

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

MARCELO GAZZANO

**UM MODELO ESPAÇO-TEMPORAL BAYESIANO PARA MEDIR A
INTERAÇÃO SOCIAL NA CRIMINALIDADE: SIMULAÇÕES E EVIDÊNCIAS
NA REGIÃO METROPOLITANA DE SÃO PAULO**

Porto Alegre

2008

MARCELO GAZZANO

**UM MODELO ESPAÇO-TEMPORAL BAYESIANO PARA MEDIR A
INTERAÇÃO SOCIAL NA CRIMINALIDADE: SIMULAÇÕES E EVIDÊNCIAS
NA REGIÃO METROPOLITANA DE SÃO PAULO**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Flavio A. Ziegelmann

Porto Alegre

2008

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Responsável: Biblioteca Gládis W. do Amaral, Faculdade de Ciências Econômicas da
UFRGS

G289m

Gazzano, Marcelo

Um modelo espaço-temporal bayesiano para medir a interação social na criminalidade: simulações e evidências na Região Metropolitana de São Paulo / Marcelo Gazzano. – Porto Alegre, 2008
146 f. : il.

Orientador: Flávio A. Ziegelmann.

Dissertação (Mestrado em Economia) - Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, 2008.

1. Criminalidade : Inferência Bayesiana. 2. Criminalidade : São Paulo, Região Metropolitana de (SP). 3. Criminosos : Interação Social : São Paulo, Região Metropolitana de (SP). I. Ziegelmann, Flávio A. II. Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas. Programa de Pós-Graduação em Economia. III. Título.

CDU 330.115

MARCELO GAZZANO

**UM MODELO ESPAÇO-TEMPORAL BAYESIANO PARA MEDIR A
INTERAÇÃO SOCIAL NA CRIMINALIDADE: SIMULAÇÕES E EVIDÊNCIAS
NA REGIÃO METROPOLITANA DE SÃO PAULO**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia.

Aprovada em: Porto Alegre, ____ de ____ de 2008.

Prof. Dr. Pedro Luiz Valls Pereira
Fundação Getulio Vargas – SP

Prof. Dr. Marcelo Savino Portugal
Universidade Federal do Rio Grande do Sul – PPGE

Prof. Dr. Thelma Sáfyadi
Universidade Federal de Lavras – Departamento de Ciências Exatas

Agradecimentos

Agradeço ao meu orientador Flavio Ziegelmann por todo o tempo despendido e comentários que elevaram o nível desse trabalho. Também agradeço as contribuições de Juan Carlos Vivar-Rojas, Dani Gamerman, Aline Trindade Figueiredo e Susana Ochoa e a CAPES pelo suporte financeiro.

Agradecimentos especiais para meus familiares, amigos e a todos que de alguma forma também ajudaram e tornaram a elaboração desse trabalho possível.

Resumo

Neste trabalho utilizamos um modelo espaço-temporal proposto em Rojas (2004) para medir a interação social da criminalidade na região metropolitana de São Paulo. Realizamos simulações de Monte Carlo para testar a capacidade de estimação do modelo em diferentes cenários. Observamos que a estimação melhora com o aumento de observações ao longo do tempo. Já os resultados empíricos indicam que a região metropolitana de São Paulo é um *hot spot* no estado, pois é encontrado um maior grau de interação social no índice de homicídio em relação aos índices de roubo e furto.

Palavras chave: Modelo espaço-temporal, MCMC, Estatística espacial, Métodos Bayesianos, Criminalidade.

Abstract

In this paper we employ a spatio-temporal model proposed in Rojas (2004) to evaluate the social interaction in crime in São Paulo metropolitan area. We carry out Monte Carlo simulations to test the model estimation capability in different scenarios. We notice that the estimation gets better as the number of observations in time raises. The results point out that São Paulo metropolitan area is a hot spot in the state since we found out a greater social interaction for the homicide index, compared to robbery and thievery.

Key words: Spatio-temporal model, MCMC, Spatial statistics, Bayesian methods, Criminality

Sumário

1	INTRODUÇÃO.....	08
2	REVISÃO DA LITERATURA: ECONOMIA/ECONOMETRIA DO CRIME....	11
3	METODOLOGIA ESTATÍSTICA.....	15
3.1	INFERÊNCIA BAYESIANA.....	15
3.2	MODELOS LINEARES DINÂMICOS.....	18
3.3	FILTRO DE KALMAN.....	18
3.4	<i>FORWARD FILTERING BACKWARD SAMPLING</i>	21
3.5	AMOSTRADOR DE GIBBS.....	22
3.6	METROPOLIS-HASTINGS.....	23
3.7	ESTATÍSTICA ESPACIAL.....	25
4	O MODELO.....	27
4.1	MODELO POLINOMIAL DE PRIMEIRA ORDEM.....	27
4.2	ESTIMAÇÃO.....	29
4.2.1	<i>Prioris</i>	29
4.2.2	Filtro de Kalman e FFBS.....	30
4.2.3	Condicionais Completas.....	30
4.2.3.1	θ_0 (“Vetor de estados”).....	30
4.2.3.2	σ^2 (“Variância da equação de medida”).....	31
4.2.3.3	p (“Correlação temporal”).....	31
4.2.3.4	κ (“Parâmetro de escala”).....	33
4.2.3.5	α (“Correlação espacial”).....	33

5	SIMULAÇÕES DE MONTE CARLO.....	35
6	CRIME NA REGIÃO METROPOLITANA DE SÃO PAULO.....	40
7	CONCLUSÃO.....	46
	REFERÊNCIAS.....	48

1 INTRODUÇÃO

O estudo da criminalidade é muito importante, descobrir seus determinantes e como reduzi-la pode ajudar na formulação de políticas e no bem estar dos indivíduos. O problema criminal no Brasil é grande, Andrade e Lisboa (2000) estimam que caso o índice de homicídios fosse erradicado da cidade do Rio de Janeiro os homens cariocas viveriam, em média, 3,42 anos a mais no ano de 1995. Esse problema no estado de São Paulo não parece ser menor. Em relatório do ILANUD (Instituto Latino Americano das Nações Unidas para Prevenção do Delito e Tratamento do Delinqüente), Khan (1999) estima os custos da violência para o estado de São Paulo. Para o ano de 1999, o instituto estima que R\$ 8,9 bilhões (cerca de 3% do PIB estadual) foram gastos pela sociedade na compra de segurança e com perdas geradas diretamente pelos incidentes.

A criminalidade afeta diretamente as decisões dos indivíduos, algumas vezes, sem perceber deixamos de visitar determinada região, estacionar o carro em certas vizinhanças, adquirir algum artigo, etc. Rondon e Andrade (2003) estimam os custos da criminalidade nos alugueis da cidade de Belo Horizonte. Os autores encontram uma redução em até 27% do valor do aluguel devido às taxas criminais em determinada região da capital mineira.

Mesmo trazendo uma grande ineficiência para a economia, economistas não prestavam muita atenção para esse assunto. A literatura sobre economia do crime só começa a aparecer com um maior destaque a partir de Becker (1968). Segundo Becker o criminoso toma decisões como qualquer outro indivíduo da sociedade, ou seja, maximizando sua utilidade. Dessa forma, utilizando um arcabouço teórico microeconômico, Becker detecta que um indivíduo cometerá um crime caso o custo esperado tenha menor valor do que o ganho esperado do crime. Posto de outra forma, o criminoso avalia os custos esperados do crime tais como probabilidade de apreensão, probabilidade de condenação, severidade da punição e custo de oportunidade de ingressar no mercado formal de trabalho; e avalia esses custos relativamente aos ganhos que ele teria com tal atividade ilícita.

A principal forma de combate ao crime por parte dos municípios do estado de São Paulo tem sido aumentar os custos do crime, principalmente com o aumento da

probabilidade do criminoso ser preso¹. Apesar de estatísticas de homicídios estarem apresentando uma tendência de queda, furtos e roubos apresentam uma evolução diferente. Os roubos no estado de 2000 a 2006 parecem estacionários ao redor de 5,83 por ano por mil habitantes do estado; já o furto no estado era de 10,92 por ano por mil em 2000, contra 13,38 por ano por mil habitantes em 2006. O esforço do estado em reduzir a criminalidade parece surtir efeito nos crimes mais violentos como homicídios, latrocínios e seqüestros, porém as estatísticas apresentadas anteriormente não apontam para uma coerção em roubos e furtos².

Glaeser, Sacerdote e Scheinkman (1996) apresentam um modelo de interação social que aparenta resolver esta interessante questão. Os autores concluem que crimes mais brandos como roubo e furto têm um alto grau de interação social, ou seja, dependem muito da interação dos indivíduos de uma determinada região, enquanto crimes como seqüestro e assassinato têm baixo grau de interação social. Os autores observam que os índices de criminalidade variam muito de região para região. Por exemplo: a cidade de Ridgewood Village em Nova Jersey tem 0,008 crimes contra pessoas per capita contra 0,384 crimes contra pessoas per capita em Atlantic City; índices de homicídios variam de 6,1 homicídios por milhão no Japão, 12,6 na Suécia e 98,0 nos Estados Unidos. Essas diferenças parecem ser muito grandes para serem explicadas apenas pelos custos e benefícios do crime, além mesmo de variáveis socioeconômicas como desigualdade de renda e desemprego. Para os autores a explicação para essa alta variância é a interação social entre os criminosos, ou seja, quando a decisão de um indivíduo ingressar no mercado criminal afetar a decisão de seu vizinho, então os índices de criminalidade das regiões vão ser diferentes dos índices previstos pelas características das mesmas.

Assim, utilizando um modelo espaço-temporal proposto por Rojas (2004) buscaremos identificar alguma interação social nos seguintes índices: homicídio doloso (doravante homicídio), roubo e furto na região metropolitana de São Paulo. Acreditamos que modelando espaço e tempo simultaneamente teremos uma melhor estimativa da interação social desses índices.

¹ Em 2000 a população carcerária do estado era de 92.186 (0,25% da população total do estado), já em 2006 a população era de 144.430 (0,35% da população total do estado).

² Resultado não esperado, pois, de acordo com a teoria, um aumento no custo do crime faria com que os índices de criminalidade caíssem.

Antes de apresentarmos os resultados para a região de interesse realizamos simulações de Monte Carlo para testar a capacidade de estimação do modelo. Observamos que mesmo com um número de observações pequeno (mesmo tamanho do conjunto de dados reais) a estimação é satisfatória, se o modelo estiver bem especificado, no sentido que temos uma idéia da direção que as variáveis estão apontando, ou seja, conseguimos saber se a interação social é alta ou baixa. Os resultados indicam que a região metropolitana de São Paulo é um *hot spot* no estado³, onde encontramos um maior grau de interação social no índice de homicídio.

Este trabalho está dividido em cinco capítulos, além da presente introdução e das considerações finais. No Capítulo 2 apresentamos uma revisão da literatura sobre economia e econometria do crime, a fim de posicionar esse trabalho em relação aos outros. O Capítulo 3 faz uma revisão do arcabouço teórico que utilizaremos em nosso processo de modelagem e estimação, a saber: uma introdução à estatística espacial, inferência bayesiana e métodos *Markov Chain Monte Carlo* (MCMC). O Capítulo 4 descreve o modelo utilizado, assim como as distribuições condicionais completas para os parâmetros. Simulações de Monte Carlo são realizadas com o objetivo de testar a capacidade de estimação do modelo no Capítulo 5. Finalmente, o Capítulo 6 apresenta os resultados para a interação social na criminalidade para região metropolitana de São Paulo.

³ O termo *hot spot* para criminalidade começou a ter um maior uso na literatura a partir de Cohen e Tita (1999) que denominaram regiões que apresentam uma alta concentração de atividades criminais como um *hot spot*. Puech (2004) utiliza o termo *hot spot* para identificar uma região com uma alta concentração de homicídios, sendo este significado que adotaremos neste trabalho.

2 REVISÃO DA LITERATURA: ECONOMIA/ECONOMETRIA DO CRIME

A literatura sobre economia do crime começa a aparecer com um maior destaque a partir de Becker (1968). O autor quebra com o pensamento de que o criminoso tem algum problema mental, é excluído da sociedade e não tem condições de competir no mercado de trabalho legal. Segundo Becker o criminoso toma decisões como qualquer outro indivíduo da sociedade, ou seja, maximizando sua utilidade.

Dessa forma, utilizando um arcabouço teórico microeconômico, Becker detecta que um indivíduo cometerá um crime caso o custo esperado tenha menor valor do que o ganho esperado do crime. Posto de outra forma, o criminoso avalia os custos esperados do crime tais como probabilidade de apreensão, probabilidade de condenação, severidade da punição e custo de oportunidade de ingressar no mercado formal de trabalho; e avalia se esses custos são menores que os ganhos que ele teria com tal atividade ilícita.

Alguns trabalhos empíricos mostram a validade do modelo econômico do crime de Becker, dentre eles destacam-se: Ehrlich (1973), Freeman (1991), Levitt (1996), Levitt (1998) e Imrohoglu, Merlo e Rubert (2001). De alguma forma esses trabalhos mostram que maiores probabilidades de apreensão e condenação diminuem as taxas de criminalidade ainda, quanto mais severa for a pena (geralmente medida em média de anos na prisão) menores são as taxas de criminalidade.

Também, a relação entre criminalidade e variáveis econômicas como desigualdade de renda, salário e desemprego tem ganhado bastante vigor principalmente após o modelo desenvolvido por Burdett, Lagos e Wright (2003). Os autores expandem o modelo de *job-search* tradicional, proposto por Mortensen (1986)⁴, assim, os autores conseguem endogeneizar o índice de criminalidade em um ambiente em que há desigualdade de renda, e chegam à conclusão que quanto maiores forem o desemprego e a desigualdade de renda maior será o índice de criminalidade.

Vale citar que a maioria dos trabalhos empíricos citados anteriormente incorporam de alguma forma a desigualdade de renda em seus estudos. Pois, era quase um consenso

⁴ No modelo original de Mortensen (1986) o indivíduo pode estar empregado em um emprego formal ou desempregado. Desta maneira, o autor tenta descrever o problema encarado pelo desempregado e propõe estratégias para uma decisão ótima de emprego. Burdett, Lagos e Wright (2003) incluem a possibilidade de o indivíduo estar cometendo algum ato criminoso além, dos estados já descritos em Mortensen (1986).

que desigualdade de renda afeta a criminalidade, mesmo sem uma teoria que conseguisse incorporar no comportamento do indivíduo a desigualdade de renda.

Contudo, Glaeser, Sacerdote e Scheinkman (1996) observam uma alta variância através do tempo e espaço nas taxas criminais. E essa variância parece ser alta demais para ser explicada apenas pelos custos e benefícios do crime, além mesmo de variáveis socioeconômicas como desigualdade de renda e desemprego. Para os autores a explicação para essa alta variância é a interação social entre os criminosos, ou seja, quando a decisão de um indivíduo ingressar no mercado criminal afetar a decisão de seu vizinho, então os índices de criminalidade das cidades vão ser diferentes dos índices previstos pelas características das cidades. Dessa forma, os autores propõem um modelo baseado no modelo de votação de Kindermann e Snell (1980). A decisão de um agente sobre entrar ou não numa atividade criminosa depende de suas características (aqui entram as características da cidade) e da decisão de seu vizinho. Existem dois tipos de indivíduos: os que influenciam e são influenciados pelos vizinhos e os que influenciam, mas não são influenciados pelos vizinhos.

Para testar o grau de interação social Glaeser, Sacerdote e Scheinkman (1996) utilizam diversas taxas de criminalidade para algumas cidades dos Estados Unidos nos anos de 1985 e 1970 e para Nova York em 1985. Ou seja, os autores trabalham com dados em *cross section* para testar a interação social na criminalidade em algumas cidades dos Estados Unidos. É encontrada uma alta interação social para roubo de carro e furto, uma interação mais baixa, mas considerável, para invasão de propriedade e roubo. Baixos índices de interação social são vistos em assassinato e estupro.

Alguns trabalhos empíricos se destacam no sentido de medir uma possível interação social nos índices de criminalidade. Case e Katz (1991), Puech (2004) e Oliveira (2005) utilizam uma abordagem semelhante. Estes autores apresentam uma estimação *cross section* para índices de criminalidade, em todos eles a referência geográfica entra como uma variável explicativa do modelo. Além disso, os autores incluem outras variáveis que possam influenciar a criminalidade.

Procurando entender a criminalidade dos jovens em Boston, Case e Katz (1991) notam que morar em uma vizinhança em que uma grande parte da população participa de atividades criminais, aumenta a probabilidade de o indivíduo ingressar em tal atividade.

Puech (2004) faz uma análise para os municípios de Minas Gerais em 2000. O autor encontra uma maior dependência espacial para crimes contra a propriedade (furto e roubo) do que para crimes contra pessoa (homicídio e tentativa de homicídio). Além disso, Puech observa conglomerados para os crimes contra pessoa, estes ficam localizados principalmente em torno de cidades grandes, como por exemplo Belo Horizonte^{5,6}.

Oliveira (2005) mede a interação social para roubo, furto e homicídio nos municípios do Rio Grande do Sul no ano 2000. Nesse trabalho é também encontrada uma menor interação social para homicídio quando comparada com a interação para roubo e furto. *Hot spots* também são observados no Rio Grande do Sul.

Utilizando testes genéricos para detectar conglomerados⁷ Beato Filho et al. (2001) analisam os 240 bairros de Belo Horizonte. Os autores encontram uma maior probabilidade de ocorrência de homicídio em 10 dessas 240 regiões⁸. Eles notam que essas regiões de maior risco estão concentradas em favelas. Porém, não parece ser a favela o responsável pelo maior risco de homicídio nessas regiões, e sim o fato desses 10 bairros serem foco do tráfico de drogas.

Contudo, notamos que muitos trabalhos empíricos trabalham ou a parte espacial ou a parte temporal. Como os efeitos temporal e espacial na criminalidade não parecem ser desprezíveis, seria interessante trabalhar com modelos que contemplem essas duas características simultaneamente. Rojas (2004) vetoriza o campo observado dos modelos lineares dinâmicos de West e Harrison (1997) e modela os erros das equações de sistema seguindo um processo de campo aleatório Markoviano Gaussiano (CAMG). Dessa forma, o autor sugere alguns modelos espaço-temporais. Neste trabalho iremos utilizar o modelo linear de primeira ordem proposto por Rojas (2004).⁹

⁵ Cohen e Tita (1999), Messner et al. (1999) e Baller et al. (2001) de alguma forma também tratam de conglomerados de homicídio em alguma região.

⁶ A literatura trata desses conglomerados de homicídios como *hot spots* de uma determinada região.

⁷ Uma referência básica para testes para detecção de conglomerados é Lawson e Kulldorf (1999). A idéia do teste genérico é testar a hipótese nula de que não existe conglomerado na região de estudo contra a hipótese alternativa de que há algum conglomerado na região sem especificar onde este conglomerado estaria.

⁸ Os autores utilizam dados agregados de homicídios para o período de 1995 a 1999.

⁹ Apesar de muitos trabalhos empíricos sobre economia do crime não contemplarem os efeitos espaço-temporais, vale citar que esses modelos estão sendo muito utilizados em várias outras aplicações como ecologia e epidemiologia. Assunção, Reis e Oliveira (2001), Nobre, Schmidt e Lopes (2004) e Lawson e Viel (1995) apresentam modelos para epidemiologia. Paez (2004) e Rojas (2004) apresentam uma aplicação ecológica desses modelos.

Com isso, buscaremos identificar alguma interação social nos seguintes índices de criminalidade: homicídio, roubo e furto na região metropolitana de São Paulo. Acreditamos que modelando espaço e tempo simultaneamente teremos uma melhor estimativa da interação social desses índices. Ainda devemos ressaltar que não serão utilizadas variáveis exógenas para explicar a criminalidade, o vetor de estados (estimado endogenamente) terá o papel de capturar os efeitos de variáveis como: probabilidade de apreensão, probabilidade de condenação, severidade da pena, desigualdade de renda, etc. Com isso, podemos contornar um problema exposto em Puech (2004), o autor nota que existem dados criminais disponíveis para cidades mineiras em outros anos além do ano 2000. Porém, só temos dados para variáveis que possam ajudar a explicar a criminalidade em nível municipal para o ano 2000 (a saber, ano do último censo no Brasil). Evidente que a adição dessas variáveis melhora a percepção do problema, no sentido que podemos mensurar o quanto um aumento na severidade da pena pode reduzir a criminalidade, ou ainda, quanto da criminalidade é explicado pela desigualdade de renda. Ainda, a adição de variáveis contemporâneas deve melhorar a previsão do modelo. Mas, devido ao já mencionado problema de falta de dados criminais no Brasil (principalmente em nível municipal) o modelo escolhido tem essa vantagem.

O próximo capítulo fará uma breve introdução aos conceitos estatísticos aplicados nesse trabalho. Começamos com as bases da inferência bayesiana, em seguida apresentamos os modelos lineares dinâmicos bem com a estimação via filtro de Kalman e FFBS. Depois apresentamos dois algoritmos MCMC: amostrador de Gibbs e Metropolis-Hastings. Por fim, uma introdução a estatística espacial é realizada.

3 METODOLOGIA ESTATÍSTICA

Neste capítulo apresentaremos uma revisão de toda a metodologia estatística que utilizaremos neste trabalho. Começamos com uma revisão da idéia da inferência bayesiana, em seguida apresentamos os modelos lineares dinâmicos e o filtro de Kalman, depois partimos para o algoritmo *Forward Filtering Backward Sampling* (FFBS). Em seguida revisamos os algoritmos, MCMC, amostrador de Gibbs e Metropolis-Hastings. Ao fim do capítulo apresentamos uma pequena introdução à estatística espacial.

Algumas referências básicas para inferência bayesiana e métodos MCMC são: Gamerman e Lopes (2006), Lopes (2008), Koop (2003), Gilks, Richardson e Spiegelhalter (1996). Já para os modelos lineares dinâmicos as referências são: West e Harrison (1997) para uma abordagem bayesiana e Kim e Nelson (1999) numa abordagem clássica e bayesiana. Destacamos para o algoritmo FBBS Cappé, Moulines e Rydén (2005). Para o leitor que queira se aprofundar em estatística espacial, algumas referências básicas são Bailey e Gatrell (1995), Cressie (1996), Ripley (1981) e Banerjee, Carlin e Gelfand (2004).

3.1 INFERÊNCIA BAYESIANA

A filosofia da estatística clássica é baseada no fato de que toda a informação a respeito do fenômeno aleatório que se deseja inferir é fundamentada apenas naquilo que for observado (i.e. dados). Além disso, o estatístico clássico considera os parâmetros do modelo como quantidades fixas. Por outro lado, a escola bayesiana considera os parâmetros de interesse (por serem desconhecidos) como quantidades aleatórias¹⁰. Assim, distribuições de probabilidade para os parâmetros são necessárias na construção de um modelo bayesiano. Com isso, estatísticos bayesianos, diferentemente dos clássicos, não consideram apenas as informações oriundas dos dados como também a informação que o pesquisador tem a respeito do fenômeno estudado. O conhecimento do pesquisador sobre os parâmetros é dito “conhecimento *a priori*” sendo representado pela distribuição *a priori*. Essa

¹⁰ Nesse sentido, para um bayesiano, não existe distinção entre quantidades observáveis e os parâmetros do modelo, pois, todos são considerados quantidades aleatórias.

informação *a priori* é combinada com a informação obtida nos dados (função de verossimilhança), essa combinação é conhecida como distribuição *a posteriori*. A distribuição *a posteriori* será a base para toda a inferência que se fizer interessante.

O arcabouço teórico para se fazer inferência bayesiana está fundado em um teorema antigo de probabilidade. Considere dois eventos aleatórios A e B definidos em \mathfrak{R} , podemos escrever a probabilidade conjunta de A e B como

$$p(A, B) = p(A | B)p(B). \quad (2.1)$$

A seguinte expressão também é válida:

$$p(A, B) = p(B | A)p(A). \quad (2.1')$$

Igualando (2.1) a (2.1') temos

$$p(A | B)p(B) = p(B | A)p(A),$$

ou ainda,

$$p(B | A) = \frac{p(A | B)p(B)}{p(A)}. \quad (2.2)$$

A equação (2.2) nada mais é que o teorema de Bayes. Denominando o vetor de parâmetros (que, como mencionado anteriormente, na abordagem bayesiana passam a ser considerados variáveis aleatórias) por $\boldsymbol{\theta}$ e o vetor de observações por \mathbf{y} , podemos trocar B por $\boldsymbol{\theta}$ e A por \mathbf{y} em (2.4), obtendo

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \quad (2.3)$$

Como estamos diretamente interessados em fazer inferência sobre parâmetros, podemos ignorar o termo $p(\mathbf{y})$ ¹¹ em (2.3) e, dessa forma, reescrever essa expressão como

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.4)$$

Denominamos $p(\boldsymbol{\theta})$ como a densidade *a priori*, $p(\mathbf{y} | \boldsymbol{\theta})$ como a função de verossimilhança e $p(\boldsymbol{\theta} | \mathbf{y})$ como a densidade *a posteriori*.

A *priori*, $p(\boldsymbol{\theta})$, não depende dos dados, e contém toda informação que o pesquisador tiver sobre os parâmetros antes de conhecer os dados. As *prioris* podem ser informativas (para o caso em que o pesquisador tem algum conhecimento da distribuição dos parâmetros antes da amostragem) ou não informativas (no caso do pesquisador não dispor de conhecimento prévio da distribuição dos parâmetros antes da amostragem)¹². A função de verossimilhança, $p(\mathbf{y} | \boldsymbol{\theta})$, antes das observações serem feitas é a densidade condicional de \mathbf{y} dados os parâmetros. Por fim, a *posteriori*, $p(\boldsymbol{\theta} | \mathbf{y})$, é a distribuição atualizada de $\boldsymbol{\theta}$ depois de se conhecer as observações.

Fica claro que toda inferência, na estatística Bayesiana, é baseada na distribuição *a posteriori*. Contudo, obter analiticamente (ou até mesmo por aproximação numérica) a distribuição *a posteriori* não é das tarefas mais fáceis, uma vez que normalmente esse cálculo envolve integrais múltiplas. Dessa forma, fazem-se necessários métodos baseados em simulação para obter-se a distribuição *a posteriori*.

Atualmente, os métodos utilizados são conhecidos como *Markov Chain Monte Carlo* (MCMC)¹³. Métodos MCMC consistem em simular observações dependentes de uma cadeia de Markov. Estes algoritmos possibilitam a geração de amostras de distribuições *a posteriori* nos casos em que ela é desconhecida ou demasiadamente complicada. Os principais métodos MCMC são o amostrador de Gibbs e o algoritmo de

¹¹ Dado que a amostra já foi observada, $p(\mathbf{y})$ passa a ser uma constante em relação à $\boldsymbol{\theta}$.

¹² Dessa forma, é natural que diferentes pesquisadores tenham diferentes graus de incerteza em relação à $\boldsymbol{\theta}$ resultando em diferentes especificações para um modelo.

¹³ Métodos MCMC foram uma verdadeira revolução na estatística Bayesiana. Uma frase traduzida de Robert e Casella (2008) indica bem esse fenômeno: “Durante o início da década de 90, pesquisadores descobriram que algoritmos de Gibbs e Metropolis-Hastings poderiam acabar com qualquer problema ao qual estivessem presos (...)”.

Metroplis-Hastings os quais serão tratados mais a frente neste capítulo. Partimos agora para uma breve introdução aos modelos lineares dinâmicos.

3.2 MODELOS LINEARES DINÂMICOS

Os modelos lineares dinâmicos numa abordagem bayesiana são apresentados inicialmente por Harrison e Stevens (1976). West e Harrison (1997) descrevem detalhadamente tais modelos. A idéia dos mesmos é relacionar uma variável observável com um vetor de estados (não observável) e um erro. Chamaremos essa equação de equação de observação ou de medida. Também teremos uma equação de transição que descreve a evolução do vetor de estados. Uma formulação geral para um problema desse tipo é

$$\mathbf{y}_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N(0, \mathbf{V}_t) \quad (2.5)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N(0, \mathbf{W}_t) \quad (2.6)$$

para $t = 1, \dots, T$, onde, (2.5) é a equação de medida ou observação e (2.6) é a equação de transição. Ainda, \mathbf{y}_t é um vetor de variáveis observáveis; $\boldsymbol{\theta}_t$ é o vetor de estados (latente); \mathbf{F}_t' é uma matriz que relaciona o processo não-observável com as observações; $\boldsymbol{\varepsilon}_t$ é o erro da equação de observação (com média zero, não correlacionado entre si e tem matriz de variância-covariância \mathbf{V}_t); \mathbf{G}_t retrata a evolução do processo não-observado no tempo; $\boldsymbol{\omega}_t$ é o erro da equação de medida (com média zero, não correlacionado entre si e tem matriz de variância-covariância \mathbf{W}_t). Além disso, os erros $\boldsymbol{\varepsilon}_t$ e $\boldsymbol{\omega}_t$ são independentes.

3.3 FILTRO DE KALMAN

A estimação do modelo proposto é feita através do filtro de Kalman, proposto em Kalman (1960). O filtro de Kalman é um procedimento recursivo para estimar o vetor de estados $\boldsymbol{\theta}_t$, assumindo que \mathbf{F}_t' , \mathbf{G}_t , \mathbf{V}_t , e \mathbf{W}_t são conhecidos. Basicamente, o filtro

consiste de dois passos. O primeiro é a previsão, em que queremos calcular o valor de \mathbf{y}_t , baseado em todas as informações até $t-1$ ($\tilde{\mathbf{y}}_{t|t-1}$), para isso precisamos calcular $\hat{\boldsymbol{\theta}}_{t|t-1}$. O próximo passo é a atualização, depois do tempo t “acontecer” devemos calcular o erro de previsão ($v_t = \mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1}$). Assim, após conhecer \mathbf{y}_t , uma melhor estimativa de $\boldsymbol{\theta}_t$ pode ser feita.

O filtro de Kalman irá nos fornecer a distribuição condicional de $\boldsymbol{\theta}_t$ dada a informação disponível D_t ($D_t = \{\mathbf{y}_t, D_{t-1}\}$). Supomos que a priori em $t=0$ é normal multivariada ($\boldsymbol{\theta}_0 | D_0$) $\sim N(\mathbf{m}_0, \mathbf{C}_0)$ para algum vetor de médias \mathbf{m}_0 e matriz de variância-covariância \mathbf{C}_0 conhecidos.

Assim, para algum vetor de médias \mathbf{m}_{t-1} e matriz de variância-covariância \mathbf{C}_{t-1} , seja a posteriori em $t-1$

$$(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}).$$

Para encontrarmos a média da distribuição a priori de $\boldsymbol{\theta}_t$ em t , basta aplicarmos o operador esperança em (2.6):

$$\begin{aligned} \boldsymbol{\theta}_{t|t-1} &= E_{t-1}(\mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t), \\ \boldsymbol{\theta}_{t|t-1} &= \mathbf{G}_t \mathbf{m}_{t-1}. \end{aligned}$$

O erro de previsão do vetor de estados será dado por

$$\begin{aligned} \boldsymbol{\theta}_{t|t-1} - \boldsymbol{\theta}_t &= \mathbf{G}_t \mathbf{m}_{t-1} - \mathbf{G}_t \boldsymbol{\theta}_{t-1} - \boldsymbol{\omega}_t, \\ \boldsymbol{\theta}_{t|t-1} - \boldsymbol{\theta}_t &= \mathbf{G}_t (\mathbf{m}_{t-1} - \boldsymbol{\theta}_{t-1}) - \boldsymbol{\omega}_t. \end{aligned}$$

A matriz de variância-covariância da distribuição a priori de $\boldsymbol{\theta}_t$ em t , será a variância do erro de previsão.

$$\mathbf{R}_t = E_{t-1} \left[(\boldsymbol{\theta}_{t|t-1} - \boldsymbol{\theta}_t)(\boldsymbol{\theta}_{t|t-1} - \boldsymbol{\theta}_t)' \right],$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t.$$

Desse modo a distribuição de $\boldsymbol{\theta}_t$ a priori em t é

$$(\boldsymbol{\theta}_t | D_{t-1}) \sim N(\mathbf{G}_t \mathbf{m}_{t-1}, \mathbf{R}_t).$$

Aplicando o operador esperança em (2.5) podemos obter a previsão de \mathbf{y}_t até $t-1$, que será dado por

$$\tilde{\mathbf{y}}_{t|t-1} = E_{t-1} \left(\mathbf{F}_t' \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \right),$$

$$\tilde{\mathbf{y}}_{t|t-1} = \mathbf{F}_t' \mathbf{G}_t \mathbf{m}_{t-1}.$$

Assim, o erro de previsão das observações um passo a frente será

$$\mathbf{v}_t = \mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1},$$

$$\mathbf{v}_t = \mathbf{F}_t' (\boldsymbol{\theta}_t - \mathbf{G}_t \mathbf{m}_{t-1}) + \boldsymbol{\varepsilon}_t.$$

A variância do erro de previsão um passo a frente é dado por

$$\mathbf{Q}_t = E_{t-1} [\mathbf{v}_t \mathbf{v}_t'],$$

$$\mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t.$$

A distribuição da previsão das observações um passo a frente é

$$(\mathbf{y}_t | D_{t-1}) \sim N \left(\mathbf{F}_t' \mathbf{G}_t \mathbf{m}_{t-1}, \mathbf{Q}_t \right).$$

Com as informações até agora podemos escrever

$$\begin{pmatrix} \boldsymbol{\theta}_t \\ \mathbf{y}_t \end{pmatrix} | D_{t-1} \sim N \left[\begin{pmatrix} \boldsymbol{\theta}_{t|t-1} \\ \boldsymbol{\theta}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_t & \mathbf{R}_t \\ \mathbf{R}_t & \mathbf{Q}_t \end{pmatrix} \right].$$

Portanto a posteriori em t é

$$(\boldsymbol{\theta}_t | D_t) \sim N(\mathbf{m}_t, \mathbf{C}_t),$$

onde, $\mathbf{m}_t = \boldsymbol{\theta}_{t|t-1} + \mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1} (\mathbf{y}_t - \mathbf{F}_t' \boldsymbol{\theta}_{t|t-1})$ e $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1} \mathbf{Q}_t (\mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1})'$.

3.4 FORWARD FILTERING BACKWARD SAMPLING

O algoritmo FFBS é uma maneira de gerar $\boldsymbol{\theta}_{1:T}$ condicionado ao conjunto completo de informações. Assim, a idéia básica é amostrar $\boldsymbol{\theta}_t$ de $N(\mathbf{m}_t, \mathbf{C}_t)$ (*Forward Filtering*) e amostrar $\boldsymbol{\theta}_{t-1}$ de $(\boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t, \mathbf{D}_T)$ (*Backward Sampling*). Podemos descrever o algoritmo em três passos:

- 1- Utilizar o filtro de Kalman para encontrar a média e variância das distribuições *a posteriori*.
- 2- Gerar um valor de $p(\boldsymbol{\theta}_T | \mathbf{D}_T)$.
- 3- Calcular média e variância de $p(\boldsymbol{\theta}_{T-j} | \boldsymbol{\theta}_{T-j+1}, \mathbf{D}_T)$, onde $j = 1, 2, \dots, T-1$. Gerar $\boldsymbol{\theta}_{T-j}$ desta distribuição.

Com isso, cada $\boldsymbol{\theta}_t$ gerado está condicionado ao conjunto completo das observações. Os estados são gerados todos de uma só vez, com a vantagem da ordenação do tempo do modelo espaço de estados.

A seguir, apresentamos os seguintes métodos MCMC: amostrador de Gibbs e Metropolis-Hastings.

3.5 AMOSTRADOR DE GIBBS

O amostrador de Gibbs foi introduzido por Geman e Geman (1984) num contexto de processamento de imagens, posteriormente Gelfand e Smith (1990) mostraram os benefícios do algoritmo para a análise bayesiana.

Assim como anteriormente a distribuição de interesse é $p(\boldsymbol{\theta} | \mathbf{y})$, onde, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$. Foi mencionado que amostrar de $p(\boldsymbol{\theta} | \mathbf{y})$ pode ser uma tarefa complicada, porém, sob escolhas adequadas, pode ser razoavelmente simples retirar amostras de $p(\theta_{(1)} | \mathbf{y}, \theta_{(2)}, \dots, \theta_{(k)})$, $p(\theta_{(2)} | \mathbf{y}, \theta_{(1)}, \theta_{(3)}, \dots, \theta_{(k)})$, ..., $p(\theta_{(k)} | \mathbf{y}, \theta_{(1)}, \dots, \theta_{(k-1)})$, ou de forma reduzida $p(\theta_{(i)} | \mathbf{y}, \boldsymbol{\theta}_{(-i)})$. Chamamos essas distribuições de distribuições condicionais completas.

Podemos descrever o processo de amostragem de Gibbs da seguinte maneira:

1. Escolher os valores iniciais dos parâmetros $\boldsymbol{\theta}^{(0)} = (\theta_{(1)}^{(0)}, \dots, \theta_{(k)}^{(0)})'$.
2. Tome uma amostra aleatória, $\theta_{(1)}^{(j)}$ de $p(\theta_{(1)} | \mathbf{y}, \theta_{(2)}^{(j-1)}, \dots, \theta_{(k)}^{(j-1)})$.
3. Tome uma amostra aleatória, $\theta_{(2)}^{(j)}$ de $p(\theta_{(2)} | \mathbf{y}, \theta_{(1)}^{(j)}, \theta_{(3)}^{(j-1)}, \dots, \theta_{(k)}^{(j-1)})$.
4. Tome uma amostra aleatória, $\theta_{(3)}^{(j)}$ de $p(\theta_{(3)} | \mathbf{y}, \theta_{(1)}^{(j)}, \theta_{(2)}^{(j)}, \theta_{(4)}^{(j-1)}, \dots, \theta_{(k)}^{(j-1)})$.
- ⋮
- k . Tome uma amostra aleatória, $\theta_{(k)}^{(j)}$ de $p(\theta_{(k)} | \mathbf{y}, \theta_{(1)}^{(j)}, \theta_{(2)}^{(j)}, \dots, \theta_{(k-1)}^{(j)})$.

Repita os passos de 2 – k até a convergência (assumimos que a convergência ocorre a partir da iteração m)^{14,15}. Aqui fica claro que essa seqüência é uma cadeia de Markov, isso porque, a iteração j depende dos valores da cadeia no passo $j - 1$.

¹⁴ A literatura apresenta algumas maneiras de determinar convergência. Podemos separá-las em dois grupos: métodos teóricos e empíricos. Robert (1995) e Cowles e Carlin (1996) fazem uma revisão das duas vertentes. Basicamente os métodos teóricos tentam medir distâncias e estabelecer fronteiras através das funções de distribuição geradas pelas cadeias. Já os métodos empíricos são construídos através de algumas estatísticas baseadas nas simulações. Devido a dificuldade de aplicação em modelos práticos dos métodos teóricos, esse trabalho irá utilizar apenas o método empírico.

¹⁵ O período até a convergência é conhecido na literatura como *burn-in*.

Existem alguns métodos para obter uma amostra de tamanho n a partir da cadeia apresentada. Uma maneira é rodar n cadeias até a convergência e retirar um único elemento de cada cadeia. Dessa forma, teremos mn gerações da cadeia para esse caso é fácil obter independência, apenas devemos definir valores iniciais diferentes e com grande dispersão para $\theta^{(0)}$. Contudo, fica claro que essa maneira é muito custosa computacionalmente.

Uma alternativa é rodar apenas uma cadeia e retirar n elementos seguidos dessa cadeia após convergência. Esse processo requer $m + n$ gerações da cadeia. Um possível problema com essa forma é o fato de a amostra ter dependência devido à dependência da cadeia. Teoremas ergódicos garantem que a inferência baseada numa amostra como essa ainda é válida, porém, se a amostra não for muito grande e se a autocorrelação da cadeia for muito alta algumas regiões serão subrepresentadas na amostra. Uma forma de contornar o problema da dependência é tomar elementos dessa cadeia a cada p iterações após o período de *burn-in*, à medida que p cresce os valores ficam menos correlacionados e “virtualmente” independentes. Aqui temos $m + pn$ gerações da cadeia.

Ainda, podemos rodar l cadeias independentes¹⁶, e retirar uma amostra n/l de cada cadeia após o *burn-in*. Esse processo necessita de $lm + n$ gerações da cadeia. Assim como foi dito anteriormente podemos extrair os elementos de cada cadeia a cada p iterações após o período de *burn-in*. Nesse caso, temos $lm + pn$ gerações da cadeia.

A próxima seção apresenta o algoritmo conhecido como Metropolis-Hastings.

3.6 METROPOLIS-HASTINGS

Se alguma das distribuições condicionais completas não tem forma fechada é possível utilizar o algoritmo de Metropolis-Hastings, proposto por Metropolis et al. (1953) e Hastings (1970). O algoritmo tem os seguintes passos.

1. Escolha o valor inicial, $\theta^{(0)}$.

¹⁶ Geralmente escolhemos um número pequeno para l , por exemplo, menor que dez.

2. Retire θ^* da densidade candidata geradora ($q(\theta^{(j-1)}; \theta)$).
3. Calcule a probabilidade de aceitação

$$\alpha(\theta^{(j-1)}; \theta^*) = \min \left[1, \frac{p(\theta = \theta^*) p(\mathbf{y} | \theta = \theta^*) q(\theta^*; \theta = \theta^{(j-1)})}{p(\theta = \theta^{(j-1)}) p(\mathbf{y} | \theta = \theta^{(j-1)}) q(\theta^{(j-1)}; \theta = \theta^*)} \right].$$
4. Defina $\theta^{(j)} = \theta^*$ com probabilidade $\alpha(\theta^{(j-1)}; \theta^*)$ e $\theta^{(j)} = \theta^{(j-1)}$ com probabilidade $1 - \alpha(\theta^{(j-1)}; \theta^*)$.

Repita os passos de 2–4 até a convergência. Vale notar que assim como o amostrador de Gibbs o algoritmo de Metropolis-Hastings é um método de MCMC. Isso porque, o candidato θ^* é retirado da variável aleatória θ em que a densidade depende de $\theta^{(j-1)}$. Devemos registrar que podemos utilizar o algoritmo de Metropolis-Hastings como um passo dentro do amostrador de Gibbs.

Claramente, a densidade candidata geradora não é igual à *posteriori*. Para que não aconteça o problema de retirar a amostra de um espaço paramétrico errado, o algoritmo corrige isso através da probabilidade de aceitação. Isso porque a probabilidade será maior nas áreas onde a probabilidade da *posteriori* for maior e menor nas áreas onde a probabilidade da *posteriori* for menor. Ou seja, se $\theta^{(j-1)}$ está numa área de baixa probabilidade da *posteriori* o algoritmo tende a se mover rapidamente de $\theta^{(j-1)}$, por outro lado, se $\theta^{(j-1)}$ está numa área de alta probabilidade da *posteriori* o algoritmo tende a ficar em $\theta^{(j-1)}$.

Porém, devemos ter atenção para a escolha da densidade candidata geradora. A literatura nessa área não é conclusiva, probabilidades de aceitação muito baixas vão fazer a cadeia “se mexer” pouco, o que pode levar a cadeia a visitar pouco regiões de alta probabilidade. Por outro lado, probabilidades de aceitação muito altas fazem a cadeia “se mexer” muito, o que pode levar a cadeia a visitar muitas vezes regiões de baixa probabilidade. Neste trabalho seguiremos a idéia de Bennet, Racine-Poon e Wakefield (1995) que consideram ótimas probabilidades de aceitação em torno de 20% e 50%.

Deve ficar claro que o processo de obtenção de uma amostra de tamanho n no algoritmo de Metropolis-Hastings é idêntico ao citado na seção anterior. A próxima seção trata das bases da estatística espacial.

3.7 ESTATÍSTICA ESPACIAL

O ramo “estatística espacial” estuda dados que possuem coordenadas geográficas. Aqui o conjunto de índices do espaço terá mais de um elemento, além disso, a referência geográfica deve ser usada explicitamente no modelo. O uso de referência geográfica parece fazer bastante sentido, pois é intuitivo pensar que duas regiões muito próximas são correlacionadas, e essa correlação deve ser menor quanto mais distante as regiões forem umas das outras.

É importante destacar os diferentes tipos de dados espaciais, temos: dados de processos pontuais, dados de superfícies aleatórias, dados de interação espacial e dados de área.

- Dados de processos pontuais são dados onde o conjunto de coordenadas geográficas representa as localizações exatas dos eventos (por exemplo: as residências que foram invadidas por ladrões), aqui o que é aleatório é a localização do evento no mapa. Problemas como: saber se existe alguma aglomeração nos casos ou se estar próximo a uma “fonte” (por exemplo, favelas) aumenta o risco de ocorrências do evento, são típicos de estudo de dados de processos pontuais.
- Dados de superfície aleatória, aqui considera-se a superfície do evento como aleatória. Esta categoria compreende o estudo da geoestatística teremos n estações de coleta de dados (acidez do solo, temperatura do solo, pH da água, etc.) em localizações fixas. Este tipo de estudo pode ter como objetivo prever a característica de uma superfície não monitorada, ou escolher uma localização para sediar uma próxima estação de coleta de dados. É importante ressaltar que ao contrario dos dados de processos pontuais, o processo que gerou o posicionamento dos dados no mapa não é de interesse.
- Dados de interação social: há pares ordenados de estações fixas, onde um elemento do par é considerado a origem e o outro é o destino, e um fluxo trafega entre os pares. Dados comuns nessa categoria são fluxo de telefonemas entre torres de captação, fluxo migratório, fluxo de passageiros entre cidades, etc. Um

estudo comum para este tipo de dado é verificar quais características das estações (tamanho, distância, PIB, etc.) determinam o volume de fluxo entre estações.

- Dados de área: são dados que representam uma agregação em determinado espaço de um mapa. Assim, o dado não está associado a uma posição específica e sim a toda a área em questão. Aqui as áreas são particionadas de acordo com uma divisão política (continentes, países, estados, cidades, etc.) ou características das regiões (zonas climáticas, culturais, econômicas, etc.). Neste trabalho estaremos trabalhando com dados de área, pois temos índices de criminalidade para cada cidade da região metropolitana de São Paulo.

Como mencionado anteriormente, a referência geográfica deve ser usada explicitamente no modelo, assim, devemos de alguma forma “traduzir” o mapa para uma linguagem acessível ao modelo. A forma mais comum de se “traduzir” o mapa para o modelo é através da matriz de pesos espaciais, essa matriz geralmente é quadrada de dimensão $n \times n$ (onde, n é o número de regiões) e denotada por \mathbf{W} , a qual mede a proximidade geográfica, ou estrutura de vizinhança, entre regiões. Tal proximidade entre as regiões i e j será medida pelo elemento w_{ij} . Existem diversas maneiras de medir a proximidade de uma região: podemos considerar w_{ij} como a distância entre os centros da região i e j , ainda, pode-se pensar em w_{ij} como sendo proporcional ao fluxo de pessoas que residem em i e trabalham em j , ou o mais comum, tomando o elemento $w_{ij} = 0$ para regiões que não fazem fronteira e $w_{ij} = k$ para regiões que fazem fronteira¹⁷.

No modelo utilizado neste trabalho a matriz \mathbf{W} “vai entrar” no erro da equação de medida. O próximo capítulo, que tratará do modelo, mostra como nossa matriz de pesos espaciais é construída.

¹⁷ Esta estrutura de vizinhança é comumente chamada de vizinhança de primeira ordem.

4 O MODELO

4.1 MODELO POLINOMIAL DE PRIMEIRA ORDEM

Vamos utilizar o modelo polinomial de primeira ordem proposto por Rojas (2004). Esse modelo é uma generalização do modelo linear dinâmico de West e Harrison (1997), com erros nas equações de observação e sistema seguindo campos aleatórios Markovianos Gaussianos (CAMG).

Consideramos que temos n regiões, indexadas pelos inteiros $1, 2, \dots, n$, formando uma grade dentro de uma área de interesse, e T observações para o tempo $(1, 2, \dots, T)$. Para cada tempo t e região s observamos a variável de interesse, y_{st} , $s = 1, 2, \dots, n$, $t = 1, 2, \dots, T$. Assim, denotando \mathbf{y}_t como $(y_{1t}, \dots, y_{nt})'$ o modelo fica:

$$\mathbf{y}_t = \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (3.1)$$

$$\boldsymbol{\theta}_t = \rho \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim \text{CAMG}(\mathbf{0}, \mathbf{W}_t^{-1}) \quad (3.2)$$

onde, $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T$ e $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_T$ os erros da equação de observação e sistemas respectivamente e são independentes; $\boldsymbol{\theta}_t$ é o processo espaço-temporal não-observado (essa variável latente irá conter o efeito de todas as variáveis que afetam criminalidade como: probabilidade de apreensão, probabilidade de condenação, severidade da pena, desigualdade de renda, etc.); ρ retrata a evolução do processo não-observado no tempo, note que caso $\rho = \pm 1$, o modelo é integrado de primeira ordem, se ρ estiver no intervalo $(-1, 1)$ o modelo é estacionário; e σ^2 representa a variância dos erros da equação de observação, note que estamos considerando que os erros da equação de observação são independentes. A matriz \mathbf{W}_t^{-1} desempenha o papel de descrever a dependência espacial (interação social) do sistema.

Antes de continuar devemos definir a estrutura de vizinhança do sistema. Desse modo, definimos $N_k = \{l : l \text{ é vizinho de } k\}$ como o conjunto de vizinhança N_k de uma

região k . Neste trabalho será considerada vizinha qualquer região que fizer fronteira com a região k .

Definimos um campo aleatório Markoviano Gaussiano da seguinte forma cada região possui um conjunto de regiões vizinhas, e para cada região k , a distribuição condicional de z_k dado z_{-k} (o conjunto de todos os z exceto o k -ésimo) depende só dos z 's em N_k , ou seja, $p(z_k | z_{-k}) = p(z_k | z_l, l \in N_k)$.

Pode-se definir a estrutura de correlação de um campo aleatório Markoviano Gaussiano como $\kappa \mathbf{M} = \Sigma^{-1}$, onde κ é um parâmetro de escala, ao invés da matriz de covariâncias Σ . Assim, definimos a matriz \mathbf{M} de dimensão $(n \times n)$ como

$$(\mathbf{M})_{k,l} = \begin{cases} m_k, & k = l \\ -g_{kl}, & k \in N_l \\ 0, & \text{caso contrário} \end{cases}, \quad (3.3)$$

onde, $g_{kl} > 0$ é uma medida de similaridade entre as regiões k e l , e $m_k = \sum_{l \in N_k} g_{kl}$. Aqui, sem perda de generalidade, consideramos $g_{kl} = 1$.

Porém, com a construção proposta, a matriz \mathbf{M} é singular, dessa forma, a distribuição conjunta de z não é própria. Como precisaremos de densidades bem definidas utilizaremos uma matriz de precisão $\mathbf{W}_t^{-1} = \kappa(\alpha \mathbf{I}_n + \mathbf{M})$, onde κ é um parâmetro de escala, $\alpha > 0$ é um parâmetro que capta o grau de correlação espacial e \mathbf{I}_n é a matriz identidade de dimensão $(n \times n)$. Deste modo, estamos inserindo α à diagonal principal de \mathbf{M} , tornando \mathbf{W}_t^{-1} uma matriz diagonal dominante estrita¹⁸, logo, \mathbf{W}_t^{-1} passa a ser positiva definida. Caso o valor de α seja pequeno a correlação espacial entre as regiões é alta, se α aumenta a correlação espacial entre as regiões cai.

¹⁸ Uma Matriz \mathbf{A} é dita diagonal dominante se $|a_{ii}| \geq \sum_{i \neq j} |a_{ij}|$ para todo i . Quando todas as inequações são estritas chamamos a matriz de diagonal dominante estrita. Pode-se provar que uma matriz diagonal dominante estrita é não singular e se todos os elementos da diagonal principal são positivos a matriz será positiva definida. Para mais informações sobre matrizes diagonal dominante ver Golub e Van Loan (1996).

Fica claro que a variável de interesse para testar a interação social nos índices de criminalidade é o α , quanto menor este parâmetro maior a interação social do índice em questão. Assim como, também é claro que ρ retrata a correlação temporal da média do índice de criminalidade.

4.2 ESTIMAÇÃO

A estimação do vetor de estados será feita via filtro de Kalman e algoritmo *Forward Filtering Backward Sampling* (FFBS), já as estimações dos parâmetros $(\rho, \alpha, \kappa, \sigma^2)$ e de θ_0 será feita através dos algoritmos MCMC: amostrador de Gibbs e Metropolis-Hastings. Devemos começar definindo as *prioris*.

4.2.1 *Prioris*

Iremos utilizar as mesmas distribuições *a priori* propostas por Rojas (2004).

- Uma gama inversa para σ^2 – $\sigma^2 \sim GI\left(\frac{n_0}{2}, \frac{n_0 s_0^2}{2}\right)$ – com hiperparâmetros n_0 e s_0^2 conhecidos.
- ρ segue uma uniforme entre -1 e 1 – $\rho \sim U(-1,1)$.
- Para α e κ utilizamos a *priori* de referência de Ferreira e Oliveira (2004)

$$p(\alpha, \kappa) \propto \kappa^{-1} \sqrt{(n-1) \sum_{i=1}^{n-1} (\lambda_i + \alpha)^{-2} - \left[\sum_{i=1}^{n-1} (\lambda_i + \alpha)^{-1} \right]^2}, \text{ onde, } \lambda_i \text{ } i = 1, \dots, n \text{ são}$$

os autovalores da matriz \mathbf{M} de vizinhança, tais que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = 0$. Fica claro na *priori* para α e κ que α e κ são independentes *a priori*.

- A distribuição *a priori* de θ_0 é $p(\theta_0 | \rho, \alpha, \kappa) = CAMG(\mathbf{0}, (1 - \rho^2) \mathbf{W}_t^{-1})$, assumindo estacionariedade temporal.

Como a distribuição conjunta *a posteriori* é proporcional à verossimilhança vezes a *priori*, teremos a seguinte distribuição *a posteriori*

$$p(\boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \sigma^2, \rho, \alpha, \kappa | D_T) \propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \sigma^2) \right] \left[\prod_{t=1}^T p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \rho, \alpha, \kappa) \right] \times p(\boldsymbol{\theta}_0 | \rho, \alpha, \kappa) p(\sigma^2) p(\rho) p(\kappa) p(\alpha). \quad (3.4)$$

As próximas subseções irão tratar das equações do filtro de Kalman, FFBS e das condicionais completas para cada parâmetro de interesse.

4.2.2 Filtro de Kalman e FFBS

As equações do filtro de Kalman ficam conforme a seguir para esse modelo:

- A distribuição de $\boldsymbol{\theta}_t$ *a priori* em t é $(\boldsymbol{\theta}_t | D_{t-1}) \sim N(\rho \mathbf{m}_{t-1}, \mathbf{R}_t)$, onde $\mathbf{R}_t = \rho^2 \mathbf{C}_{t-1} + \mathbf{W}_t$.
- A distribuição da previsão um passo a frente é $(\mathbf{y}_t | D_{t-1}) \sim N(\rho \mathbf{m}_{t-1}, \mathbf{Q}_t)$, onde $\mathbf{Q}_t = \mathbf{R}_t + \sigma^2 \mathbf{I}_n$.
- A distribuição $\boldsymbol{\theta}_t$ *a posteriori* em t é $(\boldsymbol{\theta}_t | D_t) \sim N(\mathbf{m}_t, \mathbf{C}_t)$, onde $\mathbf{m}_t = \boldsymbol{\theta}_{t|t-1} + \mathbf{R}_t \mathbf{Q}_t^{-1} (\mathbf{y}_t - \boldsymbol{\theta}_{t|t-1})$ e $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{Q}_t^{-1} \mathbf{Q}_t (\mathbf{R}_t \mathbf{Q}_t^{-1})'$.

Já as equações do algoritmo FFBS são construídas a partir da equação (3.4)

$$V(\boldsymbol{\theta}_{t-j} | \boldsymbol{\theta}_{t-j+1}, \mathbf{D}_T, \sigma^2, \rho, \alpha, \kappa, \boldsymbol{\theta}_0) = (\rho^2 \mathbf{W}^{-1} + \mathbf{C}_{t-j}^{-1})^{-1},$$

$$E(\boldsymbol{\theta}_{t-j} | \boldsymbol{\theta}_{t-j+1}, \mathbf{D}_T, \sigma^2, \rho, \alpha, \kappa, \boldsymbol{\theta}_0) = (V(\boldsymbol{\theta}_{t-j} | \boldsymbol{\theta}_{t-j+1}, \mathbf{D}_T, \sigma^2, \rho, \alpha, \kappa)) (\mathbf{C}_{t-j}^{-1} + \mathbf{m}_{t-j} + \rho \mathbf{W}^{-1} \boldsymbol{\theta}_{t-j+1})$$

4.2.3 Condicionais Completas

4.2.3.1 $\boldsymbol{\theta}_0$ (“Vetor de estados”)

Para obter a condicional completa de $\boldsymbol{\theta}_0$ combinamos os termos de (3.4) que dependem de $\boldsymbol{\theta}_0$. Assim,

$$\begin{aligned}
p(\boldsymbol{\theta}_0 | \boldsymbol{\theta}_1, \rho, \alpha, \kappa) &\propto p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_0, \rho, \alpha, \kappa) p(\boldsymbol{\theta}_0 | \rho, \alpha, \kappa), \\
p(\boldsymbol{\theta}_0 | \boldsymbol{\theta}_1, \rho, \alpha, \kappa) &\propto \exp\left\{-\frac{1}{2} \boldsymbol{\theta}_0' \mathbf{W}^{-1} \boldsymbol{\theta}_0 - 2 \boldsymbol{\theta}_0' (\rho \mathbf{W}^{-1} \boldsymbol{\theta}_1)\right\}, \\
(\boldsymbol{\theta}_0 | \boldsymbol{\theta}_1, \rho, \alpha, \kappa) &\sim N(\rho \boldsymbol{\theta}_1, \mathbf{W}).
\end{aligned}$$

Como a condicional completa de $\boldsymbol{\theta}_0$ é conhecida (normal multivariada com vetor de médias $\rho \boldsymbol{\theta}_1$ e matriz de variância-covariância \mathbf{W}), a simulação desse parâmetro será feita via amostrador de Gibbs.

4.2.3.2 σ^2 (“Variância da equação de medida”)

A condicional completa de σ^2 se dá multiplicando os termos de (3.4) que dependem de σ^2 .

$$\begin{aligned}
p(\sigma^2 | \boldsymbol{\theta}_{1:T}, \mathbf{D}_T) &\propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \sigma^2, \mathbf{D}_{t-1}) \right] (\sigma^2)^{-(n_0/2+1)} \exp\left\{-\frac{n_0 s_0^2}{2\sigma^2}\right\}, \\
p(\sigma^2 | \boldsymbol{\theta}_{1:T}, \mathbf{D}_T) &\propto (\sigma^2)^{-((nT+n_0)/2+1)} \exp\left\{-\frac{1}{2\sigma^2} \left(n_0 s_0^2 + \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\theta}_t)' (\mathbf{y}_t - \boldsymbol{\theta}_t) \right)\right\}, \\
(\sigma^2 | \boldsymbol{\theta}_{1:T}, \mathbf{D}_T) &\sim GI\left(\frac{a_{\sigma^2}}{2}, \frac{b_{\sigma^2}}{2}\right),
\end{aligned}$$

onde, $a_{\sigma^2} = nT + n_0$ e $b_{\sigma^2} = n_0 s_0^2 + \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\theta}_t)' (\mathbf{y}_t - \boldsymbol{\theta}_t)$. Também para σ^2 utilizaremos o amostrador de Gibbs para a simulação, uma vez que a condicional completa é conhecida (gama inversa).

4.2.3.3 ρ (“Correlação temporal”)

Voltamos à equação (3.4) e combinamos os termos que dependem de ρ para obter

$$p(\rho \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \kappa) \propto \left[\prod_{t=1}^T p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \alpha, \kappa, \rho) \right] p(\boldsymbol{\theta}_0 \mid \alpha, \kappa, \rho) \mathbf{1}_{(-1,1)}(\rho)^{19},$$

$$p(\rho \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \kappa) \propto \exp \left\{ -\frac{1}{2} \left[\rho^2 \left(\sum_{t=1}^{T-1} \boldsymbol{\theta}_t' \mathbf{W}^{-1} \boldsymbol{\theta}_t \right) - 2\rho \sum_{t=1}^T \boldsymbol{\theta}_{t-1}' \mathbf{W}^{-1} \boldsymbol{\theta}_t \right] \right\} (1 - \rho^2)^{\nu/2} \mathbf{1}_{(-1,1)}(\rho).$$

Como não conhecemos esta distribuição, aplicamos o algoritmo de Metropolis-Hastings para simulação desse parâmetro. Denominando ρ^p como o valor de ρ proposto em cada passo do algoritmo e $\rho^{(j-1)}$ o valor de ρ em $j-1$ (anterior ao passo atual), escolhemos como distribuição proposta $(q(\rho^{(j-1)}, \rho^p))$ uma normal truncada entre -1 e 1 , com variância $\nu_\rho = \left[\sum_{t=1}^{T-1} \boldsymbol{\theta}_t' \mathbf{W}^{-1} \boldsymbol{\theta}_t \right]^{-1}$ e média $m_\rho = \nu_\rho \left[\sum_{t=1}^T \boldsymbol{\theta}_{t-1}' \mathbf{W}^{-1} \boldsymbol{\theta}_t \right]$. Dessa forma, podemos reescrever a condicional completa de ρ da seguinte forma:

$$p(\rho \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \kappa) \propto NTr_{(-1,1)}(m_\rho, \nu_\rho) (1 - \rho^2)^{\nu/2}.$$

Assim, temos a seguinte probabilidade de aceitação,

$$\alpha(\rho^{(j-1)}; \rho^p) = \min \left[1, \frac{p(\rho^p \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \kappa) q(\rho^p, \rho^{(j-1)})}{p(\rho^{(j-1)} \mid \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \kappa) q(\rho^{(j-1)}, \rho^p)} \right],$$

$$\alpha(\rho^{(j-1)}; \rho^p) = \min \left[1, \frac{(1 - (\rho^p)^2)^{\nu/2}}{(1 - (\rho^{(j-1)})^2)^{\nu/2}} \right].$$

Teremos $\rho^{(j)} = \rho^p$ com probabilidade $\alpha(\rho^{(j-1)}; \rho^p)$ e $\rho^{(j)} = \rho^{(j-1)}$ com probabilidade $1 - \alpha(\rho^{(j-1)}; \rho^p)$.

¹⁹ $\mathbf{1}_{(-1,1)}(\rho)$ representa a função indicadora, igual a 1 se $-1 \leq \rho \leq 1$ e igual a 0 caso contrário.

4.2.3.4 κ (“Parâmetro de escala”)

A condicional completa de κ é obtida através dos termos de (3.4) que dependem do parâmetro de escala.

$$p(\kappa | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \rho) \propto \left[\prod_{t=1}^T p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \alpha, \kappa, \rho) \right] p(\boldsymbol{\theta}_0 | \alpha, \kappa, \rho) p(\kappa),$$

$$p(\kappa | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \rho) \propto \kappa^{\frac{(nT+n)}{2}-1} \exp \left\{ -\frac{\kappa}{2} \left[\sum_{t=1}^T (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1})' (\mathbf{M} + \alpha \mathbf{I}_n) (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1}) + (1 - \rho^2) \boldsymbol{\theta}_0' (\mathbf{M} + \alpha \mathbf{I}_n) \boldsymbol{\theta}_0 \right] \right\},$$

$$(\kappa | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \alpha, \rho) \sim G \left(\frac{a_\kappa}{2}, \frac{b_\kappa}{2} \right),$$

onde, $a_\kappa = n(T+1)$ e $b_\kappa = \sum_{t=1}^T (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1})' (\mathbf{M} + \alpha \mathbf{I}_n) (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1}) + (1 - \rho^2) \boldsymbol{\theta}_0' (\mathbf{M} + \alpha \mathbf{I}_n) \boldsymbol{\theta}_0$.

A distribuição da condicional completa de κ é uma gama (conhecida), assim, para o parâmetro de escala utilizaremos o amostrador de Gibbs para simulação.

4.2.3.5 α (“Correlação espacial”)

Finalmente temos a seguinte condicional completa para α , combinando os termos de (3.4) que dependem desse parâmetro.

$$p(\alpha | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \kappa, \rho) \propto \left[\prod_{t=1}^T p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \alpha, \kappa, \rho) \right] p(\boldsymbol{\theta}_0 | \alpha, \kappa, \rho) p(\alpha),$$

$$p(\alpha | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \kappa, \rho) \propto \left[\prod_{i=1}^n (\lambda_i - \alpha) \right]^{\frac{T+1}{2}} \exp \left\{ -\frac{\alpha \kappa}{2} \left[\sum_{t=1}^T (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1})' (\boldsymbol{\theta}_t - \rho \boldsymbol{\theta}_{t-1}) + (1 - \rho^2) \boldsymbol{\theta}_0' \boldsymbol{\theta}_0 \right] \right\} \times$$

$$\sqrt{(n-1) \sum_{i=1}^{n-1} (\lambda_i + \alpha)^{-2} - \left[\sum_{i=1}^n (\lambda_i + \alpha) \right]^{-2}}$$

Assim como para ρ vamos utilizar o algoritmo de Metropolis-Hastings para simular α , uma vez que não conhecemos a distribuição acima. Escolhemos como distribuição proposta o valor absoluto de uma normal do seguinte tipo $N(\alpha^{(j-1)}, \Delta_\alpha)$, onde, Δ_α é escolhido a fim de obter taxas de aceitação adequadas. Aqui escolhemos uma proposta diferente da escolhida por Rojas (2004), esta proposta parece melhorar a estimação do modelo, principalmente quando escolhemos valores iniciais distantes dos valores verdadeiros.

Com isso, a probabilidade de aceitação de α fica:

$$\alpha(\alpha^{(j-1)}; \alpha^p) = \min \left[1, \frac{p(\alpha^p | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \rho, \kappa) q(\alpha^p, \alpha^{(j-1)})}{p(\alpha^{(j-1)} | \boldsymbol{\theta}_{1:T}, \boldsymbol{\theta}_0, \rho, \kappa) q(\alpha^{(j-1)}, \alpha^p)} \right].$$

Aceitamos $\alpha^{(j)} = \alpha^p$ com probabilidade $\alpha(\alpha^{(j-1)}; \alpha^p)$ e $\alpha^{(j)} = \alpha^{(j-1)}$ com probabilidade $1 - \alpha(\alpha^{(j-1)}; \alpha^p)$.

A seguir apresentamos alguns estudos simulados para testar a capacidade de estimação do modelo apresentado.

5 SIMULAÇÕES DE MONTE CARLO

Geramos dois conjuntos de dados com o modelo apresentado anteriormente²⁰. Os dois conjuntos têm o mesmo número de regiões²¹ (i.e. 39), e os mesmos valores para ρ , α , κ e σ^2 . A única diferença entre um grupo e outro é o número de observações ao longo do tempo tempo: para o primeiro temos $T = 7$, enquanto no segundo temos $T = 100$. Assim, tentaremos distinguir o comportamento do processo de estimação para T “pequeno” (mesmo tamanho do conjunto de dados reais analisados posteriormente, a saber, $T = 7$) e T “grande” na estimação do modelo.

Os valores escolhidos para cada parâmetro foram: $\rho = 0,52$; $\alpha = 0,03$; $\kappa = 0,07$ e $\sigma^2 = 0,17$. Tais valores têm um significado especial, estes são os valores simulados das modas *a posteriori* para o índice de furto na região metropolitana de São Paulo, pelo modelo em questão.

Geramos 50 replicações (50 conjuntos de dados) com 10000 iterações MCMC para estimação de cada conjunto. A cada estimação guardamos os valores para a média e moda *a posteriori* e desvio padrão dos valores simulados para os parâmetros de interesse. Os valores iniciais para os parâmetros foram sempre (em todas replicações) os seguintes: $\rho = 0,2$; $\alpha = 0,7$; $\kappa = 1,5$ e $\sigma^2 = 1,5$, ou seja, relativamente “longe” dos valores verdadeiros.

Em ambos os casos, notamos visualmente que a convergência foi atingida em torno da iteração 2000. De maneira conservadora, utilizamos apenas as últimas 5000 iterações, ou seja, consideramos que a partir da iteração 5000 a cadeia converge. Para contornar o problema de dependência da cadeia retiramos elementos dessa cadeia a cada 5 iterações após o período de convergência estabelecido (5000), o que nos deixa com uma amostra de 1000 observações para cada parâmetro.

As tabelas 5.1 e 5.2 mostram os valores mínimos e máximos, média e desvio padrão das modas e médias *a posteriori* dos parâmetros estimados para $T = 7$. E as tabelas

²⁰ Todos os programas para simular e estimar o modelo foram escritos em R pelo autor. Para obter o programa completo, entrar em contato com o autor através do e-mail magazzano@gmail.com.

²¹ A matriz \mathbf{M} , para ambos os casos, foi construída com base na estrutura de vizinhança da região metropolitana de São Paulo.

5.3 e 5.4 mostram os valores mínimos e máximos, média e desvio padrão das modas e médias *a posteriori* dos parâmetros estimados para $T = 100$. Já as figuras 5.1 e 5.2 mostram os gráficos das iterações e o histograma das distribuições marginais *a posteriori* para $T = 7$ e $T = 100$, respectivamente, para uma repetição escolhida ao acaso.

Tabela 5.1: Valores mínimo e máximo, média e desvio padrão das modas *a posteriori* dos parâmetros para $T = 7$ (entre parênteses está o valor verdadeiro dos parâmetros).

	ρ (0,52)	α (0,03)	κ (0,07)	σ^2 (0,17)
Mínimo	0,4379	0,0188	0,0672	0,2080
Máximo	0,4489	0,0256	0,0792	0,3179
Média	0,4412	0,0220	0,0735	0,2316
Desvio Padrão	0,0025	0,0024	0,0043	0,0179

Fonte: Elaborada pelo autor.

Tabela 5.2: Valores mínimo e máximo, média e desvio padrão das médias *a posteriori* dos parâmetros para $T = 7$ (entre parênteses está o valor verdadeiro dos parâmetros).

	ρ (0,52)	α (0,03)	κ (0,07)	σ^2 (0,17)
Mínimo	0,4280	0,0196	0,0680	0,2403
Máximo	0,4320	0,0229	0,0806	0,2908
Média	0,4302	0,0209	0,0744	0,2582
Desvio Padrão	0,0011	0,0008	0,0041	0,0105

Fonte: Elaborada pelo autor.

Tabela 5.3: Valores mínimo e máximo, média e desvio padrão das modas *a posteriori* dos parâmetros para $T = 100$ (entre parênteses está o valor verdadeiro dos parâmetros).

	ρ (0,52)	α (0,03)	κ (0,07)	σ^2 (0,17)
Mínimo	0,5066	0,0277	0,0679	0,1673
Máximo	0,5282	0,0335	0,0710	0,1797
Média	0,5174	0,0292	0,0696	0,1751
Desvio Padrão	0,0047	0,0013	0,0007	0,0040

Fonte: Elaborada pelo autor.

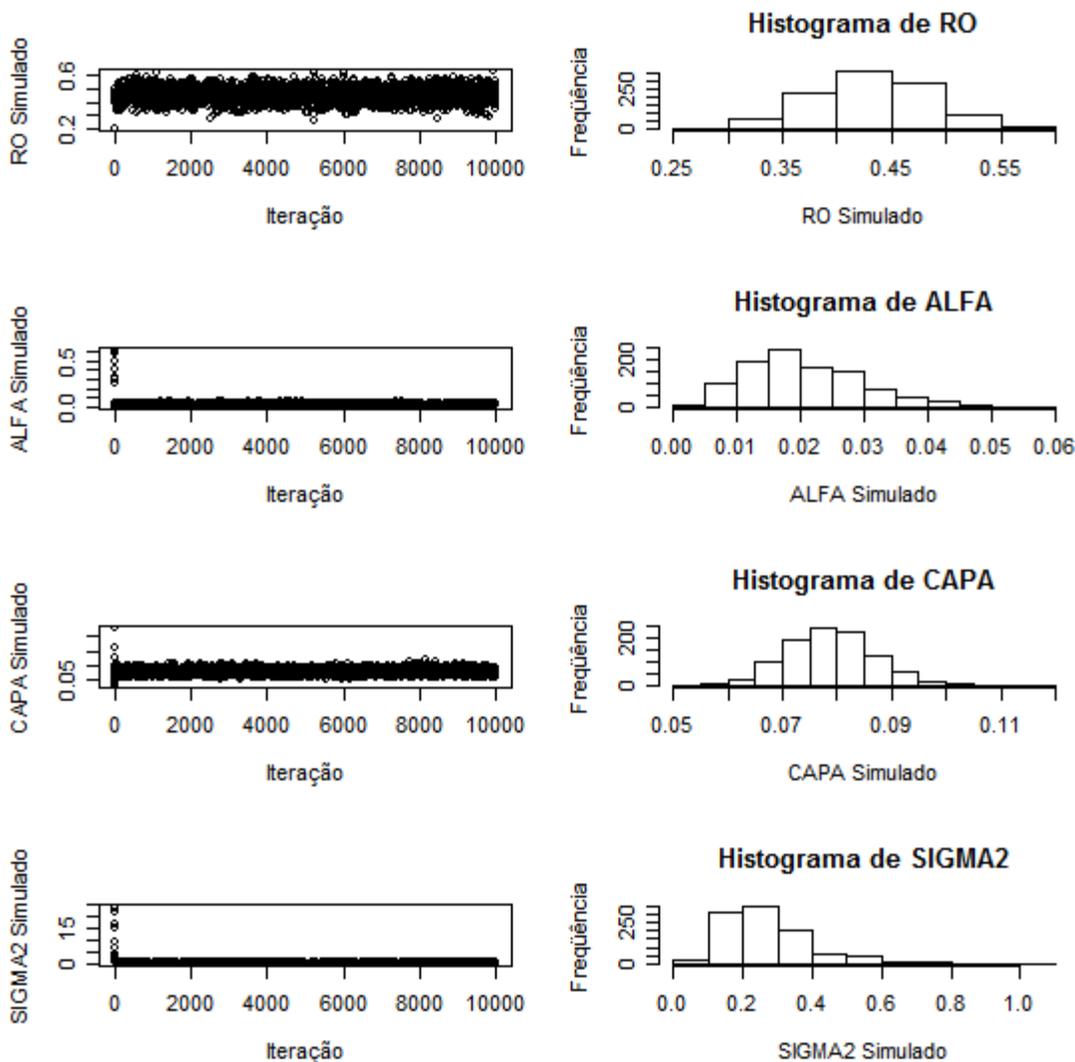
Tabela 5.4: Valores mínimo e máximo, média e desvio padrão das médias *a posteriori* dos parâmetros para $T = 100$ (entre parênteses está o valor verdadeiro dos parâmetros).

	ρ (0,52)	α (0,03)	κ (0,07)	σ^2 (0,17)
Mínimo	0,5140	0,0257	0,0689	0,1736
Máximo	0,5266	0,0290	0,0710	0,1840
Média	0,5196	0,0265	0,0698	0,1796
Desvio Padrão	0,0035	0,0008	0,0006	0,0032

Fonte: Elaborada pelo autor.

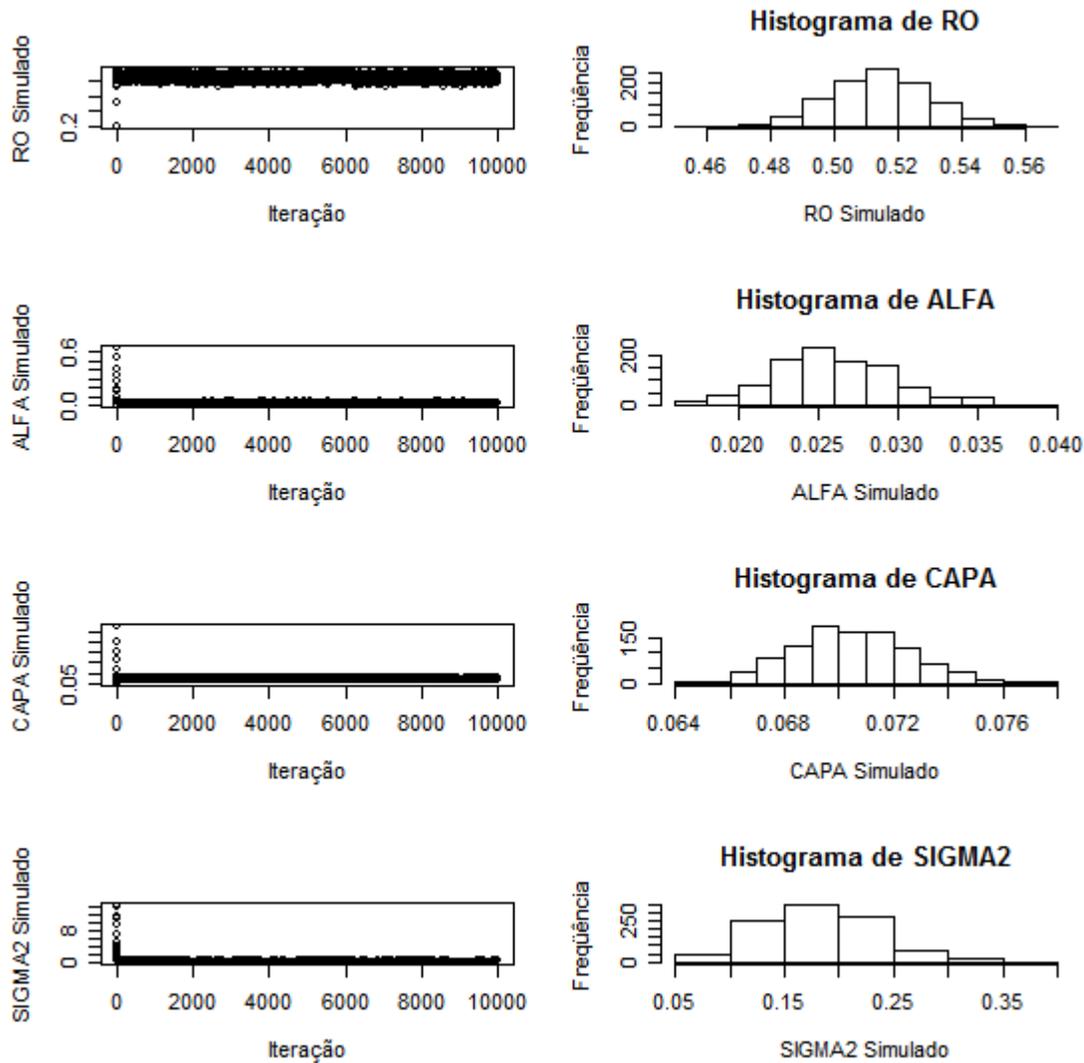
Como era de se esperar, quanto maior o número de observações melhor é a estimação. Analisando as tabelas 5.1 e 5.2, percebemos que o único parâmetro em que o valor verdadeiro está dentro dos valores máximo e mínimo das modas e médias *a posteriori* é o κ . Porém, a Figura 5.1 mostra que a distribuição *a posteriori* dá alguma probabilidade para o valor verdadeiro dos parâmetros. Note que, a distribuição *a posteriori* para ρ varia de algo em torno de 0,2 e 0,6 enquanto o verdadeiro valor desse parâmetro é 0,56, e isso vale para todos os parâmetros em questão.

Figura 5.1: Gráficos das iterações e o histograma das distribuições marginais *a posteriori* para $T = 7$.



Fonte: Elaborada pelo autor.

Figura 5.2: Gráficos das iterações e o histograma das distribuições marginais *a posteriori* para $T = 100$.



Fonte: Elaborada pelo autor.

Quando aumentamos a amostra para $T = 100$, vemos uma clara melhora tanto nos valores das modas quanto das médias *a posteriori* (tabelas 5.3 e 5.4), em todas as repetições modas e médias *a posteriori* ficaram muito próximas dos valores verdadeiros dos parâmetros. Para ter uma idéia de viés percentual estimado podemos medir a diferença percentual entre o verdadeiro valor e a média das modas *a posteriori*. No caso de T “pequeno” a média das modas *a posteriori* está: 17,86%, 36,36%, 4,76% e 26,59%

“longe” do verdadeiro valor de ρ , α , κ e σ^2 , respectivamente²². Quando aumentamos T para 100 esses valores caem para: 0,52%, 2,73%, 0,57% e 2,91% para ρ , α , κ e σ^2 , respectivamente. O desempenho das estimações para T “pequeno” carece de precisão porém, como notado anteriormente, as distribuições *a posteriori* contemplam os verdadeiros valores dos parâmetros e, de certa forma, ignoram valores que levariam a conclusões opostas em termos de interpretação das estimativas. Isso significa que nossas conclusões na aplicação serão no sentido de para onde os valores estarão apontando.

Com exceção de ρ o desvio padrão das modas e médias *a posteriori* foi menor em todos os outros parâmetros quando comparamos com o caso de tamanho de amostra menor. Outro ponto a destacar é a melhor precisão da distribuição *a posteriori* de T “grande”. Para uma repetição escolhida ao acaso temos um desvio padrão dos valores simulados em torno de: 0,05 para ρ , 0,01 para α , 0,07 para κ e 0,19 para σ^2 numa amostra com $T = 7$. Para uma repetição escolhida ao acaso em que $T = 100$ esses números são 0,01 para ρ , 0,03 para α , 0,02 para κ e 0,05 para σ^2 . Notamos, apenas em α um aumento do desvio padrão²³. A maior precisão também pode ser notada comparando os histogramas das figuras 5.1 e 5.2, enquanto a distribuição *a posteriori* de ρ para $T = 7$ variava em torno de 0,2 e 0,6 no caso de $T = 100$ vemos que os valores simulados de ρ variam entre 0,46 e 0,56.

A próxima seção irá expor os resultados obtidos para os índices de criminalidade escolhidos.

²² Repetimos esse mesmo exercício para a média das médias *a posteriori*, e encontramos os seguintes valores para ρ , α , κ e σ^2 , respectivamente, 20,37%, 43,48%, 5,97% e 34,15%.

²³ Fato também notado em Rojas (2004).

6 CRIME NA REGIÃO METROPOLITANA DE SÃO PAULO

Para medir a interação social da criminalidade na região metropolitana de São Paulo, coletamos dados anuais de 2000 a 2006, da Secretaria de Segurança Pública de São Paulo, dos índices de furto, roubo e homicídio dos 39 municípios que compõem a região de interesse. Assim como no caso dos dados simulados, geramos 10000 iterações para cada índice e começamos a estimação com os mesmos valores iniciais para os parâmetros que na seção anterior.

Nos três casos, notamos visualmente que a convergência foi atingida em torno da iteração 2000, de maneira conservadora utilizamos apenas as últimas 5000 iterações. Para contornar o problema de dependência da cadeia retiramos elementos dessa cadeia a cada 5 iterações após o período de convergência estabelecido (5000).

As tabelas 6.1, 6.2 e 6.3 exibem moda e média *a posteriori* e desvio padrão dos valores simulados para cada parâmetro para os dados de furto, roubo e homicídio, respectivamente. As figuras 6.1, 6.2 e 6.3 mostram os gráficos das iterações e o histograma das distribuições marginais *a posteriori* para os dados de furto, roubo e homicídio, respectivamente.

Tabela 6.1: Moda e média *a posteriori* e desvio padrão dos valores simulados dos parâmetros para os dados de furto.

	ρ	α	κ	σ^2
Moda	0,5270	0,0234	0,0772	0,1694
Média	0,5977	0,0347	0,0771	0,1833
Desvio Padrão	0,0302	0,0149	0,0087	0,0755

Fonte: Elaborada pelo autor

Tabela 6.2: Moda e média *a posteriori* e desvio padrão dos valores simulados dos parâmetros para os dados de roubo.

	ρ	α	κ	σ^2
Moda	0,5467	0,0707	0,1326	0,1229
Média	0,6053	0,0804	0,1330	0,1316
Desvio Padrão	0,0263	0,0449	0,0163	0,0486

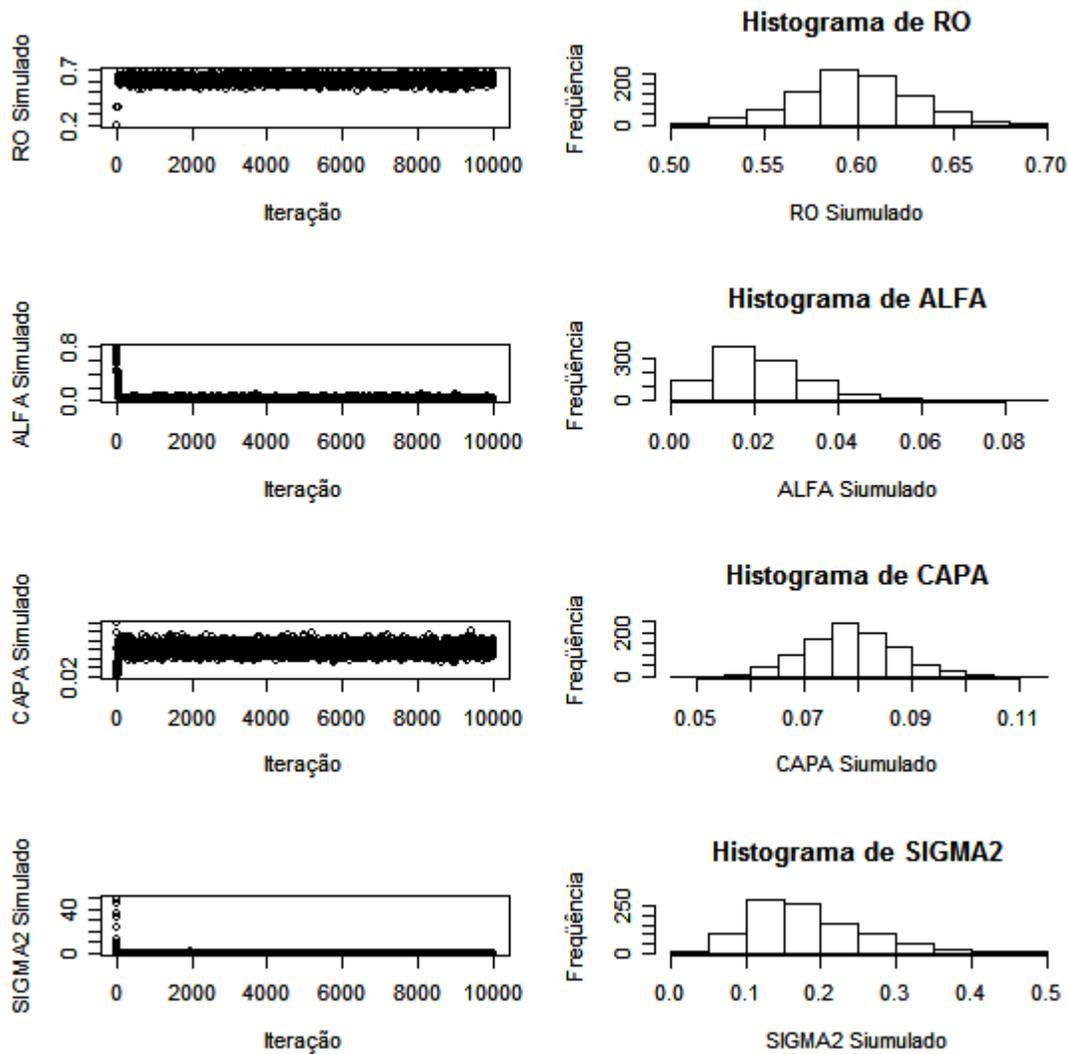
Fonte: Elaborada pelo autor

Tabela 6.3: Moda e média a posteriori e desvio padrão dos valores simulados dos parâmetros para os dados de homicídio.

	ρ	α	κ	σ^2
Moda	0,2022	0,0015	51,3189	0,0185
Média	0,1656	0,0068	55,7230	0,0186
Desvio Padrão	0,1139	0,0049	19,9024	0,0028

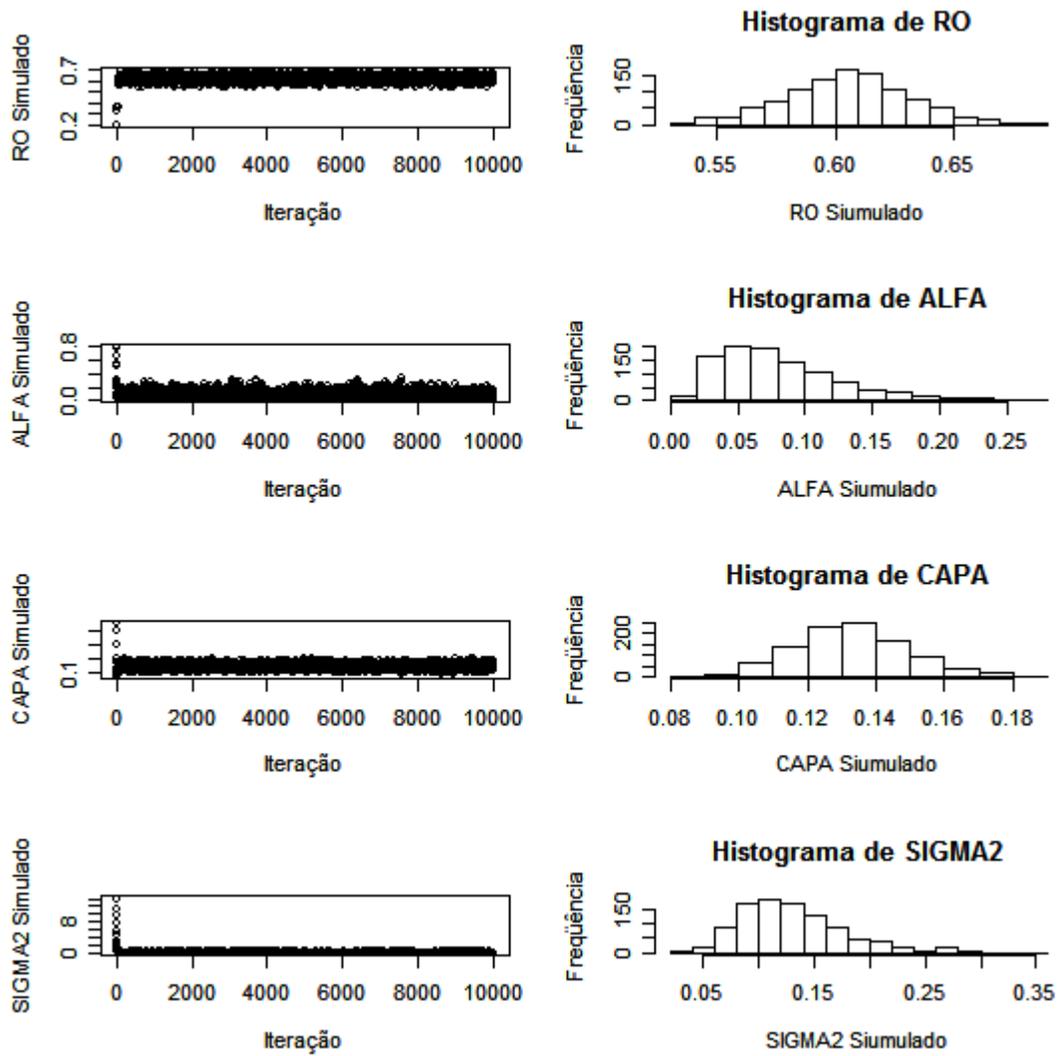
Fonte: Elaborada pelo autor

Figura 6.1: Gráficos das iterações e o histograma das distribuições marginais *a posteriori* para furto.



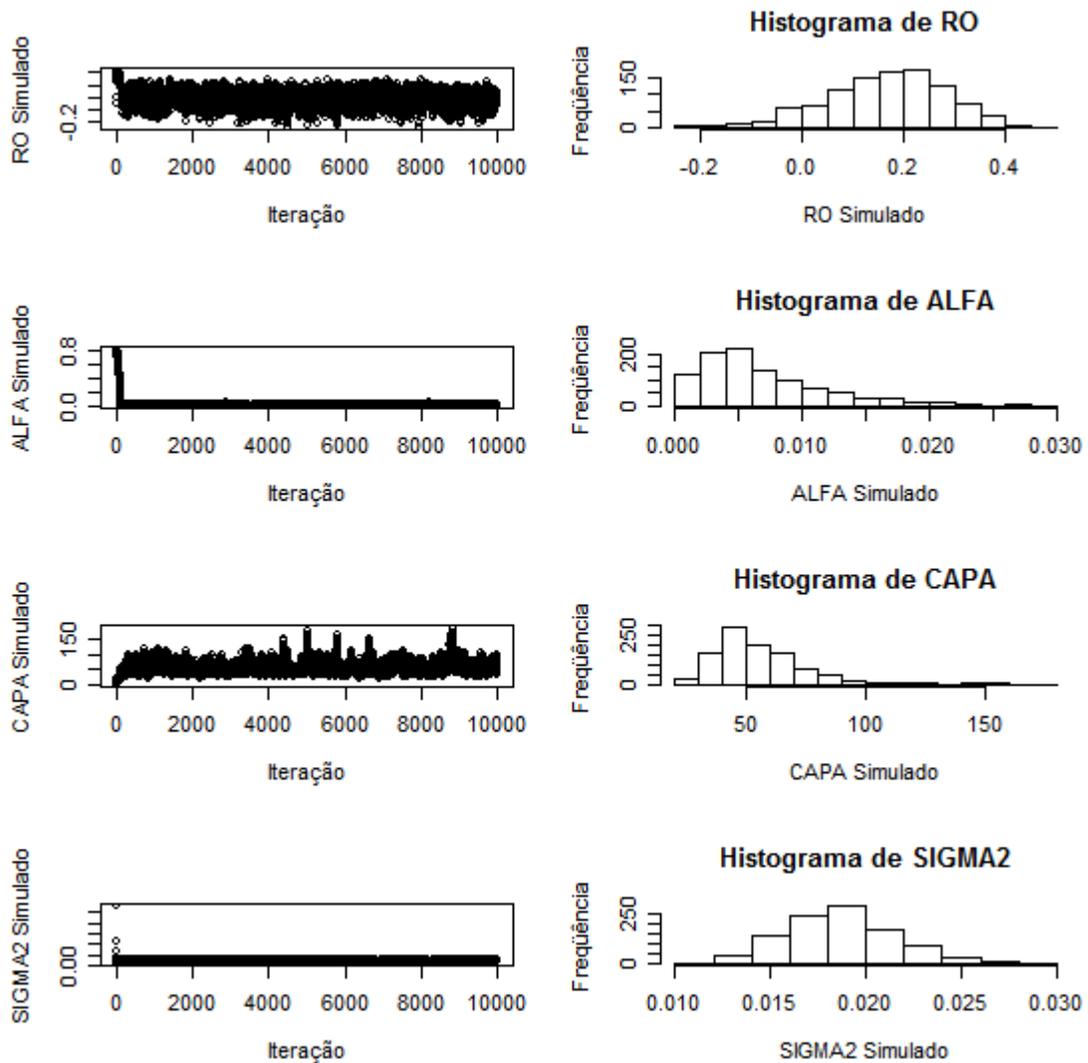
Fonte: Elaborada pelo autor

Figura 6.2: Gráficos das iterações e o histograma das distribuições marginais *a posteriori* para roubo.



Fonte: Elaborada pelo autor

Figura 6.3: Gráficos das iterações e o histograma das distribuições marginais *a posteriori* para homicídio.



Fonte: Elaborada pelo autor

Não podemos começar a analisar os resultados sem lembrar dos resultados do capítulo anterior, que alertou para o fato das estimações com um menor número de observações não serem tão precisas. Porém, a distribuição *a posteriori* dos parâmetros sempre continha os verdadeiros valores dos parâmetros. Posto de outra forma temos dificuldades em ter uma estimativa para os graus de correlação temporal e espacial, porém, temos indícios para onde esses valores apontam.

Notamos que a correlação temporal dos índices de roubo e furto são semelhantes e maiores que a de homicídio (média *a posteriori* de 0,5977 e 0,6053 para furto e roubo, respectivamente, e média *a posteriori* igual a 0,1656 para homicídio), fato que ao analisarmos os dados pudemos perceber. Observando a figura 6.3 percebemos que existe alguma probabilidade de a correlação temporal dos homicídios ser negativa, ou seja, que os homicídios estão caindo na região analisada, fato observado anteriormente neste trabalho. A interação social no furto apresentou-se maior que a interação ocorrida no caso de roubo (média *a posteriori* de 0,0347 e 0,0804 para furto e roubo, respectivamente), ainda que a correlação espacial para os dados de roubo seja alta.

Porém, a maior correlação espacial foi encontrada nos índices de homicídio moda *a posteriori* de 0,0015 e média de 0,0068. A análise dos dados vai na mesma direção do que é encontrado no modelo, isso porque, a variância espacial dos índices de furto e roubo é maior do que a variância espacial de homicídio. Ou seja, estamos tratando de uma região em que os índices de homicídio são muito parecidos entre as cidades o que nos faz acreditar numa alta interação social desse tipo de atividade criminosa. Aparentemente esse fato parece ir contra o que a literatura tem encontrado. Contudo, devemos lembrar dos trabalhos de Puech (2004), Cohen e Tita (1999), Messner et al. (1999) e Baller et al. (2001) que tratam dos já mencionados *hot spots*. A análise dos dados nos faz acreditar que a região metropolitana de São Paulo seja um *hot spot* no estado, principalmente porque essa região fica entorno da maior cidade do estado (São Paulo).

O valor de κ também parece ser bastante condizente com a análise dos dados. Como κ é um parâmetro de escala na matriz de precisão \mathbf{W}_t^{-1} quanto maior esse parâmetro menor será a variância do vetor de estados ($\mathbf{\theta}_t$). Mencionamos anteriormente que encontramos uma variância espacial para os dados de furto e roubo muito maior que a encontrada para homicídio e observando as tabelas 6.1, 6.2 e 6.3 vemos que o valor de κ para o índice de homicídio na região é muito maior (média *a posteriori* de 55,7230 para homicídio contra 0,1330 e 0,0771 para roubo e furto, respectivamente).

Os valores de σ^2 também vão na mesma direção, porém, com uma idéia de variância temporal. E inclusive nesse caso a variância do índice de homicídio é menor do que no caso para roubo e furto. As médias *a posteriori* para σ^2 0,1833, 0,1316 e 0,0186 para os índices de furto, roubo e homicídio, respectivamente.

O resultado também é interessante para possíveis políticas públicas a serem tomadas para coibir essas atividades. Devemos recordar que em 2003 a Secretaria Nacional de Segurança Pública passou a incentivar as administrações municipais a implementar projetos voltados ao combate à violência e criminalidade. Em 2005, os municípios da região metropolitana de São Paulo gastaram R\$ 74,9 milhões em despesas com segurança nacional e defesa pública²⁴. Porém, oito cidades dessa região declaram que tiveram gasto zero com segurança. Aparentemente, os municípios que compõe a região metropolitana de São Paulo não estão tendo uma ação homogênea ao combate a tais atividades. Nossos resultados indicam que um combate eficiente às atividades criminais será aquele que levar em consideração a interação social na criminalidade. Ou seja, políticas que tenham como objetivo reduzir a criminalidade devem ser tomadas com certa homogeneidade entre cidades vizinhas. Assim, políticas como aumentar o número de policiais na rua em determinada cidade²⁵ podem ser incompletas, uma vez que o que conta para explicar a criminalidade não são apenas as características das cidades e sim a interação social entre elas. Desse modo, parece inadequado que cidades vizinhas tenham um gasto com segurança muito discrepante entre elas, mesmo quando levamos em consideração a população do município.

Esse ponto é de grande interesse uma vez que estamos em ano de eleições municipais e observamos que grande parte dos debates e temas de discussões entre candidatos e eleitores tem sido a questão da segurança. Nossas evidências indicam que um simples aumento nos gastos municipais com segurança pública talvez não seja suficiente a resposta completa para reduzir a criminalidade do município, fato que seria frustrante para um prefeito com a melhor das intenções. A eficácia ao combate a atividade criminal será maior quando ações forem tomadas de forma coordenada com os vizinhos.

Devemos ressaltar que a região metropolitana de São Paulo apresentou, para o período em questão, uma alta interação social entre todos os índices de criminalidade estudados. Assim, as evidências sugerem que para que exista uma boa compreensão da criminalidade em determinada cidade, não basta apenas “olhar” para as características da cidade, devemos levar em conta a região em que a cidade está inserida.

²⁴ Dados do Atlas Eletrônico do Estado de São Paulo (EATLAS) do portal do Governo do Estado de São Paulo.

²⁵ Fato que deveria reduzir a criminalidade.

7 CONCLUSÃO

Notamos que em termos da teoria econômica do crime, o trabalho de Becker (1968) foi crucial para entender o fato dos agentes ingressarem no mercado criminal. Vários trabalhos empíricos mostraram a validade da teoria de Becker (1968). Em seguida a literatura buscou a relação entre criminalidade e variáveis socioeconômicas, como por exemplo, desemprego e desigualdade de renda. Atualmente, o principal interesse parece ser o de medir a interação social na criminalidade, ou seja, o papel que a vizinhança tem na decisão de entrada ou não na atividade criminal. Percebemos que muitos trabalhos empíricos que tentam medir a interação social na criminalidade não levavam em conta o efeito temporal da criminalidade. Dessa forma, este trabalho apresentou um modelo espaço-temporal Bayesiano para medir a interação social da criminalidade apresentado em Rojas (2004). O modelo é construído especificamente para dados de área, caso clássico de dados criminais.

Notamos que uma vantagem do modelo, especificamente para o caso brasileiro de escassos dados criminais, é o fato do vetor de estados latente ter o papel de capturar os efeitos de variáveis que afetam a criminalidade como: probabilidade de apreensão, probabilidade de condenação, severidade da pena, variáveis socioeconômicas, etc.

Antes de explorarmos os resultados empíricos, apresentamos um interessante trabalho com dados simulados que mostrou como um maior número de observações faz diferença na estimação do modelo. Percebemos que quando temos um número pequeno de observações, a saber sete (mesmo número do conjunto de dados criminais analisados), o desempenho das estimações deixa a desejar. Porém, notamos que o valor verdadeiro do parâmetro está sempre contido dentro da distribuição *a posteriori* dos parâmetros.

Os resultados empíricos mostram que o efeito temporal não pode ser descartado. Principalmente nos índices de furto e roubo, onde esse efeito é maior relativamente ao índice de homicídio. Os três índices (furto, roubo e homicídio) apresentam um alto grau de interação social. Além disso, encontramos uma maior interação social no índice de homicídio. Trabalhos que estudam o estado como um todo, ao invés de uma região, encontram um menor grau de interação social para o índice de homicídio, porém, esses trabalhos observam que esse determinado índice apresenta conglomerados principalmente

na região de alguma grande cidade do estado. Fato que nos faz acreditar que a região metropolitana do estado de São Paulo é um *hot spot* dentro do estado de São Paulo. Fica claro que para uma boa compreensão dos determinantes da criminalidade em determinada cidade a interação social deve ser incluída.

Notamos que as autoridades das cidades que compõem a região metropolitana da São Paulo não parecem levar em conta a interação social na criminalidade em seu processo decisório. Note que essa região gastou ao todo R\$ 74,9 milhões em despesas com segurança e defesa pública, porém oito cidades dessa região declaram que tiveram gasto zero com segurança. Nossos resultados indicam que um combate eficiente às atividades criminais será aquele que levar em consideração a interação social na criminalidade. Ou seja, políticas que tenham como objetivo reduzir a criminalidade devem ser tomadas com certa homogeneidade entre cidades vizinhas.

Portanto, autoridades que lidam com coerção criminal não devem menosprezar esses resultados, pois, ações que deveriam coibir esse tipo de atividade podem surtir efeito inferior ao desejado, já que grande parte do que explica esse comportamento dos indivíduos é a interação social e não somente as características das cidades. Dessa forma, é de se esperar que medidas que forem tomadas no sentido de reduzir a criminalidade devam ser tomadas em conjunto e não apenas pensando nas características individuais de uma cidade.

Trabalhos futuros poderiam focar na tentativa de obter um maior número de observações ao longo do tempo e também, se possível, inserir uma variável exógena ao modelo (o que não estava disponível em nosso caso).

REFERÊNCIAS

ANDRADE, Monica Viegas; LISBOA, Marco de Barros. Desesperança de vida: homicídio em Minas Gerais, Rio de Janeiro e São Paulo no período 1981/97. In: HENRIQUES, Ricardo (Org.). **Desigualdade e pobreza no Brasil**. Ipea - Instituto de Pesquisa Econômica Aplicada, Rio de Janeiro, v. 01, p. 347-384, 2001.

ASSUNÇÃO, Renato Martins; REIS, Ilka; OLIVEIRA, Claudia di Lorenzo. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. **Statistics In Medicine**, Chichester, v. 20, p. 2319-2335, 2001.

BAILEY, Trevor; GATRELL, Tony. **Interactive Spatial Data Analysis**. Londres: Prentice Hall, 1995, 414 p.

BALLER, Robert, MESSNER, Steven; ANSELIN, Luc; HAWKINS, Darnell; DEANE, Glenn. Structural covariates of US county homicide rates, incorporating spatial effects. **Criminology**, Boston, v. 39, n. 3, p. 201-232, 2001.

BANERJEE, Sudipto; CARLIN, Bradley; GELFAND, Alan. **Hierarchical modeling and analysis of spatial data**. Londres, Chapman & Hall, 2004, 452 p.

BEATO FILHO, Claudio Chaves; ASSUNÇÃO, Renato Martins; SILVA, Bráulio Figueiro da; MARINHO, Frederico Couto; REIS, Ilka Afonso; ALMEIDA, Maria Cristina Mattos. Conglomerados de homicídios e o tráfico de drogas em Belo Horizonte, Minas Gerais, Brasil, de 1995 a 1999. **Cadernos de Saúde Pública**. Rio de Janeiro, v. 17, n. 5, p. 1163-1171, 2001.

BECKER, Gary Stanley. Crime and punishment: an economic approach. **The Journal of Political Economy**, Chicago, v. 2, p. 169-217, 1968.

BENNETT, A; RACINE-POON, Amy; WAKEFIELD, Jon. MCMC for nonlinear hierarchical models. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, David. **Markov Chain Monte Carlo in Practice**. Londres: Chapman & Hall/CRC, 1996, 520 p.

BURDETT, Kenneth; LAGOS, Ricardo; WRIGHT, Randall; Crime, inequality, and unemployment. **The American Economic Review**; Vol. 35, n. 05, pp. 1764-1777, 2003.

CAPPÉ, Olivier; MOULINES, Eric; RYDÉN, Tobias. **Inference in Hidden Markov Models**. Nova York: Springer, 2005, 678 p.

CASE, Anne; KATZ, Lawrence. The company you keep: the effects of family and neighborhood on disadvantaged youths. **Harvard Institute of Economic Research Working Papers 1555**, Harvard, 1991. Disponível em: <http://www.nber.org/papers/w3705.pdf>. Acesso em: 16 jun. 2007.

COHEN, Jacqueline; TITA, George. Spatial diffusion in homicide: exploring a general method of detecting spatial diffusion processes. **Journal of Quantitative Criminology**, Nova York, v. 15, p. 451-493, 1999.

COWLES, Mary Kathryn; CARLIN, Bradley. Markov chain Monte Carlo convergence diagnostics: a comparative review. **Journal of the American Statistical Association**, v. 91, p. 883-904, 1996.

CRESSIE, Noel. **Statistics for Spatial Data**. 2^a edição. Nova York: Wiley, 1996, 928 p.

EHRLICH, Isaac. Participation in illegitimate activities: a theoretical and empirical investigation. **The Journal of Political Economy**, Chicago, v. 81, p. 521-565, 1973.

FERREIRA, Marco Antonio Rosa; OLIVEIRA, Victor de. Bayesian analysis for a class of Gaussian Markov random fields. In: ISBA 2004 World Meeting, 1, 2004, Viña del Mar.

FREEMAN, Richard. Crime and the employment of disadvantaged youths. **NBER Working Papers 3875, National Bureau of Economic Research, Inc**, 1991.

GAMERMAN, Dani; LOPES, Hedibert. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2^a edição. Londres: Chapman & Hall/CRC, 2006, 336 p.

GELFAND, Alan; SMITH, Adrian. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, n. 410, p. 398-409, 1990.

GEMAN, Stuart; GEMAN, Donald. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Trans. Pattern Analysis and Machine Intelligence**, v. 6, p. 721-741, 1984.

GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, David. **Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics**. Boca Raton: Chapman & Hall/CRC, 1996, 520 p.

GLAESER, Edward; SACERDOTE, Bruce; SCHEINKMAN, José Alexandre. Crime and social interactions. **The Quarterly Journal of Economics**, Cambridge, v. 111, p. 507-548, 1996.

GOLUB, Gene; VAN LOAN, Charles. **Matrix computations**. 3^a edição. Baltimore: Johns Hopkins University Press, 1996. 694 p.

HARRISON, Jeff; STEVENS, C. Bayesian Forecasting. **Journal of the Royal Statistical Society**, Boston, v. 38, p. 205-247, 1976.

HASTINGS, W. Monte Carlo Sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97-109, 1970.

IMROHOGLU, Ayse; MERLO, Antonio; RUPERT, Peter, What accounts for the decline in crime. **USC Finance & Business Econ. Working Paper No. 01-15**, 2001. Disponível em: < <http://www.econ.nyu.edu/cvstarr/working/2000/RR00-11.PDF> >. Acesso em: 12 de agosto de 2007.

KALMAN, R. A New Approach to Linear Filtering and Prediction Problems. **Journal of Basic Engineering**, Nova Jersey, v. 82, p. 35-45, 1960.

KHAN, Tulio. Os custos da violência: quanto se gasta ou deixa de ganhar por causa do crime no Estado de São Paulo. In: **São Paulo em Perspectiva**. São Paulo, v.13, n.4, out/dez 1999.

KIM, C.; NELSON, C. **State-Space Models with Regime Switching**. Cambridge: The MIT Press, 1999.

KINDERMAN, R; SNELL, J L. On the relation between markov random fields and social networks. **Journal of Mathematical Sociology**, Nova York, VII, p. 1-13, 1980.

KOOP, Gary. **Bayesian Econometrics**. Nova York: Wiley, 2003, 359 p.

LAWSON, Andrew; KULLDORF, Martin, A review of cluster detection methods. In: LAWSON, Andrew; et at. **Disease Mapping and Risk Assessment for Public Health**. Chichester: Wiley, 1999, p. 99-110.

LAWSON, Andrew; VIEL, Jean Francois. Tests for directional space-time interaction in epidemiological data. **Statistics in Medicine**, Nova Jersey, v. 14, p. 21-22, 1995.

LEVITT, Steven. The effect of prison population size on crime rates: evidence from prison overcrowding litigation. **The Quarterly Journal of Economics**, Cambridge v. 111, n. 2, p. 319-351, 1996.

LEVITT, Steven. The relationship between crime reporting and police: implications for the use of uniform crime reports. **Journal of Quantitative Criminology**, Nova York, v. 14, p. 61-81, 1998.

LOPES, Hedibert. Modern Bayesian econometrics. In: Encontro Brasileiro de Estatística Bayesiana, 9, 2008, Maresias.

MESSNER, Steven; ANSELIN, Luc; BALLER, Robert; HAWKINS, Darnell; DEANE, Glenn; TOLNAY, Stewart. The spatial patterning of county homicide rates: an application of exploratory spatial data analysis. **Journal of Quantitative Criminology**, Nova York, v. 15, n. 4, p. 423-450, 1999.

METROPOLIS, Nicholas; ROSENBLUTH, Arianna; ROSENBLUTH, Marshall; TELLER, Augusta, TELLER, Edward. Equation of state calculations by fast computing machine. **Journal of Chemical Physics**, v. 21, p. 1087-1091, 1953.

MORTENSEN, Dale. Job search and labor market analysis. In: **Handbook in labor economics**, O. Ashenfelter e R. Layard (eds). Amsterdam: North Holland, 1986.

NOBRE, Aline; SCHIMDT, Alexandra; LOPES, Hedibert. Spatio-temporal models for mapping the incidence of malaria in Pará. **Environmetrics**, Nova Jersey, v. 16, p. 291-304, 2005.

OLIVEIRA, Cristiano Aguiar de. Análise espacial da criminalidade no Rio Grande do Sul. **Texto para Discussão do CEPEAC (Centro de Pesquisa e Extensão da FEAC)**, 2005. Disponível em <http://www.upf.br/cepeac/download/td_15_2005.pdf>. Acesso em: 04 de maio de 2007.

PAEZ, Marina Silva. **Análise de Modelos para a Estimção e Previsão de Processos Espaço-Temporais**. Rio de Janeiro, 2004, 147 f. Dissertação (Mestrado em estatística) – Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

PUECH, Frédéric. How do criminals locate? Crime and spatial dependence in Minas Gerais. **Working Paper do CERDI (Centre d'Etudes et de Recherches sur le Développement International)**, 2004 Disponível em: <<http://www4.fe.uc.pt/30years/papers/68.pdf>>. Acesso em: 04 maio 2007.

RIPLEY, Brian. **Spatial Statistics**. Chichester: Wiley, 1981, 252 p.

ROJAS, Juan Carlos Vivar. **Uma nova classe de modelos espaço-temporais para dados de área**. Rio de Janeiro, 2004, 114 f. Dissertação (Mestrado em estatística) – Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

ROBERT, Christian. Convergence control methods for Markov chain Monte Carlo algorithms. **Statistical Science**, v. 10, n. 3, p. 231-253, 1995.

ROBERT, Christian; CASELLA, George. A history of Markov Chain Monte Carlo – Subjective recollections from incomplete data. 2008. Disponível em: <http://arxiv.org/PS_cache/arxiv/pdf/0808/0808.2902v1.pdf> acesso em 25 de ago. 2008.

RONDON, Vinícius Velasco; ANDRADE, Mônica Viegas. Impactos da criminalidade no valor dos aluguéis em Belo Horizonte. **Texto para discussão 194 - UFMG/Cedeplar**, Belo Horizonte, 2003. Disponível em: <<http://www.cedeplar.ufmg.br/pesquisas/td/TD%20194.pdf>>. Acesso em: 10 de jul. de 2008.

WEST, Mike; HARRISON, Jeff. **Bayesian Forecasting and Dynamic Models**. 2a.edição. Nova York: Springer, 1997, 701 p.