1, ..., n)$. In particular, the aspects that are common to individuals in a given region i can be similar to individuals in neighboring regions. This is considered assuming that the interaction vector θ presents a spatial autoregressive structure, given by

$$\theta_i = \rho \sum_{j=1}^{m} w_{ij} \theta_j + u_i, \qquad i = 1, ..., m, \tag{2.3}$$

where the w_{ij} elements are measures of spatial proximity between the regions i and j. The u_i error are considered independent and identically distributed (iid) normal random variable with zero mean and precision ϕ. Writing in matrix form, where $\theta = (\theta_i : i = 1, ..., m)$ and $u = (u_i : i = 1, ..., m)$,we have

$$\theta = \rho W \theta + u, \qquad u \sim N(0, \sigma^2 I_m),$$

where W is the spatial proximity matrix with $w_{ii} = 0$ and I_m is the identity matrix of dimension m. Assuming that $B_\rho = I_m - \rho W$ is non-singular, we have that $\theta = B_\rho^{-1} u$ and

$$\theta | \rho, \sigma^2 \sim N(0, \sigma^2 [B_\rho^t B_\rho]^{-1}). \tag{2.4}$$

The individual effects, ε_{ik} are assumed normal with zero mean and precision v_i independent of the regional effects θ_i. The precisions v_i, indexed by region i, imply a intra-region homocedasticity and inter-region heterocedasticity . If we denote the vector of individual effects in the region i by $\varepsilon_i = (\varepsilon_{ik} : k = 1, ..., n_i)$, and then $\varepsilon_i \sim N(0, v_i^{-1} I_{n_i})$. Or, in global terms, $\varepsilon = (\varepsilon_i^t : k = 1, ..., n_i)^t$, then $\varepsilon \sim N(0, V)$ where V has the form, $V = diag\{v_1^{-1} I_{n_1}, ..., v_m^{-1} I_{n_m}\}$

Now, the likelihood is given by

$$L(\beta, y^* | y) = \prod_{k=1}^{m} \prod_{i=1}^{n_k} \{\mathbf{1}_{(y_{ik}^* > 0)} \mathbf{1}_{(y_{ik} = 1)} + \mathbf{1}_{(y_{ik}^* \leq 0)} \mathbf{1}_{(y_{ik} = 0)}\} \times \phi(y_{ik}^*; x_{ik}^t \beta, t_i). \tag{2.5}$$

We can also express (2.2) in matrix form where $Y_i^* = (Y_{ik}^* : k = 1, ..., n_i)^t$ and $X_i = (x_{ik} : k = 1, ..., n_i)^t$, and difference in utility for region i is then given by

$$Y_i^* = X_i \beta + \theta_i \mathbf{1}_i + \varepsilon_i, \qquad i = 1, ..., m, \tag{2.6}$$

where $\mathbf{1}_i$ is a $n_i \times 1$ unitary vector. Defining $n = \sum_i^m n_i$, $Y^* = (Y_i^* : i = 1, ..., m)$ and $X = (X_i^t : i = 1, ..., m)^t$ the regional equation (2.6) can be reduced in the form

$$Y^* = X\beta + \Delta\theta + \varepsilon, \tag{2.7}$$

where $\Delta = diag\{\mathbf{1}_1, \ldots, \mathbf{1}_m\}$

When $\rho = 0$, the model just presented became a probit model with non-structured random effect (MCCULLOCH; SEARLE, 2001), where

$$\theta_i = u_i,$$

and u_i follows an iid normal distribution with zero mean and precision ϕ. Consequently, the model equation (2.7) became,

$$Y^* = X\beta + \Delta U + \varepsilon,$$

where $U = (u_1, \ldots, u_n)^t$.

3 Bayesian Hierarchical Model

The model presented in (2.1) with a latent variable given by (2.7) can be formulated as a hierarchical model with the following structure,

(I) Conditional on $\beta, \rho, \phi, \theta$, and $V = diag\{v_1^{-1} I_{n_1}, ..., v_m^{-1} I_{n_m}\}$,the latent variables Y_{ik}^* follow normal distributions, $N(x_{ik}^t \beta + \theta_i, v_i^{-1})$ for $i = 1, \ldots, m; k = 1, \ldots, n_i$.

(II) Conditional on ρ and ϕ, the spatial effects vector θ follows a multivariate normal distribution $N(0, \phi^{-1}[B_\rho^t B_\rho]^{-1})$.

(III) Conditional on matrix V, the error vector ε has a multivariate normal distribution $N(0, V)$.

(IV) Conditional conjugate prior distribution for β, ρ, ϕ and $V = diag\{v_1^{-1} I_{n_1}, ..., v_m^{-1} I_{n_m}\}$

$$\beta \sim N(\nu_0, \Sigma_0) \qquad \rho \sim Beta(a, b)$$
$$v_i \overset{iid}{\sim} Gamma\left(\frac{r}{2}, \frac{r}{2}\right) \qquad \phi \sim Gamma(\alpha, \eta)$$

where α, η and r are the gamma precision hiperparameters, ν_0 and Σ_0 are the normal regression coefficients hiperparameters and a, b are the *Beta* spatial coefficient hiperparameters ($a = b = 1$ corresponds to a uniform non-informative prior distribution). These prior distributions considered above are very close to the ones used in (SMITH; LESAGE, 2003), except for a detail in the specification of ρ. With the more flexible Beta distribution, we are considering the restriction $0 < \rho < 1$, differently from (SMITH; LESAGE, 2003), since they consider a uniform distribution in the interval defined by the inverse of largest and smallest eigenvalue of the matrix W.

3.1 Gibbs sampler

The Bayesian estimation of the above hierarquical model is implemented via the Gibbs sampler and related algorithms through the random sampling of the following set of (complete) conditional distributions:

(i) $\beta | \rho, \phi, \theta, V^*, y^*, y \sim N(\hat{\beta}, (X^t V^* X + \Sigma_0^{-1})^{-1})$ where $\hat{\beta} = [X^t V^* X + \Sigma_0^{-1}]^{-1} [X^t V^* (y^* - \Delta\theta) + \Sigma_0^{-1} \nu_0]$.

(ii) $\theta | \beta, \rho, \phi, V^*, y^*, y \sim N(\hat{\theta}, M^{-1})$, where $M = \Delta^t V^{-1} \Delta + \phi B_\rho^t B_\rho$ and $\hat{\theta} = M^{-1} \Delta^t V^{-1} (y^* - \Delta\theta)$

(iii) $\rho | \beta, \phi, \theta, V^*, y^*, y \propto \phi^{\frac{m}{2}} |B_\rho| exp\left\{-\frac{\phi}{2} \theta^t (B_\rho^t B_\rho) \theta\right\} \rho^a (1-\rho)^{b-1}$

(iv) $\phi | \beta, \rho, \theta, V^*, y^*, y \sim Gamma(\alpha', \eta')$, where $\alpha' = \frac{m+2\alpha}{2}$ and $\eta' = \frac{\theta^t (B_\rho^t B_\rho) \theta + 2}{2\eta}$

(v) $v_i^* | \beta, \rho, \phi, \theta, v_{-i}^*, y^*, y \sim Gamma(g, h)$, where $g = \frac{n_i + r}{2}$ and $h = \frac{\sum_k^{n_i} e_{ik} + r}{2}$, with $e_{ik} = y_{ik}^* - \theta_i - x_{ik}^t \beta$.

(vi) $Y_{ik}^* | \beta, \rho, \phi, \theta, V^*, y \sim \begin{cases} N(x_{i1}^t \beta + \theta_i, v_i^*) & \text{left truncated in 0 if } y_i = 1 \\ N(x_{i1}^t \beta + \theta_i, v_i^*) & \text{right truncated in 0 if } y_i = 0 \end{cases}$

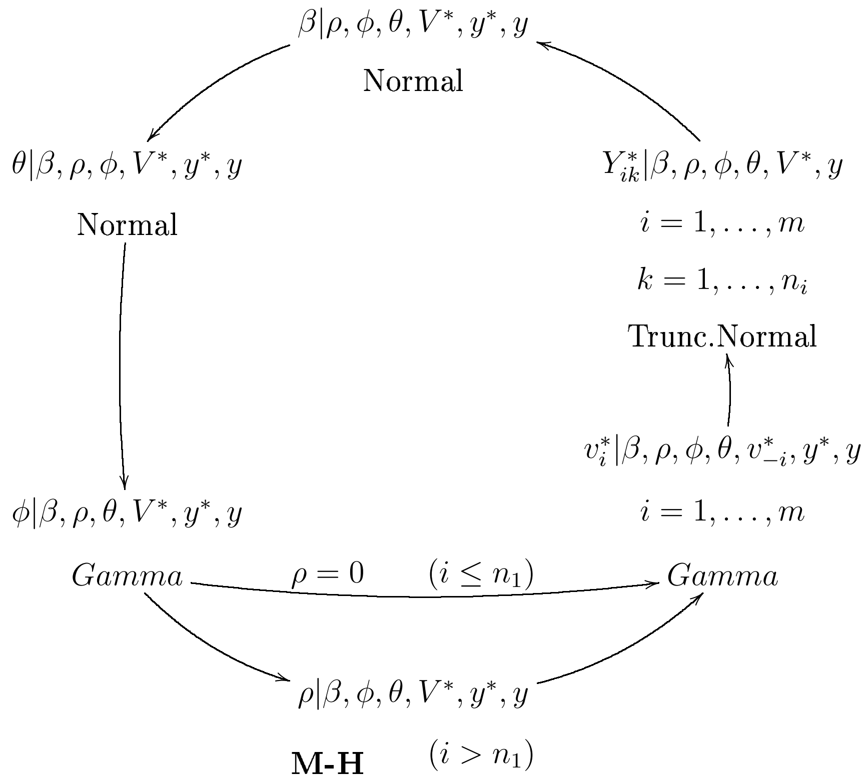

Figure 1: Plan 2 - Gibbs sampling for MCMC analysis

(LESAGE, 1997, 2000) notes that the conditional distribution for ρ has no known form and suggests a Metropolis-Hastings step (METROPOLIS, 1953; HASTINGS, 1970) inside the Gibbs procedure. Some suggestions for the proposal of the transition kernel's chain in the MCMC are the normal distribution and the t distribution with 3 degree of freedom (LESAGE, 1997). Here in this paper, the proposed transition is formulated independently of the actual position ρ^j of the chain, and this proposal is generated by normal distribution with mean μ and variance τ^2. For the mean μ, it is used here the numerical maximum of the conditional log-likelihood function for ρ. (CHIB; NARDARI; N.SHEPARD, 2002)suggest also that the variance τ^2 could be calculated as minus the numerical second derivative of this conditional log-likelihood. However, it is used here the first samples for getting the posterior variance estimate (CARLIN; LOUIS, 2000).

In this paper it is considered a modification in the Gibbs sampler implementation, which we refer here as *Plan 2*, where it is fixed the spatial coefficient with a reference value $\rho=0$ for a certain number n_1 of iteration in the Gibbs sampler and only after that number of iteration we include this parameter in the sampling and updating process (Figure 1).

4 Simulation Study

The spatially structured random effects probit (SSREP) model simulation is made using the neighborhood map given by the administrative region of Campinas, Brazil, composed by 90 counties, shown at figure 2.

The data are generated through a SSREP model with two covariates, and latent structure for the k-th individual in i-th region given by

$$\begin{aligned} Y_{ik}^* &= \beta_1 x_{1ik} + \beta_2 x_{2ik} + \theta_i + \varepsilon_{ik} \\ \theta_i &= \rho \sum_j w_{ij}\theta_j + u_i, \end{aligned}$$

where the generation values are $\beta_1 = 3,14$, $\beta_2 = -1,62$, $\rho = 0,70$ and $w_{ij} = \frac{w_{ij}^*}{\sum_j w_{ij}^*}$ with $w_{ij}^* = 1$ if counties are contiguous and zero otherwise. For simplicity, the error terms ε_{ik} and u_i are generated as standard normals. The covariate x_{1ik} is uniform variable in [-5,5] and x_{2ik} is generated using a normal variable with zero mean and variance 3. Finally, by discretization rule, it is generated the response data as

$$Y_{ik} = \begin{cases} 1 & \text{if } Y_{ik}^* > 0, \\ 0 & \text{if } Y_{ik}^* \leq 0. \end{cases} \tag{4.8}$$

As mentioned earlier in this paper, the simulation study is formed by two parts or experiments. In the first one , it is studied the number of individuals or observation per area. In this part, it is considered three situations. In case 1, which is the more difficult to estimate, there is only one observation or individual per area. This case is very common for area data, where information is aggregated in order to reflect the average behavior of a given area. In the cases 2 and 3 it is simulated a larger number of individuals per area, $n_i = 5$ and $n_i = 15$, respectively.

In the second part of the simulation experiment are considered different values for the spatial dependence parameter ρ and it is studied the model estimation performance (difference in mean square error of point estimates relative to generation values).

4.1 Varying the number n_i of observations per area (ρ fixed)

In this section it is studied the model estimation performance considering different numbers of agents or individuals per area. We consider the following cases: $n_i = 1$, 5 and 15.

The spatially structured random effect model is fitted to the simulated data in two versions: the original version presented by (SMITH; LESAGE, 2003) (computer routine *semip_ g*) and the alternative one proposed in this paper which we call Plan 2. In order to implement this procedure, we consider the following prior settings,

(a) gamma prior for ϕ: $\alpha = 0,001$ and $\eta = 1000$

(b) gamma prior for v: $r = 100$

(c) Beta prior for ρ: $a = 3$ and $b = 1$.

(d) Normal prior for β: $\mu_0 = (0,0)^t$ and $\Sigma_0 = 10^6 I_2$

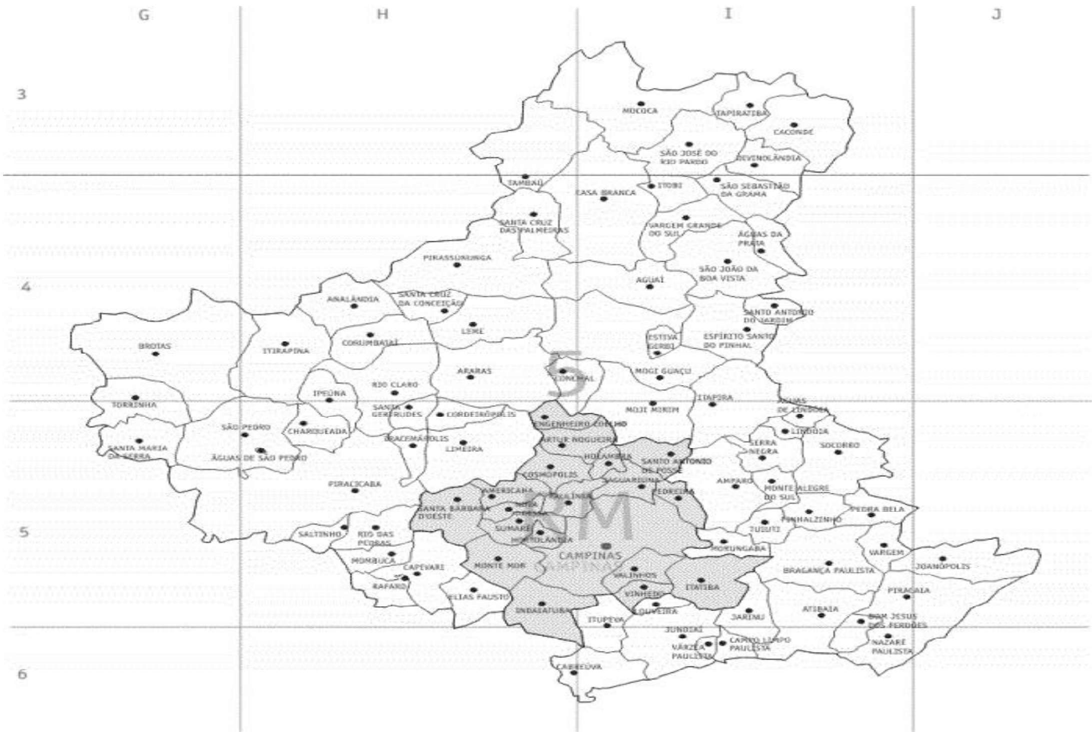

Figure 2: Administrative (geographical) Region of Campinas

It is also considered a burn-in of 2000 iterations and the following 2000 iteration are sampled. In particular, at Plan 2, it is fixed $\rho = 0$ in the first 2000 iterations.

In order to obtain a point estimate of each of the components (β, ρ, ϕ, v) it is necessary to define a resume measure for the marginal posterior distribution such as mean, median or mode. When the posterior distribution is symmetric, the mean and the median will be equal, and for unimodal symmetric posterior, all three measures will be the same. For non-symmetric posterior distributions however, the choice is not so clear. Although the median is eventually preferable since it is an intermediate measure between the mode (just considers the value corresponding to the maximum of the density) and the mean (it can eventually give large weight to extreme values) (CARLIN; LOUIS, 2000). (SMITH; LESAGE, 2003) adopt the sampled posterior mean as a point estimate for the parameters. In order to study better this model and its estimation procedure, there were generated (simulated) 150 samples of data and, for each one, it was fitted the hierarchical model using both proposed Plan 2 (implemented in MATLAB) and the former procedure considered by (SMITH; LESAGE, 2003) through the *semip_ g* routine of the Econometric Toolbox, in MATLAB.

4.1.1 Case 1: $n_i = 1$

In a simulation study similar to the one presented here, (SMITH; LESAGE, 2003), say that the hierarchical model estimation procedure via MCMC estimates the regression coefficients very well but it underestimate the spatial coefficient ρ. These authors however do not explain why or suggest alternatives to approach this problem.

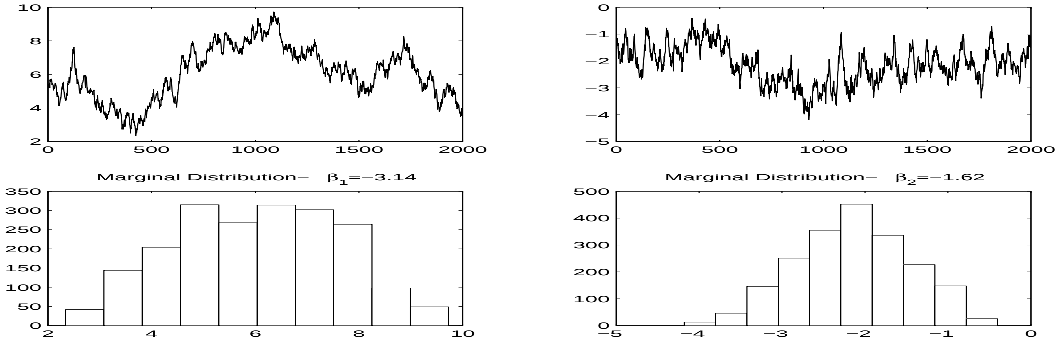

Figure 3: MCMC chains for β - *semip_ g*

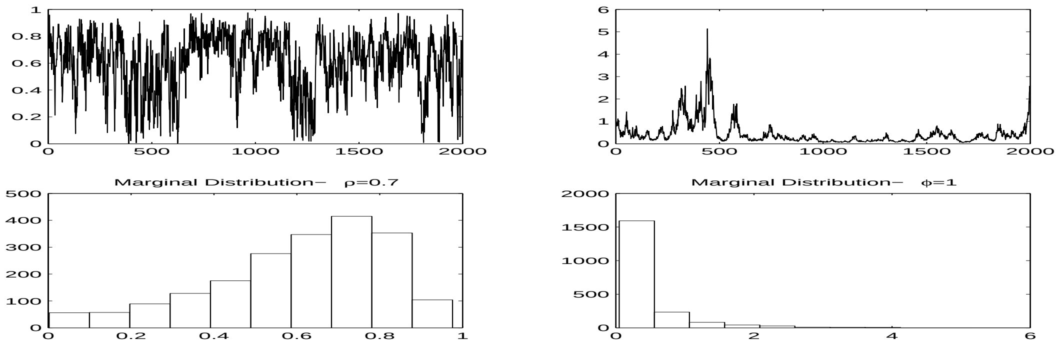

Figure 4: MCMC chains for ρ and ϕ - *semip_ g*

The posterior distribution of the spatial coefficient ρ, in Plan 2, shows some non-symmetry (Figure 4 and Figure 6), but the posterior mode is very close to the parameter generation value($\rho = 0,70$). When (SMITH; LESAGE, 2003) conclude that there is underestimation of ρ, in fact they are considering the mean as a resume for a non-symmetric distribution instead of considering the other alternative measures such as the median and mode.

The results from the 150 generated samples, given at Table 4.1, show that (considering the 3 main point estimators: posterior mean, median or mode) the median is the measure of better performance considering the set of all 3 parameters of interest, with the smallest mean square error - MSE (taking the generation values). Also, it is clear from this table that the point estimates from (SMITH; LESAGE, 2003) are very bad because their estimates are very far from the generated values taken as references ($\beta_1 = 3,14$, $\beta_2 = -1,62$ and $\rho = 0,70$). On the other hand, the results obtained from the proposed procedure (Plan 2) give point estimates much closer to the reference values and with smaller MSE values.

4.1.2 Case 2: $n_i = 5$ and Case 3: $n_i = 15$

In these cases, when there are more agents for each area, the sample has more information, particularly about the regional effects θ_i, and consequently, more information about the spatial dependence parameter ρ. Consequently, we have more information about the spatial

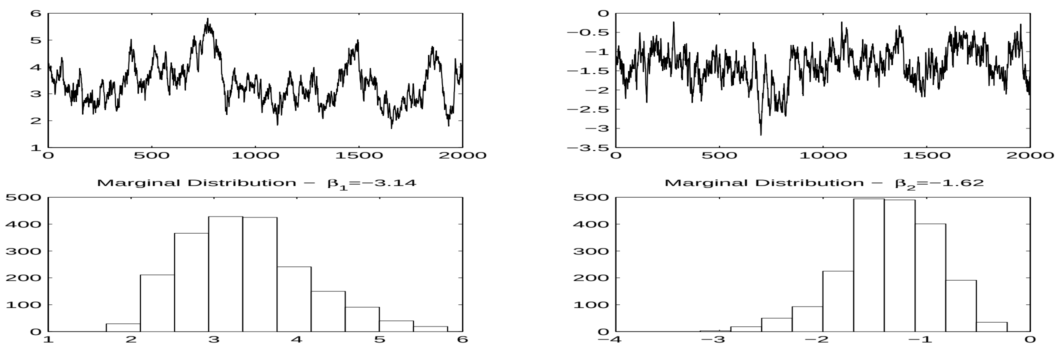

Figure 5: MCMC chains for β - Plan 2

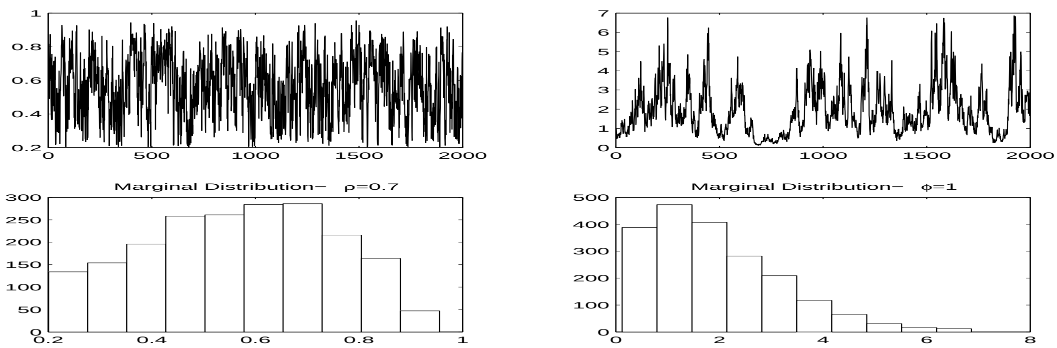

Figure 6: MCMC chains for ρ and ϕ - Plan 2

parameter, ρ. The posterior distribution of ρ in case 2 (Figure 7) is more well behaved than in case 1.

In this case, the median was the better point estimate for the parameter of interest according to the mean square error criterion. For the β regression coefficients, the estimates by the three different point estimators (mean, median and mode) are reasonably close because of the posterior distribution symmetry for these parameters. The spatial dependence parameter ρ however presents larger differences between the different point estimates as a consequence of its non-symmetric posterior distribution (Figure 7). It is worth to notice that with the *semip_g* routine, as the sample size n_i increases, their estimates get closer to the

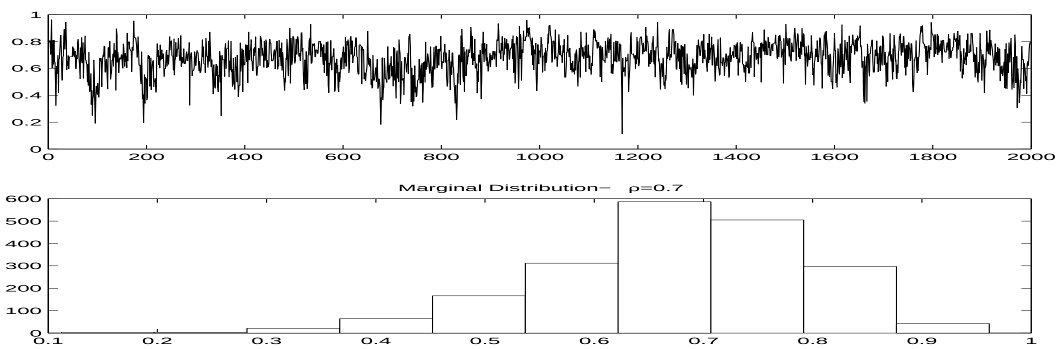

Figure 7: Posterior distribution for ρ - $n_i = 15$

respective generation values. These estimates however are not so good (in relation to the MSE measure) as the ones obtained through the proposed Plan 2 procedure.

In general, when the number of observation per area increases, the three posterior resume measures get closer, as can be seem from Table 4.1. The non-symmetry of the spatial dependence coefficient ρ tend to decrease with the increase of information about regional effects θ_i. In particular, for the case 3, any of the three measures could be used since all of them are quite close. In this sense we suggest the use of the posterior median as point estimate for the quantities of interest.

4.2 Varying the ρ values (n_i fixed)

In this section it is studied the model estimation process considering different settings for the spatial dependence parameter. This is made simulating the spatially structured random effects probit model (SSREP) in (4.8) using small, moderate and large values for the ρ parameter. It is introduced in this analysis not only the implementation of the hierarchical model but also the estimation results from the non-structured random effect probit model (NSREP) and the standard probit model taken as reference. This simpler random effects model, even being not spatially structured, it introduces correlation or dependence between the observations through common effects (MCCULLOCH; SEARLE, 2001). Therefore, we expect that in the presence of small or moderate dependence, this non spatial model could be a simple alternative to deal with this problem, since without the estimation of ρ, there is no Metropolis step in the Gibbs sampler.

It is considered in this study six increasing values for the ρ parameter: 0, 0,05, 0,10, 0,20, 0,50 and 0,70. One hundred simulated samples were generated under the spatial structure shown at figure 2, with $n_i = 5$ observations per geographical area and the same values for the regressions coefficients and other parameters considered in the previous section. Also, it was considered vague prior distributions in the model implementation.

The results are presented at Tables 4.2 and 4.3. For the samples generated with small spatial dependence (ρ between 0 and 0,20) there is little difference between the structured and the non-structured model. As ρ increases, as expected, the estimates obtained by the spatial model tend to improve in relation to the simpler models, as shown in terms of mean square error at Table 4.2. For instance, when the spatial dependence is high ($\rho = 0,70$) the estimates from the non-structured model and from the standard probit model, differ very much from the corresponding generation values, as measured through the MSE, and therefore, it is preferable in this case to use the hierarchical spatial model.

Table 4.1: Simulation Results from the 150 samples

		Plan 2			(SMITH; LESAGE, 2003)		
		Estimates	Std.Dev	MSE	Estimates	Std.Dev	MSE
Case 1 - $n_i = 1$							
β_1	Mean	3,438	0,775	0,650	7,44	4,980	41,650
	Median	3,062	0,6249	0,3706	7,292	5,177	42,250
	Mode	2,724	0,569	0,476	6,857	5,562	42,680
β_2	Mean	-1,758	0,536	0,290	- 3,771	2,547	10,720
	Median	-1,580	0,470	0,207	- 3,728	2,772	11,660
	Mode	-1,401	0,480	0,259	- 3,564	3,144	13,04
ρ	Mean	0,624	0,066	0,043	0,501	0,439	0,186
	Median	0,681	0,077	0,034	0,530	0,536	0,268
	Mode	0,695	0,080	0,034	0,642	0,649	0,268
Case 2 - $n_i = 5$							
β_1	Mean	3,346	0,761	0,602	3,688	0,972	1,214
	Median	3,327	0,834	0,707	3,691	1,088	1,447
	Mode	3,286	0,882	0,773	3,714	1,140	1,585
β_2	Mean	- 1,739	0,411	0,179	- 1,917	0,551	0,388
	Median	- 1,728	0,431	0,193	- 1,923	0,604	0,450
	Mode	- 1,712	0,494	0,247	- 1,943	0,632	0,497
ρ	Mean	0,622	0,215	0,051	0,521	0,190	0,066
	Median	0,646	0,219	0,049	0,528	0,201	0,068
	Mode	0,684	0,230	0,049	0,539	0,229	0,068
Case 3 - $n_i = 15$							
β_1	Mean	3,234	0,279	0,083	2,958	0,292	0,114
	Median	3,234	0,281	0,084	2,957	0,294	0,115
	Mode	3,213	0,301	0,091	2,958	0,310	0,124
β_2	Mean	- 1,650	0,159	0,025	- 1,507	0,154	0,033
	Median	- 1,650	0,158	0,025	- 1,505	0,153	0,033
	Mode	- 1,640	0,154	0,023	- 1,504	0,152	0,033
ρ	Mean	0,673	0,109	0,012	0,650	0,137	0,020
	Median	0,681	0,108	0,011	0,656	0,137	0,019
	Mode	0,700	0,109	0,011	0,667	0,138	0,019

Table 4.2: Point estimates with MSE of ρ for 100 simulated samples

Generation Value		$\beta_1 = 3,14$	$\beta_2 = -1,62$	ρ
A. Estimates for $\rho = 0$				
Standard probit		2.251(0.844)	- 1.149(0.235)	-
NSREP	Mean	3.362(0.445)	- 1.716(0.135)	-
	Median	3.355(0.461)	- 1.708(0.129)	-
	Mode	3.357(0.578)	- 1.689(0.119)	-
SSREP - Plan 2	Mean	3.374(0.454)	- 1.728(0.161)	0.173(0.289)
	Median	3.347(0.459)	- 1.719(0.169)	0.142(0.325)
	Mode	3.290(0.488)	- 1.697(0.200)	0.085(0.325)
B. Estimates for $\rho = 0,05$				
Standard probit		2.255(0.833)	- 1.144(0.236)	-
NSREP	Mean	3.399(0.478)	- 1.716(0.122)	-
	Median	3.370(0.436)	- 1.703(0.114)	-
	Mode	3.283(0.317)	- 1.662(0.097)	-
SSREP - Plan 2	Mean	3.387(0.408)	- 1.713(0.133)	0.202(0.2603)
	Median	3.344(0.391)	- 1.693(0.134)	0.172(0.2960)
	Mode	3.278(0.408)	- 1.663(0.156)	0.106(0.2960)
C. Estimates for $\rho = 0,10$				
Standard probit		2.286(0.787)	- 1.186(0.207)	-
NSREP	Mean	3.444(0.466)	- 1.785(0.161)	-
	Median	3.424(0.449)	- 1.775(0.160)	-
	Mode	3.397(0.486)	- 1.773(0.185)	-
SSREP - Plan 2	Mean	3.427(0.474)	- 1.776(0.171)	0.192(0.268)
	Median	3.393(0.452)	- 1.761(0.164)	0.161(0.305)
	Mode	3.352(0.528)	- 1.743(0.184)	0.097(0.305)

Table 4.3: Point estimates with MSE of ρ for 100 simulated samples

Generation Value		$\beta_1 = 3,14$	$\beta_2 = -1,62$	ρ
		D. Estimates for $\rho = 0,20$		
Standard probit		2,231 (0,856)	-1,125(0,246)	-
	Mean	3,324 (0,174)	-1,694 (0,059)	-
NSREP	Median	3,317 (0,163)	-1,686 (0,053)	-
	Mode	3,327 (0,201)	-1,663 (0,055)	-
	Mean	3,243 (0,116)	-1,650 (0,048)	0,259 (0,204)
SSREP - Plan 2	Median	3,210 (0,101)	-1,633 (0,046)	0,242 (0,225)
	Mode	3,218 (0,229)	-1,626 (0,058)	0,166 (0,225)
		E. Estimates for $\rho = 0,5$		
Standard probit		2,071 (1,182)	-1,093 (0,280)	-
	Mean	3,351 (0,131)	-1,749 (0,049)	-
NSREP	Median	3,284 (0,090)	-1,737 (0,040)	-
	Mode	3,088 (0,233)	-1,693 (0,046)	-
	Mean	3,184 (0,036)	-1,645 (0,013)	0,477 (0,053)
SSREP - Plan 2	Median	3,208 (0,053)	-1,653 (0,013)	0,498 (0,045)
	Mode	3,383 (0,162)	-1,725 (0,0511)	0,541 (0,031)
		F. Estimates for $\rho = 0,7$		
Standard probit		2,010 (1,307)	-1,118 (0,259)	-
	Mean	3,677 (0,648)	-2,039 (0,299)	-
NSREP	Median	3,694 (0,669)	-2,053 (0,318)	-
	Mode	3,665 (0,795)	-2,092 (0,391)	-
	Mean	3,156 (0,115)	-1,730 (0,078)	0,687 (0,013)
SSREP - Plan 2	Median	3,138 (0,115)	-1,716 (0,074)	0,704 (0,011)
	Mode	3,237 (0,244)	-1,698 (0,064)	0,705 (0,011)

5 A Short example with real data

In order to give one more confirmation or illustration of the usefulness of the models and procedures studied and proposed in this paper, it is presented in this section a short example of application of these tools with real data.

The data set considered here is about the Brazilian presidential election of 2002. In fact, the data are the results of the second turn of the election in the Administrative Region of Campinas, Brazil (Figure 2), composed by 90 counties. The dependent variable is defined as $Y_i = 1$ if the main candidate Luíz Inácio Lula da Silva (the most voted in the first turn election) has obtained the majority of votes in county i, and $Y_i = 0$ otherwise. As a final result, Lula won in 55 of the 90 counties of the region.

As a covariate or regression variable it was considered one component of Human Development Index related to education, calculated in the year 2000 for each of the counties in the region. This variable was chosen based on empirical grounds since it it the component of the HDI with better association with the response variable and also because it is a prediction with known value previously to the electin.

5.1 Models estimation

As an illustration of the spatial econometrics tools for discrete choice, it is fitted the following models with the data described above.

(A) Standard probit model;

(B) Random effects probit model;

(C) Spatial random effect probit model.

Although these models could be estimated by other methods sucha as maximum likelihood or generalized method of moments (FLEMING, 2002; TAKEYAMA; BARBOSA, 2004), it is considered here the same Bayesian approach developed previously in this paper for the implementation of all the three models, since the first two of them are particular cases of the last (spatial random effects model formulated as a Bayesian hierarchical model).

The hiperparameter values considered for the vague prior distributions are

(a) $\mu_0 = 0$ and $\Sigma_0 = 10^6$ for β in the models A, B and C.

(b) $\alpha = 0.001$ and $\eta = 1000$ for ϕ and $r = 7$ for v in the models B and C.

(c) $a = 1$ and $b = 1$ for ρ in the model C.

The number of iteration considered here is different for each model. The standard probit model model (A) has a quicker convergence since it has less parameters; it was considered a burn-in of 500 iterations followed by sampling 2000 posteriors. For model B it was considered the same procedure. For model C, it was fixed $\rho = 0$ for the first 500s iteration followed by a burn-in of 2000 iterations, and the sampling of the next 2000 iterations.

Table 5.4: Estimation Results

A. Standard probit

	2,5%	25%	median	mode	75%	97,5%
β_1	12,959	20,271	24,009	24,287	27,894	35,829

B. Random effects probit

	2,5%	25%	median	mode	75%	97,5%
β_1	13,432	21,622	27,135	27,863	33,135	47,228
ϕ	0,724	2,591	4,826	3,787	9,525	24,910

C1. Spatial random effect probit - Plan 2

	2,5%	25%	median	mode	75%	97,5%
β_1	11,609	20,055	25,703	23,144	33,292	48,165
ρ	0,044	0,393	0,585	0,627	0,713	0,872
ϕ	0,300	1,199	4,458	3,702	10,899	30,663

C2. Spatial random effect probit - semip_g

	2,5%	25%	median	mode	75%	97,5%
β_1	19,870	48,022	65,939	55,077	87,787	112,373
ρ	0,243	0,994	0,998	0,925	0,998	0,999
ϕ	0,214	3,465	6,673	1,853	11,065	18,864

5.2 Fitting results

The main estimation results from the four fitted models (the last two are different implementations of the same spatial model) are presented at Table 5.4.

In general terms, the models A, B and C1 present reasonably similar results in the estimation of β. However, when the results are analyzed taking the deviance as a measure of performance for each model (Table 5.5), it is clear that the introduction of the spatial random effect, considering Plan 2, improves considerably the model fitting quality. The same also occurs (with less intensity) with model B which is a particular case of C when $\rho = 0$, and presents a better fitting than the standard probit model A. In Figure 9, it is shown the sampled chains for the spatial dependence parameter ρ (considering both Plan 2 and semip_g routine), where it is visible the instability of the chain produced by the Lesage routine. It is all noted that the posterior for ρ under this routine is too much concentrated near 1, while the one given by Plan 2 is more smooth and less asymmetric.

It is also clear from Table 5.5 that the proportion of correct prediction from the spatial model with the Plan 2 procedure is the highest, followed by the one given by the simple random effects model, that confirm and illustrates once more the usefulness of the proposed tools.

Table 5.5: Goodness of fit measures

Models	Deviance	Prop
A. Standard probit	103,603	0.677
B. Random effects probit	88,773	0.744
C1. Spatial random effect probit - Plan 2	76,529	0.766
C2. Spatial random effect probit - semip_g	148,799	0.723

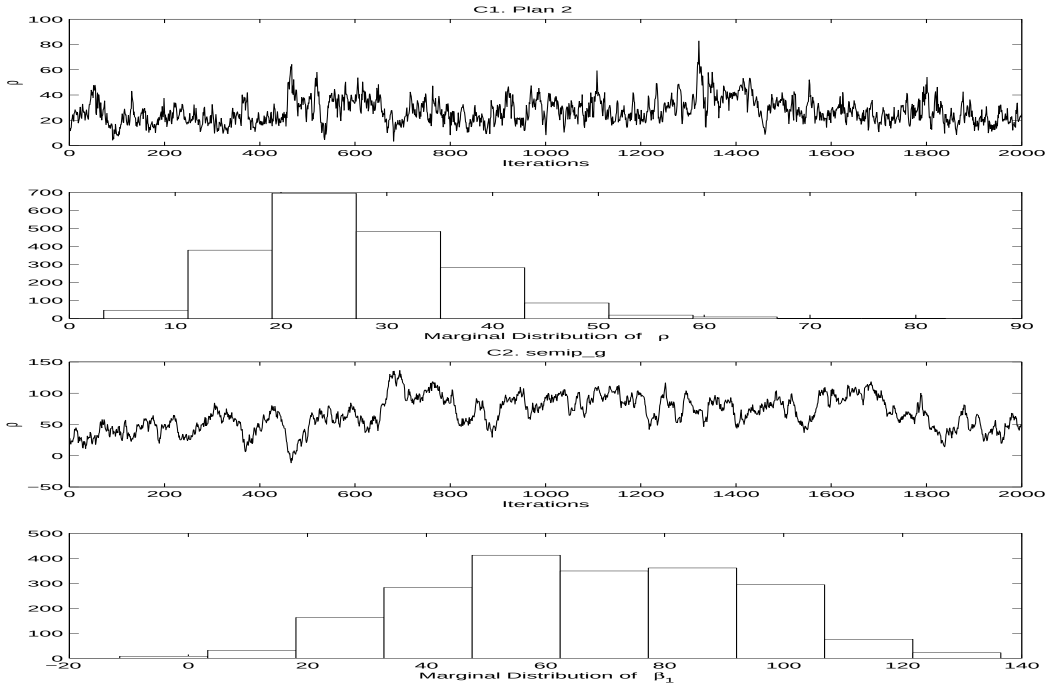

Figure 8: MCMC chains for β generated by Plan 2 and semip_g

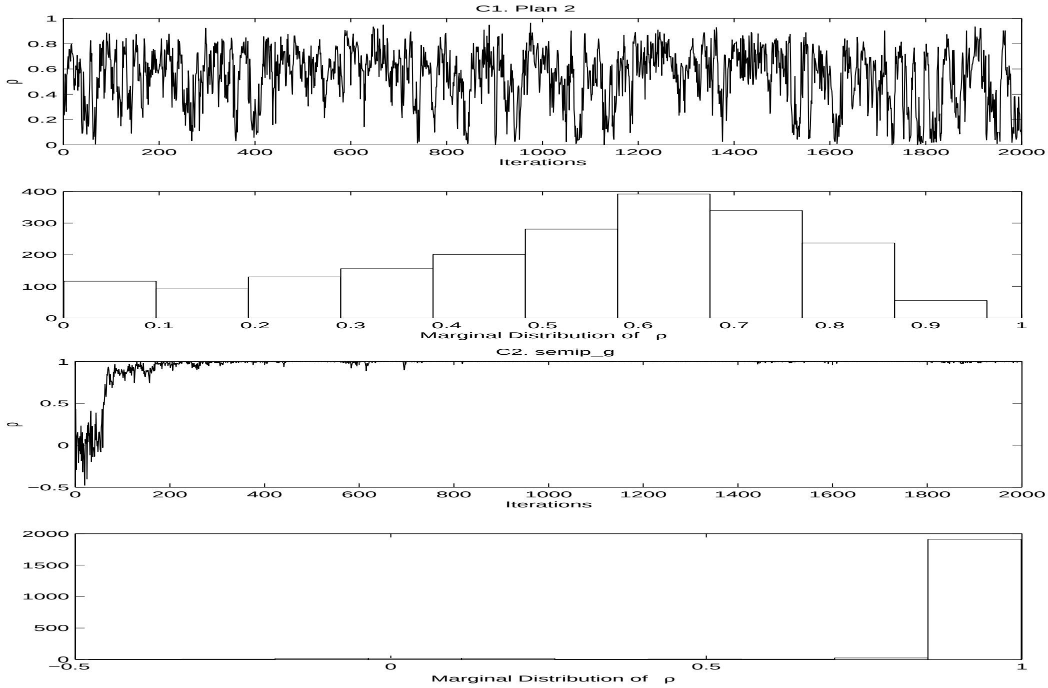

Figure 9: MCMC chains for ρ generated by Plan 2 and semip_g

6 Final Conclusion

It has been studied in this paper two possible ways of analysis and implementation of a probit regression model for binary data with spatial dependence (useful in the econometric analysis of discrete choice problems) through a simulation study. It is also included in this analysis the study of a few related models of simpler structure and easier implementation, as well as the fitting of all models to a binary choice real data set.

The main results and conclusions from the study are the following. (i) As expected but not obvious, the standard probit model is not appropriate for analysis of spatial binary data, even with small spatial dependence. Or, in other words, it is necessary to consider non-standard models in such cases in order to have quality in the data analysis. (ii) The Bayesian hierarchical model has shown to be a powerful tool for investigation of spatial dependencies in discrete choice. This is particularly the case of the spatially structured random effects model considered in the study. (iii) A particular case of this model where the random effect is non structured, which is simpler and more known in the literature, has shown to be very effective even in cases of moderate spatial dependence. One of the advantages of this model is that the estimation procedure is simpler, using Gibbs sampling (it is not necessary Metropolis in Gibbs), what shorten considerably the computational burden.

Finally, (iv) the proposed modification for the implementation of the spatial model have improved the parameter estimates in relation to others algorithm found in the econometric literature. In particular, the modification in the Gibbs sampling procedure has produced more stable chains in the MCMC with better final estimates, as presented in the simulation study and also illustrated with a real spatial data set of binary choice.

References

ANSELIN, L. *Spatial Econometrics: Methods and Models.* Dodrecht: Kluwer Academic Publishers, 1988.

ANSELIN, L. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, v. 27, n. 3, p. 247–267, 2002.

BESAG, J.; YORK, C.; MOLLIE, A. Bayesian image restoration, with two principle applications in spatial statisstics. *Annals of Institute of Statistical Mathematics*, v. 43, p. 1–59, 1991.

CARLIN, B.; LOUIS, T. *Bayes and empirical bayes methods for data analysis. 2 ed.* London: Chapman and Hall, 2000.

CHIB, S.; NARDARI, F.; N.SHEPARD. Markov chains monte carlo methods for stochastic volatility models. *J. Econometrics*, v. 108, p. 281–316, 2002.

CRESSIE, N. *Statistics for spatial data.* New York: Wiley, 1993.

DIGGLE, P.; TAWN, J.; MOYEED, R. Models based geostatistics. *JRSS serie C Applied Statistics*, n. 47, p. 299–350, 1998.

FLEMING, M. Techniques for estimating spatially dependent discrete choice models. *Advances in Spatial Econometrics*, Springer-Verlag, 2002.

GAMERMAN, D. *Monte Carlo Markov Chains: Sthocastic Simulation for Bayesian Inference.* London: Chapman Hall, 1997.

GELMAN, A. et al. *Bayesian data analyis.* London: Chapman and Hall, 1997.

GREENE, W. H. *Econometric Analysis, 5 ed.* New Jersey: Prentice Hall, 2003.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, v. 57, p. 97–109, 1970.

HOLLOWAY, G.; SHANKAR, B.; RAHMAN, S. Bayesian spatial probit estimation: a primer and an application to hyv rice adoption. *Agricultural Economics*, n. 27, p. 383–402, 2002.

KELEJIAN, H. H.; PRUCHA, I. R. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, v. 40, n. 2, p. 509–533, 1999.

LESAGE, J. Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, v. 20, p. 113–129, 1997.

LESAGE, J. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, v. 32, n. 1, p. 19–35, 2000.

MCCULLOCH, C. E.; SEARLE, S. R. *Generalized, linear and mixed models.* New York: Wiley, 2001.

MCMILLEN, D. Probit with spatial autocorrelation. *J. Regional Science*, v. 32, n. 3, p. 335–348, 1992.

METROPOLIS, N. et al. Equations of state calculations by fast computing machines. *J. Chemical Physics*, v. 21, p. 1087–1092, 1953.

PINKSE, J.; SLADE, M. E. Contracting in space: An application of spatial statistics to discrete-choice models. *J. Econometrics*, v. 85, p. 125–154, 1998.

SMITH, T. E.; LESAGE, J. P. A bayesian probit model with spatial dependencies. *working paper*, 2003.

TAKEYAMA, R.; BARBOSA, E. Modelos econométricos de regressão probito com estrutura espacial. *Relatório de Pesquisa - RP15/04 Imecc/UNICAMP*, 2004.