

# Modelos Econométricos de Regressão Probita com Estrutura Espacial

Ricardo T. Takeyama\*  
tadashi@ime.unicamp.br

Emanuel P. Barbosa  
emanuel@ime.unicamp.br

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

Março de 2004

## Resumo

São analisados dois modelos econométricos para a regressão probita linear em respostas binárias espacialmente dependentes ou correlacionadas, os quais tem aplicação em problemas de escolha discreta para dados de área. A incorporação desta correlação, via um processo latente, tem como consequência uma função de verossimilhança, envolvendo várias integrais, que exceto em casos especiais, não podem ser resolvidas analiticamente.

O primeiro modelo insere a estrutura espacial na forma de defasagens ou erros espacialmente correlacionados na variável latente. Para este modelo apresentamos uma comparação entre três métodos de estimação, a saber: máxima verossimilhança (MV) via algoritmo EM e método dos momentos generalizados (MMG) e análise Bayesiana via MCMC. Para tal foram geradas 2500 amostras.

O segundo, introduz a espacialidade ao modelo como um efeito aleatório espacialmente estruturado, na forma de um processo autoregressivo espacial. Analisamos este modelo através de um modelo hierárquico Bayesiano, sugerindo prioris adequadas, bem como, usando Metropolis em Gibbs, estudando duas possíveis distribuições para a transição. Foram feitas simulações, gerando amostras de vários cenários possíveis e discutidos os resultados encontrados

**Palavras Chaves:** Econometria espacial, modelo de escolha discreta, regressão probita, análise Bayesiana, estudo de simulação

---

\* bolsista CAPES

# 1 Introdução

Modelos econométricos para variáveis endógenas contínuas que lidam com a dependência espacial são largamente difundidos, como em Anselin (1988, 2002), Kelejian e Prucha (1999), LeSage (1997) entre outros. Por outro lado, modelos discretos, particularmente no contexto do modelo probito de escolha discreta, embora de reconhecida importância prática, tem recebido menos atenção. Em parte este fato é devido a complexidade adicional oriunda da dependência espacial quando esta é incorporada ao modelo (Fleming, 2001). Algumas técnicas de estimação foram propostas para o modelo discreto na presença da estrutura espacial. McMillen (1992) usou o modelo probito espacial num exemplo ilustrativo com dados de criminalidade apresentados em Anselin (1988). Pinkse e Slade (1998) usaram o modelo probito para estudar um problema de concorrência em postos de gasolina. Vários outros exemplos de problemas de escolha discreta num contexto espacial podem ser encontrados na literatura econométrica. Um dos mais interessantes, talvez, seja o trabalho de Smith e LeSage (2003) em que se estuda a influência do espaço, no caso os estados dos EUA, na eleição presidencial daquele país em 1992. Aqui, vamos nos basear primordialmente nestes últimos trabalhos citados.

Este artigo é organizado da seguinte maneira. Na Seção 1, definimos o modelo de regressão probito e fazemos uma revisão de maneiras de como pode ser interpretado. Na Seção 3, mostramos os modelos de regressão probito sob duas perspectivas diferentes. Uma, com a influência espacial incorporada no modelo via variável latente espacialmente defasada ou com erros autorregressivos espaciais. A segunda via um efeito aleatório espacialmente estruturado, interpretado como um fator de perturbação nos dados. Na seção 4, mostramos um estudo de simulação, em que comparamos métodos de estimação clássicos com Bayesianos no primeiro contexto espacial apresentado acima e estudamos também, como o modelo de efeitos aleatórios se comporta diante de diferentes cenários de simulação. Uma discussão geral sobre os resultados é apresentada na Seção 5.

## 2 Modelos Probitos

Suponha que  $N$  variáveis aleatórias independentes binárias  $Y_1, \dots, Y_N$  são observadas, e cada  $Y_i$  segue uma distribuição de Bernoulli com probabilidade de sucesso  $p_i$ . Tal probabilidade, por sua vez, é relacionada com um conjunto de covariáveis discretas ou contínuas por meio de uma função de ligação  $H$ . Assim definimos o modelo de regressão binária como

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(p_i), \\ p_i &= H(x_i^t \beta), \quad i = 1, \dots, N, \end{aligned} \tag{2.1}$$

em que  $\beta : k \times 1$  é vetor de parâmetros desconhecidos,  $x_i^t = (x_{i1}, \dots, x_{iN})$  é vetor de covariáveis conhecido e  $H$  é uma função de distribuição acumulada (fda) conhecida ligando as probabilidades  $p_i$  à estrutura linear  $x_i^t \beta$ . O modelo probito é obtido quando  $H$  é a fda da normal padrão, do mesmo modo que o modelo logit é obtido se  $H$  é a fda da logística<sup>1</sup>.

Albert e Chib (1993) apresentam este modelo como tendo uma estrutura de regressão padrão implícita em variáveis latentes contínuas normalmente distribuídas. Isto é, introduzindo  $N$  variáveis aleatórias independentes latentes  $Y_1^*, \dots, Y_N^*$  em que  $Y_i^* \sim N(x_i^t \beta, 1)$  e definindo

$$Y_i = \begin{cases} 1 & \text{se } Y_i^* > 0, \\ 0 & \text{se } Y_i^* \leq 0. \end{cases} \tag{2.2}$$

Obviamente os  $Y_i^*$  são desconhecidos, isto é, não observáveis. Porém, dado  $Y_i$ , a variável latente segue uma distribuição normal truncada de acordo com (2.2). Esta abordagem então, junta o modelo probito em  $Y_i$  com uma regressão linear padrão em  $Y_i^*$ , isto permite também que elaboremos o modelo probito de modo mais flexível, por exemplo, usando uma mistura de distribuições normais para o modelo de variáveis latentes [Albert e Chib (1993)]. Uma das aplicações mais conhecidas do modelo probito se refere aos chamados

---

<sup>1</sup>Para uma discussão mais geral destes modelos, ver McCulloch e Searle (2001)

experimentos de dose-resposta em toxicologia, onde a variável latente tem a interpretação de tolerância dos indivíduos a substâncias tóxicas<sup>2</sup>.

A interpretação para este modelo aqui considerada é através dos modelos de escolha discreta em econometria [Pinkse e Slade (1998), Conley (1999)]. Assuma que existam  $N$  agentes ou indivíduos e que estes escolham uma entre um conjunto  $a$  de escolhas. Em particular, assumamos que  $a = a_0, a_1$ <sup>3</sup>, então o dado observado  $y_i$  relativo ao agente  $i$  é binário assumindo o valor 1 se o agente escolhe a alternativa  $a_1$  e 0 caso contrário. Seja então,  $U_{ia_0}$  a função utilidade individual para a escolha de  $a_0$  e  $U_{ia_1}$  a função utilidade individual para a escolha de  $a_1$  em que  $i = 1, \dots, N$ . A escolha observada indica qual delas produziu a maior utilidade. Daí,

$$Y_i = \begin{cases} 1 & \text{se } U_{i1} > U_{i0} \\ 0 & \text{se } U_{i1} \leq U_{i0} \end{cases} \quad (2.3)$$

onde a mais comum formulação é o modelo utilidade aleatória linear,  $U_{i1} = x_i^t \beta_1 + \varepsilon_1$  e  $U_{i0} = x_i^t \beta_0 + \varepsilon_0$  [Greene (2003)]. Assim, fazendo  $Y_i^* = U_{i1} - U_{i0} = x_i^t (\beta_1 - \beta_0) + (\varepsilon_1 - \varepsilon_0) = x_i^t \beta + \varepsilon_i$  tem-se,

$$P[Y_i = 1 | x_i] = P[Y_i^* > 0 | x_i] \quad (2.4)$$

Assumindo que  $\varepsilon_i \sim N(0, 1)$  chegamos ao modelo probito em (2.2).

A função de verossimilhança para o modelo probito acima pode ser escrita como [Albert e Chib (1993)],

$$L(\beta, Y^* | y) = \prod_i^n \{ \mathbf{1}_{(y_i^* > 0)} \mathbf{1}_{(y_i=1)} + \mathbf{1}_{(y_i^* \leq 0)} \mathbf{1}_{(y_i=0)} \} \times \phi(Y_i^*; x_i^t \beta, 1), \quad (2.5)$$

onde  $\mathbf{1}_{(Z \in A)}$  é a função indicadora que assume o valor 1 se a variável aleatória  $Z$  está contida no conjunto  $A$ , 0 caso contrário e  $\phi(\cdot; \mu, \sigma^2)$  é a função densidade de probabilidade da  $N(\mu, \sigma^2)$ .

### 3 Modelo probito com dependência espacial

Dois modelos probito serão abordados aqui. A diferença entre eles está em como a influência do espaço é incorporada. O primeiro modelo usa a variável latente espacialmente defasada e também erros espacialmente correlacionados. Este tipo de modelo probito com dependência espacial foi primeiro estudado na literatura econometrica por McMillen (1992), em que a estimação por máxima verossimilhança via algoritmo EM [Dempster, Laird e Rubin (1977)] foi proposta para produzir estimativas consistentes para o modelo. Já Pinkse e Slade (1998) e Conley (1999), trabalhando com um modelo econômico de escolha discreta, usou o método dos momentos generalizados para estimar um modelo probito com erros espacialmente correlacionados. Como destacado por McMillen (1992), tais procedimentos de estimação sustentam-se sobre propriedades assintóticas. Uma alternativa, proposta por LeSage (2000) é um modelo hierárquico bayesiano, que é computacionalmente mais oneroso, mas é mais robusto nos casos em que o tamanho da amostra é pequeno. Esta abordagem permite que a dependência espacial e a heterogeneidade espacial sejam tratadas simultaneamente. No segundo modelo, o espaço é incorporado através de um efeito aleatório espacialmente estruturado descrito por Smith e LeSage (2003), que será discutido com mais detalhes na Seção 3.2.

#### 3.1 Modelo 1

Seja o modelo em (2.2), cuja estrutura da variável latente pode ser formulada de duas formas diferentes. A primeira, na forma vetorial

$$Y^* = \rho W Y^* + X \beta + \epsilon, \quad (3.6)$$

<sup>2</sup>Ver Collet (2003) para detalhes quanto a análise de tais experimentos

<sup>3</sup>Para  $a = 1, \dots, c$  teríamos o modelo probito multinomial. Ver Natarajan, McCulloch e Kiefer (2000)

onde  $\epsilon \sim N(0, \sigma^2 I_n)$ , o qual denotaremos de *SAR*. E, para o segundo caso,

$$Y^* = X\beta + \epsilon, \quad \epsilon = \delta V\epsilon + u, \quad (3.7)$$

onde  $u \sim N(0, \sigma^2 I_n)$ , que chamaremos de *SEM*.

A autocorrelação espacial está geralmente associada à heterocedasticidade, dificultando a estimação por máxima verossimilhança. Isto se dá pois esta correlação produz uma função de verossimilhança envolvendo várias integrais, tornando o processo de estimação muito difícil [McMillen (1992), Fleming (2002)].

### 3.1.1 Estimação de MV via algoritmo EM

Acima foi descrito um caso particular de como os efeitos espaciais podem ser incorporados em uma regressão probito. Este modelo probito espacial, entretanto, tem uma estrutura de variância e covariância complicada, fato que dificulta a estimação por máxima verossimilhança direta. Todavia, o algoritmo EM dá um meio para se obter estas estimativas. No procedimento de estimação por este algoritmo, a variável dependente discreta é substituída pelo valor esperado da variável latente e o valor esperado da função de log-verossimilhança é maximizado usando os valores obtidos como se fosse a própria variável latente. Iterativamente, dadas as novas estimativas dos parâmetros retorna-se ao primeiro passo. Este processo continua até a convergência. A seguir descreveremos o algoritmo EM para o caso do modelo SAR. McMillen (1992) e Fleming (2002) mostram o algoritmo para o modelo SEM.

Usaremos o algoritmo EM com as seguintes definições. A variável latente  $Y^*$  como sendo os dados completos. Então a log-verossimilhança para os dados completos é dado por,

$$l(\rho, \beta, \sigma_\epsilon/Y^*) \propto -\frac{n}{2} \ln|\Sigma_1| - \frac{1}{2} (Y^* - X_1\beta)^t \Sigma_1 (Y^* - X_1\beta), \quad (3.8)$$

onde  $X_1 = (I_n - \rho W)^{-1} X$  e  $\Sigma_1 = \sigma^2 [(I - \rho W)^t (I - \rho W)]^{-1}$ .

O passo E requer que calculemos as esperanças condicionais  $E[Y^*|Y]$ , o que é relativamente simples já que é o cálculo da esperança de uma normal truncada. Seguindo o algoritmo proposto por McMillen (1992), seja  $\lambda_{ij}^a$  um elemento da matriz  $(I - \rho W)^{-1}$ . Então, da forma reduzida apresentada em (3.6) temos,

$$Y_i^* = \sum_j^N \lambda_{ij}^a x_j^t \beta + \sum_j^N \lambda_{ij}^a \epsilon_j = x_i^* \beta + \nu_i^a \quad (3.9)$$

Note-se que  $\text{var}(\nu_i^a) = \sigma_\epsilon^2 \sum_j^N \lambda_{ij}^a \equiv \sigma_{\nu_i^a}^2$ . McMillen (1992) definiu  $\sigma_\epsilon^2 = 1$  argumentando que, desta forma, elimina o problema de identificabilidade do modelo probito, que não permite a estimação simultânea de  $\sigma_\epsilon$  e  $\beta$ .

Tomando a esperança condicional em (3.9) temos,

$$\hat{Y}_i^* = E[Y_i^*|Y_i = 1] = x_i^* \beta + \sigma_{\nu_i^a} \phi(x_i^* \beta / \sigma_{\nu_i^a}) / \Phi(x_i^* \beta / \sigma_{\nu_i^a}), \quad (3.10)$$

onde  $\Phi(\cdot)$  é a fda da normal padrão. A maximização da log-verossimilhança neste caso é feita em duas etapas. Primeiro é feita a maximização com respeito a  $\beta$  condicionada ao parâmetro  $\rho$ . A segunda etapa é a maximização sobre o parâmetro  $\rho$  condicionada aos valores de  $\beta$ . Este passo M condicional [Natarajan, McCulloch e Kiefer (2000)] é mais simples pois a estimativa de máxima verossimilhança (MV) condicional de  $\beta$  é simplesmente o estimador de mínimos quadrados generalizados,

$$\hat{\beta}_{SAR} = [X_1^t \Sigma_1^{-1} X_1]^{-1} X_1^t \Sigma_1^{-1} \hat{Y}^*,$$

onde  $\hat{Y}^* = E[Y^*|Y]$ . A estimativa de MV de  $\rho$  é obtida maximizando

$$-\frac{n}{2} \ln|\Sigma_1| - \frac{1}{2} (Y^* - X_1 \hat{\beta}_{SAR})^t \Sigma_1 (Y^* - X_1 \hat{\beta}_{SAR}), \quad (3.11)$$

Funções não-lineares tais como (3.11), são relativamente bem estudadas e existem vários algoritmos muito eficientes para maximizá-las. Aqui, usaremos essencialmente as rotinas de otimização do programa MATLAB.

*O algoritmo*

- (i) Passo 0: Dê valores iniciais para  $\beta^{(0)}$  e para  $\rho^{(0)}$  no modelo SAR. Faça contador  $m = 0$ ;
- (ii) Passo 1: (passo E) Para cada  $i$ , calcule as esperanças condicionais (3.10). Obtenha  $\hat{Y}^{*(m)}$ ;
- (iii) Passo 2a: (Passo M condicional 1) Faça  $\hat{\beta}_{\text{SAR}}^{(m+1)} = [X_1^t \Sigma_1^{-1} X_1]^{-1} X_1^t \Sigma_1^{-1} \hat{Y}^{*(m)}$
- (iv) Passo 2b: (Passo M condicional 2) Usando os valores de  $\beta$  encontrados no passo anterior, maximize (3.11). Obtenha  $\rho^{(m+1)}$ .
- (v) Passo 3: Se a convergência é atingida faça  $\hat{\beta}_{\text{EM}} = \hat{\beta}_{\text{SAR}}^{(m+1)}$  e  $\hat{\rho}_{\text{EM}} = \rho^{(m+1)}$ . Caso contrário, faça  $m = m + 1$  e retorne ao passo 1.

Resta então obter a estimativa da matriz de variância e covariâncias. Para isto, define-se  $\eta_i \equiv Y_i - \Phi(x_i^* \beta / \sigma_{\nu i})$  e  $\Omega = E(\eta \eta^t)$ . Para um dado  $\rho$ , o estimador via mínimos quadrados ponderados não linear é obtido minimizando  $\eta^t \Omega^{-1} \eta$  com respeito a  $\beta$ . Note que  $E(\eta \eta^t) = E[(Y - E(Y))(Y - E(Y))^t]$  e que

$$\text{cov}(Y_i, Y_j) = \Phi_2(x_i^* \beta / \sigma_{\nu i}, x_j^* \beta / \sigma_{\nu j}, \tau_{ij}) - \Phi(x_i^* \beta / \sigma_{\nu i}) \Phi(x_j^* \beta / \sigma_{\nu j}), \quad (3.12)$$

onde  $\Phi_2$  é função de distribuição da normal bivariada e  $\tau_{ij}$  o coeficiente de correlação. Condicionado em  $\rho$  a matriz de covariância de  $\beta$  é  $(F^t \Omega^{-1} F)$ , onde  $F_i = \phi(x_i^* \beta / \sigma_{\nu i}) x_i^*$  e  $\Omega$  tem como um elemento típico (3.12).

Os coeficientes de correlação  $\tau_{ij}$  são calculados usando a matriz  $[(I - \rho W)^t (I - \rho W)]^{-1}$ , que é obtida depois do modelo ter sido estimado pelo algoritmo EM.

### 3.1.2 Método dos Momentos Generalizados

Uma alternativa para o algoritmo EM acima é usar o método dos momentos generalizado [Conley (1999), Greene (2003)] que, embora ineficiente, é assintoticamente consistente. Pinkse e Slade (1998) consideram inicialmente o modelo de erros espaciais (SEM). Entretanto, fazem uma generalização que acrescenta pouco para a complexidade do modelo. Considere o modelo em (2.2) cuja variável latente é definida como

$$Y_i^* = g(\theta, x_i) + \varepsilon_i, \quad (3.13)$$

onde  $\theta = (\beta, \rho)$ . Assumimos que o vetor de erros  $\varepsilon$  tem média 0 e uma matriz positiva definida de variância  $V_n$ , onde  $n$  indica que esta matriz pode variar com o tamanho da amostra. Ainda, denotando a variância de  $\varepsilon_i$  em um tamanho de amostra  $n$  por  $v_{ni}^2$ . Se os verdadeiros  $v_{ni}$  são conhecidos, podemos dividir cada observação pelo seu respectivo desvio padrão e aplicar o modelo probito padrão. Quando  $v_{ni}$  é desconhecido, temos que estimar o vetor  $\theta$  todo. Para isto, Pinkse e Slade (1998), Fleming (2002) usam o *método dos momentos generalizados* (MMG).

Uma vez que os resíduos de (3.13) não são observáveis, usa-se o conceito de resíduos generalizados definido como

$$\tilde{u}_i(\theta) = E[u_i | y_i, \theta] = [Y_i - \Phi(G(\theta))] \frac{\phi(G(\theta))}{\Phi(G(\theta)) [1 - \Phi(G(\theta))]}, \quad (3.14)$$

onde  $\phi(\cdot)$  e  $\Phi(\cdot)$  são função densidade e distribuição acumulada da normal padrão, respectivamente. Ainda  $G_n(\theta) = g(\theta, x_i) v_{ni}$

A condição de momento usada é a ortogonalidade entre os regressores e os erros, isto é,

$$E[Z_n^t u(\theta)] = 0, \quad S_n(\theta) = n^{-1} Z_n^t u(\theta) = 0,$$

onde  $Z$  é a matriz de variáveis instrumentais, que deve incluir as variáveis endógenas  $X$ . E seja  $S(\theta) = \lim_{n \rightarrow \infty} E[S_n(\theta)]$ . O estimador pelo método dos momentos generalizado é dado por

$$\hat{\theta} = \arg \min_{\theta} Q_n(\theta), \quad (3.15)$$

onde  $Q_n(\theta) = [S_n(\theta)]^t M_n S_n(\theta)$  e  $M_n$  alguma matriz definida-positiva. Finalmente, definimos  $Q(\theta) = [S(\theta)]^t M S(\theta)$ , com  $M$  um matriz definida-positiva para a qual a sequência  $M_n$  converge. Pinkse e Slade (1998) dá o teorema e sua prova, para a consistência assintótica do estimador  $\hat{\theta}$ , bem como a normalidade assintótica e a estimação da matriz de covariância. Aqui usaremos  $M_n = I_n$ , onde  $I_n$  é a matriz identidade de dimensão  $m$ , como sugerem Pinkse e Slade (1998) e Kelejian e Prucha (1999).

### 3.1.3 Modelo Bayesiano

Modelos espaciais de escolha discreta Bayesianos foram propostos em detalhe por LeSage (2000) sendo uma extensão para o modelo não espacial apresentado por Albert e Chib (1993). LeSage (2000), baseado em Geweke (1993), estende os modelos SAR e SEM incorporando a heterocedasticidade separadamente da dependência espacial [Fleming (2002)]. Isto é importante pois a heterocedasticidade causa inconsistência nos modelos de escolha discreta [Greene (2003)].

As priors do modelo probito SAR em (3.6) são,

- (a)  $\beta \sim N(c, T)$ ,
- (b)  $v_i^{-1} \stackrel{iid}{\sim} Gama\left(\frac{r}{2}, \frac{r}{2}\right)$ ,
- (c)  $\sigma^{-2} \sim Gama(\alpha, \eta)$ ,
- (d)  $\rho \sim U[\lambda_{min}^{-1}, \lambda_{max}^{-1}]$

No modelo SEM a única mudança é para a priori de  $\delta$ , isto é, (d')  $\delta \sim U[\lambda_{min}^{-1}, \lambda_{max}^{-1}]$ , já que neste caso  $\rho = 0$ .

Em ambos os casos temos que  $c$ ,  $V$ ,  $\alpha$ ,  $\eta$  e  $r$  são conhecidos.  $\lambda_{min}$  e  $\lambda_{max}$  são, respectivamente, o menor e o maior autovalor da matriz  $W$

A estimação é feita via Métodos de Monte Carlo em Cadeia de Markov (MCMC), usando o amostrador de Gibbs. Para tanto, a condicional da variável latente dado os demais parâmetros deve ser calculada. Conforme Albert e Chib (1993) e LeSage (2000), temos as seguintes condicionais completas para o modelo SAR,

- (i)  $\beta | (\delta, \sigma^2, V, Y^*, Y) \sim N(\hat{\beta}, \sigma^2 B^{-1} V B^{-1})$ , onde  
 $\hat{\beta} = [X_1^t (\sigma^2 B^{-1} V B^{-1})^{-1} X_1]^{-1} [X_1^t (\sigma^2 B^{-1} V B^{-1})^{-1} y^* + T^{-1} c]$
- (ii)  $\sigma^{-2} | (\beta, \delta, V, Y^*, Y) \sim Gama\left(\frac{n}{2} + \alpha, \frac{H}{2} + \eta\right)$ ,  
onde  $H = [y^* - X_1 \beta]^t (\sigma^2 B^{-1} V B^{-1})^{-1} [y^* - X_1 \beta]$
- (iii)  $\rho | (\beta, \sigma^2, V, Y^*, Y) \propto |V'|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} (y^* - X_1 \beta)^t (V')^{-1} (y^* - X_1 \beta)\right\}$ ,  
onde  $V' = (I - \delta W)^{-1} V (I - \delta W)^{-1}$ .
- (iv)  $v_i^{-1} | (\beta, \delta, \sigma^2, v_{-i}, Y^*, Y) \propto Gama\left(\frac{r+1}{2}, \frac{\xi_i^t \xi_i / \sigma^2 + r}{2}\right)$ ,  
onde  $\xi_i$  é a  $i$ -ésima coluna da matriz  $(I - \delta W)(y^* - X_1 \beta)$  e  $v_{-i} = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ .
- (v)  $Y_i^* | (\beta, \rho, \sigma^2, V, Y) \sim \begin{cases} N(x_{i1}^t \beta, v_i) & \text{truncada à esquerda em 0 se } Y_i = 1 \\ N(x_{i1}^t \beta, v_i) & \text{truncada à direita em 0 se } Y_i = 0 \end{cases}$

onde  $V = \text{diag}(v_i)$  e  $X_1 = (I_n - \rho W)^{-1} X$ . Em resumo, o esquema de estimação começa com valores iniciais arbitrários  $\beta^0, \rho^0, \theta^0, \sigma^0, v^0$  e  $Y^{*0}$ . Amostra-se das distribuições de (i) à (v) atualizando os valores a cada etapa. Repetindo este até a convergência.

Aqui a escolha dos valores para o hiperparâmetro  $r$  é de grande importância, pois sua magnitude permite ou não o modelo a lidar eficientemente com a variância não constante e valores aberrantes. Valores pequenos de  $r$ , em torno de 2 a 7, permite que a heterocedasticidade seja modelada e torna o modelo mais robusto em relação a valores aberrantes. Valores grandes de  $r$ , 20 a 50, produzirão estimativas de  $v_i$  próximos de 1, tornando o modelo homocedástico.

## 3.2 Modelo 2

### 3.2.1 Escolhas envolvendo agentes distribuídos espacialmente

Suponha que existam dados sobre as escolhas de um conjunto de agentes ou indivíduos distribuídos em um sistema de  $m$  regiões no espaço. Em particular, suponha que apenas duas escolhas mutuamente exclusivas sejam relevantes as quais denominaremos '0' e '1'. A escolha observada para cada indivíduo ( $k = 1, \dots, n_i; i = 1, \dots, m$ ) na  $i$ -ésima região é tratada como uma realização de uma variável aleatória,  $Y_{ik}$ , em que

$$Y_{ik} = \begin{cases} 1 & \text{se o indivíduo } k \text{ da região } i \text{ escolhe 1} \\ 0 & \text{se o indivíduo } k \text{ da região } i \text{ escolhe 0} \end{cases} \quad (3.16)$$

Podemos postular que as escolhas são baseadas em funções de utilidade aleatórias [Smith e LeSage (2003), Greene (2003)], em que a  $k$ -ésima utilidade, para cada uma das duas alternativas, é assumida como tendo a forma

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

, em que  $w_{ika}$  é um vetor de dimensão  $w$  de atributos observados para a alternativa  $a = 0, 1$  e  $s_{ik}$  é um vetor  $s$ -dimensional de atributos observados relativos ao indivíduo  $k$ . O termo  $\theta_{ia} + \varepsilon_{ika}$ , representa a contribuição para a utilidade de todos outros atributos não observáveis da região  $i$ , indivíduo  $k$  e alternativa  $a$ . Separando em efeito regional,  $\theta_{ia}$ , representando componente não observável de utilidade relativa à escolha  $a$  comum a todos indivíduos da região  $i$ , e um efeito individual,  $\varepsilon_{ika}$ , representando todos os outros componentes não observáveis. Tomando a diferença de utilidades para o indivíduo  $k$ ,

$$Y_{ik}^* = U_{ik1} - U_{ik0} = x_{ik}^t \beta + \theta_i + \varepsilon_{ik}, \quad (3.17)$$

em que  $\beta = (\gamma^t, \alpha_0^t - \alpha_1^t)^t$  é o vetor de parâmetros, o vetor de atributos  $x_{ik} = [(w_{ik1} - w_{ik0})^t, s_{ik}^t]^t$ , o efeito regional  $\theta_i = (\theta_{i1} - \theta_{i0})$  e os efeitos individuais  $\varepsilon_{ik} = (\varepsilon_{ik1} - \varepsilon_{ik0})$ , independentes do efeito regional. Então, seguindo a hipótese de maximização de utilidades, similar a (2.4),

$$P[Y_{ik} = 1] = P[U_{ik1} > U_{ik0}] = P[Y_{ik}^* > 0]. \quad (3.18)$$

Para as quantidades não observáveis do modelo, é postulado que toda dependência entre a diferença de utilidades para os indivíduos em regiões diferentes são capturadas pela dependência entre os efeitos regionais ( $\theta_i, i = 1, \dots, m$ ). Em particular, os aspectos comuns a indivíduos em uma dada região  $i$  podem ser similares a indivíduos de regiões vizinhas. Isto é feito assumindo que o vetor de interação,  $\theta$ , apresenta a seguinte estrutura autorregressiva espacial<sup>4</sup>.

$$\theta_i = \rho \sum_{j=1}^m w_{ij} \theta_j + u_i, \quad i = 1, \dots, m, \quad (3.19)$$

onde os elementos  $w_{ij}$  são medidas de proximidade espacial entre as regiões  $i$  e  $j$ . Os erros  $u_i$  são assumidos como sendo normais independentes identicamente distribuídos (iid) com média 0 e variância  $\sigma^2$ . Escrevendo na forma matricial, em que  $\theta = (\theta_i : i = 1, \dots, m)$  e  $u = (u_i : i = 1, \dots, m)$ , temos

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

sendo  $W$  a matriz de proximidade espacial com  $w_{ii} = 0$  e  $I_m$  matriz identidade. Assumindo que  $B_\rho = I_m - \rho W$  é não singular temos  $\theta = B_\rho^{-1} u$  e

$$\theta | \rho, \sigma^2 \sim N(0, \sigma^2 [B_\rho^t B_\rho]^{-1}) \quad (3.20)$$

<sup>4</sup>Esta é especificação simultânea para as dependências regionais e segue a abordagem econométrica, apresentada em [Anselin (1988, 2002)]

Os efeitos individuais,  $\varepsilon_{ik}$  são assumidos como sendo normais (iid) com média 0 e variância  $v_i$  independentes dos efeitos regionais  $\theta_i$ . As variâncias  $v_i$ , indexada na região, implica que estamos postulando a homocedasticidade intra-região e heterocedasticidade inter-região. Se denotarmos o vetor de efeitos individuais da  $i$ -ésima região por  $\varepsilon_i = (\varepsilon_{ik} : k = 1, \dots, n_i)$ , então nossa suposição implica que  $\varepsilon_i \sim N(0, v_i I_{n_i})$ . Expressando em termos globais, se  $\varepsilon = (\varepsilon_i^t : k = 1, \dots, n_i)^t$  temos que  $\varepsilon \sim N(0, V)$  em que  $V$  é da forma,

$$V = \begin{pmatrix} v_1 I_{n_1} & & \\ & \ddots & \\ & & v_m I_{n_m} \end{pmatrix}.$$

A verossimilhança para (3.17) fica então

$$L(\beta, y^* | y) = \prod_{k=1}^m \prod_{i=1}^{n_k} \{ \mathbf{1}_{(y_{ik}^* > 0)} \mathbf{1}_{(y_{ik}=1)} + \mathbf{1}_{(y_{ik}^* \leq 0)} \mathbf{1}_{(y_{ik}=0)} \} \times \phi(y_{ik}^*; x_{ik}^t \beta, t_i). \quad (3.21)$$

Podemos ainda expressar (3.17) na forma matricial em que  $Y_i^* = (Y_{ik}^* : k = 1, \dots, n_i)^t$  e  $X_i = (x_{ik} : k = 1, \dots, n_i)^t$ , então a diferença de utilidade para a região  $i$  tem a forma

$$Y_i^* = X_i \beta + \theta_i \mathbf{1}_i + \varepsilon_i, \quad i = 1, \dots, m, \quad (3.22)$$

onde  $\mathbf{1}_i$  é um vetor  $n_i \times 1$  unitário. Definindo  $n = \sum_i^m n_i$ ,  $Y^* = (Y_i^* : i = 1, \dots, m)$  e  $X = (X_i^t : i = 1, \dots, m)^t$  a equação regional (3.22) pode ser reduzida na forma

$$Y^* = X \beta + \Delta \theta + \varepsilon, \quad (3.23)$$

em que

$$\Delta = \begin{pmatrix} \mathbf{1}_1 & & \\ & \ddots & \\ & & \mathbf{1}_m \end{pmatrix}.$$

### 3.2.2 Modelo hierárquico Bayesiano

Por simplicidade e sem perda de generalidade, adotaremos prioris condicionalmente conjugadas ou difusas na maioria dos casos [Gelman (1997), Carlin e Louis (2000)]. Então para o modelo descrito em (3.16) com a variável latente em (3.23), temos a seguinte estrutura para o modelo hierárquico,

- (I) Condicionalmente a  $\beta, \rho, \sigma^2, \theta$ , e  $V$ , as variáveis latentes  $Y_{ik}^*$  tem distribuição normal  $N(x_{ik}^t \beta + \theta_i, v_i)$  com  $i = 1, \dots, m; k = 1, \dots, n_i$ .
- (II) Condicionalmente a  $\rho$  e  $\sigma^2$ , o vetor de efeitos espaciais  $\theta$  tem distribuição normal multivariada  $N(0, \sigma^2 [B_\rho^t B_\rho]^{-1})$ .
- (III) Condicionalmente a  $V$ ,  $\varepsilon$  tem distribuição normal multivariada  $N(0, V)$ .

Para completar o modelo hierárquico, designamos as seguintes prioris para  $\beta, \rho, \sigma^2$  e os elementos da diagonal de  $V$ .

- (IV) Sejam as prioris, similares a Smith e LeSage (2003),

$$\begin{aligned} \beta &\sim N(\nu_o, \Sigma_0) \\ \rho &\sim \text{Beta}(a, b) \\ v_i &\stackrel{iid}{\sim} \text{InvGama}\left(\frac{r}{2}, \frac{r}{2}\right) \\ \sigma^2 &\sim \text{InvGama}(\alpha, \eta) \end{aligned}$$

Reparametrizando  $\sigma^2 = 1/\phi$  e  $v_k = 1/v_k^*$ , para facilitar a notação na seção seguinte, redefinimos novamente as duas últimas priors,

$$v_i^* \stackrel{iid}{\sim} Gama\left(\frac{r}{2}, \frac{r}{2}\right) \quad \text{e} \quad \phi \sim Gama(\alpha, \eta)$$

Smith e LeSage (2003) usam uma distribuição uniforme para a priori de  $\rho$ , conforme apresentado na Subseção 3.1.3, com a restrição do valor de  $\rho$  entre o inverso no menor autovalor e o inverso do maior autovalor da matriz de  $W$  [Anselin (2002), Cressie (1993)]. Aqui, usamos a restrição de  $0 < \rho < 1$  e também adotamos a distribuição *Beta* para a priori de  $\rho$ . É uma forma de generalizar a distribuição uniforme proposta por Smith e LeSage (2003), LeSage (2000), já que esta distribuição é um caso particular da distribuição *Beta*, em que os parâmetros tem valores  $a = 1$  e  $b = 1$ . Esta priori permite mais flexibilidade, quando da suspeita de que os dados tem alta correlação espacial ou mesmo quando tem correlação baixa.

### 3.2.3 Implementação computacional

A estimação será feita via amostrador de Gibbs que amostra de um conjunto de distribuições condicionais completas a posteriori dos parâmetros. Para implementar a abordagem de amostragem MCMC, precisamos das distribuições condicionais completas, para todos os parâmetros do modelo. Estas são apresentados a seguir, em que consideramos modelo hierárquico apresentado acima, com  $B_\rho = (I - \rho W)$  e  $V^* = \text{diag}(v_1^* I_{n_1}, \dots, v_m^* I_{n_m})$ , onde  $I_{n_i}$  é matriz identidade de dimensão  $n_i$ .

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

(ii)  $\theta | \beta, \rho, \phi, V^*, y^*, y \sim N(\hat{\theta}, M^{-1})$ ,

onde  $M = \Delta^t V^{-1} \Delta + \phi B_\rho^t B_\rho$  e  $\hat{\theta} = M^{-1} \Delta^t V^{-1} (y^* - \Delta \theta)$

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

LeSage (1997, 2000) mostrou que esta distribuição condicional de  $\rho$  não tem uma forma de uma distribuição conhecida. Sugere então um passo Metropolis-Hasting [Metropolis (1953), Hastings (1970)] dentro do procedimento de amostragem do amostrador de Gibbs, frequentemente chamado de Metropolis em Gibbs [Carlin e Louis (2000), Gamerman (1997)]. Algumas sugestões para a chamada proposta do núcleo de transição da cadeia do MCMC são a distribuição normal e a distribuição  $t$  com três graus de liberdade [LeSage (1997)]. Neste trabalho são apresentadas duas propostas diferentes para o núcleo de transição, discutidas na seção 3.2.4

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

(v)  $v_i^* | \beta, \rho, \phi, \theta, v_{-i}^*, y^*, y \sim Gama(g, h)$ , onde  $g = \frac{n_i+r}{2}$  e  $h = \frac{\sum_k^i e_{ik} + r}{2}$ .

(vi)  $Y_{ik}^* | \beta, \rho, \phi, \theta, V^*, y \sim \begin{cases} N(x_{i1}^t \beta + \theta_i, v_i^*) & \text{truncada à esquerda em 0 se } y_i = 1 \\ N(x_{i1}^t \beta + \theta_i, v_i^*) & \text{truncada à direita em 0 se } y_i = 0 \end{cases}$

Em resumo, o esquema do amostrador de Gibbs envolve atribuir valores iniciais arbitrários para os parâmetros que chamaremos de  $\beta^0, \theta^0, \rho^0, \phi^0, V^{*0}$  e a variável latente  $Y^{*0}$ . Dadas as distribuições condicionais a posteriori descritas acima, o procedimento de amostragem é definido como

1.  $\pi(\beta | \rho^j, \phi^j, \theta^j, V^{*j}, y^{*j}, y)$ , que é uma distribuição normal multivariada com parâmetros definidos em (i). Isto atualiza os valores do vetor  $\beta$  que denotaremos por  $\beta^{j+1}$ ;
2.  $\pi(\theta | \beta^{j+1}, \rho^j, \phi^j, V^{*j}, y^{*j}, y)$ , que também é uma distribuição normal multivariada definida em (ii). Obtemos, então  $\theta^{j+1}$ ;

3.  $\pi(\rho|\beta^{j+1}, \phi^j, \theta^{j+1}, V^{*j}, y^{*j}, y)$ , que é obtido usando o algoritmo M-H, que será descrito na Subseção 3.2.4. Obtém-se  $\rho^{j+1}$ ;
4.  $\pi(\phi|\beta^{j+1}, \rho^{j+1}, \theta^{j+1}, V^{*j}, y^{*j}, y)$ , que tem uma distribuição Gama, como descrito em (iv), amostramos  $\phi^{j+1}$ ;
5.  $\pi(v_i^*|\beta^{j+1}, \rho^{j+1}, \phi^{j+1}, \theta^{j+1}, v_{-i}^{*j}, y^{*j}, y)$ , uma distribuição Gama conforme (v), com  $k = 1, \dots, m$ . Atualizamos  $V^{*(j+1)}$ ;
6.  $\pi(y^*|\beta^{j+1}, \rho^{j+1}, \phi^{j+1}, \theta^{j+1}, V^{*(j+1)}, y)$  cuja distribuição normal truncada é descrita em (vi), obtemos  $Y^{*(j+1)}$ .

O amostrador de Gibbs prossegue até que a convergência seja atingida. Note que o amostrador de Gibbs não precisa necessariamente seguir esta ordem de amostragem. Adotamos arbitrariamente esta configuração do algoritmo.

### 3.2.4 Metropolis em Gibbs

A seguir são apresentadas as densidades normais propostas para o núcleo de transição do algoritmo M-H.

#### (A) Cadeias independentes

Neste caso, a transição proposta é formulada independentemente da posição atual  $\rho^{(j)}$  da cadeia. A proposta é gerada por uma distribuição normal com média  $\mu$  e variância  $\tau^2$ . Para a média do núcleo de transição, usamos o máximo numérico da função de log-verossimilhança condicional de  $\rho$ , seguindo Chib, Nardari e N. Shepard (2002), usando uma rotina de otimização para obtê-la. Estes autores sugerem ainda que a variância seja menos a segunda derivada numérica desta função (uma estimativa da informação observada de Fisher). Entretanto, no intuito de economizar em tempo de geração, adotamos a sugestão de Carlin e Louis (2000) e usamos as primeiras amostras para obter uma estimativa da variância a posteriori.

#### (B) Cadeia passeio aleatório

Novamente a caracterização da cadeia se refere à transição proposta. Um passeio aleatório é uma cadeia de Markov onde a evolução é dada por  $\rho^{(j)} = \rho^{(j-1)} + z_i$  em que  $z_i$  é uma variável aleatória com uma distribuição independente da cadeia [Gamerman (1997)]. Algumas distribuições foram propostas para este problema particular de estimação do coeficiente de dependência espacial. LeSage (1997, 2000), por exemplo, propõe uma distribuição normal ou uma  $t - Student$  com 3 graus de liberdade, adicionando ainda a restrição  $\lambda_{min}^{-1} < \rho^{j+1} < \lambda_{max}^{-1}$  nos valores que as propostas podem assumir, em que  $\lambda_{min}$  e  $\lambda_{max}$  são, respectivamente, o menor e o maior autovalor da matriz de proximidade  $W$ . A restrição que usamos tanto aqui quanto no caso de cadeias independentes é  $0 < \rho^{j+1} < 1$ , conforme descrito na Seção 3.2.2. Usamos a distribuição normal com parâmetro de escala obtido usando as primeiras iterações do algoritmo.

#### Escolha para o algoritmo M-H

O tempo gasto para gerar uma amostra de tamanho 500 no MATLAB usando um processador Athlon XP 1GHZ é de aproximadamente 1.10 segundos para M-H com cadeias independentes e de 1.04 segundos para cadeia passeio aleatório. Neste trabalho, optamos pela proposta usando cadeias independentes geradas pela distribuição normal, que apesar de ser ligeiramente mais lenta devida ao procedimento de otimização usado para encontrar o parâmetro de escala, a convergência é mais rápida que no caso de cadeia passeio aleatório usada por LeSage (1997, 2000).

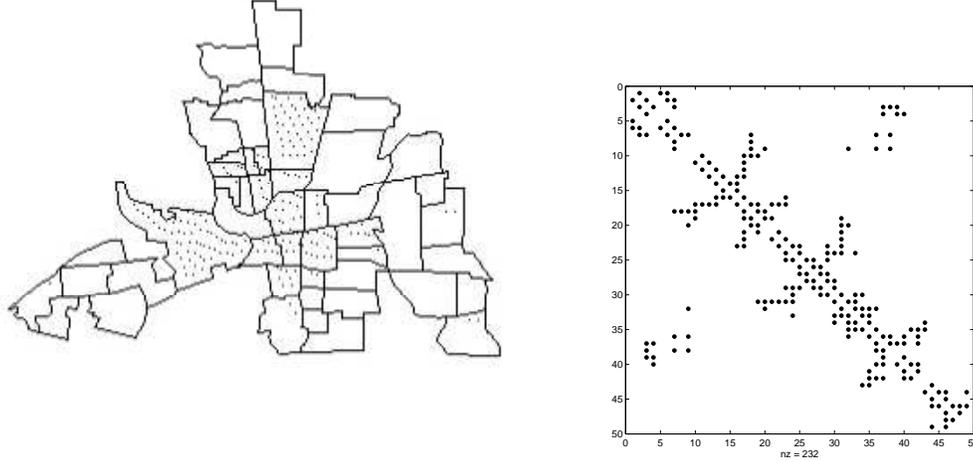


Figura 1: Estrutura espacial usada nas simulações - Mapa e sua matriz de contiguidade

## 4 Estudo de simulação

### 4.1 Simulação do modelo 1

Anselin (1988) ilustra vários exemplos de modelos espaciais usando um conjunto de dados de crimes em 49 setores censitários de Columbus, Ohio, EUA. A variável dependente  $Y^*$ , combina o total de arrombamento de casas e roubos de carros na vizinhança por mil famílias agregadas nestes setores censitários. McMillen (1992) discretiza esta variável com a regra,

$$Y_i = \begin{cases} 1 & \text{se } Y_i^* > 40, \\ 0 & \text{se } Y_i^* \leq 40. \end{cases} \quad (4.24)$$

Isto é,  $Y = 1$  caso o setor tenha uma alta taxa de criminalidade. Como em Anselin (1988) e McMillen (1992), usamos estes dados para ilustrar a técnica e não para adquirir conhecimento substantivo sobre o padrão espacial do crime. Para as variáveis independentes, Anselin (1988) usou  $X_1$  a renda média e  $X_2$  o valor médio dos imóveis dos setores censitários. A distribuição espacial dos setores pode ser vista na Figura 1.

A partir destas covariáveis, geramos 2500 amostras segundo o modelo probito SAR em (3.6), usando como valores de geração para os parâmetros as estimativas encontradas por McMillen (1992). Então, o modelo gerado é o seguinte,

$$Y_i^* = 0.429 \sum_j w_{ij} Y_j^* + 2.58\beta_0 - 0.128x_{1i}\beta_1 - 0.029x_{2i}\beta_2 + \varepsilon_i. \quad (4.25)$$

Onde  $\varepsilon_i \sim N(0, 1)$  e  $w_{ij} = 1$  se os setores são contíguos e 0 caso contrário. A discretização foi feita usando (4.24).

Para cada amostra, ajustamos o modelos usando as três técnicas apresentadas na Seção 3.1: máxima verossimilhança (MV) via algoritmo EM [McMillen (1992)], método dos momentos generalizados (MMG) [Pinkse e Slade (1998)] e a abordagem bayesiana [LeSage (2000)]. Na estimação dos parâmetros por MV, usamos como valores iniciais para o algoritmo EM,  $(\beta_0, \beta_1, \beta_2)^t = (0, 0, 0)^t$  e  $\rho = 0.5$ . Estes são os mesmos valores iniciais usados por McMillen (1992). Para a maximização da verossimilhança, usamos a função *sar* da *Econometric Toolbox* que basicamente ajusta o modelo SAR gaussiano na variável latente gerada pelo algoritmo EM, que é equivalente a maximizar a verossimilhança em (3.11). A condição de parada para o algoritmo é quando a diferença das estimativas do passo anterior com o passo atual for menor que 0.001 .

A estimação pelo MMG proposta por Pinkse e Slade (1998) é uma otimização da forma quadrática,

$$[S_n(\beta, \rho)]^t M_n S_n(\beta, \rho)$$

definida na Seção 3.1.2. A matriz  $M_n$  deve ser definida positiva. Neste trabalho, como em Pinkse e Slade (1998), usamos  $M_n = I_n$ , onde  $I_n$  é matriz identidade. Outro aspecto importante do MMG é a escolha das variáveis instrumentais. Pinkse e Slade (1998) sugerem que, dentre as variáveis instrumentais, devem estar as próprias regressoras  $X$  e a variável latente espacialmente "defasada",  $WY^*$ . Usamos os algoritmos de otimização disponíveis no programa MATLAB para encontrar as estimativas dos parâmetros por MMG. Os valores iniciais usados para o procedimento iterativo de otimização são  $(\beta_0, \beta_1, \beta_2)^t = (0, 0, 0)^t$  e  $\rho = 0.3$

Uma das vantagens da utilização do MMG é a simplicidade de implementação computacional. Este procedimento envolve cálculo simples dos resíduos generalizados definidos em (3.14) e dado o grande número de algoritmo de otimização já implementados de formas quadráticas, torna-se um problema relativamente simples.

Na abordagem Bayesiana, utilizamos os comando "*sarp - g*" da *Econometric Toolbox*. A escolha do hiperparâmetro  $r$  da distribuição a priori de  $v$  é um ponto importante. Se escolhermos  $r$  pequeno estamos supondo heterocedasticidade e/ou a presença de algum valor aberrante e  $r$  grande admitimos que os dados são homocedásticos sem valores aberrantes. LeSage (1997, 2000) sugere a escolha de valores entre  $4 < r < 7$ . Aqui, usamos  $r = 4$ , supondo que temos heterocedasticidade e que existam possíveis valores aberrantes nos dados. Os hiperparâmetros para a priori de  $\beta$  foram escolhidos de tal forma que esta fique vaga, porém própria. Assim, dada que a priori de  $\beta$  é  $N(c, T)$ , definimos  $c = (0, 0, 0)^t$  e  $T = 10^6 I_p$ , onde  $I_p$  é matriz identidade de dimensão  $p$ . Ajustamos o modelo com um *burn-in* de 100 e amostramos as 1000 iterações seguintes, conforme sugere LeSage (1997, 2000).

Tabela 4.1: Valor real e as estimativas médias, erro padrão e erroquadrático médio de 2500 amostras simuladas

Parâmetros		$\beta_0$	$\beta_1$	$\beta_2$	$\rho$
	Real	2.5887	-0.128	-0.029	0.429
	Média	1.8822	-0.1051	-0.0181	0.1817
EM	Erro Padrão	0.6779	0.0456	0.0131	0.0997
	EQM	0.9561	0.0026	$0.2899 \times 10^{-3}$	0.0711
	Média	1.6630	-0.1342	-0.0131	0.2373
MMG	Erro Padrão	1.3309	0.1107	0.0186	0.3245
	EQM	2.6243	0.0123	$0.5975 \times 10^{-3}$	0.1420
	Média	2.3876	-0.1366	-0.0237	0.2334
Bayes	Erro Padrão	0.5195	0.0409	0.0140	0.1086
	EQM	0.1420	0.0017	$0.2234 \times 10^{-3}$	0.0500

As estimativas produzidas pelos três métodos foram resumidos na Tabela 4.1. A abordagem Bayesiana do modelo probito SAR obteve os melhores resultados dentre os métodos testados, produzindo estimador com o menor erro quadrático médio (EQM) em relação ao verdadeiro valor do parâmetro. As estimativas dos coeficientes de regressão  $\beta_1$  e  $\beta_2$ , em média, foram que mais se aproximaram dos valores de geração com um EQM das 2500 amostras de 0.0017 e  $0.22 \times 10^{-3}$ , respectivamente. Os resultados obtidos pelo MMG apresentaram a maior dispersão e os maiores EQM das estimativas dos parâmetros nas amostras simuladas (Figura 2).

Note que todos os métodos subestimaram o coeficiente de dependência espacial  $\rho$  ficando longe do valor de geração (Figura 3). Uma possível causa para a subestimação é o tamanho da amostra limitada pela estrutura de vizinhança.

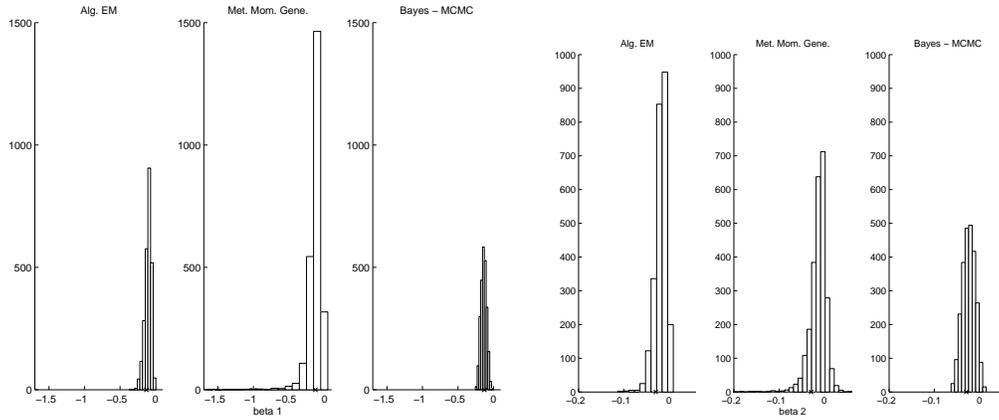


Figura 2: Histograma das Simulações -  $\beta_1 = -0.128$  e  $\beta_2 = -0.029$

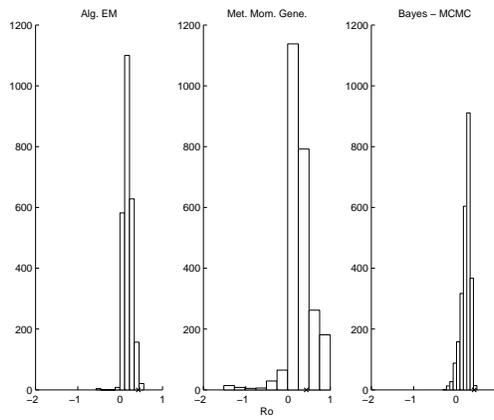


Figura 3: Histograma das Simulações -  $\rho = 0.429$

## 4.2 Simulação do modelo 2

As simulações do modelo 2 são feitas usando a estrutura de vizinhança da Região Administrativa de Campinas, SP, composta por 90 municípios, dada na figura 4. Geramos os dados segundo um modelo relativamente simples com apenas duas covaráveis, de acordo com (3.17) e (3.19)

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

onde os valores de geração são  $\beta_1 = 3.14$ ,  $\beta_2 = -1.62$  e  $\rho = 0.70$ . Por simplicidade, os erros  $\varepsilon_i$  e  $u_i$  foram gerados de normais padrão. A  $x_{1ik}$  é variável uniforme com média zero e variância 8.33 (uma variável uniforme em  $[-5,5]$ ) e  $x_{2ik}$  é uma variável normal com média zero e variância 3. Note que a média zero das covaráveis, em aplicações, pode ser obtida subtraindo-se suas respectivas médias. Finalmente, temos a regra para discretização,

$$Y_{ik} = \begin{cases} 1 & \text{se } Y_{ik}^* > 0, \\ 0 & \text{se } Y_{ik}^* \leq 0. \end{cases} \quad (4.26)$$

Dividimos as simulações em três casos. A primeira, a situação mais crítica para o modelo 2, quando temos apenas uma observação ou um indivíduo por área. Este caso é comum no que se refere a dados de área, quando a informação está agregada (resumida) para refletir o comportamento médio de uma determinada área. No segundo caso, simulamos um número um pouco maior de indivíduos por área. Aqui, usamos  $n_i = 5$  para todas as áreas. Finalmente no Caso 3, consideramos um número maior de observações, fixando  $n_i = 15$ . O objetivo é verificar como o modelo é estimado quando o tamanho da amostra aumenta.

## 4.3 Caso 1

Smith e LeSage (2003), em um estudo de simulação similar ao que é apresentado aqui, afirmam que a estimação via MCMC do modelo hierárquico, apesar de estimar bem os parâmetros de regressão  $\beta$ , subestimam o coeficiente autorregressivo  $\rho$ . Os autores, porém, não explicam nem sugerem alternativas para lidar com este problema.

Primeiro, ajustamos os dados gerados com um algoritmo similar ao usado no comando "*semip - g*" da *Econometric Toolbox* de James LeSage, cujo amostrador de Gibbs foi definido no seção anterior. Para a priori de  $\phi = 1/\sigma^2$ , fizemos  $\alpha = 0.001$  e  $\eta = 1000$ ,  $r = 100$  para a priori  $v_i^* = 1/v$ . Para a priori *Beta* de  $\rho$  usamos  $a = 3$  e  $b = 1$ . Definimos um *burn-in* de 2000 e amostramos as 2000 iterações seguintes.

Para obtermos a estimação pontual de cada um dos componentes do vetor  $(\beta, \rho, \sigma^2, v)$ , precisamos selecionar uma medida que resuma a distribuição a posteriori marginal, tal como média, mediana ou moda. Quando a distribuição a posteriori é simétrica, a média e a mediana serão iguais; para posteriori simétrica unimodal, as três medidas coincidirão. Para posterioris assimétricas, a escolha é menos clara, embora a mediana é eventualmente preferida desde que é uma medida intermediária entre moda (que considera apenas o valor correspondente ao máximo valor da densidade) e a média (que pode, eventualmente, dar muito peso a valores extremos) [Carlin e Louis (2000)]. Sabemos que Smith e LeSage (2003) adotam a média da posteriori amostrada para a estimação pontual dos parâmetros. Na Tabela 4.2 apresentamos os resultados da estimação.

Ao analisar o gráfico das cadeias (Figura 5), temos que as cadeias parecem convergir muito lentamente. Modelos como este, com muitos parâmetros em relação aos dados disponíveis tipicamente, conduzem a altas correlações a posteriori entre parâmetros que reduz dramaticamente o movimento do amostrador de Gibbs através do espaço paramétrico [Carlin e Louis (2000)]. A variância  $\sigma^2$  do processo  $u_i$ , que é uma variável aleatória, tem uma correlação forte com as demais cadeias relativas aos outros parâmetros do modelo, em particular com o  $\beta$  da regressão (Tabela 4.3), o que pode estar ocasionando a lentidão na convergência do amostrador de Gibbs. A distribuição a posteriori de  $\rho$  exibe assimetria (Figura 6) com a moda da distribuição

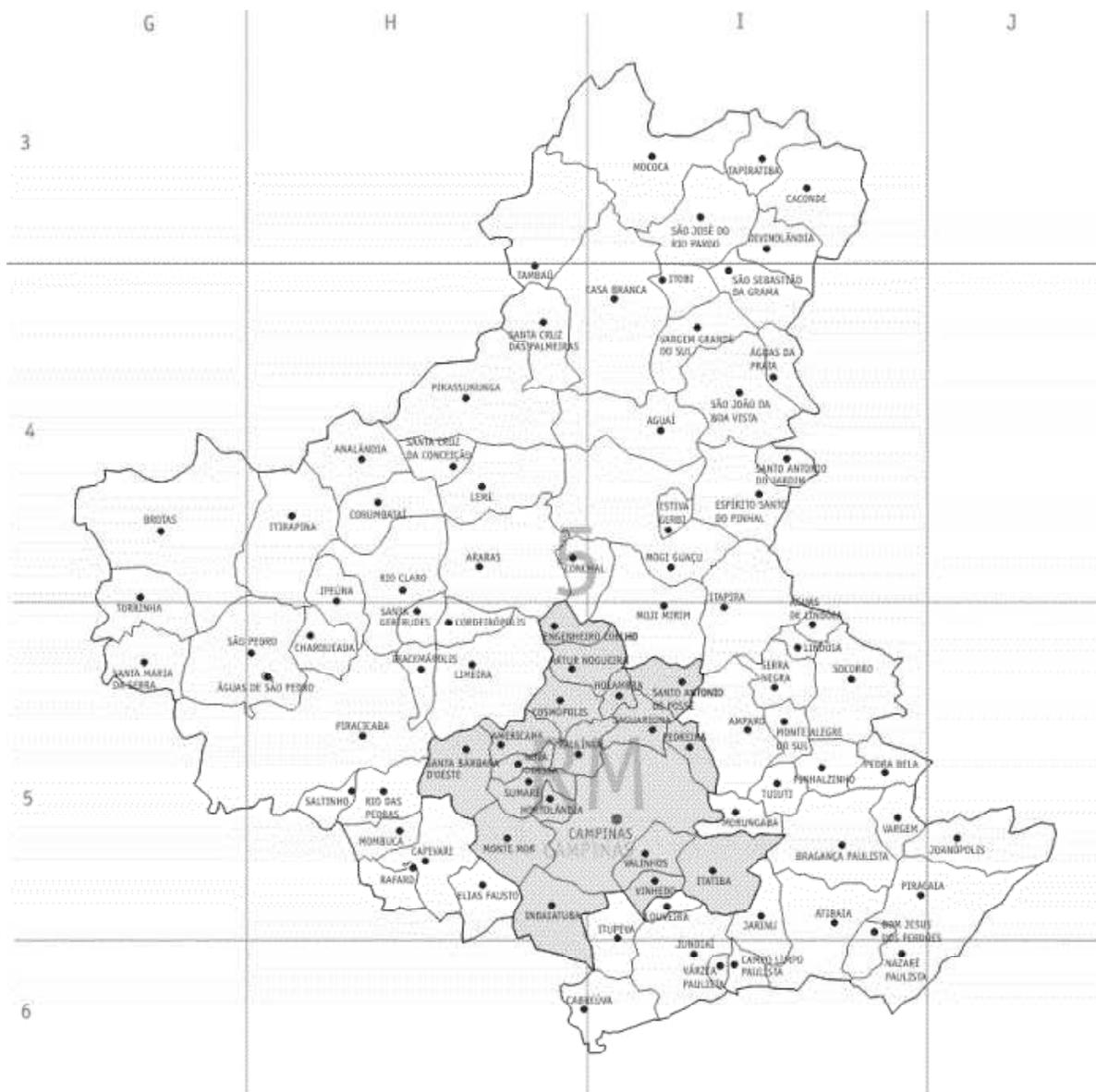


Figura 4: Região administrativa de Campinas

Tabela 4.2: Valor real e as estimativas médias, mediana e moda

	$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$
Real	3.14	-1.62	0.70	1
Média	6.053	-2.144	0.6078	0.4398
Mediana	6.119	-2.141	0.6526	0.2740
Moda	4.928	-2.101	0.7324	0.3031
Desv.Pad	1.576	0.6837	0.2114	0.5244
2.50%	3.145	-3.490	0.09393	0.07739
97.50%	8.963	-0.8767	0.9023	2.093

Tabela 4.3: Correlações cruzadas a posteriori dos parâmetros

	$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$	v
$\beta_1$	1.0				
$\beta_2$	-0.7483	1.0			
$\rho$	0.08026	0.02015	1.0		
$\sigma^2$	-0.4226	0.2918	0.1727	1.0	
v	-0.08331	0.02886	-0.06072	-0.1021	1.0

situando-se ao redor do valor de geração do parâmetro. Quando Smith e LeSage (2003) chegaram a conclusão que para este caso especial ocorre subestimação, usando a média como estimador do coeficiente de dependência espacial  $\rho$ , não levaram em consideração que neste caso, em que a distribuição a posteriori é assimétrica e devemos levar em consideração as demais alternativas de medidas que resumam a distribuição, tais como mediana e moda. Como dito acima, no exemplo, a moda é a medida que mais se aproxima do valor de geração de  $\rho$ .

Neste trabalho, propomos um artifício do amostrador que Gibbs, o qual denominaremos Esquema 2, para tentar contornar este problema da subestimação e da correlação entre os parâmetros, sem partir para a reparametrização do modelo. Sabemos que uma cadeia de Markov não precisa seguir necessariamente uma seqüência com é apresentado na Seção 3.2.3, amostrando-se sequencialmente a cada passo de todas a condicionais completas. Podemos fazer com que a cadeia se desloque, por um determinado período de tempo,

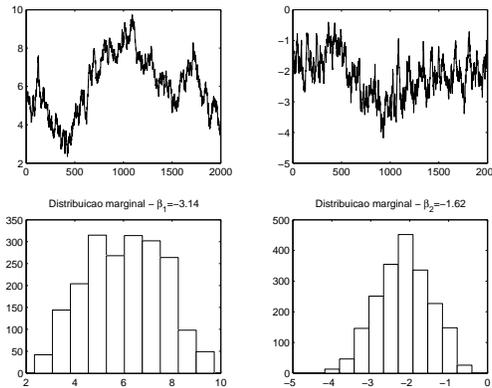


Figura 5: Cadeias dos parâmetros de regressão  $\beta$

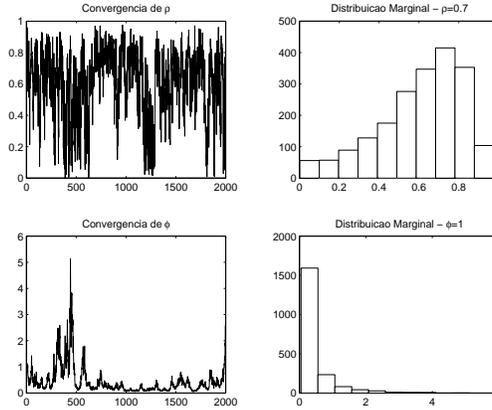


Figura 6: Cadeias dos parâmetros  $\rho$  e  $\phi$

Tabela 4.4: Valor real e as estimativas médias, mediana e moda

		$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$
		3.14	-1.62	0.70	1
Esquema 2	Média	3.383	-1.352	0.5669	1.945
	Mediana	3.321	-1.340	0.5771	1.686
	Moda	3.143	-1.556	0.6914	1.133
	Desv.Pad	0.7448	0.4560	0.1809	1.264
	2.50 %	2.204	-2.389	0.2274	0.2712
	97.50%	5.129	-0.5501	0.8788	4.979
Smith e LeSage (2003) semip-g	Média	11.30	-4.22	0.649	3.086
	Mediana	13.42	-4.302	0.9940	2.432
	Moda	16.38	-1.502	0.8991	1.491
	Desv.Pad	5.877	2.217	0.6478	2.411
	2.50%	2.757	-7.743	-0.9930	0.5199
	97.50%	19.97	-0.9238	0.9980	9.850

com alguns parâmetros fixos. Neste caso, fixamos o coeficiente espacial  $\rho = 0$  por um número de iterações do amostrador de Gibbs para então começar a amostrar e atualizar este parâmetro e os demais. Usando os mesmos valores para os hiperparâmetros citados acima, ajustamos o modelo aos dados gerados, com  $\rho = 0$  e  $\phi$  fixos por 2000 iterações. Depois deste período, definimos um *burn-in* de 2000 iterações e amostramos as 2000 seguintes.

Com esta estratégia para o amostrador de Gibbs, a cadeia amostrada dos parâmetros de regressão  $\beta$  mostrou-se mais bem comportada que no caso anterior (Figura 8). Isto é constatado também no resultado da estimação (Tabela 4.4) em que os coeficientes da regressão foram bem estimados. O coeficiente  $\rho$ , devido a assimetria de sua distribuição a posteriori, foi melhor estimado pela moda da distribuição.

Para estudar melhor este modelo, simulamos 50 amostras e para cada uma ajustamos o modelo usando o Esquema 2 e também usando *semip-g* da Econometric Toolbox.

Os resultados das 50 simulações (Tabela 4.5) mostram que a mediana foi a medida de melhor desempenho para todas as quantidades de interesse, com o menor erro quadrático médio (dado que conhecemos os valores reais dos parâmetros).

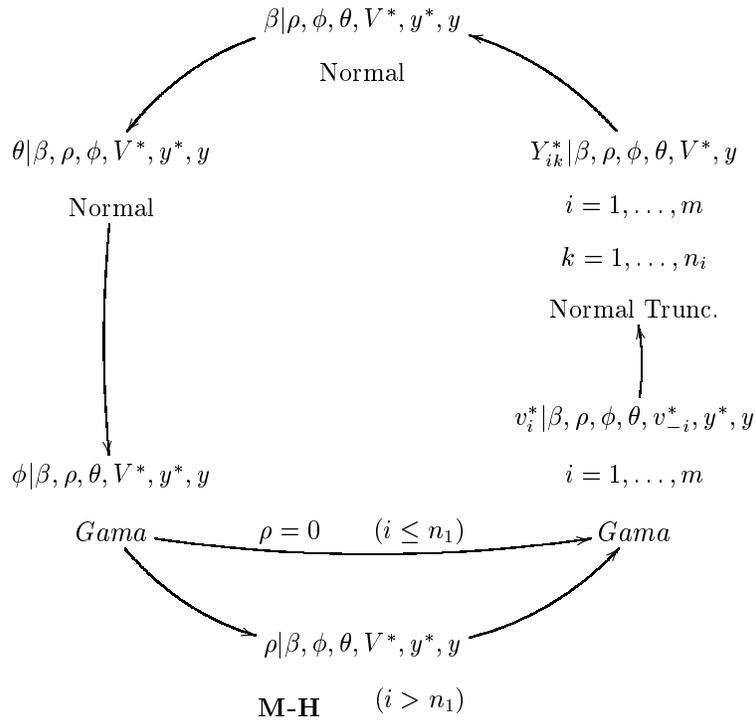


Figura 7: Esquema 2 - Amostrador de Gibbs para a análise MCMC, onde  $n_1$  é período em que  $\rho$  é fixo

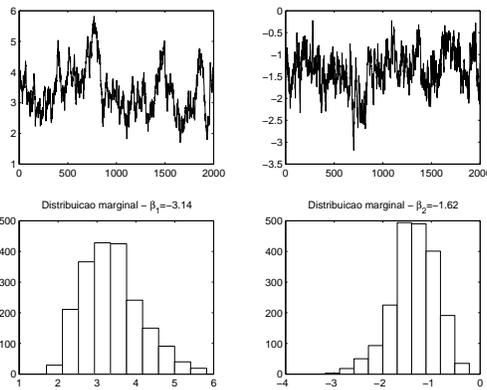


Figura 8: Cadeias dos parâmetros de regressão  $\beta$

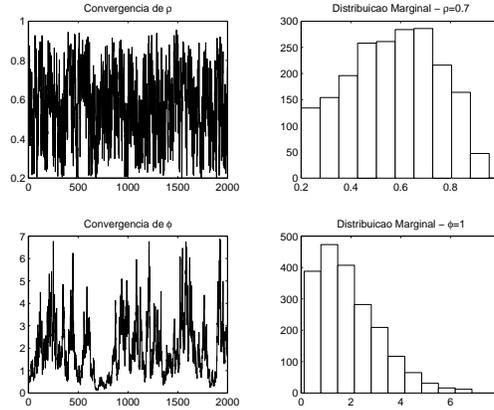


Figura 9: Cadeias dos parâmetros  $\rho$  e  $\phi$

Tabela 4.5: Caso 1 - Resultado das simulações de 50 amostras

		Esquema 2			Smith e LeSage (2003)		
		Estimativa	Dp	EQM	Estimativa	Dp	EQM
$\beta_1$	Média	3.438	0.7759	0.6508	7.441	4.980	41.65
	Mediana	3.062	0.6249	0.3706	7.292	5.177	42.25
	Moda	2.724	0.5699	0.4764	6.857	5.562	42.68
$\beta_2$	Média	-1.758	0.5365	0.2906	- 3.771	2.547	10.72
	Mediana	-1.580	0.4701	0.2071	- 3.728	2.772	11.66
	Moda	-1.401	0.4804	0.2592	- 3.564	3.144	13.04
$\rho$	Média	0.6241	0.06627	0.04370	0.5010	0.4397	0.1862
	Mediana	0.6815	0.07795	0.03430	0.5308	0.5362	0.2687
	Moda	0.6950	0.08063	0.03430	0.6424	0.6490	0.2687
$\sigma^2$	Média	1.395	0.3644	0.2799	4.614	4.919	35.65
	Mediana	0.9810	0.3924	0.1441	3.940	4.561	28.06
	Moda	0.6258	0.1091	0.1511	3.435	4.384	23.87

#### 4.4 Caso 2

Apresentamos primeiro o exemplo de uma amostra gerada, em que mostramos que a cadeia amostrada de  $\rho$  é estável, ao contrário do que acontecia no Caso 1. Isso é devido a, quando temos mais indivíduos dentro das áreas a amostra contém mais informação sobre os efeitos regionais  $\theta_i$ . Conseqüentemente, mais informação temos sobre o coeficiente de dependência espacial,  $\rho$ . A Tabela 4.6, mostra o resultado da estimação usando também comando *semip-g*. A distribuição a posteriori de  $\rho$ , na Figura 10, como no Caso 1 exibe assimetria. As estimativas pelo Esquema 2, são mostradas na Tabela 4.6.

Completando o estudo do Caso 2, geramos 25 amostras, para as quais usamos o Esquema 2 e *semip-g* para o ajuste do modelo. Os resultados são mostrados na Tabela 4.7

Neste caso, a moda foi a melhor estimativa para os parâmetros de interesse. No caso dos coeficiente de regressão  $\beta$ , as estimativas pelas três medidas se aproximam. Isto é devido a simetria da distribuição a posteriori de  $\beta$ . Para o coeficiente de dependência espacial,  $\rho$ , a diferença entre as estimativas ainda é consequência da assimetria da distribuição a posteriori (Figura 10).

Tabela 4.6: Valor real e as estimativas médias, mediana e moda

		$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$
		3.14	-1.62	0.70	1
Esquema 2	Média	3.107	- 1.717	0.6761	0.6690
	Mediana	3.077	- 1.706	0.6876	0.6203
	Moda	2.985	- 1.762	0.6643	0.6130
	Desv.Pad	0.3695	0.2383	0.1195	0.2538
	2.50 %	2.487	- 2.251	0.4148	0.3196
	97.50%	3.919	- 1.316	0.8671	1.315
Smith e LeSage (2003) semip-g	Média	3.342	- 1.846	0.8209	1.330
	Mediana	3.247	- 1.799	0.8340	1.214
	Moda	2.926	- 1.776	0.8205	1.245
	Desv.Pad	0.5189	0.2987	0.09896	0.5307
	2.50%	2.579	- 2.613	0.5955	0.6869
	97.50%	4.785	- 1.389	0.9770	2.761

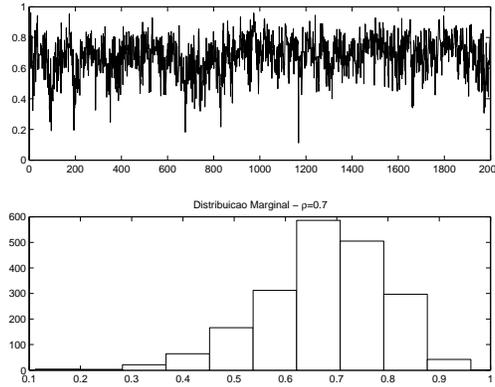


Figura 10: Distribuição a posteriori de  $\rho$

Tabela 4.7: Caso 2 - Resultado das simulações de 25 amostras

		Esquema 2			Smith e LeSage (2003)		
		Estimativa	Dp	EQM	Estimativa	Dp	EQM
$\beta_1$	Média	3.346	0.7615	0.6029	3.688	0.9723	1.214
	Mediana	3.327	0.8343	0.7078	3.691	1.088	1.447
	Moda	3.286	0.8821	0.7733	3.714	1.140	1.585
$\beta_2$	Média	- 1.739	0.4110	0.1799	- 1.917	0.5509	0.3880
	Mediana	- 1.728	0.4312	0.1936	- 1.923	0.6039	0.4504
	Moda	- 1.712	0.4947	0.2470	- 1.943	0.6321	0.4973
$\rho$	Média	0.6219	0.2155	0.05100	0.5212	0.1901	0.06689
	Mediana	0.6463	0.2189	0.04922	0.5282	0.2007	0.06845
	Moda	0.6842	0.2305	0.04922	0.5394	0.2289	0.06845
$\sigma^2$	Média	1.206	0.6076	0.3993	1.540	1.345	2.041
	Mediana	1.086	0.5584	0.3088	1.474	1.382	2.072
	Moda	0.9668	0.4987	0.2415	1.382	1.530	2.409

Tabela 4.8: Valor real e as estimativas médias, mediana e moda

		$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$
		3.14	-1.62	0.70	1
Esquema 2	Média	3.094	-1.645	0.7906	1.293
	Mediana	3.066	-1.642	0.7962	1.242
	Moda	2.993	-1.626	0.8086	1.166
	Desv.Pad	0.1746	0.1115	0.06284	0.3470
	2.50 %	2.792	-1.870	0.6595	0.7573
	97.50%	3.486	-1.436	0.8985	2.058
Smith e LeSage (2003) semip-g	Média	3.393	-1.819	0.7488	1.046
	Mediana	3.378	-1.819	0.7550	1.008
	Moda	3.322	-1.818	0.7580	0.9123
	Desv.Pad	0.2650	0.1584	0.09038	0.2552
	2.50%	2.912	-2.093	0.5490	0.6469
	97.50%	3.873	-1.518	0.9020	1.633

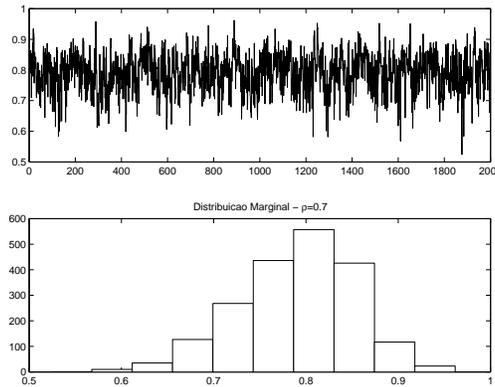


Figura 11: Distribuição a posteriori de  $\rho$

#### 4.5 Caso 3

Finalmente consideramos um número grande de indivíduos por área,  $n_i = 15$  para todo  $i$ . O custo computacional cresce muito neste caso, pois lidamos com várias operações matriciais (decomposição de Cholesky e autovalores) com grande quantidade de dados. Primeiro, como nos Caso 1 e Caso 2, vamos gerar uma amostra e ajustar o algoritmo do Esquema 2 e *semip-g*.

As estimativas obtidas usando *semip-g* superestimam os parâmetros, em comparação com as estimativas obtidas pelo Esquema 2 (Tabela 4.8).

A assimetria da distribuição a posteriori de  $\rho$  diminuiu quando aumentamos o número de observações por área. Os valores da média, mediana e moda amostrais estão bem próximos nas duas alternativas de estimação. Finalizando o estudo deste caso, geramos 25 amostras, e para as quais ajustamos com os dois procedimentos.

As estimativas dos parâmetros produzidas mostram que quando aumentamos o número de observações por área, as três medidas de resumo da posteriori ficam muito próximas, como pode ser visto na Tabela 4.9. Como foi constatado no estudo preliminar com uma amostra, a assimetria do coeficiente de dependência espacial tende a diminuir com o aumento da informação disponível sobre os efeitos regionais  $\theta_i$ . Neste caso,

Tabela 4.9: Caso 3 - Resultado das simulações de 25 amostras

		Esquema 2			Smith e LeSage (2003)		
		Estimativa	Dp	EQM	Estimativa	Dp	EQM
$\beta_1$	Média	3.234	0.2797	0.08321	2.958	0.2923	0.1144
	Mediana	3.234	0.2817	0.08418	2.957	0.2942	0.1157
	Moda	3.213	0.3010	0.09145	2.958	0.3100	0.1243
$\beta_2$	Média	- 1.650	0.1592	0.02571	- 1.507	0.1542	0.03322
	Mediana	- 1.650	0.1584	0.02542	- 1.505	0.1538	0.03351
	Moda	- 1.640	0.1549	0.02372	- 1.504	0.1520	0.03311
$\rho$	Média	0.6734	0.1096	0.01213	0.6504	0.1378	0.02051
	Mediana	0.6819	0.1088	0.01156	0.6562	0.1375	0.01988
	Moda	0.7004	0.1099	0.01156	0.6671	0.1381	0.01988
$\sigma^2$	Média	1.141	0.2319	0.07097	1.227	0.3795	0.1885
	Mediana	1.111	0.2257	0.06080	1.178	0.3591	0.1542
	Moda	1.053	0.2046	0.04262	1.124	0.3492	0.1313

temos 15 indivíduos dentro de cada área dando informação sobre o efeito da área sobre eles, enquanto, no Caso 1, temos apenas um indivíduo para estimar o efeito regional.

Neste último caso, qualquer uma das medidas poderia ser usada como medida de resumo da posteriori, já que estão todas próximas. Numericamente, sugerimos o uso da moda como estimador para as quantidade de interesse.

## 5 Discussão

Os modelos Bayesianos são uma ferramenta poderosa na investigação da dependência espacial em fenômenos econômicos de escolha discreta ou de variáveis truncadas. Aqui, mostramos dois possíveis modelos.

No caso do Modelo 1, o estudo de simulação mostrou que a abordagem Bayesiana é mais efetiva na estimação dos coeficientes de regressão se comparada com métodos frequentistas, tais como, máxima verossimilhança via EM e o método dos momentos generalizados (MMG).

O modelo hierárquico Bayesiano (Modelo 2), é flexível o bastante para lidar com vários cenários de dependência espacial. Nas simulações feitas aqui, este modelo teve um bom desempenho, tanto na estimação dos coeficientes da regressão quanto no coeficiente espacial  $\rho$ .

Este trabalho foi ilustrativo para mostrar as técnicas existentes e propor algumas alterações para a melhoria na estimação dos modelos descritos, em particular, no Modelo 2, em que as mudanças propostas melhoraram as estimativas.

Tais estudos de simulação deverão ser continuados, assim como a utilização desses modelos, notadamente o modelo hierárquico, na análise de dados econômicos reais.

## Referências

- ALBERT, J.; CHIB, S. Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.*, v. 88, n. 422, p. 516–522, 1993.
- 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.

- 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.
- COLLET, D. *Modelling binary data 2.ed.* London: Chapman Hall, 2003.
- CONLEY, T. Gmm estimation with cross sectional dependence. *J. Econometrics*, v. 92, p. 1–45, 1999.
- CRESSIE, N. *Statistics for spatial data.* New York: Wiley, 1993.
- DEMPSTER, A.; LAIRD, N.; RUBIN, D. Maximum likelihood estimation from incomplete observations. *J. Royal Statist. Soc. B*, v. 39, p. 1–38, 1977.
- FLEMING, M. Techniques for estimating spatially dependent discrete choice models. *Advances in Spatial Econometrics*, Springer-Verlag, 2002.
- GAMERMAN, D. *Monte Carlo Markov Chains: Stochastic Simulation for Bayesian Inference.* London: Chapman Hall, 1997.
- GELMAN, A. et al. *Bayesian data analysis.* London: Chapman and Hall, 1997.
- GEWEKE, J. Bayesian treatment of the independent student t linear model. *Journal of applied econometrics*, v. 8, p. 33–50, 1993.
- 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.
- 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.
- MCMLLEN, 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.
- NATARAJAN, R.; MCCULLOCH, C. E.; KIEFER, N. A monte carlo em method for estimating multinomial probit models. *Computational statistics and data analysis*, v. 34, p. 19–40, 2000.
- 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.