

**Modelagem Conjunta de Média e Variância:
Uma Aplicação ao Problema de Estimação
em Pequenos Domínios**

por

Joaquim Henriques Vianna Neto

DME - IM - UFRJ

2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Modelagem Conjunta de Média e Variância: Uma Aplicação ao Problema de Estimação em Pequenos Domínios

Joaquim Henriques Vianna Neto

Dissertação submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Aprovada por:

Prof. Alexandra M. Schmidt.

PhD - IM - UFRJ - Orientadora.

Prof. Fernando A. S. Moura

PhD - IM - UFRJ.

Prof. Helio S. Migon

PhD - IM - UFRJ.

Prof. Jorge A. Achcar.

PhD - CCET - UFSCAR.

Rio de Janeiro, RJ - Brasil

2007

FICHA CATALOGRÁFICA

Neto, Joaquim Henriques Vianna.

Modelagem Conjunta de Média e Variância:

Uma Aplicação ao Problema de Estimação em Pequenos Domínios \

Joaquim Henriques Vianna Neto.

Rio de Janeiro: UFRJ, IM, DME, 2007.

Dissertação - Universidade Federal do Rio de Janeiro, IM, DME.

1. Introdução. 2. Revisão de Literatura.

3. Conjunto de Dados e Estudo de Referência. 4. Modelagem Conjunta.

5. Análise de dados. 6. Conclusões e Trabalhos Futuros.

(Mestrado-UFRJ/IM/DME) I. Schmidt, Alexandra

II. Universidade Federal do Rio de Janeiro III. Título.

À minha mãe e aos meus filhos.

“Our world, our life, our destiny, are dominated by uncertainty; this is perhaps the only statement we may assert without uncertainty”.

de Finetti.

Agradecimentos

Aos meus pais, Maria Julia e Tibiriça, por estarem sempre ao meu lado me dando tranquilidade e segurança.

À minha amada esposa, Isabelle, por todo carinho, apoio e compreensão.

Ao Rafinha por tentar desligar meu computador de 2 em 2 minutos cronometrados, tarefa realizada com muito profissionalismo e competência (inclusive na minha ausência). Rafa, sem suas investidas, meu trabalho teria sido muito triste. Ao Caizinho pelo companheirismo, por sua axila tranquilizante e por todos os conselhos superficialmente incompreensíveis. Meus amores, vocês são tudo para mim. Amo vocês demais.

Aos professores Alexandra Schmidt e Fernando Moura, pelos conhecimentos transmitidos, pela orientação e principalmente por me incentivarem constantemente.

Aos demais professores e colegas do departamento de estatística da UFRJ pela amizade e por tudo que me ensinaram.

À toda família de minha mãe, por estar sempre ao meu lado. Em especial à tia Odaléia, tio Helvécio, tia Camila, tia Vera, tia Marisa, tia Tadéia, ao tio Tufi, tio Zizito e ao Bigode.

Aos meus primos Junim, Sávio, Sandro, Igor, Alexandre e Anderson, por serem muito

legais.

Aos amigos Diogo, Lyano, Leimar e Michel, que apesar da distância são sempre lembrados.

Ao Glauber pelos incentivos na época da graduação e por ter acreditado sempre em mim.

À Leda e à Zoraide pela hospedagem durante o ano de 2005.

Ao Lucas por me deixar usar seu computador para rodar os algoritmos MCMC.

Por fim, agradeço ao financiamento da Capes através da minha bolsa de mestrado.

Resumo

O objetivo geral deste trabalho é aplicar modelos capazes de tratar conjuntamente informações de pesquisas com diferentes precisões, incorporando desvios mensuráveis ao processo de estimação e promovendo a troca de informação entre períodos de tempo e localizações no espaço. Particularmente, o interesse é obter estimativas intercensitárias atualizadas a partir de informações para pequenos domínios.

Apresentamos uma aplicação à dados de três pesquisas com diferentes precisões: o Censo Demográfico Brasileiro, a Contagem Populacional e a Pesquisa Nacional por Amostra de Domicílios (PNAD), todas divulgadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Os municípios do Estado de São Paulo formam a base de dados do estudo e o objetivo é obter estimativas anuais da densidade populacional combinando os dados das três pesquisas. Informações obtidas a partir da PNAD caracterizam-se por apresentar baixa precisão a nível municipal, por isto, o município será tratado aqui como pequeno domínio. Adotando uma abordagem Bayesiana, iremos utilizar conjuntamente estas pesquisas para obter a distribuição a posteriori das densidades populacionais municipais e, conseqüentemente, produzir estimativas desta densidade apresentando o seu erro associado.

Em pesquisas por amostragem, a quantidade efetiva de informação de um estimador é traduzida pela estimativa de sua variância. Propomos neste trabalho a modelagem destas estimativas para regular a troca de informação temporal e espacial, conduzida aqui por Modelos Dinâmicos e Campos Aleatórios Gaussianos.

Palavras Chaves: Amostragem, Modelos de Crescimento, Modelagem de Variância, Pequenas Áreas, Pesquisas Populacionais e Processos Gaussianos.

Sumário

1	Introdução	2
2	Revisão de Literatura	6
2.1	Noções de Inferência Bayesiana	6
2.1.1	Estimação de Parâmetros	7
2.2	Modelo de Crescimento Exponencial Generalizado	8
2.3	Modelos Dinâmicos	10
2.3.1	Fatores de Desconto	12
2.3.2	Modelos Dinâmicos de Crescimento Exponencial	13
2.4	Geoestatística	14
2.4.1	Dados Espacialmente Referenciados	14
2.4.2	Estacionariedade e Isotropia	15
2.4.3	Processos Gaussianos	16
2.4.4	Famílias de Funções de Correlação	16
3	Conjunto de Dados e Estudo de Referência	18
3.1	Descrição das Pesquisas	21
3.1.1	Censos no Brasil	22
3.1.2	Contagem Populacional	22
3.1.3	Pesquisa Nacional por Amostra de Domicílios	23
3.1.4	Estimativas Populacionais Oficiais	24
3.2	Estudo de Referência	24

4	Modelagem Conjunta	28
4.1	Modelo Dinâmico de Crescimento Exponencial	28
4.2	Modelagem Conjunta	32
4.2.1	Análise Exploratória das Estimativas da Variância	32
4.3	Previsões	33
4.4	Modelagens de β e γ	36
4.4.1	Modelagem Hierárquica	36
4.4.2	Modelagem Hierárquica Espacialmente Estruturada	36
4.5	Prioris	38
4.6	Distribuições a posteriori	40
4.6.1	Distribuição a posteriori para o Modelo Hierárquico	40
4.6.2	Distribuição a posteriori para o Modelo Espacial	41
4.7	Comparação de Modelos	41
4.7.1	Critério de Informação Baseado na <i>Deviance</i>	42
4.7.2	Desvio Preditivo Esperado	43
5	Análise de Dados	44
5.1	Aspectos Computacionais	44
5.1.1	Amostrador de Gibbs	45
5.1.2	Condicionais Completas	46
5.1.3	Metropolis-Hastings	49
5.1.4	Sintonizando a variância das distribuições propostas	51
5.1.5	<i>Forward Filtering Backward Sampling</i>	53
5.2	Verificação de Convergência	55
5.3	Estudo Simulado	58
5.4	Ajuste dos Dados Reais	62
5.4.1	Ajuste com $W_{i,t} = \tau^2$	63
5.4.2	Ajuste com Fatores de Desconto	64
5.4.3	Resultados do DIC e do EPD para os Dados Reais	69
5.5	Previsão para municípios não selecionados pela PNAD	71

5.6	Comparação com Dados Externos	72
6	Conclusões e Trabalhos Futuros	75
A	Estimadores da Média e Variância	78
B	Taxas de Aceitação	80
C	Cálculo das Condicionais Completas	81
D	FFBS - Distribuição de $(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$	89
E	Autocorrelações e Histogramas	91
F	Ajustes com fator 0,50	93
G	Ajustes com fator 0,80	95

Lista de Tabelas

3.1	Anos em que foram realizadas as pesquisas, tipo de pesquisa e numeração utilizada para identificar cada ano.	19
3.2	Modelagens da variância em Souza (2004), onde $\eta_0, \eta_1, \eta_2, \eta_3, \eta_4$ e η_5 são parâmetros a estimar e f_i é a fração de amostragem do município i	25
5.1	Resultados do DIC para o Modelo Hierárquico e para o Modelo Espacial com diferentes fatores de desconto.	70
5.2	Resultados do EPD para o Modelo Hierárquico e para o Modelo Espacial com diferentes fatores de desconto.	71

Lista de Figuras

2.1	Curva de crescimento exponencial $\alpha + \beta \exp(\gamma t)$ para $\lambda = 1$ e diferentes valores de α , β e γ	10
3.1	Municípios da PNAD selecionados para análise.	20
3.2	Região onde é sugerida a aplicação de modelos CAR (Souza 2004).	26
4.1	Gráfico de dispersão de $\log(s_i^*)$ versus $\log(s_i')$	33
4.2	Localização dos municípios selecionados para análise e para previsão.	35
4.3	Diferença entre a densidade populacional em 2000 e em 1991 (em habitantes por quilometro quadrado).	37
5.1	Verossimilhanças de $(\beta_{14}, \gamma_{14})$ (linha a) e $(\beta_{17}, \gamma_{17})$ (linha b) calculadas a menos de uma constante de proporcionalidade e fixando os demais parâmetros nos valores usados para simular dados.	50
5.2	Taxas de aceitação obtidas a cada 50 ciclos do Amostrador de Gibbs durante o processo de sintonia do Modelo Espacial com fator 0,90.	52
5.3	Cadeias a posteriori para o Modelo Espacial com fator de desconto de 0,90, <i>burn-in</i> 0 e <i>thinning</i> 150.	57
5.4	Cadeias a posteriori para o Modelo Hierárquico com fator de desconto de 0,80, <i>burn-in</i> 0 e <i>thinning</i> 100.	58
5.5	Estimação dos parâmetros γ_i (a) e β_i (b) para o Modelo Hierárquico ajustado com conjunto de dados artificiais, onde “●” representa a média a posteriori, “×” o valor usado para simular os dados e a linha informa o intervalo de 95% de credibilidade da média a posteriori.	59

5.6	Estimação dos parâmetros γ_i (a) e β_i (b) para o Modelo Espacial ajustado com conjunto de dados artificiais, onde “•” representa a média a posteriori destes parâmetros, “×” o valor usado para simular os dados e a linha informa o intervalo de 95% de credibilidade da média a posteriori.	60
5.7	Histogramas de amostras a posteriori sob o Modelo Espacial com fator 0,95 e valores usados para gerar os dados (linha azul).	61
5.8	$\mu_{i,t}$ usado para simular os dados do modelo Hierárquico versus suas estimativas a posteriori.	62
5.9	Ajustes do Modelo Espacial com a estratégia $W_{i,t} = \tau^2$ para tratar a variância dos erros de evolução.	63
5.10	Ajuste do Modelo Espacial com fator de desconto 0,90.	64
5.11	Gráfico de dispersão do i (município) versus estatística Q_i (precisão relativa entre PNAD e censo).	65
5.12	Ajustes do Modelo Espacial com fator 0,90 para os municípios marcados por “×” na figura 5.11.	66
5.13	Ajustes do Modelo Espacial com fator 0,90 para os municípios marcados por “•” na figura 5.11.	66
5.14	Ajustes do Modelo Espacial com fator 0,90 para alguns municípios que apresentaram indícios de subcobertura censitária.	68
5.15	Ajustes do Modelo Espacial com fatores de desconto 0,50 (linha a) e 0,90 (linha b).	69
5.16	Estimativas da densidade populacional publicadas pelo IBGE (pontos verdes), previsões para o nível (linha vermelha) e intervalo de 95% de credibilidade associado (linha tracejada) obtidos sob o Modelo Espacial com fator 0,50.	72
5.17	Ajuste do Modelo Espacial com fator 0,50 (em vermelho) e estimativas publicadas pelo IBGE (pontos verdes).	73
5.18	Ajuste do Modelo Espacial com fator 0,80 (em vermelho) e estimativas publicadas pelo IBGE (pontos verdes).	73

5.19	Resultados apresentados em Souza (2004), com médias a posteriori para o nível e intervalo de credibilidade em vermelho e em azul as observações $y_{i,t}$.	74
2.1	Taxas de aceitação obtidas a cada 50 ciclos do Amostrador de Gibbs sob o Modelo Espacial com fator de desconto de 0,90.	80
5.1	Histogramas e funções de autocorrelação de algumas cadeias a posteriori para o Modelo Espacial com fator de desconto 0,90, <i>burn-in</i> 2000 e <i>thin</i> 150.	91
5.2	Histogramas e funções de autocorrelação de algumas cadeias a posteriori para o Modelo Espacial com fator de desconto 0,90, <i>burn-in</i> 2000 e <i>thin</i> 150.	92
6.1	Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.	93
6.2	Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.	94
7.1	Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.	95
7.2	Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.	96

Capítulo 1

Introdução

O termo pequenas áreas é freqüentemente utilizado para denominar áreas geográficas pequenas, tais como municípios, distritos e bairros. Num sentido mais amplo, é utilizado também para descrever um pequeno subconjunto da população, podendo ser, inclusive, resultado da classificação cruzada de duas ou mais variáveis, como por exemplo, uma subpopulação obtida a partir da seleção de pessoas do sexo feminino com idades entre 20 e 30 anos (Souza, 2004). Por isso, muitos autores preferem a expressão “pequeno domínio” ao invés de “pequena área”.

As informações para pequenos domínios podem ser obtidas a partir de censos e enumerações populacionais, com a população sendo inteiramente investigada, ou através de pesquisas por amostragem. Para conhecer algumas características de uma população, é comum observar apenas um subconjunto de seus elementos, denominado de amostra e, a partir dos resultados dessa amostra, obter estimativas para as características de interesse da população. Neste caso, a amostra deve ser representativa da população, ou seja, deve ter características similares às da população de onde foi retirada. Em geral, as razões para realizar uma amostragem se devem à economia de custos para investigar uma população, à redução do tempo e de mão-de-obra para a realização da coleta dos dados, à confiabilidade e qualidade dos dados e à facilidade na realização dos trabalhos. Quando a população é pequena, quando a característica é de fácil mensuração ou quando há necessidade de alta precisão, pode não ser interessante a realização de uma amostragem.

No Brasil, um exemplo de pesquisa por amostragem de múltiplos propósitos é a Pes-

quisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Como exemplo da necessidade de estimativas confiáveis para pequenas áreas, podemos citar a distribuição de verbas federais pelo Fundo de Participação dos Municípios, definida a partir de critérios que levam em consideração os tamanhos das populações municipais (Souza, 2004).

Nas últimas décadas, as pesquisas por amostragem com múltiplos propósitos, como a PNAD, tomaram grande impulso por apresentarem menor custo em relação aos censos e contagens, e por possibilitarem a obtenção de informações de interesse com periodicidade menor que a dos censos. No entanto, pesquisas por amostragem geralmente apresentam tamanhos de amostras reduzidos nos pequenos domínios, ou até mesmo a ausência de unidades amostrais em algumas áreas, não sendo possível obter estimativas com boa precisão a partir de estimadores diretos, baseados unicamente no desenho da amostra. Com raras exceções, o desenho e o tamanho da amostra são escolhidos de modo que a amostra forneça estimativas com boa precisão para subpopulações maiores, formadas pela agregação dos pequenos domínios, optando-se por exemplo, por não apresentar estimativas por sexo, município ou faixa de idade, mas apenas para o agregado.

Assim, a estimação em pequenas áreas vem recebendo atenção cada vez maior de pesquisadores, com diversas propostas de modelos que objetivam o aumento da precisão das estimativas e, conseqüentemente, permitem decisões mais corretas e conscientes por parte de administradores públicos e privados. Este trabalho se insere neste contexto, utilizando modelos e dados obtidos a partir de censos e pesquisas por amostragem para estimação em pequenas áreas, sendo que a variável de interesse é a densidade populacional municipal.

Como uma amostra jamais será a representação perfeita da população, todos os estudos por amostragem têm desvios em relação aos valores reais do universo. Esses desvios podem arrumar-se em dois grandes grupos: erro amostral e erro não amostral. O primeiro é mensurável e varia na ordem inversa da dimensão da amostra ou subamostra em análise. O segundo, não mensurável, advém de fatores como perguntas mal desenhadas, entrevistas mal aplicadas ou erros de introdução dos dados. Assim, destacam-se duas questões importantes ao considerarmos o problema de estimação em pequenas áreas: a

primeira diz respeito a como produzir estimativas confiáveis das características de interesse investigadas baseando-se na informação proveniente de amostras muito pequenas; a segunda está relacionada a como tratar o erro de estimação.

Existindo somente uma amostra pequena para um particular domínio, uma solução possível para o problema de estimação é pedir emprestada informação de outros dados relacionados, tais como: dados obtidos para a característica de interesse em outros domínios considerados similares e em outros períodos de tempo. Neste trabalho, iremos intercalar periodicamente informações censitárias e amostrais, promovendo a troca de informação entre períodos de tempo e pontos no espaço. Com isto, espera-se um significativo ganho de informação, uma vez que o censo é extremamente preciso.

Apresentaremos uma aplicação que usa dados de três pesquisas populacionais com diferentes variabilidades: o Censo Demográfico Brasileiro, a Contagem Populacional e a Pesquisa Nacional por Amostra de Domicílios (PNAD). Todas divulgadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Os municípios do Estado de São Paulo formam a base de dados do estudo. Informações obtidas a partir da PNAD caracterizam-se por apresentar baixa precisão a nível municipal, por isto, o município será tratado aqui como pequeno domínio.

Adotando uma abordagem Bayesiana, iremos utilizar conjuntamente estas pesquisas para obter a distribuição a posteriori das densidades populacionais municipais e, conseqüentemente, produzir estimativas desta densidade apresentando o erro associado. A proposta é semelhante à apresentada em Souza (2004), no sentido de fazer uso de modelos que permitam a troca de informação entre tempo e espaço, e com isso, aumentar a precisão das estimativas.

Ao utilizar conjuntamente pesquisas com diferentes variabilidades e tipos de erros, é essencial tratar de modo diferenciado as precisões das informações levantadas. Neste trabalho, propomos a modelagem conjunta de estimativas diretas, baseadas unicamente no desenho da amostra, e de suas estimativas da variância. A intenção é fazer com que as estimativas da variância funcionem como um filtro, regulando a troca de informação temporal e espacial. Para tratar a evolução temporal, optou-se pelo uso de Modelos Dinâmicos de Crescimento Exponencial e, para captar a dependência espacial

desta evolução, serão usados Processos Gaussianos.

Em síntese, o objetivo principal deste trabalho é aplicar modelos capazes de tratar conjuntamente informações de pesquisas com diferentes precisões, incorporando desvios mensuráveis ao processo de estimação e promovendo a troca de informação entre períodos de tempo e localizações no espaço. Particularmente, o interesse é obter estimativas intercensitárias atualizadas a partir de informações para pequenos domínios.

Esta dissertação está organizada da seguinte forma. No capítulo 2 revisaremos alguns conceitos abordados ao longo do trabalho. No capítulo 3 descreveremos o conjunto de dados utilizado e apresentaremos as principais idéias de Souza (2004), estudo que foi referência para esta dissertação. No capítulo 4 apresentaremos os modelos propostos, discutindo suas motivações e características. No capítulo 5 apresentaremos a análise dos resultados obtidos e, como o procedimento de inferência adotado segue o Paradigma de Bayes, apresentaremos também neste capítulo os métodos de simulação estocástica e algoritmos computacionais utilizados. Finalmente, no capítulo 6 apresentaremos algumas conclusões e perspectivas futuras deste trabalho.

Capítulo 2

Revisão de Literatura

Neste capítulo, faremos uma breve revisão de alguns conceitos abordados ao longo deste trabalho. Apresentaremos os principais fundamentos da inferência estatística sob o enfoque Bayesiano. Definiremos o Modelo de Crescimento Exponencial Generalizado e discutiremos suas principais características. Apresentaremos uma introdução à modelagem dinâmica, assim como alguns Modelos Dinâmicos de Crescimento Exponencial encontrados na literatura. Por fim, dissertaremos sobre algumas técnicas e conceitos de Geoestatística, a área da estatística que modela observações feitas ao longo de uma região geográfica.

2.1 Noções de Inferência Bayesiana

A inferência Bayesiana, assim como a frequentista, trabalha na presença de observações \mathbf{Y} , que são descritas através de uma distribuição de probabilidade com densidade $p(\mathbf{Y} | \theta)$. A quantidade θ é desconhecida (parâmetro), podendo ser um escalar, um vetor ou uma matriz. Podemos dizer que o principal objetivo de um estudo de inferência é a determinação do valor de θ . É bastante provável que o pesquisador saiba caracterizá-lo de alguma maneira. Neste caso, é possível e recomendável que esse conhecimento prévio a respeito da quantidade seja incorporado à análise. É nesse ponto que o método Bayesiano se diferencia do frequentista. Enquanto o segundo não admite o uso dessa informação, por não ser observável, o primeiro a incorpora à análise através de

uma densidade $p(\theta)$. Esta densidade é conhecida como densidade a priori e representa a informação disponível sobre θ antes de observar \mathbf{Y} . Para uma leitura extensiva sobre o assunto veja Gamerman & Migon (1999).

A função de verossimilhança de θ , geralmente denotada por $l(\theta, \mathbf{Y})$, é a função que associa cada θ ao valor $p(\mathbf{Y} | \theta)$. Considerando Θ como o espaço paramétrico de θ , podemos defini-la da seguinte maneira:

$$\begin{aligned} l(\theta, \mathbf{Y}) : \Theta &\rightarrow R^+ \\ \theta &\rightarrow l(\theta, \mathbf{Y}) = p(\mathbf{Y} | \theta) \end{aligned}$$

O processo de inferência Bayesiano se baseia na distribuição de probabilidade de θ após observar \mathbf{Y} , que passa a fazer parte do conjunto de informações disponíveis. Esta distribuição é conhecida como distribuição a posteriori de θ .

O Teorema de Bayes fornece uma regra de atualização de probabilidades, que parte de $p(\theta)$ e resulta na posteriori $p(\theta | \mathbf{Y})$ da seguinte maneira:

$$\begin{aligned} p(\theta | \mathbf{Y}) &= \frac{p(\mathbf{Y} | \theta)p(\theta)}{p(\mathbf{Y})}, \text{ onde} \\ P(\mathbf{Y}) &= \int_{\Theta} p(\mathbf{Y}, \theta)d\theta = \int_{\Theta} p(\mathbf{Y} | \theta)p(\theta)d\theta. \end{aligned}$$

Note que $p(\mathbf{Y})$ não depende de θ e, portanto, pode ser considerado constante. Assim, pode-se reescrever o teorema de Bayes como

$$p(\theta | \mathbf{Y}) \propto p(\mathbf{Y} | \theta)p(\theta).$$

2.1.1 Estimação de Parâmetros

Depois de obtida a distribuição a posteriori, é necessária a estimação do parâmetro θ . Um estimador, $\hat{\theta}$, pode ser definido como uma regra de decisão ótima a respeito de uma dada função perda esperada, $L(\delta, \theta)$ (Gamerman & Migon, 1999). A estimação pode ser pontual ou intervalar.

- **Estimação Pontual:** Na metodologia Bayesiana, a estimação pontual lida sempre com a minimização de uma função perda, $L(\delta, \theta) = h(\theta - \delta)$, para alguma função h e estimador $\delta = \hat{\theta}$. Três funções perda são geralmente utilizadas,

- a função perda quadrática $L(\delta, \theta) = (\theta - \delta)^2$,
- a função perda absoluta $L(\delta, \theta) = |\theta - \delta|$ e
- a função perda zero-um $L(\delta, \theta) = \lim_{\varepsilon \rightarrow 0} I_{|\theta - \delta|}(|\varepsilon, \infty|)$.

Os estimadores obtidos minimizando estas funções são, respectivamente,

- a média a posteriori $(\hat{\theta} = E(\theta | \mathbf{Y}))$,
- a mediana a posteriori $\hat{\theta} \left(\int_{-\infty}^{\hat{\theta}} p(\theta | \mathbf{Y}) d\theta = 0.5 \right)$ e
- a moda a posteriori $\hat{\theta} \left(p(\hat{\theta} | \mathbf{Y}) = \sup_{\theta} p(\theta | \mathbf{Y}) \right)$.

- **Estimação por Intervalo:** Sintetizar toda a informação da distribuição a posteriori em um único valor pode não ser apropriado. O uso de intervalos de credibilidade a posteriori permite dimensionar a precisão deste valor. Uma região em $C \subset \Theta$ é uma região com $100(1 - \alpha)\%$ de credibilidade se $P(\theta \in C | \mathbf{Y}) \geq 1 - \alpha$. Neste caso, $1 - \alpha$ é chamado de nível de credibilidade. Desta forma, um intervalo de 95% de credibilidade a posteriori pode ser obtido a partir dos quantis 2.5% (q_1) e 97.5% (q_2) da posteriori, ou seja,

$$\int_{-\infty}^{q_1} p(\theta | \mathbf{Y}) d\theta = 0.025 \quad \text{e} \quad \int_{-\infty}^{q_2} p(\theta | \mathbf{Y}) d\theta = 0.975.$$

2.2 Modelo de Crescimento Exponencial Generalizado

Suponhamos que as variáveis aleatórias y_t , observadas ao longo dos períodos $t = 1, 2, \dots, T$, têm distribuição de probabilidade tal que

$$E(y_t | \boldsymbol{\theta}_t) = \mu_t,$$

onde θ_t é um conjunto de parâmetros. Modelos caracterizados pela parametrização $\theta = [\alpha \ \beta \ \gamma \ \lambda]$ e por uma função de ligação h tal que

$$h(\mu_t) = \alpha + \beta \exp(\gamma t) \quad \text{e}$$

$$h(\mu_t) = \begin{cases} \mu_t^\lambda & \text{se } \lambda \neq 0 \\ \log(\mu_t) & \text{se } \lambda \cong 0 \end{cases}$$

são chamados de Modelos de Crescimento Exponencial Generalizados (MCEG) (Migon & Gamerman, 1993).

Algumas funções de ligação especiais conhecidas na literatura são:

- Logística, quando $\lambda = -1$,
- Gompertz, quando $\lambda = 0$ e
- Exponencial Modificada, quando $\lambda = 1$.

Em particular, um MCEG com função de ligação Exponencial Modificada em que $(y_t|\theta_t)$ tem distribuição normal é descrito pelas equações

$$y_t = \mu_t + v_t, \text{ onde } v_t \sim N(0, \sigma_t^2) \quad \text{e} \tag{2.1}$$

$$\mu_t = \alpha + \beta \exp(\gamma t).$$

Faremos a seguir uma breve análise dos parâmetros α , β e γ para o caso específico $\lambda = 1$.

1. Na figura 2.1, apresentamos algumas curvas de crescimento exponencial com $\lambda = 1$ e diferentes valores de α , β e γ . Comparando o gráfico (a) com os demais gráficos, podemos visualizar a influência destes parâmetros no comportamento da curva. O parâmetro γ está relacionado com a velocidade de crescimento/decrescimento (ou curvatura), o parâmetro β com a intensidade do crescimento/decrescimento e o α com a localização da curva. Derivando a expressão $\alpha + \beta \exp(\gamma t)$ em relação a t , podemos concluir que a curva será crescente se α e β tiverem o mesmo sinal e decrescente caso contrário. Pela derivada segunda, podemos concluir que a curva tem concavidade voltada para cima se $\beta > 0$ e para baixo se $\beta < 0$. Vale notar ainda que se $\gamma < 0$ então a curva tem um comportamento não explosivo, convergindo para α quando $t \rightarrow \infty$.

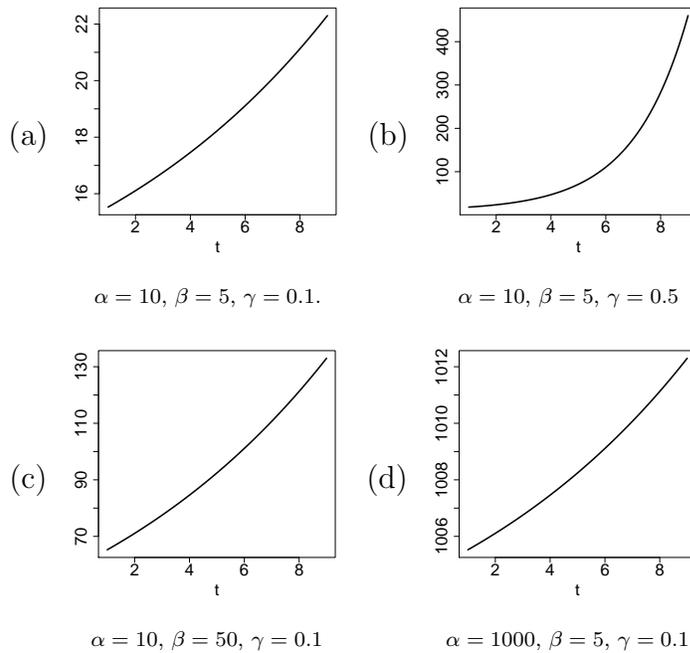


Figura 2.1: Curva de crescimento exponencial $\alpha + \beta \exp(\gamma t)$ para $\lambda = 1$ e diferentes valores de α , β e γ .

A principal vantagem dos modelos abordados nesta seção é a possibilidade de manter as medições y_t na escala original, transformando apenas a trajetória de μ_t , o que torna mais simples a interpretação dos resultados.

Uma crítica a estes modelos questiona o fato de μ_t , a média do processo, seguir exatamente uma função exponencial ao longo do tempo. Sendo μ_t geralmente associado ao verdadeiro valor de uma determinada característica de interesse, parece mais natural adicionar uma perturbação aleatória na evolução temporal de μ_t . Nas seções 2.3.2 e 4.1, abordaremos modelagens mais flexíveis, que procuram manter a tendência exponencial descrita em (2.1) e permitem ainda captar oscilações do nível μ_t em torno desta tendência.

2.3 Modelos Dinâmicos

Uma forma flexível de tratar a variação temporal é através de Modelos Dinâmicos (MD), introduzidos por (Harrison & Stevens, 1976) e bem documentados em West & Harrison (1997). O princípio fundamental dessa abordagem é o uso de modelos proba-

bilísticos sequenciais, definidos através de estruturas paramétricas. Estes modelos são caracterizados pela distribuição de uma variável aleatória observada condicionalmente a um vetor de parâmetros e por distribuições que definem a evolução temporal destes parâmetros.

De um modo bem geral, um MD é caracterizado pelas distribuições

$$\begin{aligned} p(\mathbf{y}_t | \boldsymbol{\theta}_t, \lambda) & \text{ (observacionais),} \\ p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \lambda) & \text{ (de evolução ou de transição) e} \\ p(\boldsymbol{\theta}_0 | D_0) & \text{ (a priori),} \end{aligned}$$

onde, para cada tempo t , \mathbf{y}_t é um vetor de observações e $\boldsymbol{\theta}_t$ é um vetor de parâmetros, chamado de vetor de estado. D_0 representa a informação a priori disponível antes de observar os dados e λ é um conjunto de parâmetros estáticos, geralmente referenciados como hiperparâmetros.

O procedimento de inferência para estes modelos segue os passos usuais da inferência Bayesiana, combinando duas operações principais: evolução e atualização. A evolução consiste no cálculo sequencial das distribuições a priori $p(\boldsymbol{\theta}_t | D_{t-1})$ e a atualização no cálculo das distribuições $p(\boldsymbol{\theta}_t | D_t)$.

Seja $D_t = D_{t-1} \cup \{\mathbf{y}_t\}$ a informação disponível até o tempo t . As distribuições a priori, preditivas e a posteriori para cada período de tempo são dadas, respectivamente, por

$$\begin{aligned} p(\boldsymbol{\theta}_t | D_{t-1}) &= \int p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1} | D_{t-1})d\boldsymbol{\theta}_{t-1}, \\ p(\mathbf{y}_t | D_{t-1}) &= \int p(\mathbf{y}_t | \boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t | D_{t-1})d\boldsymbol{\theta}_t \quad \text{e} \\ p(\boldsymbol{\theta}_t | D_t) &= p(\mathbf{y}_t | D_{t-1})p(\boldsymbol{\theta}_t | D_{t-1}). \end{aligned}$$

Sendo que a última é obtida via Teorema de Bayes. Mais especificamente, um Modelo Linear Dinâmico (MLD), como é comumente conhecido, é caracterizado por um conjunto de quádruplas $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ para cada tempo t , onde

a) \mathbf{F}_t é uma matriz $p_1 \times p_2$,

b) \mathbf{G}_t é uma matriz $p_1 \times p_1$,

c) \mathbf{V}_t é uma matriz de covariâncias $p_2 \times p_2$ e

d) \mathbf{W}_t é uma matriz de covariâncias $p_1 \times p_1$.

Estas quádruplas definem o modelo através das distribuições sequenciais

$$\begin{aligned} (\mathbf{y}_t \mid \boldsymbol{\theta}_t, \mathbf{V}_t) &\sim N(\mathbf{F}_t' \boldsymbol{\theta}_t, \mathbf{V}_t) \\ (\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) &\sim N(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t). \end{aligned} \tag{2.2}$$

Para $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ conhecidos, a inferência nesta classe de modelos pode ser feita utilizando o filtro de Kalman de maneira sequencial (West & Harrison, 1997), o que fornece de maneira eficiente a distribuição condicional de $\boldsymbol{\theta}_t$ dado D_t . No entanto, em aplicações a dados reais, as quádruplas $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ são geralmente desconhecidas, ou parcialmente desconhecidas. Para estes casos e para modelos mais complexos, os métodos de simulação estocástica que serão apresentados na seção 5.1 permitem obter amostras de $p(\boldsymbol{\theta}_t \mid D_t)$ e, conseqüentemente, sumarizar a informação desta distribuição.

2.3.1 Fatores de Desconto

Consideremos o MLD

$$\begin{aligned} (\mathbf{y}_t \mid \boldsymbol{\theta}_t, \mathbf{V}_t) &\sim N(\mathbf{F}_t' \boldsymbol{\theta}_t, \mathbf{V}_t) \\ (\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) &\sim N(\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t). \end{aligned}$$

Para evitar a difícil tarefa de especificar ou estimar as variâncias \mathbf{W}_t dos erros de evolução, pode-se usar uma estratégia de descontos. A seguir, iremos descrever esta estratégia.

Aplicando o filtro de Kalman ao modelo acima de maneira sequencial (West & Harrison, 1997), a variância a priori $\mathbf{R}_t = \text{Var}(\boldsymbol{\theta}_t \mid D_{t-1})$ é obtida por

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t, \tag{2.3}$$

onde \mathbf{C}_{t-1} é a variância de $(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$.

Usando um Fator de Desconto δ , pode-se reescrever (2.3) como

$$\mathbf{R}_t = \frac{\mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'}{\delta},$$

o que mostra claramente uma relação entre \mathbf{W}_t e δ . Esta relação é dada por $\mathbf{W}_t = (\delta^{-1} - 1) \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t$. Fazendo isto, a perda de informação sobre $\boldsymbol{\theta}_t$ ao passar de um tempo $t - 1$ ao tempo t é proporcional à variância da distribuição a posteriori no tempo $t - 1$. Assim, δ é um número entre zero e um que descreve a perda de informação ao passar do tempo. Por exemplo, se $\delta = 0,90$, apenas 90% da informação passa de um tempo a outro.

Esta abordagem será utilizada nos modelos propostos no capítulo 4.

2.3.2 Modelos Dinâmicos de Crescimento Exponencial

Migon & Gamerman (1993) sugerem dois Modelos Dinâmicos capazes de descrever a evolução temporal em (2.1). Nesta seção, apresentaremos estes modelos.

O primeiro é um MD descrito por

$$\begin{aligned} (y_t | \boldsymbol{\theta}_t, \mathbf{V}_t) &\sim N(\mathbf{F}'\boldsymbol{\theta}_t, \mathbf{V}_t) \text{ e} \\ (\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) &\sim N(\mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}_t), \end{aligned}$$

onde $\boldsymbol{\theta}_t = [\theta_{t,1} \ \theta_{t,2}]'$, $\mathbf{F} = [1 \ 0]$ e

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 0 & \exp(\gamma) \end{bmatrix}.$$

Neste modelo, o parâmetro $\theta_{t,1}$ descreve o nível do processo. Já γ e $\theta_{t,2}$ descrevem a evolução exponencial deste nível.

O segundo é um MD descrito por

$$\begin{aligned} (y_t | \boldsymbol{\theta}_t, \mathbf{V}_t) &\sim N(\mathbf{F}'(\boldsymbol{\theta}_t), \mathbf{V}_t) \text{ e} \\ (\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \mathbf{W}_t) &\sim N(g(\boldsymbol{\theta}_{t-1}), \mathbf{W}_t), \end{aligned}$$

onde $\boldsymbol{\theta}_t = [\theta_{t,1} \ \theta_{t,2} \ \theta_{t,3}]'$, $\mathbf{F} = [1 \ 0 \ 0]'$ e $g(\boldsymbol{\theta}_t)$ é uma função tal que

$$g(\boldsymbol{\theta}_t) = \begin{bmatrix} \theta_{t,1} + \theta_{t,2} \\ \theta_{t,2} \ \theta_{t,3} \\ \theta_{t,3} \end{bmatrix}.$$

Agora, $\theta_{t,2}$ e $\theta_{t,3}$ descrevem a evolução exponencial do nível e $\theta_{t,1}$ o nível. A inferência para este modelo pode ser feita usando a expansão em série de Taylor para a função g (Migon & Gamerman, 1993).

Uma característica destes modelos é a evolução dinâmica dos parâmetros associados ao crescimento exponencial ($\theta_{t,2}$ e $\theta_{t,3}$ no segundo modelo e $\theta_{t,2}$ no primeiro modelo).

Neste trabalho, iremos propor um Modelo Dinâmico de Crescimento Exponencial semelhante aos apresentados nesta seção. Esse modelo será descrito detalhadamente na seção 4.1

2.4 Geoestatística

2.4.1 Dados Espacialmente Referenciados

Ultimamente, temos observado um grande desenvolvimento de novas técnicas para o estudo de dados espacialmente referenciados, ou seja, observados ao longo de uma região geográfica. Na análise deste tipo de dado, agregamos as observações às suas respectivas localizações e a análise é feita levando em consideração essas localizações.

Em estudos envolvendo observações espacialmente referenciadas, a caracterização da estrutura de correlação espacial é de grande importância. É intuitivo pensar que observações próximas tendem a ser mais correlacionadas do que observações mais distantes entre si. Desta forma, quando modelamos dados deste tipo, desejamos identificar estas estruturas. Para simplificar o processo de inferência, é comum na literatura de geoestatística, assumir estacionariedade e isotropia do processo sob estudo. A seguir, descreveremos estas duas propriedades.

Cressie (1993) divide a estatística espacial de acordo com os tipos de observações associadas ao espaço em que elas são observadas. De uma forma geral, a estatística espacial contém três grandes áreas: geoestatística, dados de área e processos pontuais. Neste trabalho abordaremos apenas o primeiro tipo. Seja $s \in \mathbb{R}^p$ um ponto no espaço euclidiano p -dimensional e seja $Y(s)$ o valor da variável aleatória Y no ponto s . Consideremos que $\{y(s), s \in G\}$ é uma realização parcial do campo aleatório $\{Y(s), s \in G\}$, onde G é um

subconjunto fixo de \mathbb{R}^p com volume p-dimensional positivo (Cressie, 1993). Em geral, vamos nos referir aos elementos de G como estações ou localizações monitoradoras.

2.4.2 Estacionariedade e Isotropia

Um campo aleatório $\{Y(s), s \in G\}$ é estritamente estacionário se para qualquer coleção finita de estações $s_1, s_2, \dots, s_n \in G$ e qualquer $u \in G$, a distribuição conjunta de $(Y(s_1), Y(s_2), \dots, Y(s_n))$ é idêntica à distribuição conjunta de $(Y(s_1 + u), Y(s_2 + u), \dots, Y(s_n + u))$.

Uma definição menos restritiva de estacionariedade exige que as funções de média e covariância sejam invariantes sob translações, ou seja,

$$m(s) = E(Y(s)) = m \quad \text{e} \quad Cov(Y(s), Y(s + u)) = c(u), \quad \forall s, s + u \in G.$$

Este tipo de estacionariedade é chamada de estacionariedade fraca. A função $c(\cdot)$ é chamada de covariograma.

Note que estacionariedade no sentido estrito implica em estacionariedade fraca, mas em geral a recíproca não é verdadeira. Note ainda que qualquer campo aleatório estacionário precisa ter variância constante. Assim, podemos considerar então $Var(s) = \varepsilon^2 \quad \forall s \in G$.

Um campo aleatório é intrinsecamente estacionário se

$$E(Y(s + u) - Y(s)) = 0 \quad \text{e} \quad Var(Y(s + u) - Y(s)) = 2\gamma(u), \quad \forall s, s + u \in G.$$

A função $2\gamma(\cdot)$ é chamada de variograma e $\gamma(\cdot)$ de semi-variograma. Note que se um processo é fracamente estacionário, então é intrinsecamente estacionário.

Um campo aleatório estacionario é isotrópico se a função de covariância depende apenas da distância, ou seja, $Cov(Y(s), Y(s')) = c(\|s - s'\|)$, onde $\|s - s'\|$ denota a distância entre s e s' .

Usualmente na literatura de geoestatística, processos espaciais estacionários e isotrópicos são chamados de homogêneos. Caso contrário, são conhecidos como heterogêneos.

Quando um processo é homogêneo, podemos escrever a função de covariância como $Cov(Y(s), Y(s')) = \varepsilon^2 \rho(\|s - s'\|; \phi^*)$, onde $\rho(\cdot; \phi^*)$ denota uma função de correlação válida (positiva definida) em \mathbb{R}^p (Schmidt et al., 2002), que depende possivelmente de um

vetor paramétrico ϕ^* . A seção a seguir descreve duas funções de correlação encontradas frequentemente na literatura de geoestatística.

Dessa forma, a conveniência dos processos homogêneos fica clara, já que a estrutura de covariância do processo pode ser modelada apenas através dos parâmetros ε^2 e ϕ^* .

2.4.3 Processos Gaussianos

Um campo aleatório $\{Y(s), s \in G\}$ segue um Processo Gaussiano (PG) com função de média $m(\cdot)$ e função de covariância $Cov(\cdot, \cdot)$, denotando por $Y(\cdot) \sim PG(m(\cdot), Cov(\cdot, \cdot))$, se para quaisquer $s_1, s_2, \dots, s_z \in G$ e qualquer $z \in \{1, 2, \dots\}$, a distribuição conjunta de $(Y(s_1), Y(s_2), \dots, Y(s_z))$ é normal multivariada com parâmetros dados por $E(Y(s_i)) = m(s_i)$ e $Cov(Y(s_i), Y(s_j))$, para $i, j \in \{1, 2, \dots, z\}$.

2.4.4 Famílias de Funções de Correlação

Existem na literatura algumas famílias de funções de correlação. Geralmente a função de correlação $\rho(\cdot)$ deve traduzir algumas de nossas intuições. Por exemplo, espera-se que a correlação entre duas medidas decresça ao aumentarmos a distância entre suas localizações, e também que $\lim_{d \rightarrow \infty} \rho(d) = 0$, isto é, a correlação entre duas localizações muito distantes tende a 0. Diggle & Ribeiro Jr (2000) apresentam alguns dos principais modelos paramétricos de funções de correlação, assim como simulações de processos Gaussianos univariados mostrando o efeito do uso das diferentes famílias. Dois dos principais exemplos de função de correlação são:

a) Família Exponencial Potência:

$$\rho(d; \phi, \kappa) = \exp(-(\phi d)^\kappa),$$

onde $\phi > 0$ e $\kappa \in (0, 2]$. O parâmetro ϕ é de escala e d é a distância euclidiana entre dois pontos quaisquer em G . Quando $\kappa = 1$ temos o caso particular da função de correlação exponencial. Quando $\kappa = 2$, temos a função de correlação Gaussiana.

b) Família Matérn:

$$\rho(d; \phi, \lambda) = \frac{1}{2^{\lambda-1} \Gamma(\lambda)} \left(2\sqrt{\lambda} d \phi \right)^{\lambda-1/2} K_{\lambda-1/2} \left(2\sqrt{\lambda} d \phi \right),$$

onde $\phi > 0$ é o parâmetro de escala, $\lambda > 0$ é o parâmetro de forma. A função $\Gamma(\cdot)$ é a função Gama usual e κ_λ é a função modificada de Bessel do terceiro tipo de ordem λ .

Um aspecto importante de superfícies espaciais é o seu grau de suavidade. Matematicamente, essa propriedade é descrita através do grau de diferenciabilidade do processo. A especificação da família de função de correlação é de grande importância, pois em Processos Gaussianos, a suavidade do processo está diretamente relacionada à diferenciabilidade da sua estrutura de covariância. Por exemplo, a função de correlação Gaussiana resulta em processos infinitamente diferenciáveis, ou seja, em processos extremamente suaves, o que na prática é difícil de se observar. Recentemente, a função de correlação em (b) tem sido a mais usada na literatura, pois além das funções exponencial (quando $\lambda = 0.5$) e Gaussiana (quando $\lambda \rightarrow \infty$) serem seus casos particulares, o parâmetro λ controla a diferenciabilidade do processo (Schmidt et al., 2002).

Neste trabalho, usaremos a função de correlação exponencial $\rho(d; \phi) = \exp(-\phi d)$. Nela, o parâmetro ϕ nos diz quão rapidamente a correlação decai pra zero.

Os tópicos aqui revisados serão base dos modelos propostos (no capítulo 4) para ajustar os dados apresentados no próximo capítulo).

Capítulo 3

Conjunto de Dados e Estudo de Referência

Neste trabalho, apresentaremos uma aplicação a dados de três pesquisas com diferentes variabilidades: o Censo Demográfico Brasileiro, a Contagem Populacional e a Pesquisa Nacional por Amostra de Domicílios (PNAD), todas divulgadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Por simplicidade, a Contagem Populacional será tratada aqui como pesquisa censitária.

Os municípios do estado de São Paulo formam a base de dados do estudo. O objetivo é fazer inferência sobre a densidade populacional dos municípios paulistas, combinando os dados das três pesquisas. Para isto, iremos usar modelos probabilísticos que permitem a troca de informação entre pontos no espaço e períodos de tempo.

Dentre os dados disponíveis, temos as populações dos municípios do estado de São Paulo apuradas pela contagem populacional de 1996 e pelos censos demográficos de 1991 e de 2000. Nos períodos entre censos, dispõe-se de estimativas do total populacional dos municípios selecionados para a amostra da PNAD. Como o nível município não faz parte do domínio de divulgação da PNAD, esta estimativa foi baseada no estimador apresentado em Klein & Moura (1998) e é descrita na equação (A1.2). As estimativas calculadas dessa forma caracterizam-se por levar em conta somente o desenho amostral e por apresentar baixa precisão. Para dimensionar a precisão destas estimativas, foram calculadas também as estimativas da variância descritas pela equação (A1.3). Foram utilizados os dados da

PNAD de 1992 a 1999, exceto o ano de 1994, para o qual a pesquisa não foi realizada. A tabela 3.1 apresenta os anos utilizados, o tipo de pesquisa realizada e a numeração que usaremos para identificá-los. Esta numeração será utilizada para simplificar a notação ao longo do texto.

Anos utilizados	Tipo de pesquisa realizada	Numeração
1991	Censo	1
1992	PNAD	2
1993	PNAD	3
1995	PNAD	4
1996	Contagem Populacional	5
1997	PNAD	6
1998	PNAD	7
1999	PNAD	8
2000	Censo	9

Tabela 3.1: Anos em que foram realizadas as pesquisas, tipo de pesquisa e numeração utilizada para identificar cada ano.

Ao longo do período houve modificações na malha municipal do estado, ocorrendo criação de novos municípios a partir de alguns existentes. Este fato levou à decisão de seguir a base territorial vigente em 1991, descartando os municípios com diferenças significativas entre a área geográfica apurada no censo de 1991 e no censo de 2000. Assim, dentre os 111 municípios selecionados pela PNAD na década de 90, apenas 91 foram usados. A figura 3.1 apresenta os municípios usados. A opção de descartar alguns municípios foi influenciada também pelo tempo computacional exigido pelos métodos de simulação estocástica que serão usados para implementar o processo de inferência (veja a seção 5.1 para maiores detalhes). Assim, seja $\Psi = \{1, 2, \dots, 91\}$ o conjunto que representa os municípios da PNAD utilizados e $\Omega = \{1, 2, \dots, 572\}$ o conjunto que representa todos os 572 municípios paulistas.

- Para $i \in \Psi$ (figura 3.1) e t ano de PNAD, $Y_{i,t}$ representará a estimativa do total

populacional obtida usando o estimador (1.3), apresentado no apêndice A.

- Já para $i \in \Omega$ e t ano de censo ou contagem populacional, $Y_{i,t}$ representará a população total informada no respectivo censo ou contagem.

Além disso, $S_{i,t}$ representará a estimativa da variância de $Y_{i,t}$.

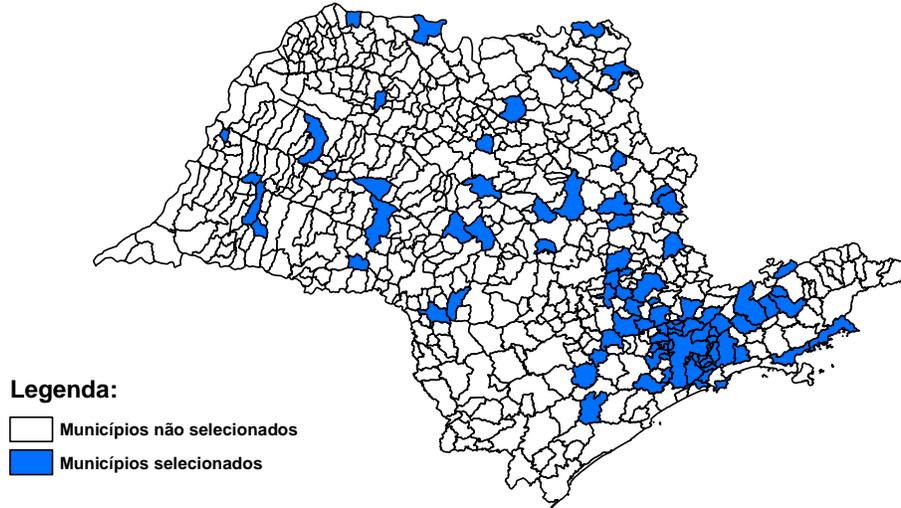


Figura 3.1: Municípios da PNAD selecionados para análise.

Nos censos demográficos brasileiros, apesar do objetivo ser investigar inteiramente a população, ocorrem omissões ou inclusões indevidas, que se constituem num erro de cobertura, considerado como sendo de 5% para todos os municípios (Souza, 2004). Estas falhas de cobertura são comuns em pesquisas como o censo demográfico, sobretudo quando realizadas em países como o Brasil, que apresenta, em seu vasto território, situações bastante difíceis para a coleta de informações. Algumas fontes que levam à ocorrência de falhas de cobertura são: a interpretação incorreta de definições ou procedimentos estabelecidos nos manuais técnicos, problemas na definição de limites geográficos, omissão de unidades residenciais ou não-residenciais e omissão de pessoas residentes em domicílios particulares ocupados (Souza, 2004). Portanto, para t ano de censo, iremos atribuir incerteza à $Y_{i,t}$ assumindo para esta medida uma distribuição de probabilidade

Normal e fixaremos a variância desta distribuição em

$$S_{i,t} = \left(\frac{0.05 \times Y_{i,t}}{2} \right)^2.$$

Com isto, a intenção é manter o intervalo de 95% de credibilidade do erro de cobertura em aproximadamente $(95\% \times Y_{i,t}, 105\% \times Y_{i,t})$.

A partir dos dados descritos acima, optamos por modelar a densidade populacional dos municípios ao invés do total populacional. O uso da densidade foi uma forma de tornar mais similares os parâmetros dos modelos ajustados (Souza, 2004).

Estimativas $y_{i,t}$ das densidades populacionais foram calculadas dividindo os totais populacionais estimados por suas respectivas áreas geográficas em quilômetros quadrados, ou seja,

$$y_{i,t} = \frac{Y_{i,t}}{A_i},$$

onde A_i é a área geográfica do município i informada no censo de 1991.

Estimativas $s_{i,t}$ da variância de $y_{i,t}$ foram então obtidas dividindo as variâncias do total populacional pelo quadrado da área geográfica do município, ou seja,

$$s_{i,t} = \frac{S_{i,t}}{A_i^2},$$

Afim de comparar os resultados obtidos neste trabalho com resultados obtidos por outros pesquisadores, iremos utilizar: as estimativas populacionais oficiais, publicadas pelo IBGE, e os resultados publicados em Souza (2004).

3.1 Descrição das Pesquisas

Apresentaremos nesta seção uma breve descrição das pesquisas populacionais utilizadas. As estimativas populacionais municipais, os registros da Contagem Populacional de 1996 e dos censos de 1991 e de 2000 estão disponíveis no site do IBGE (<http://www.ibge.gov.br>), onde é possível visualizar algumas tabelas, ou mesmo construí-las utilizando a base de dados SIDRA (<http://www.sidra.ibge.gov.br>). Já dados da PNAD podem ser obtidos a partir do CD de Microdados da pesquisa. Através da

“Série Relatórios Metodológicos”, também disponível no site, o IBGE documenta e divulga as metodologias empregadas nas diversas fases do planejamento e execução de suas pesquisas.

3.1.1 Censos no Brasil

O primeiro censo no Brasil foi realizado em 1872. Depois desse, vieram os de 1890, 1900 e 1920. Com a criação do IBGE, em 1936, inaugurou-se uma moderna fase censitária no país. Sendo assim, a experiência do IBGE em censos demográficos remonta a 1940, ano em que foi realizado o primeiro levantamento desse tipo pelo Instituto. A partir de então, o IBGE realiza a cada 10 anos o Censo Demográfico, que é a operação estatística mais importante para retratar a realidade sócio-demográfica do país. O censo de 1990 atrasou um ano, rompendo assim a sequência histórica.

Censos demográficos são operações de levantamento de dados fundamentais para a formulação de políticas públicas e para tomada de decisões de investimentos privados e governamentais. Dos resultados do Censo Demográfico e das Contagens Populacionais são obtidas as tendências e parâmetros indispensáveis à elaboração de projeções e estimativas populacionais.

3.1.2 Contagem Populacional

Em 1996, o IBGE realizou uma contagem da população para todo o território nacional. Desde a década de 1980, havia um plano generalizado entre instituições brasileiras no sentido do IBGE realizar, ao meio da década, uma contagem da população, ou um minicenso. Tais solicitações não foram atendidas naquela década. Contudo, com a Constituição de 1988, criaram-se os mecanismos legais e os instrumentos que reforçavam a posição tributária e administrativa dos municípios na vida nacional, fazendo com que as administrações municipais necessitassem de informações demográficas e sócio-econômicas para a implantação de programas e projetos públicos. O IBGE tem a obrigação legal de fornecer anualmente estimativas da população municipal para a classificação dos municípios na distribuição dos Fundos de Participação Municipal (FPM). Após a Consti-

tuição de 1988, estas estimativas geraram demandas e contestações por parte das administrações municipais. Considerando ainda o grande número de municípios criados após o censo de 1991, o IBGE julgou que a necessidade de uma contagem populacional no meio da década de 1990 passava a ser uma questão prioritária.

3.1.3 Pesquisa Nacional por Amostra de Domicílios

A Pesquisa Nacional por Amostra de Domicílios (PNAD) tem como finalidade a produção de informações básicas para o estudo do desenvolvimento sócio-econômico do país. A unidade de investigação da PNAD é o domicílio, onde são investigadas características sócio-econômicas e outros temas que variam de acordo com as necessidades de informação do país (Souza, 2004). A vantagem de realizar pesquisas domiciliares, como a PNAD, é a obtenção de informações mais atualizadas sobre a população em períodos intercensitários. Uma desvantagem é a possibilidade do surgimento de problemas de representatividade e erros de estimação elevados decorrentes do tamanho reduzido das amostras e da falta de cobertura de algumas áreas (Souza, 2004).

A pesquisa só não é realizada em anos de censos e não foi realizada excepcionalmente no ano de 1994. A amostra de domicílios da PNAD é obtida em três estágios de seleção, considerando os municípios como unidades primárias, os setores censitários como unidades secundárias e os domicílios como unidades terciárias (Souza, 2004). Nas pesquisas realizadas em uma mesma década, são mantidos na amostra os mesmos municípios e setores, enquanto as unidades domiciliares são renovadas anualmente.

As regiões metropolitanas formam o nível mais desagregado para o qual o IBGE considera que as estimativas fornecidas possuem boa precisão.

Resumindo, a PNAD é uma pesquisa de múltiplos propósitos e sua amostra é construída de maneira que: baixos custos de obtenção e apuração de cada variável estudada sejam atingidos. Para uma descrição detalhada do plano amostral da PNAD, ver Souza (2004).

3.1.4 Estimativas Populacionais Oficiais

Desde 1973 que o IBGE, através de seu Departamento de População e Indicadores Sociais (DEPIS) da Diretoria de Pesquisas (DPE), elabora as projeções populacionais para o total do País pelo método das componentes demográficas. A partir de 1989, o IBGE passou a informar, anualmente, estimativas da população residente para todos os municípios brasileiros instalados. Estas estimativas são publicadas no Diário Oficial da União, até 31 de agosto de cada ano. O método adotado pelo IBGE para estimar os totais populacionais dos municípios brasileiros emprega metodologia desenvolvida por Madeira & Simões (1972). Esta metodologia observa a tendência de crescimento populacional do município em relação a mesma tendência de uma área geográfica maior (Unidade da Federação, Grande Região ou País) e leva em consideração ainda, a evolução das componentes demográficas fecundidade, mortalidade e migração. Uma descrição detalhada de métodos para projeções populacionais pode ser encontrada em Souza (2004).

3.2 Estudo de Referência

Fay & Herriot (1979) apresentam um modelo básico em nível de área para obter estimativas da renda per capita dos municípios norte americanos. Souza (2004) estende o trabalho de Fay & Herriot (1979) para situações em que o parâmetro de interesse evolui segundo um modelo de crescimento exponencial e faz uma aplicação ao conjunto de dados aqui utilizado.

Nesta seção, procuramos sintetizar as principais idéias de Souza (2004), principalmente no que diz respeito aos modelos utilizados, permitindo assim um paralelo com a abordagem aqui apresentada.

Souza (2004) propõe o seguinte modelo hierárquico de crescimento exponencial para

as densidades populacionais $y_{i,t}$:

$$\begin{aligned}
 y_{i,t} &= \mu_{i,t} + v_{i,t}, \\
 \mu_{i,t} &= \alpha_i + \beta_i \exp(\gamma_i t), \\
 (\alpha_i \mid \mu_\alpha, \sigma_\alpha^2) &\sim N(\mu_\alpha, \sigma_\alpha^2), \\
 (\beta_i \mid \mu_\beta, \sigma_\beta^2) &\sim N(\mu_\beta, \sigma_\beta^2) \quad \text{e} \\
 (\gamma_i \mid \mu_\gamma, \sigma_\gamma^2) &\sim N(\mu_\gamma, \sigma_\gamma^2),
 \end{aligned} \tag{3.1}$$

onde $\mu_{i,t}$ representa a densidade populacional real do município i no tempo t e $v_{i,t} \sim N(0, \sigma_{i,t}^2)$ representa o erro de estimação.

O objetivo do modelo (3.1) é descrever relações entre os períodos de tempo e entre as áreas, explorando suas similaridades e aumentando a precisão das estimativas. Além disso, este modelo permite apresentar o erro de estimação associado, o que não é possível a partir de técnicas de projeção populacional bastante difundidas, como: o Método das Componentes Demográficas, o Método dos Coeficientes e Método de Relação de Coortes (Souza, 2004). Estes métodos fornecem estimativas pontuais, sem informações sobre o erro associado, fazendo uso apenas dos dados de censos recentes.

Para tratar as diferentes variabilidades das pesquisas, foram consideradas quatro modelagens para $\sigma_{i,t}^2$, que são apresentadas na tabela 3.2.

Modelagem	Lei de Variância
1.1	$\sigma_{i,t}^2 = \sigma^2$
1.2	$\log(\sigma_{i,t}^2) = \eta_0 + \eta_1 f_i$ na PNAD e $\log(\sigma_{i,t}^2) = \log(s_{i,t})$ no censo
1.3	$\log(\sigma_{i,t}^2) = \eta_2 + \eta_3 f_i$ na PNAD e $\log(\sigma_{i,t}^2) = \eta_4 \log(s_{i,t})$ no censo
1.4	$\sigma_{i,t}^{-2} = \eta_5 (1/s_{i,t})$

Tabela 3.2: Modelagens da variância em Souza (2004), onde $\eta_0, \eta_1, \eta_2, \eta_3, \eta_4$ e η_5 são parâmetros a estimar e f_i é a fração de amostragem do município i .

Para avaliar a dependência espacial na região selecionada de municípios vizinhos apre-

sentada na figura 3.2, Souza (2004) sugere também o uso de modelos Condicionalmente Auto-Regressivos (CAR), ou seja, ao invés das modelagens hierárquicas $\beta_i \sim N(\mu_\beta, \sigma_\beta^2)$ e $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$, foram testadas prioris CAR para os parâmetros β_i e α_i . No entanto, o processo de inferência para este modelo apresentou problemas de convergência e o modelo (3.1) foi usado para estimar a população dos municípios desta região. Como há muitas observações faltantes na região selecionada e como a priori CAR é imprópria, a posteriori pode não ser própria, o que pode ser um motivo para este problema de convergência.



Figura 3.2: Região onde é sugerida a aplicação de modelos CAR (Souza 2004).

Como já mencionamos, uma crítica a este tipo de modelo questiona o fato de $\mu_{i,t}$, o nível do processo, seguir exatamente uma função exponencial ao longo do tempo, como em (3.1). Neste caso, sendo $\mu_{i,t}$ o verdadeiro valor da densidade populacional, parece mais natural adicionar uma perturbação aleatória em sua evolução temporal. Diante disto, optamos aqui pelo uso de um Modelo Dinâmico de Crescimento Exponencial como uma forma de permitir oscilações do nível em torno de uma tendência de crescimento exponencial.

Para modelar a variância $\sigma_{i,t}^2$, propomos o uso das estimativas da variância $s_{i,t}$ ao invés das modelagens apresentadas na tabela 3.2 e, para tratar a dependência espacial, optamos pelo uso de Processos Aleatórios Gaussianos. Como este processo contempla

distribuições a priori próprias, sua principal vantagem está na possibilidade de fazer também previsões para localizações não vizinhas, o que sob a hipótese de uma priori CAR é feito de forma ad-hoc (para maiores detalhes, veja Banerjee et al. (2004)).

No próximo capítulo, apresentaremos detalhadamente nossa proposta.

Capítulo 4

Modelagem Conjunta

Neste capítulo, iremos propor modelos para estimar a densidade populacional real a partir das estimativas diretas $y_{i,t}$ e das estimativas da variância $s_{i,t}$. Ao considerar a modelagem conjunta destas estimativas, apresentada na seção 4.2, nossa intenção é fazer com que as estimativas da variância $s_{i,t}$ informem o quão precisas são as estimativas $y_{i,t}$.

Para tratar a correlação temporal das densidades populacionais, propomos na seção 4.1 um Modelo Dinâmico de Crescimento Exponencial semelhante àqueles apresentados em Migon & Gamerman (1993). E, na seção 4.3, iremos descrever o procedimento utilizado para fazer previsões da densidade populacional para municípios fora da amostra da PNAD. Por fim, apresentaremos dois critérios que serão utilizados para usados para comparar modelos.

4.1 Modelo Dinâmico de Crescimento Exponencial

Uma maneira de melhorar a eficiência do processo de estimação consiste em intercalar periodicamente informações censitárias e amostrais, promovendo a troca de informação entre os períodos de tempo. Com isto, espera-se que a elevada precisão dos dados censitários aumente consideravelmente a quantidade de informação disponível, conduzindo assim à estimativas mais confiáveis e informativas.

Enquanto modelos lineares supõem o aumento ou decréscimo constante da população, modelos exponenciais presumem uma porcentagem de aumento ou de diminuição. Por

isto, a escolha destes modelos para descrever o comportamento de populações é extremamente comum. Nesta seção, apresentaremos o Modelo Dinâmico de Crescimento Exponencial que usaremos para tratar a relação temporal entre as estimativas $y_{i,t}$ das densidades populacionais municipais. Nosso interesse é estimar a densidade populacional real $\mu_{i,t}$ do município i no tempo t .

Como o objetivo agora é tratar apenas a relação temporal entre os dados e não a relação espacial, consideremos inicialmente, para simplificar a notação, um conjunto qualquer de observações unidimensionais y_t coletadas ao longo dos tempos $t = 1, 2, \dots, T$. Migon & Gamerman (1993) sugerem como alternativa para o modelo em (2.1), o Modelo Dinâmico descrito pelas equações

$$\begin{aligned} y_t &= F'\boldsymbol{\theta}_t + v_t \quad \text{e} \\ \boldsymbol{\theta}_t &= G\boldsymbol{\theta}_{t-1}, \quad \text{onde} \end{aligned} \tag{4.1}$$

$$v_t \sim N(0, \sigma_t^2), \quad F = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{e} \quad G = \begin{bmatrix} 1 & 1 \\ 0 & \exp(\gamma) \end{bmatrix}.$$

Note que o erro de evolução em (4.1) foi omitido. Nossa intenção é descrever exatamente a curva $\mu_t = \alpha + \beta \exp(\gamma t)$ usando a equação de evolução $\boldsymbol{\theta}_t = G\boldsymbol{\theta}_{t-1}$.

Considerando $\boldsymbol{\theta}_t = [\theta_{t,1} \quad \theta_{t,2}]'$, temos que

- $y_t = F'\boldsymbol{\theta}_t + v_t \Rightarrow y_t = \theta_{t,1} + v_t \Rightarrow \theta_{t,1} = \mu_t \quad \text{e}$
- $\boldsymbol{\theta}_t = G\boldsymbol{\theta}_{t-1} \Rightarrow \begin{cases} \mu_t = \mu_{t-1} + \theta_{t-1,2} \\ \theta_{t,2} = \exp(\gamma)\theta_{t-1,2} \end{cases}.$

Para que a equação de evolução em (4.1) descreva exatamente a curva $\mu_t = \alpha + \beta \exp(\gamma t)$, devemos ter

$$\begin{aligned} \mu_t &= \mu_{t-1} + \theta_{t-1,2} \Rightarrow \\ \alpha + \beta \exp(\gamma t) &= \alpha + \beta \exp(\gamma(t-1)) + \theta_{t-1,2} \Rightarrow \\ \beta \exp(\gamma t) - \beta \exp(\gamma(t-1)) &= \theta_{t-1,2} \Rightarrow \\ \beta \exp(\gamma(t-1))(\exp(\gamma) - 1) &= \theta_{t-1,2} \Rightarrow \\ \beta \exp(\gamma t)(\exp(\gamma) - 1) &= \theta_{t,2}. \end{aligned}$$

Portanto, ao reparametrizar o modelo (4.1) fazendo $\boldsymbol{\theta}_t = [\mu_t \ \beta \exp(\gamma t)(\exp(\gamma) - 1)]'$, temos as equações

$$\begin{aligned} y_t &= \mu_t + v_t, \text{ onde } v_t \sim N(0, \sigma_t^2) \text{ e} \\ \mu_t &= \mu_{t-1} + \beta \exp(\gamma(t-1))(\exp(\gamma) - 1). \end{aligned}$$

Podemos ainda acrescentar ao modelo acima um erro de evolução w_t , fazendo

$$\begin{aligned} y_t &= \mu_t + v_t, \text{ onde } v_t \sim N(0, \sigma_t^2) \text{ e} \\ \mu_t &= \mu_{t-1} + \beta \exp(\gamma(t-1))(\exp(\gamma) - 1) + w_{i,t}, \text{ onde } w_t \sim N(0, W_t). \end{aligned} \tag{4.2}$$

Enquanto os modelos apresentados na seção 2.3.2 evoluem os parâmetros relacionados à tendência de crescimento exponencial, o modelo (4.2) mantém estes parâmetros (β e γ) fixos ao longo do tempo. Como em nosso conjunto de dados há poucas observações no tempo, espera-se que a tendência de crescimento exponencial se mantenha fixa ao longo do período considerado. Por isto, optamos por usar a evolução temporal em (4.2) para modelar as estimativas $y_{i,t}$ das densidades populacionais municipais, ou seja, para estimar a verdadeira densidade populacional $\mu_{i,t}$, nossa proposta é usar as equações

- $y_{i,t} = \mu_{i,t} + v_{i,t}$, com
 - $v_{i,t} \sim N(0, s_{i,t})$ para t ano de censo e
 - $v_{i,t} \sim N(0, \sigma_{i,t}^2)$ para t ano de PNAD, e
- $\mu_{i,t} = \mu_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\exp(\gamma_i) - 1) + w_{i,t}$, com
 - $w_{i,t} \sim N(0, W_{i,t})$ para $t = 1, 2, \dots, 9$.

Através da distribuição $N(0, s_{i,t})$, que estabelece variâncias conhecidas, estamos fixando em anos censitários um erro de cobertura de 5% para todos os municípios.

Serão consideradas duas estratégias para tratar as variâncias $W_{i,t}$. A primeira consiste em fazer $W_{i,t} = \tau^2$ e estimar o parâmetro τ^2 . A segunda consiste em usar um fator de desconto δ para especificar as variâncias $W_{i,t}$, promovendo o decaimento da informação sobre $\mu_{i,t}$ ao longo do tempo. Serão testados os fatores 0,40, 0,50, 0,80, 0,85, 0,90, 0,95 e 0,99.

Como o plano amostral da PNAD não sofre modificações significativas ao longo dos anos, é razoável assumir que a cobertura destas pesquisas não se altera de um ano para outro. Assim, para os anos t associados à PNAD, usaremos uma única variância para todos os tempos, ou seja, faremos $\sigma_{i,t}^2 = \sigma_{PNAD,i}^2$. A baixa precisão das estimativas $s_{i,t}$ foi também um fator importante para descartarmos a opção de modelar a evolução temporal de $\sigma_{i,t}^2$.

Vale notar ainda, que a partir das densidades $\mu_{i,t}$, o total populacional $\pi_{i,t}$ pode ser calculado por $\pi_{i,t} = \mu_{i,t} \times A_i$.

Este modelo é bastante flexível, pois além de ter uma estrutura de evolução temporal que possibilita explorar possíveis movimentos dinâmicos, também nos permite incorporar estrutura espacial em qualquer nível da hierarquia. A partir desta modelagem, podemos desenvolver modelos bem estruturados no espaço e no tempo, contemplando interações entre estas duas dimensões.

Para simplificar a notação, iremos supor ainda que

$$\mathbf{y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,9} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ y_{91,1} & y_{91,1} & \cdots & y_{91,9} \end{bmatrix}$$

é a matriz de observações para os 91 municípios da PNAD utilizados (ver figura 3.1),

$$\mathbf{s} = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,9} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ s_{91,1} & s_{91,1} & \cdots & y_{91,9} \end{bmatrix}$$

é a matriz com as estimativas da variância para estes municípios, $C = \{1, 5, 9\}$ é o conjunto que representa os anos de censo, $P = \{2, 3, 4, 6, 7, 8\}$ é o conjunto que representa os anos de PNAD e

$$\boldsymbol{\sigma}_{PNAD}^2 = \left[\sigma_{PNAD,1}^2 \quad \sigma_{PNAD,2}^2 \quad \cdots \quad \sigma_{PNAD,91}^2 \right]'$$

4.2 Modelagem Conjunta

Todos os estudos por amostragem têm desvios em relação aos valores reais do universo. Ao utilizar dados de pesquisas com diferentes variabilidades e tipos de erros, é preciso considerar de modo diferenciado as precisões das informações levantadas. Nesta seção, propomos uma modelagem que filtra a quantidade de informação de um estimador a partir da estimativa de sua variância.

Afim de incorporarmos as estimativas $s_{i,t}$ ao processo de estimação de $\mu_{i,t}$, nossa proposta é considerar, para t ano de PNAD, a modelagem conjunta

$$p(y_{i,t}, s_{i,t} \mid \mu_{i,t}, \sigma_{PNAD,i}^2) = p(y_{i,t} \mid \mu_{i,t}, \sigma_{PNAD,i}^2) p(s_{i,t} \mid \sigma_{PNAD,i}^2).$$

Com isto, espera-se que $p(s_{i,t} \mid \sigma_{PNAD,i}^2)$ funcione como um filtro de informação, permitindo que as estimativas $s_{i,t}$ informem o quão precisas são as estimativas $y_{i,t}$.

Assumiremos que $(s_{i,t} \mid \sigma_{PNAD,i}^2)$ segue uma distribuição Gamma de modo que

$$\begin{aligned} E(s_{i,t} \mid \sigma_{PNAD,i}^2) &= \sigma_{PNAD,i}^2 \quad \text{e} \\ \text{Var}(s_{i,t} \mid \sigma_{PNAD,i}^2) &= f^*(\sigma_{PNAD,i}^2), \end{aligned}$$

onde $f^*(\sigma_{PNAD,i}^2)$ é uma função que relaciona $\sigma_{PNAD,i}^2$ à $\text{Var}(s_{i,t})$. Sugerimos que a função f^* deve ser investigada através de uma análise exploratória dos dados. Esta análise será feita na próxima subseção.

4.2.1 Análise Exploratória das Estimativas da Variância

Nesta seção, apresentaremos uma análise exploratória das estimativas $s_{i,t}$ obtidas em anos de PNAD. A intenção é investigar padrões que permitam definir a função f^* .

Para investigar a relação entre $\text{Var}(s_{i,t})$ e $\sigma_{PNAD,i}^2$, comparamos as médias e variâncias amostrais

$$s'_i = \frac{\sum_{t \in P} s_{i,t}}{6} \quad \text{e} \quad s_i^* = \frac{\sum_{t \in P} (s_{i,t} - s'_i)^2}{6 - 1},$$

calculadas usando os 6 anos, $P = \{2, 3, 4, 6, 7, 8\}$, de PNAD.

A figura 4.1 mostra o gráfico de dispersão de $\log(s_i^*)$ versus $\log(s'_i)$, assim como a reta regressora $\log(s_i^*) = 2,046\log(s'_i) - 1,228$, estimada pelo método de mínimos quadráticos ordinários. Desta regressão, obtemos a relação

$$s_i^* \cong \frac{s_i'^{2,046}}{\exp(1,228)} \cong 0,3s_i'^2$$

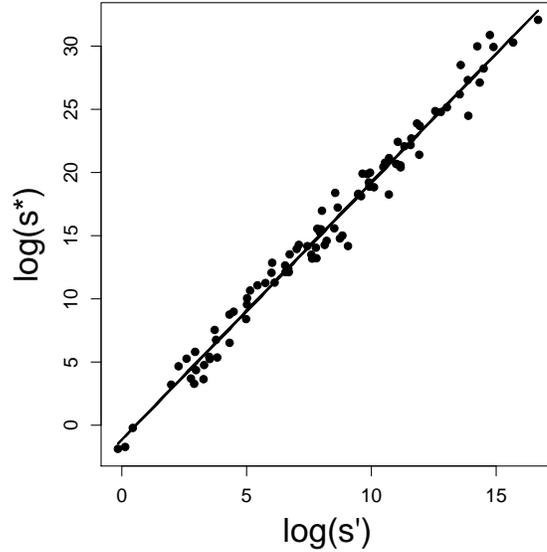


Figura 4.1: Gráfico de dispersão de $\log(s_i^*)$ versus $\log(s'_i)$

Como a normalidade do estimador $y_{i,t}$ é geralmente assumida devido ao Teorema Central do Limite (Rao, 2003) e, como a variância da estimativa da variância de uma distribuição normal depende de sua variância ao quadrado (Gamerman & Migon, 1999), a relação proporcional entre s_i^* e $s_i'^2$ já era esperada. Assim, a partir da relação acima, definimos

$$\text{Var}(s_{i,t} \mid \sigma_{PNAD,i}^2, \eta) = \eta(\sigma_{PNAD,i}^2)^2,$$

onde η é um parâmetro a ser estimado.

4.3 Previsões

Para os municípios não selecionados pela PNAD, temos a disposição apenas as informações censitárias. Assim, para t' ano intercensitário, a previsão da densidade popula-

cional $\mu_{p,t'}$ de um município p não selecionado pela PNAD será feita usando as densidades populacionais $y_{p,t}$ obtidas em anos censitários e as equações de evolução

$$\mu_{p,t} = \mu_{p,t-1} + \beta_p \exp(\gamma_p(t-1))(\exp(\gamma_p) - 1) + w_{p,t}, \text{ com } w_{p,t} \sim N(0, W_{p,t}), \text{ para } t = 1, \dots, 9.$$

A observação $y_{p,t}$ obtida em anos censitários será incorporada ao modelo assumindo

$$(y_{p,t} \mid \mu_{p,t}, s_{p,t}) \sim N(\mu_{p,t}, s_{p,t}). \quad (4.3)$$

Ao utilizar modelos que permitem a troca de informação entre municípios, a informação censitária de municípios não selecionados pela PNAD, incorporada por (4.3), também é compartilhada.

Sendo assim, a abordagem aqui proposta permite, de forma natural, a estimação da densidade populacional para os municípios não selecionados pela PNAD. Para exemplificar, foram escolhidos aleatoriamente seis municípios não selecionados pela PNAD para previsão: Guará, Quintana, Valinhos, Vargem Grande Paulista, Várzea Paulista e Votorantim. Estes municípios serão identificados aqui pela numeração: 92, 93, 94, 95, 96 e 97. A figura 4.2 mostra a localização destes municípios.

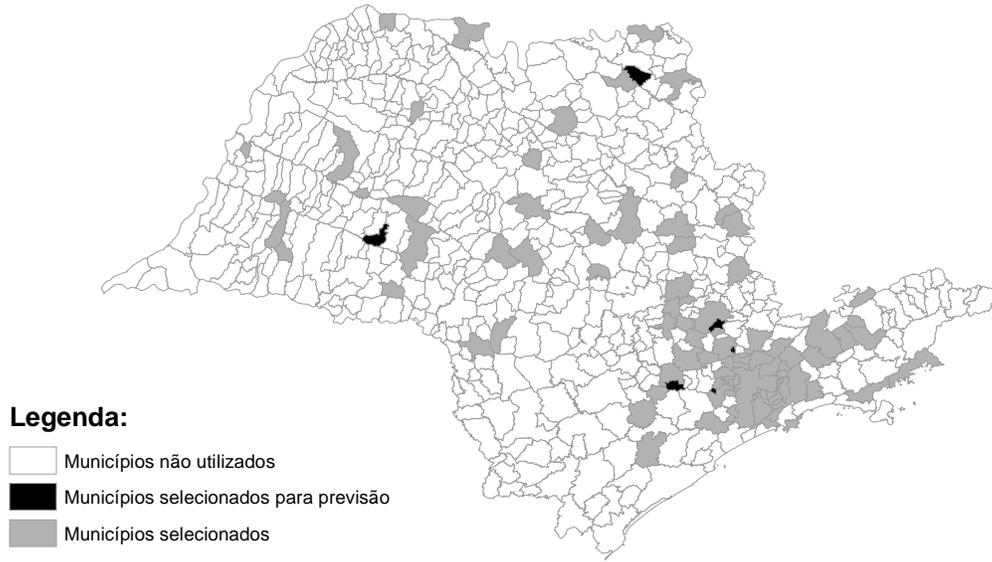


Figura 4.2: Localização dos municípios selecionados para análise e para previsão.

Dada a definição dos municípios para previsão, sejam

- \mathbf{y}_t o vetor de observações disponíveis para o tempo t , dado por

$$\mathbf{y}_t = \begin{cases} \begin{bmatrix} y_{1,t} & y_{2,t} & \cdots & y_{91,t} \end{bmatrix}, & \text{para } t \text{ ano de PNAD} \\ \begin{bmatrix} y_{1,t} & y_{2,t} & \cdots & y_{97,t} \end{bmatrix}, & \text{para } t \text{ ano de censo} \end{cases},$$

- $\boldsymbol{\mu} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,9} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{97,1} & \mu_{97,1} & \cdots & \mu_{97,9} \end{bmatrix},$

- $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \cdots \ \beta_{97}]'$ e

- $\boldsymbol{\gamma} = [\gamma_1 \ \gamma_2 \ \cdots \ \gamma_{97}]'$.

4.4 Modelagens de β e γ

4.4.1 Modelagem Hierárquica

Dados com estrutura hierárquica aparecem em diversas situações, tais como em ciências sociais, onde as medições de uma dada variável de interesse são realizadas em diferentes níveis de agregação, definidos por exemplo, pelo local de residência, pelo grupo social ou pela raça do indivíduo. Modelos hierárquicos proporcionam uma forma de compartilhar a informação entre diferentes grupos, explorando suas similaridades e aumentando a precisão das estimativas. Exemplos de dados com estrutura hierárquica podem ser vistos em Goldstein (1995).

Para os dados aqui utilizados, iremos considerar o município como nível de agregação. A estrutura hierárquica abordada nesta seção assume que os fatores sócio-econômicos que influenciam o crescimento populacional afetam a sociedade como um todo. A partir desta premissa, iremos assumir que a evolução temporal das densidades populacionais é similar entre os municípios, fazendo com que os parâmetros β_i e γ_i compartilhem uma mesma média e variância (Souza, 2004), ou seja, $\forall i \in \Omega$,

$$\begin{aligned}(\beta_i | \mu_\beta, \sigma_\beta^2) &\sim N(\mu_\beta, \sigma_\beta^2) \quad \text{e} \\(\gamma_i | \mu_\gamma, \sigma_\gamma^2) &\sim N(\mu_\gamma, \sigma_\gamma^2).\end{aligned}\tag{4.4}$$

Chamaremos o modelo com esta estrutura de Modelo Hierárquico.

4.4.2 Modelagem Hierárquica Espacialmente Estruturada

A figura 4.3 mostra a diferença entre as densidades populacionais obtidas no censo de 2000 e de 1991. Pela figura, pode-se observar que municípios próximos geograficamente tendem a ter uma evolução populacional similar. Isto sugere que um modelo de componentes espaciais pode ser usado para explicar esta evolução.

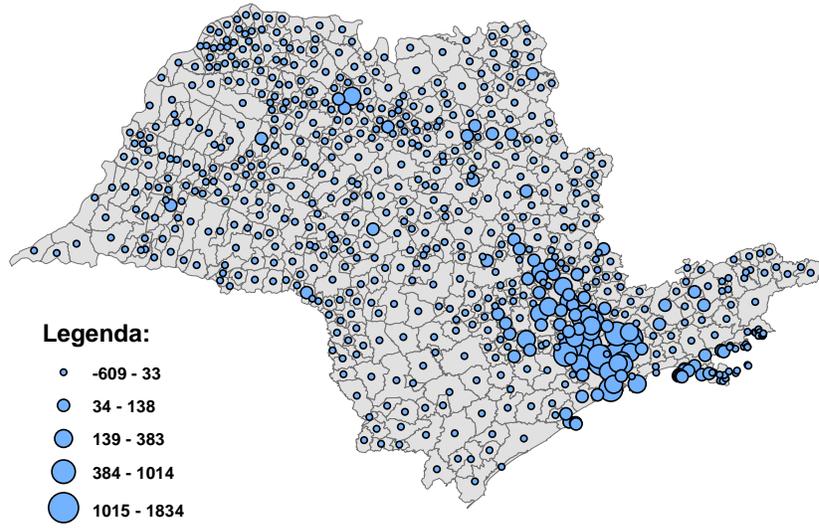


Figura 4.3: Diferença entre a densidade populacional em 2000 e em 1991 (em habitantes por quilometro quadrado).

Para tratar a dependência espacial da evolução populacional, optamos pelo uso de Processos Aleatórios Gaussianos. Como já dissemos, a vantagem destes processos está na possibilidade de fazer também previsões para localizações não vizinhas.

Propomos então a seguinte modelagem hierárquica espacialmente estruturada:

$$\begin{aligned}
 (\gamma_i | \mu_\gamma, \sigma_\gamma^2) &\sim N(\mu_\gamma, \sigma_\gamma^2) \quad \text{e} \\
 \beta_i &= \mu_\beta + \xi(i),
 \end{aligned}
 \tag{4.5}$$

onde $\xi(i)$ segue um Processo Gaussiano que assume valores no conjunto Ω de todos os municípios do estado de São Paulo, tem média zero, função de correlação

$$\rho(d_{r,r'}, \phi) = \exp(-\phi d_{r,r'})$$

e função de covariância

$$c(r, r') = \varepsilon^2 \rho(d_{r,r'}, \phi),$$

onde $d_{r,r'}$ é a distância entre os centróides dos municípios r e r' . Chamaremos o modelo com esta estrutura de Modelo Espacial.

Deste modo, estamos assumindo que a intensidade de crescimento/decrescimento (β_i) da densidade populacional de um determinado município sofre influência maior de municípios mais próximos e que a velocidade deste crescimento/decrescimento (γ_i) é similar entre os municípios.

Como em nosso conjunto de dados as observações para o primeiro ano (1991) são extremamente precisas, julgamos previamente que o uso de modelagens hierárquicas ou espaciais para o vetor de densidades populacionais iniciais μ_0 não traria contribuições significativas. Porém, para dados em que o primeiro ano não é censitário, a troca de informação entre as componentes de μ_0 constitui uma fonte de informação valiosa.

Para simplificar a notação, iremos supor ainda que

$$\mathbf{d} = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,9} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ d_{97,1} & d_{97,2} & \cdots & d_{97,9} \end{bmatrix}$$

é a matriz de distâncias entre os centróides dos municípios, $\Sigma_{\varepsilon, \phi}$ é uma matriz de covariâncias 97×97 tal que

$$\Sigma_{\varepsilon, \phi}[r, r'] = \exp(-\phi d_{r, r'}), \text{ para } r, r' \in \{1, 2, \dots, 97\},$$

e Σ_{ϕ} é uma matriz 97×97 com as correlações $\rho(d_{r, r'}, \phi)$, ou seja,

$$\Sigma_{\phi}[r, r'] = \exp(-\phi d_{r, r'}), \text{ para } r, r' \in \{1, 2, \dots, 97\}.$$

4.5 Priors

Discutiremos nesta seção a especificação de priors para os modelos descritos nas seções anteriores.

Como os parâmetros β_i e γ_i podem assumir tanto valores positivos quanto negativos, assumiremos Distribuição Normal para as médias μ_{β} e μ_{γ} , ou seja,

$$\mu_{\beta} \sim N(m_{\beta}, v_{\beta}) \text{ e}$$

$$\mu_{\gamma} \sim N(m_{\gamma}, v_{\gamma})$$

Como $Var(s_{i,t} | \sigma_{PNAD,i}^2, \eta) = \eta(\sigma_{PNAD,i}^2)^2$, então η deve ser necessariamente positivo, pois caso contrário teríamos variância negativa. Assim, assumimos para este parâmetro uma distribuição Gamma Invertida com parâmetros a_η e b_η , $\eta \sim GI(a_\eta, b_\eta)$.

Para os parâmetros de variância σ_γ^2 , σ_β^2 e ε^2 e para o parâmetro ϕ , que é necessariamente positivo, também assumimos distribuição Gamma Invertida, ou seja,

$$\begin{aligned}\sigma_\gamma^2 &\sim GI(a_\gamma, b_\gamma), \\ \sigma_\beta^2 &\sim GI(a_\beta, b_\beta), \\ \varepsilon^2 &\sim GI(a_\varepsilon, b_\varepsilon) \text{ e} \\ \phi &\sim GI(a_\phi, b_\phi).\end{aligned}$$

Já para os parâmetros $\sigma_{PNAD,i}^2$, assumimos uma distribuição Gamma com parâmetros a_σ e b_σ , ou seja,

$$\sigma_{PNAD,i}^2 \sim Ga(a_\sigma, b_\sigma)$$

Fixamos $m_\beta = 0$, $v_\beta = 100000$, $m_\gamma = 0$ e $v_\gamma = 1000$. Dada a natureza populacional dos dados ajustados, estamos sendo bem abrangentes, pois os valores de μ_β e μ_γ não contemplados (com baixa probabilidade a priori) representariam acréscimos (ou decréscimos) populacionais absurdos, impossíveis na prática.

Para as variâncias σ_β^2 , σ_γ^2 e ε^2 , optamos por prioris não informativas fixando $a_\beta = 0.1$, $b_\beta = 0.1$, $a_\gamma = 0.1$ e $b_\gamma = 0.1$.

Para $i = 1, 2, \dots, 91$, as prioris de $\sigma_{PNAD,i}^2$ foram localizadas em s'_i , porém admitindo uma incerteza bem grande, contemplando diversos cenários. Ou seja, a_σ e b_σ foram escolhidos de modo que $E(\sigma_{PNAD,i}^2) = s'_i$ e $Var(\sigma_{PNAD,i}^2) = 100000$.

A priori para η também foi localizada a partir da análise exploratória da seção 4.2.1. Os parâmetros a_η e b_η foram escolhidos de modo que $E(\eta) = 0.3$ e $Var(\eta) = 1000000$.

Berger et al. (2001) citam que deve-se tomar cuidado ao usar prioris próprias muito vagas para o parâmetro ϕ . Uma sugestão encontrada na literatura (Schmidt et al., 2002) é atribuir uma média tal que a correlação é 0.05 quando a distância é igual à metade da distância máxima entre as localizações. Em outras palavras, essa priori reflete o fato de esperarmos que para distâncias maiores que $dmax/2$, a correlação espacial é próxima de

0. Assim, a_ϕ e b_ϕ foram escolhidos de modo que $E(\phi) = 6/dmax$ e $Var(\phi) = 1000$, onde $dmax$ é a distância máxima entre os 97 municípios selecionados.

A ausência de informação a priori sobre $\mu_{i,0}$ foi traduzida pela priori $\mu_{i,0} \sim N(0, 10^8)$.

4.6 Distribuições a posteriori

Sob a perspectiva bayesiana, sabemos que o procedimento de inferência é baseado na distribuição a posteriori dos parâmetros. Nesta seção, combinaremos as funções de verossimilhança dos modelos propostos com as distribuições a priori descritas na seção anterior para gerar a distribuição a posteriori conjunta dos parâmetros.

4.6.1 Distribuição a posteriori para o Modelo Hierárquico

A Distribuição a Posteriori para o Modelo Hierárquico é dada por

$$\begin{aligned}
p(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu_\beta, \sigma_\beta^2, \mu_\gamma, \sigma_\gamma^2, \eta, \boldsymbol{\sigma}_{PNAD}^2 \mid \mathbf{y}, \mathbf{s}) &= \left(\prod_{t \in C} \prod_{i=1}^{97} p(y_{i,t} \mid \mu_{i,t}, s_{i,t}) \right) \\
&\times \left(\prod_{t \in P} \prod_{i=1}^{91} p(y_{i,t} \mid \mu_{i,t}, \sigma_{PNAD,i}^2) \right) \\
&\times \left(\prod_{t \in P} \prod_{i=1}^{91} p(s_{i,t} \mid \sigma_{PNAD,i}^2, \eta) \right) \\
&\times \left(\prod_{t=1}^9 \prod_{i=1}^{97} p(\mu_{i,t} \mid \mu_{i,t-1}, \beta_i, \gamma_i) \right) \\
&\times \left(\prod_{i=1}^{97} p(\beta_i \mid \mu_\beta, \sigma_\beta^2) p(\gamma_i \mid \mu_\gamma, \sigma_\gamma^2) p(\mu_{i,0}) \right) \\
&\times \left(\prod_{i=1}^{91} p(\sigma_{PNAD,i}^2) \right) p(\mu_\beta) \\
&\times p(\sigma_\beta^2) p(\mu_\gamma) p(\sigma_\gamma^2) p(\eta),
\end{aligned}$$

onde

- $(y_{i,t} \mid \mu_{i,t}, s_{i,t}) \sim N(\mu_{i,t}, s_{i,t})$,
- $(y_{i,t} \mid \mu_{i,t}, \sigma_{PNAD,i}^2) \sim N(\mu_{i,t}, \sigma_{PNAD,i}^2)$,
- $(s_{i,t} \mid \sigma_{PNAD,i}^2, \eta) \sim Ga\left(\frac{1}{\eta}, \frac{1}{\eta \sigma_{PNAD,i}^2}\right)$,
- $(\mu_{i,t} \mid \mu_{i,t-1}, \beta_i, \gamma_i) \sim N(\mu_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\gamma_i - 1), W_{i,t})$,

- $(\gamma_i | \mu_\gamma, \sigma_\gamma^2) \sim N(\mu_\gamma, \sigma_\gamma^2)$ e
- $(\beta_i | \mu_\beta, \sigma_\beta^2) \sim N(\mu_\beta, \sigma_\beta^2)$.

4.6.2 Distribuição a posteriori para o Modelo Espacial

A Distribuição a Posteriori para o Modelo Espacial é dada por

$$\begin{aligned}
p(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mu_\beta, \varepsilon^2, \phi, \mu_\gamma, \sigma_\gamma^2, \eta, \boldsymbol{\sigma}_{PNAD}^2 | \mathbf{y}, \mathbf{s}, \mathbf{d}) &= \left(\prod_{t \in C} \prod_{i=1}^{97} p(y_{i,t} | \mu_{i,t}, s_{i,t}) \right) \\
&\times \left(\prod_{t \in P} \prod_{i=1}^{91} p(y_{i,t} | \mu_{i,t}, \sigma_{PNAD,i}^2) \right) \\
&\times \left(\prod_{t \in P} \prod_{i=1}^{91} p(s_{i,t} | \sigma_{PNAD,i}^2, \eta) \right) \\
&\times \left(\prod_{t=1}^9 \prod_{i=1}^{97} p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \right) \\
&\times p(\boldsymbol{\beta} | \mu_\beta, \varepsilon^2, \phi, d) \left(\prod_{i=1}^{97} p(\gamma_i | \mu_\gamma, \sigma_\gamma^2) p(\mu_{i,0}) \right) \\
&\times \left(\prod_{i=1}^{91} p(\sigma_{PNAD,i}^2) \right) p(\mu_\beta) p(\varepsilon^2) p(\phi) \\
&\times p(\mu_\gamma) p(\sigma_\gamma^2) p(\eta),
\end{aligned}$$

onde

- $(y_{i,t} | \mu_{i,t}, s_{i,t}) \sim N(\mu_{i,t}, s_{i,t})$,
- $(y_{i,t} | \mu_{i,t}, \sigma_{PNAD,i}^2) \sim N(\mu_{i,t}, \sigma_{PNAD,i}^2)$,
- $(s_{i,t} | \sigma_{PNAD,i}^2, \eta) \sim Ga\left(\frac{1}{\eta}, \frac{1}{\eta \sigma_{PNAD,i}^2}\right)$,
- $(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \sim N(\mu_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\gamma_i - 1), W_{i,t})$,
- $(\gamma_i | \mu_\gamma, \sigma_\gamma^2) \sim N(\mu_\gamma, \sigma_\gamma^2)$ e
- $(\boldsymbol{\beta} | \mu_\beta, \phi, \varepsilon^2) \sim N(\mu_\beta, \Sigma_{\varepsilon, \phi})$.

4.7 Comparação de Modelos

Na análise de dados que será apresentada no próximo capítulo, utilizaremos dois critérios para comparar os modelos propostos. Nesta seção, apresentaremos a descrição destes critérios.

Uma das dificuldades ao propor modelos diferentes para analisar um certo fenômeno é como escolher o melhor modelo. Há na literatura muitos critérios diferentes para apontar o “melhor” modelo ajustado. No entanto, não há um consenso quanto a qual critério funciona melhor. Por esta razão, é comum usar diferentes critérios de comparação de modelos (Schmidt et al., 2006). Neste contexto, insere-se a aplicação de critérios para seleção de modelos baseados em medidas que tentam quantificar sua qualidade de ajuste e aplicam alguma punição a modelos complexos, com muitos parâmetros.

Apresentaremos nesta seção os dois critérios utilizados para comparar os modelos propostos: o Critério de Informação Baseado na *Deviance* e o Desvio Preditivo Esperado. Mais detalhes sobre estes critérios podem ser encontrados em Banerjee et al. (2004).

4.7.1 Critério de Informação Baseado na *Deviance*

Spiegelhalter et al. (2002) propõem um método de comparação de modelos com o objetivo de superar as dificuldades encontradas na utilização de certos critérios, como o AIC (Akaike Information Criterion), em modelos complexos e com estrutura hierárquica. Este critério, chamado de *Deviance Information Criterion* (DIC), baseia-se na distribuição a posteriori da estatística *deviance*, isto é:

$$D(\theta) = -2\log(p(Y | \theta)) + 2\log f(Y),$$

onde $f(Y | \theta)$ é a verossimilhança condicionada ao conjunto de parâmetros θ e $f(Y)$ é a distribuição marginal da variável Y . Este método combina uma parte que mede o ajuste do modelo com outra que avalia sua complexidade.

O DIC pode ser calculado por

$$DIC = \bar{D} + pD,$$

onde \bar{D} é a esperança a posteriori da estatística *deviance*, $E_{\theta|Y}[D(\theta)]$, que é justamente a parte que avalia o ajuste do modelo. A componente pD , pode ser interpretada como sendo o número efetivo de parâmetros no modelo, que mede a complexidade do mesmo. Podemos calculá-la a partir de

$$pD = \bar{D} - \hat{D},$$

onde $\widehat{D} = -2\log(p(Y | \widehat{\theta}))$, ou seja, uma estimativa pontual da deviance utilizando a média a posteriori $\widehat{\theta}$ de θ . Portanto, o DIC também pode ser reescrito como

$$DIC = 2\overline{D} - \widehat{D}.$$

4.7.2 Desvio Preditivo Esperado

Uma alternativa para esta escolha de modelos é a utilização de métodos que fazem esta comparação via distribuições preditivas. Ou seja, considera-se como o melhor modelo aquele que faz as melhores previsões para os valores observados. Gelfand & Ghosh (1998) propõem um método em que a posteriori de uma função perda é minimizada. Na prática, calcula-se uma medida que é baseada em replicações $y_{rep,l}$ dos dados observados $y_{obs,l}$, $l = 1, 2, \dots, L$, e considera-se duas quantidades, uma de bondade de ajuste e uma de capacidade preditiva dos modelos. Este critério é chamado *Expected Prediction Deviation* (EPD) e é obtido através da estatística D_c , obtida por

$$D_c = P + \frac{c}{c+1}G, \text{ com}$$

$$G = \sum_{l=1}^n (\mu_l - y_{obs})^2 \text{ e}$$

$$P = \sum_{l=1}^n \sigma_l^2,$$

onde $\mu_l = E(y_{rep,l})$ e $\sigma_l^2 = Var(y_{rep,l})$ são a média e a variância da distribuição de $y_{rep,l}$ condicionada aos dados observados. Banerjee et al. (2004) mencionam que a classificação dos modelos não é sensível à escolha do valor de c , assim, aqui fixamos $c = 1$.

Capítulo 5

Análise de Dados

Neste capítulo, faremos uma análise dos resultados obtidos e das técnicas de inferência utilizadas. Como o procedimento de inferência é realizado sob uma perspectiva bayesiana, apresentaremos também neste capítulo todos os aspectos computacionais ligados a esse procedimento. Apresentaremos ainda, na seção 5.6, uma comparação dos resultados obtidos em nossa análise com resultados oficiais do IBGE e com os resultados de Souza (2004).

5.1 Aspectos Computacionais

Seguindo o paradigma de Bayes, sabemos que a distribuição a posteriori é proporcional ao produto da função de verossimilhança pela priori. Porém, como podemos ver na seção 4.6, tanto para o Modelo Hierárquico quanto para o Modelo Espacial, quaisquer que sejam as distribuições a priori associadas aos parâmetros, não é possível fazer nenhuma sumarização da posteriori de forma analítica. Portanto, é preciso fazer uso de métodos de simulação estocástica para obter amostras da distribuição a posteriori de interesse. Nos últimos quinze anos, a inferência Bayesiana vem experimentando um grande avanço devido a introdução de métodos de Monte Carlo via Cadeias de Markov (MCMC) e também devido a disponibilidade de computadores velozes (Schmidt et al., 2002). O MCMC é uma técnica poderosa que permite a análise de modelos altamente estruturados.

A idéia básica do MCMC é gerar amostras da distribuição de interesse a partir de dis-

tribuições que constituam uma cadeia de Markov, chamadas de distribuições de transição. Tais distribuições devem ser adequadamente escolhidas de forma que a cadeia convirja para uma distribuição estacionária que corresponda a própria distribuição de interesse, neste caso, a distribuição a posteriori conjunta. Dessa forma, após atingir a convergência, as amostras estarão sendo geradas dessa distribuição estacionária. O objetivo, então, é gerar uma amostra grande o suficiente desta distribuição estacionária que aproxime bem a distribuição a posteriori conjunta exata. Em geral, para considerar apenas amostras obtidas após a cadeia ter atingido sua distribuição estacionária, descarta-se as primeiras iterações. A quantidade de iterações descartadas é chamada de *burn-in* e é peculiar a cada estudo. Também, para evitar a autocorrelação destas amostras, guarda-se apenas os valores entre um intervalo específico de iterações, chamado *thinning*. Para maiores detalhes, veja Gamerman & Lopes (2006).

Neste trabalho, o Amostrador de Gibbs com passos de Metropolis-Hastings foi o método de simulação estocástica utilizado para gerar amostras das distribuições a posteriori. No entanto, deve-se tomar cuidado ao amostrar parâmetros que variam no tempo, devido à alta autocorrelação que eles apresentam. Assim, um algoritmo eficiente foi usado para gerar amostras de $\mu_{i,t}$ a cada passo do Amostrador de Gibbs. Este algoritmo será descrito na seção 5.1.5.

A linguagem Ox versão 3.40 (Doornik, 2002) foi utilizada para implementar o processo de estimação.

5.1.1 Amostrador de Gibbs

O Amostrador de Gibbs é um algoritmo iterativo com origem no contexto de processamento de imagens e introduzido no contexto estatístico por Geman & Geman (1984) com grande contribuição de Gelfand & Smith (1990). Este algoritmo permite gerar amostras de uma distribuição conjunta $p(\theta_1, \theta_2, \dots, \theta_q)$ a partir das distribuições de cada parâmetro (ou de bloco de parâmetros) condicionadas aos demais parâmetros do modelo, chamadas de distribuições condicionais completas, $p(\theta_i | \cdot)$. O método consiste em tomar as distribuições condicionais completas como as distribuições de transição de uma cadeia de Markov.

Supondo que $\theta_i^{(k)}$ representa a amostra de θ_i na iteração k (para $i \in \{1, 2, \dots, q\}$), os passos a seguir descrevem resumidamente o algoritmo:

1. Inicializar o contador de iterações da cadeia $k = 0$.
2. Especificar valores iniciais para $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_{q-1}^{(0)}$ e $\theta_q^{(0)}$.
3. Obter um novo valor de $\theta_i^{(k)}$ a partir de $\theta_i^{(q-1)}$ através da geração sucessiva dos valores

$$\begin{aligned}\theta_1^{(k)} &\sim p\left(\theta_1|\theta_2^{(k-1)}, \theta_3^{(k-1)}, \dots, \theta_q^{(k-1)}\right), \\ \theta_2^{(k)} &\sim p\left(\theta_2|\theta_1^{(k-1)}, \theta_3^{(k-1)}, \dots, \theta_q^{(k-1)}\right), \\ &\vdots \\ \theta_q^{(k)} &\sim p\left(\theta_q|\theta_1^{(k-1)}, \theta_2^{(k-1)}, \dots, \theta_{q-1}^{(k-1)}\right).\end{aligned}$$

4. Incrementar o contador de k para $k+1$ e retornar ao passo 2 até obter convergência.

Como as distribuições a posteriori apresentadas na seção 4.6 não têm forma analítica fechada, usaremos o Amostrador de Gibbs com passos de Metropolis-Hastings para gerar amostras da posteriori. Para isto, precisamos calcular as condicionais completas dos modelos propostos. Na próxima subseção, apresentaremos esses cálculos.

A partir das condicionais completas do Modelo Espacial (veja a subseção a seguir), nota-se que o procedimento de inferência para este modelo envolve sucessivas inversões da matriz $\Sigma_{\varepsilon, \phi}$. Assim, o aumento do número de municípios considerados implica no aumento da dimensão de $\Sigma_{\varepsilon, \phi}$ e, conseqüentemente, o tempo computacional exido para gerar as cadeias deste modelo aumenta consideravelmente. Já para o Modelo Hierárquico, o tempo computacional não sofre tanto a influência da quantidade de municípios, pois este envolve apenas inversões de matrizes diagonais.

5.1.2 Condicionais Completas

Nesta subseção, serão apresentadas as distribuições condicionais completas dos modelos propostos. O cálculo detalhado destas distribuições pode ser visto no apêndice C.

Condicionalis Completas do Modelo Hierárquico:

Supondo que I é a matriz identidade 97×97 e $\vec{1}$ é o vetor unitário 97×1 , temos que

- $(\mu_\gamma | \gamma, \sigma_\gamma^2) \sim N(M, V)$, com

$$V = \left(\vec{1}' \sigma_\gamma^{-2} I \vec{1} + v_\gamma^{-2} \right)^{-1} \quad \text{e} \quad M = V \left(\vec{1}' \sigma_\gamma^{-2} I \gamma + v_\gamma^{-2} m_\gamma \right);$$

- $(\sigma_\gamma^2 | \gamma, \mu_\gamma) \sim GI(A, B)$, com

$$\begin{aligned} A &= \frac{97}{2} + a_\gamma, \\ B &= b_\gamma + \left(\frac{1}{2} K \right) \quad \text{e} \\ K &= \left(\gamma - \vec{1} \mu_\gamma \right)' I \left(\gamma - \vec{1} \mu_\gamma \right); \end{aligned}$$

- $(\mu_\beta | \beta, \sigma_\beta^2) \sim N(M, V)$, com

$$V = \left(\vec{1}' \sigma_\beta^{-2} I \vec{1} + v_\beta^{-2} \right)^{-1} \quad \text{e} \quad M = V \left(\vec{1}' \sigma_\beta^{-2} I \beta + v_\beta^{-2} m_\beta \right);$$

- $(\sigma_\beta^2 | \beta, \mu_\beta) \sim GI(A, B)$, com

$$\begin{aligned} A &= \frac{97}{2} + a_\beta, \\ B &= b_\beta + \left(\frac{1}{2} K \right) \quad \text{e} \\ K &= \left(\beta - \vec{1} \mu_\beta \right)' I \left(\beta - \vec{1} \mu_\beta \right); \end{aligned}$$

- $(\beta_i | \gamma_i, \mu_\beta, \sigma_\beta^2) \sim N(M, V)$, com

$$\begin{aligned} V &= \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right) + (\sigma_\beta^2)^{-1} \right)^{-1}, \\ M &= V \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right) + (\sigma_\beta^2)^{-1} \mu_\beta \right), \\ A_t &= \mu_{i,t} - \mu_{i,t-1} \quad \text{e} \\ B_t &= \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1). \end{aligned}$$

As condicionais completas de η , $\sigma_{PNAD,i}^2$ e γ_i não tem forma fechada.

Condicionais Completas do Modelo Espacial:

Supondo que $\vec{1}_{96}$ é o vetor unitário 96×1 e $\boldsymbol{\beta}_{-i} = [\beta_1 \ \cdots \ \beta_{i-1} \ \beta_{i+1} \ \cdots \ \beta_{97}]'$, temos que

- $(\beta_i \mid \boldsymbol{\beta}_{-i}, \gamma_i, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \sim N(M, V)$, onde

- $V = \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right) + D^{-1} \right)^{-1}$,
- $M = V \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right) + D^{-1} C \right)$,
- $A_t = \mu_{i,t} - \mu_{i,t-1}$,
- $B_t = \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1)$,
- $C = \mu_\beta + \Sigma_{1,i} \Sigma_{2,i}^{-1} (\boldsymbol{\beta}_{-i} - \mu_\beta \vec{1}_{96})$,
- $D = \varepsilon^2 - \Sigma_{1,i} \Sigma_{2,i}^{-1} \Sigma'_{1,i}$,
- $\Sigma_{1,i}$ é o vetor de covariâncias entre β_i e $\boldsymbol{\beta}_{-i}$ e
- $\Sigma_{2,i} = \text{Var}(\boldsymbol{\beta}_{-i} \mid \mu_\beta, \varepsilon^2, \phi, \mathbf{d})$;

- $(\mu_\beta \mid \boldsymbol{\beta}, \phi, \varepsilon^2, \mathbf{d}) \sim N(M, V)$, com

$$V = \left(\vec{1}' \Sigma_{\varepsilon, \phi}^{-1} \vec{1} + v_\beta^{-2} \right)^{-1} \quad \text{e} \quad M = V \left(\vec{1}' \Sigma_{\varepsilon, \phi}^{-1} \boldsymbol{\beta} + v_\beta^{-2} m_\beta \right);$$

- $(\varepsilon^2 \mid \boldsymbol{\beta}, \mu_\beta, \phi, \mathbf{d}) \sim GI(A, B)$, com

$$\begin{aligned} A &= \frac{97}{2} + a_\varepsilon, \\ B &= b_\varepsilon + \frac{1}{2} C \quad \text{e} \\ C &= \left(\boldsymbol{\beta} - \vec{1} \mu_\beta \right)' \Sigma_\phi^{-1} \left(\boldsymbol{\beta} - \vec{1} \mu_\beta \right). \end{aligned}$$

As condicionais completas de γ_i , μ_γ , σ_γ^2 , $\sigma_{PNAD,i}$ e η para o Modelo Espacial são idênticas às do Modelo Hierárquico e a condicional completa de ϕ não tem forma fechada.

5.1.3 Metropolis-Hastings

Inicialmente proposto por Metropolis et al. (1953) e estendido por Hastings (1970), o algoritmo de Metropolis-Hastings (MH) é um esquema MCMC que utiliza uma distribuição proposta para obter um estado proposto para a cadeia. Este estado então é aceito com uma probabilidade α .

Supondo que $p(\theta)$ é a distribuição para a qual se deseja gerar amostras sucessivas e admitindo que $\theta^{(k)}$ representa a amostra na iteração k , os passos a seguir descrevem resumidamente o algoritmo:

1. Inicializar o contador de iterações $k = 0$ e especificar um valor inicial $\theta^{(0)}$.
2. Gerar um valor proposto θ' da distribuição proposta $q(\theta'|\theta^{(k)})$.
3. Calcular a probabilidade de aceitação $\alpha(\theta^{(k)}, \theta') = \min \left\{ 1, \frac{p(\theta')q(\theta^{(k)}|\theta')}{p(\theta^{(k)})q(\theta'|\theta^{(k)})} \right\}$.
4. Gerar $u \sim U(0, 1)$.
5. Se $u \leq \alpha(\theta^{(k)}, \theta')$ então aceitar o novo valor e fazer $\theta^{(k+1)} = \theta'$, caso contrário rejeitar e fazer $\theta^{(k+1)} = \theta^{(k)}$.
6. Incrementar o contador de k para $k + 1$ e voltar ao passo 2.

Como dissemos na subseção 5.1.2, as distribuições condicionais completas a posteriori dos parâmetros ϕ , η , $\sigma_{PNAD,i}^2$ e γ_i não seguem uma distribuição de probabilidade conhecida. Assim, utilizamos o algoritmo Metropolis-Hastings para obter amostras destas distribuições.

Porém, métodos MCMC tornam-se mais complicados no caso de alta correlação entre parâmetros do modelo. Quando isto acontece, as cadeias tendem a ser muito autocorrelacionadas e a convergir lentamente para a distribuição de equilíbrio. Neste caso, uma alternativa é amostrar os parâmetros correlacionados em bloco. Os gráficos da figura 5.1 foram gerados a partir de dados simulados para o Modelo Espacial (a simulação de dados artificiais será descrita na seção 5.3) e apresentam a função de verossimilhança conjunta, a menos de uma constante de proporcionalidade, de dois pares de parâmetros extremamente correlacionados: $(\beta_{14}, \gamma_{14})$ e $(\beta_{17}, \gamma_{17})$.

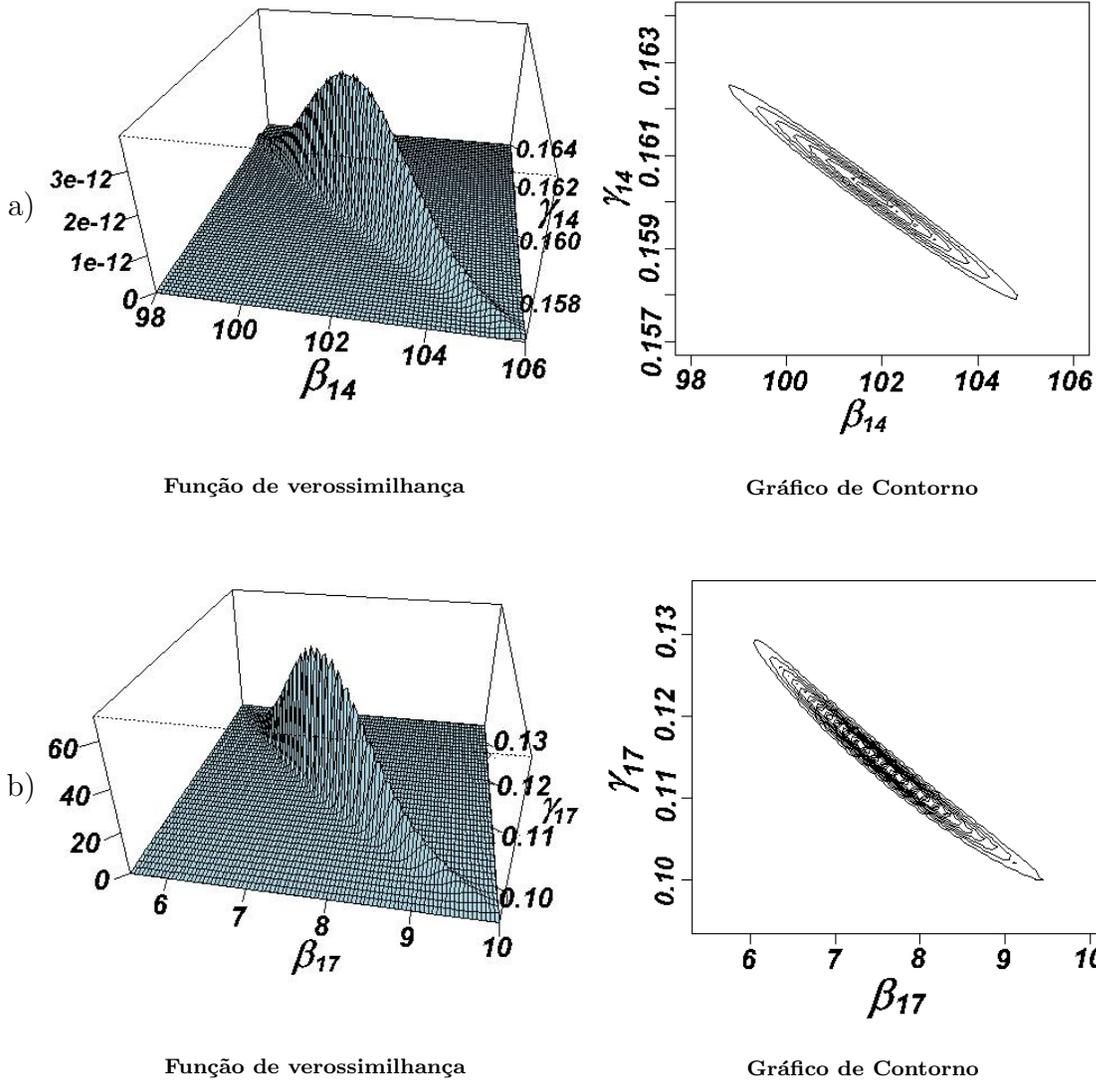


Figura 5.1: Verossimilhanças de $(\beta_{14}, \gamma_{14})$ (linha a) e $(\beta_{17}, \gamma_{17})$ (linha b) calculadas a menos de uma constante de proporcionalidade e fixando os demais parâmetros nos valores usados para simular dados.

Como indica a figura 5.1, os parâmetros β_i e γ_i são extremamente correlacionados a posteriori e, portanto, optamos por amostrar estes parâmetros conjuntamente. Como a eficiência de um algoritmo MH melhora quando a densidade proposta se aproxima da distribuição alvo, utilizamos como densidade proposta

$$q(\beta'_i, \gamma'_i \mid \beta_i^{(k)}, \gamma_i^{(k)}, \cdot) = p(\beta'_i \mid \gamma'_i, \cdot)q(\gamma'_i \mid \gamma_i^{(k)}),$$

onde $q(\gamma'_i | \gamma_i^{(k)})$ é um passeio aleatório, $N(\gamma_i^{(k)}, \tau_{\gamma_i}^2)$, e $p(\beta'_i | \gamma'_i, \cdot)$ é a condicional completa de β_i condicionada em γ'_i e nos valores correntes da cadeia. Note que a escolha desta fatoração foi conveniente, uma vez que β_i possui condicional completa conhecida e γ_i não. Assim, a probabilidade de aceitação utilizada para gerar amostras da condicional completa conjunta $(\beta_i, \gamma_i | \cdot)$ é dada por

$$\alpha \left(\left(\beta_i^{(k)}, \gamma_i^{(k)} \right), \left(\beta'_i, \gamma'_i \right) \right) = \min \left\{ 1, \frac{p(\beta'_i, \gamma'_i | \cdot) p(\beta_i^{(k)} | \gamma_i^{(k)}, \cdot) q(\gamma_i^{(k)} | \gamma'_i, \cdot)}{p(\beta_i^{(k)}, \gamma_i^{(k)} | \cdot) p(\beta'_i | \gamma'_i, \cdot) q(\gamma'_i | \gamma_i^{(k)}, \cdot)} \right\}.$$

Para ϕ , η e $\sigma_{PNAD,i}^2$, usamos passeios aleatórios na escala logarítmica como proposta, ou seja, proposta $Lognormal(\log(\phi^{(k)}), \tau_\phi^2)$ para ϕ , $Lognormal(\log((\sigma_{PNAD,i}^2)^{(k)}), \tau_{PNAD,i}^2)$ para $\sigma_{PNAD,i}^2$ e $Lognormal(\log(\eta^{(k)}), \tau_\eta^2)$ para η .

O método utilizado para sintonizar as variâncias τ_γ^2 , τ_ϕ^2 , $\tau_{PNAD,i}^2$ e τ_η^2 será apresentado na próxima seção.

5.1.4 Sintonizando a variância das distribuições propostas

O algoritmo Metropolis-Hastings é uma importante ferramenta ao implementar métodos de Monte Carlo via Cadeias de Markov. No entanto, sua eficiência depende crucialmente da escala da distribuição proposta. Se a variância da proposta é muito pequena, a Cadeia de Markov irá convergir lentamente, uma vez que seus incrementos serão pequenos. Ao contrário, se a variância é muito grande, a taxa de rejeição dos valores propostos será alta e a cadeia tenderá a não se mover. A adaptação, ou sintonia, desta variância foi feita utilizando o método apresentado em Roberts & Rosenthal (2006), que procura manter a taxa de aceitação em torno de 0,44, valor ótimo para propostas unidimensionais (Roberts et al. (1997) e Roberts & Rosenthal (2001)).

Supondo que τ^2 é a variância de uma distribuição proposta para um parâmetro qualquer θ , os passos a seguir resumem o método.

1. Inicializar o contador de iterações $k = 0$ e fixar a variância proposta inicial $(\tau^2)^{(0)}$.
2. Rodar 50 iterações, ou seja, 50 ciclos do Amostrador de Gibbs.

3. Se a taxa de aceitação de θ for maior que 0,44, fazer $\log((\tau^2)^{(k+1)}) = \log((\tau^2)^{(k)}) + \delta(k+1)$, se for menor, fazer $\log((\tau^2)^{(k+1)}) = \log((\tau^2)^{(k)}) - \delta(k+1)$, onde $\delta(k+1) = \min(0.01, (k+1)^{-\frac{1}{2}})$.
4. Incrementar o contador de k para $k+1$ e voltar ao passo 2.

A figura 5.2 apresenta as taxas de aceitação dos parâmetros ϕ e η , obtidas pelo método de sintonia descrito acima aplicado ao Modelo Espacial com fator de desconto 0,90. A linha azul marca o valor 0,44.

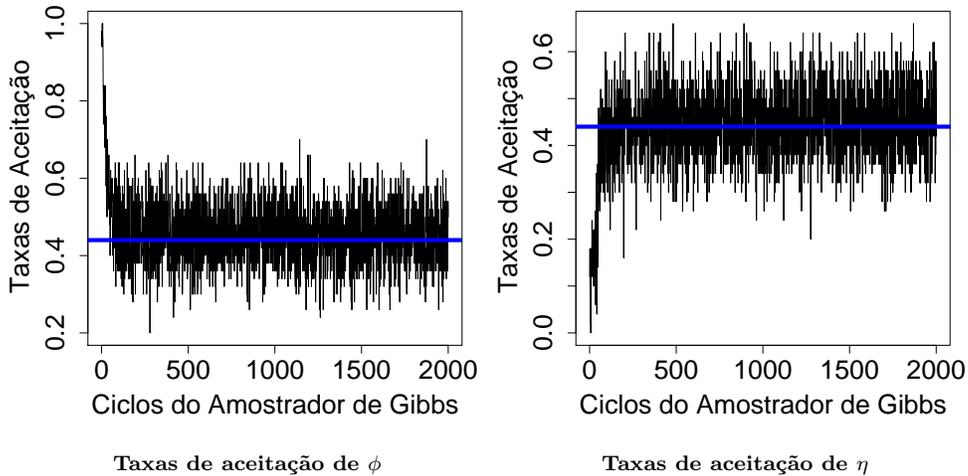


Figura 5.2: Taxas de aceitação obtidas a cada 50 ciclos do Amostrador de Gibbs durante o processo de sintonia do Modelo Espacial com fator 0,90.

Para todos os parâmetros amostrados por Metropolis-Hastings, as taxas de aceitação ficaram em torno de 0,44 a partir de $k = 500$. Assim, as variâncias propostas foram fixadas nas médias amostrais, tomadas a partir de $k = 500$, das variâncias geradas pelo algoritmo acima.

A sintonia de variâncias propostas se mostrou eficiente e foi extremamente válida devido ao grande número de parâmetros gerados por Metropolis-Hastings: 287 para o Modelo Espacial e 286 para o Modelo Hierárquico. A evolução das taxas de aceitação de outros parâmetros podem ser vistas no apêndice B.

5.1.5 *Forward Filtering Backward Sampling*

Proposto independentemente por Frühwirth-Schnatter (1994) e Carter & Kohn (1994), o algoritmo *Forward Filtering Backward Sampling* (FFBS) foi um dos primeiros métodos MCMC desenvolvidos para Modelos Dinâmicos.

Dado o Modelo Linear Dinâmico $\{F_t, G_t, V_t, W_t\}$ descrito pelas equações

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, \text{ onde } v_t \sim N(0, \mathbf{V}_t) \text{ e} \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + w_t, \text{ onde } w_t \sim N(0, \mathbf{W}_t), \text{ para } t = 1, 2, \dots, T, \end{aligned}$$

a idéia básica consiste em gerar amostras da distribuição conjunta a posteriori de $[\boldsymbol{\theta}_1 \ \boldsymbol{\theta}_2 \ \dots \ \boldsymbol{\theta}_T]$ dado todo o conjunto de informações disponíveis (distribuição suavizada).

O FFBS é mais simples de implementar que o amostrador estado por estado. A principal diferença é que os estados são gerados conjuntamente, tomando vantagem da estrutura de correlação temporal do modelo.

A extensão do algoritmo FFBS a modelos de espaço de estados não lineares e não normais não é direta. Para esses modelos, o amostrador de Gibbs geralmente é uma melhor opção. Mas, para modelos de espaço de estados lineares e normais o FFBS é superior, pois explora a estrutura condicionalmente linear do modelo.

Para modelos gaussianos em que não há conjugação, o FFBS pode ser usado como parte de um algoritmo MCMC para amostrar o vetor espaço de estados condicionalmente aos valores correntes de $\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t$ (Stroud et al., 2001).

Neste trabalho, a cada passo do Amostrador de Gibbs, o algoritmo FFBS foi usado para gerar amostras de $\mu_{i,t}$. Segue abaixo uma descrição detalhada deste processo.

Seja D_0 a informação inicial e $D_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\} \cup D_0$, para $t = 1, 2, \dots, 9$, a informação disponível até o ano t . Explorando a estrutura Markoviana da equação de sistema, podemos escrever

$$\begin{aligned} p(\mu_{i,0}, \mu_{i,1}, \dots, \mu_{i,9} | D_9, \Phi) &\propto p(\mu_{i,9} | D_9, \Phi) \\ &\times \prod_{t=0}^8 p(\mu_{i,t} | D_9, \mu_{i,t+1}, \Phi), \end{aligned}$$

onde Φ representa os demais parâmetros, fixados a cada passo do amostrador de Gibbs.

A seguir, apresentaremos as distribuições de $(\mu_{i,9} \mid D_9, \Phi)$ e $(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$. Começaremos pela distribuição de $(\mu_{i,9} \mid D_9, \Phi)$, que pode ser calculada a partir da aplicação sequencial do Filtro de Kalman.

Posteriori no instante $t - 1$:

Seja $(\mu_{i,t-1} \mid D_{t-1}, \Phi) \sim N(m_{i,t-1}, C_{i,t-1})$ a posteriori no instante $t - 1$.

Priori no instante t :

Como $\mu_{i,t} = \mu_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\exp(\gamma_i) - 1) + w_{i,t}$ e a combinação linear de normais é normal, então $(\mu_t \mid D_{t-1}, \Phi)$ terá distribuição normal com média

$$\begin{aligned} a_{i,t} &= E(\mu_{i,t} \mid D_{t-1}, \Phi) = E(\mu_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\exp(\gamma_i) - 1) + w_{i,t} \mid D_{t-1}, \Phi) \\ &= E(\mu_{i,t-1} \mid D_{t-1}, \Phi) + \beta_i \exp(\gamma_i(t-1))(\exp(\gamma_i) - 1) \\ &= m_{i,t-1} + \beta_i \exp(\gamma_i(t-1))(\exp(\gamma_i) - 1) \end{aligned}$$

e variância

$$R_{i,t} = Var(\mu_{i,t} \mid D_{t-1}, \Phi) = Var(\mu_{i,t-1} \mid D_{t-1}, \Phi) + Var(w_{i,t} \mid D_{t-1}, \Phi) = C_{i,t-1} + W_{i,t}.$$

Previsão um passo a frente:

Como $y_{i,t} = \mu_{i,t} + v_{i,t}$, $(y_{i,t} \mid D_{t-1}, \Phi)$ também terá distribuição normal, com média

$$f_{i,t} = E(y_{i,t} \mid D_{t-1}, \Phi) = E(\mu_{i,t} \mid D_{t-1}, \Phi) + E(v_{i,t} \mid D_{t-1}, \Phi) = a_{i,t}$$

e variância

$$\begin{aligned} Q_{i,t} &= Var(y_{i,t} \mid D_{t-1}, \Phi) = Var(\mu_{i,t} \mid D_{t-1}, \Phi) + Var(v_{i,t} \mid D_{t-1}, \Phi) \\ &= R_{i,t} + \sigma_{PNAD,i}^2, \quad \text{para } t \text{ ano de PNAD e} \\ Q_{i,t} &= Var(y_{i,t} \mid D_{t-1}, \Phi) = Var(\mu_{i,t} \mid D_{t-1}, \Phi) + Var(v_{i,t} \mid D_{t-1}, \Phi) \\ &= R_{i,t} + s_{i,t}, \quad \text{para } t \text{ ano de censo.} \end{aligned}$$

Posteriori no instante t :

- Para os municípios da PNAD utilizados:

$$(\mu_{i,t} \mid D_t, \Phi) \mid N(m_{i,t}, C_{i,t}), \quad \text{onde}$$

$$m_{i,t} = a_{i,t} + A_{i,t}e_{i,t},$$

$$C_{i,t} = R_{i,t} - A_{i,t}^2 Q_{i,t},$$

$$A_{i,t} = R_{i,t} Q_{i,t}^{-1} \quad \text{e}$$

$$e_{i,t} = y_{i,t} - f_{i,t}.$$

O cálculo da posteriori no tempo t pode ser encontrado em West & Harrison (1997).

- Para os municípios selecionados para previsão: Como nos municípios selecionados para previsão só há observações censitárias, se t for ano censitário, o cálculo da distribuição de $(\mu_{i,t} \mid D_t, \Phi)$ é idêntico ao apresentado no item anterior e, se t for ano de PNAD, a distribuição de $(\mu_{i,t} \mid D_t, \Phi)$ será idêntica a priori no tempo t , pois não há observações para a atualização.

Já a distribuição de $(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$ é dada por

$$(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi) \sim N(M_{i,t}, V_{i,t}), \quad \text{onde}$$

$$V_{i,t} = (W_{i,t+1}^{-1} + C_{i,t}^{-1})^{-1},$$

$$M_{i,t} = V_{i,t}(W_{i,t+1}^{-1} U_{i,t} + C_{i,t}^{-1} m_{i,t}) \quad \text{e}$$

$$U_{i,t} = \mu_{i,t+1} - \beta_i \exp(\gamma_i t)(\exp(\gamma_i) - 1).$$

O cálculo da distribuição de $(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$ pode ser encontrado no apêndice D.

Como podemos notar, a distribuição $p(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$ não depende diretamente das observações $y_{i,t}$, uma vez que estas observações já foram incorporadas por $m_{i,t}$ e $C_{i,t}$. Portanto, para os municípios p selecionados para previsão, a geração de $\mu_{p,t}$, para qualquer ano $t = 1, 2, \dots, 9$, pode ser feita diretamente da distribuição de $(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$.

5.2 Verificação de Convergência

A verificação de convergência dos métodos MCMC utilizados será baseada na idéia de que as trajetórias de cadeias com diferentes valores iniciais deve ser a mesma após a

convergência.

Convergência do Modelo Espacial: Para verificar a convergência do Modelo Espacial, a figura 5.3 apresenta algumas cadeias de amostras a posteriori obtidas com fator de desconto 0,90, *burn-in* 0 e *thinning* 150. Para este modelo, os resultados obtidos para os demais parâmetros, assim como os resultados obtidos com outros fatores de desconto (0,40, 0,50, 0,80, 0,85, 0,95 e 0,99) apresentaram comportamento similar e, por isto, foram omitidos. Os resultados obtidos com a estratégia $W_{i,t} = \tau^2$ para tratar os erros de evolução também foram similares e omitidos.

Como mostra a figura 5.3, há indícios de convergência do Modelo Espacial, que foi assumida com *burn-in* 2000 e *thinning* 150. O apêndice E apresenta as funções de autocorrelação e os histogramas das amostras a posteriori. O tempo computacional para rodar os algoritimos MCMC sob o Modelo Espacial foi em média de 24 horas.

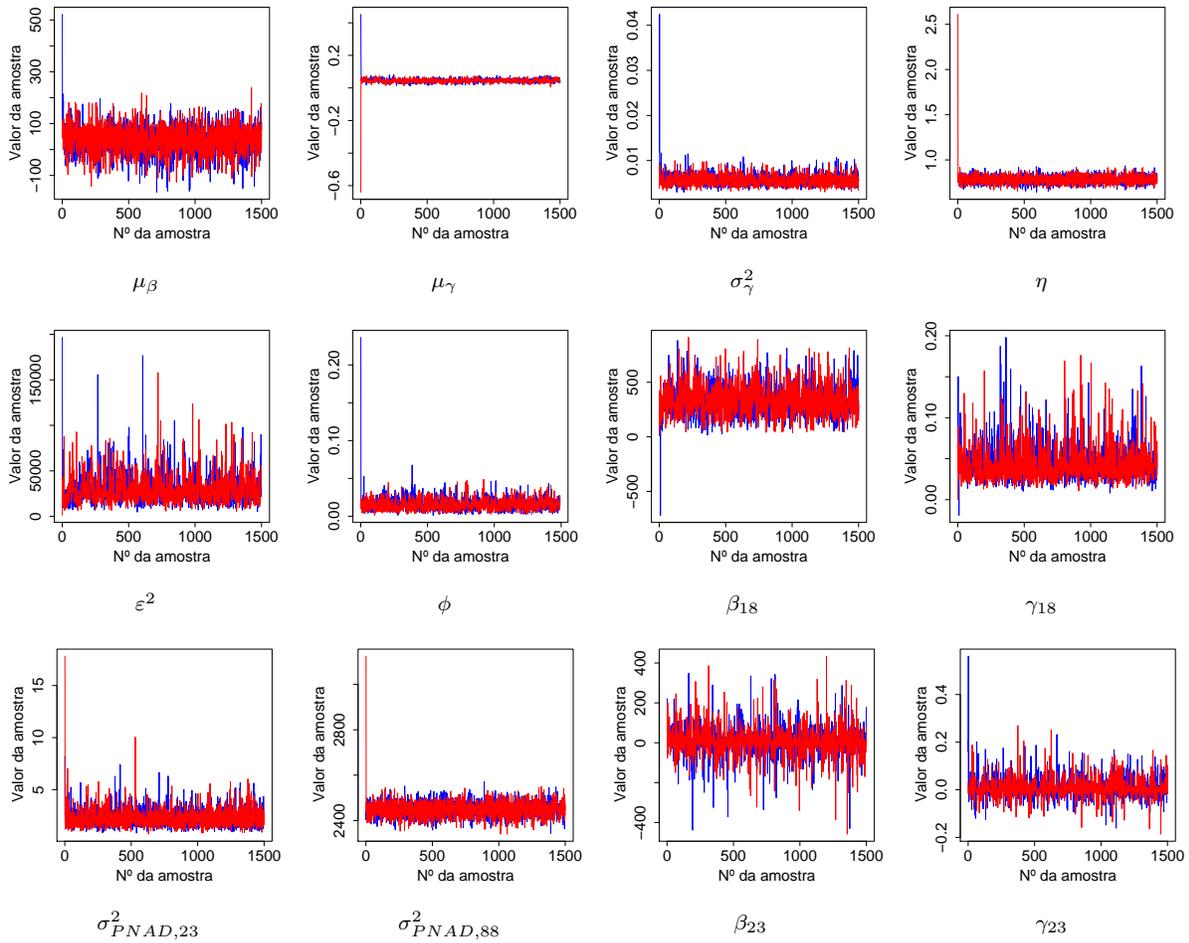


Figura 5.3: Cadeias a posteriori para o Modelo Espacial com fator de desconto de 0,90, *burn-in* 0 e *thinning* 150.

Convergência do Modelo Hierárquico: A figura 5.4 apresenta algumas cadeias a posteriori obtidas no processo de inferência para o Modelo Hierárquico com fator de desconto igual a 0,80, *burn-in* 0 e *thinning* 100. Também para este modelo, os resultados obtidos para os demais parâmetros, assim como os resultados obtidos com outros fatores de desconto apresentaram comportamento similar e foram omitidos. Os resultados obtidos com a estratégia $W_{i,t} = \tau^2$ para tratar os erros de evolução também foram similares e omitidos.

Como mostra a figura 5.4, há indícios de convergência para o Modelo Hierárquico, que foi assumida com *burn-in* 1000 e *thinning* 100. Também para este modelo, o apêndice

E apresenta as funções de autocorrelação e os histogramas das amostras a posteriori. O tempo computacional para rodar os algoritimos MCMC sob o Modelo Hierárquico foi em média de 10 horas.

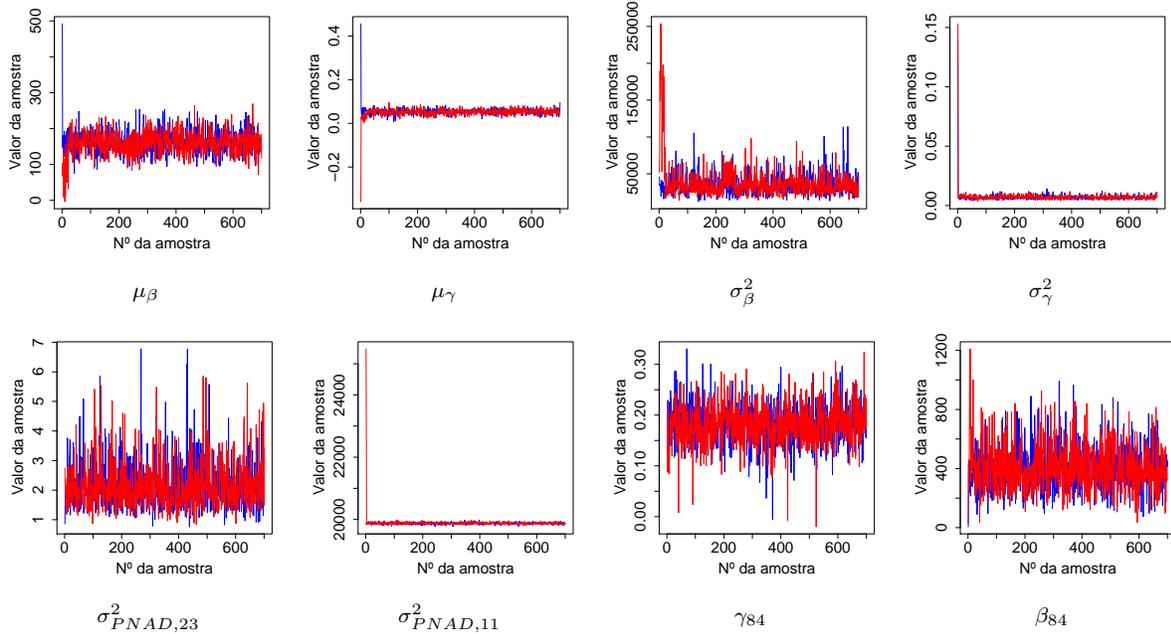


Figura 5.4: Cadeias a posteriori para o Modelo Hierárquico com fator de desconto de 0,80, *burn-in* 0 e *thinning* 100.

5.3 Estudo Simulado

Analisaremos nesta seção um conjunto artificial de dados simulados a partir dos modelos citados. O objetivo é verificar a eficiência da metodologia adotada para fazer inferência sobre os parâmetros.

Foram simulados dados artificiais a partir do Modelo Hierárquico com fator de desconto de 0,80 e do Modelo Espacial com fator de desconto de 0,95. Nesta simulação, foram considerados 91 pontos no espaço com localização idêntica à dos dados reais. Para especificar as variâncias $W_{i,1}$ do primeiro erro de evolução, tomamos $W_{i,1} = s_{i,1}$ e, para ser fiel aos dados reais, os parâmetros $\mu_{i,0}$, $\sigma_{PNAD,i}^2$, μ_β , σ_β^2 , μ_γ , σ_γ^2 , ϕ , ε^2 e η foram fixados nas estimativas obtidas através do ajuste dos dados reais. Na simulação, foram usadas

as priors descritas na seção 4.5, pois são priors pouco informativas.

Para avaliar a inferência sobre os parâmetros β_i e γ_i ($i = 1, 2, \dots, 91$), as figuras 5.5 e 5.6 apresentam as médias a posteriori destes parâmetros, os intervalos de 95% de credibilidade associados e os valores utilizados para gerar os dados.

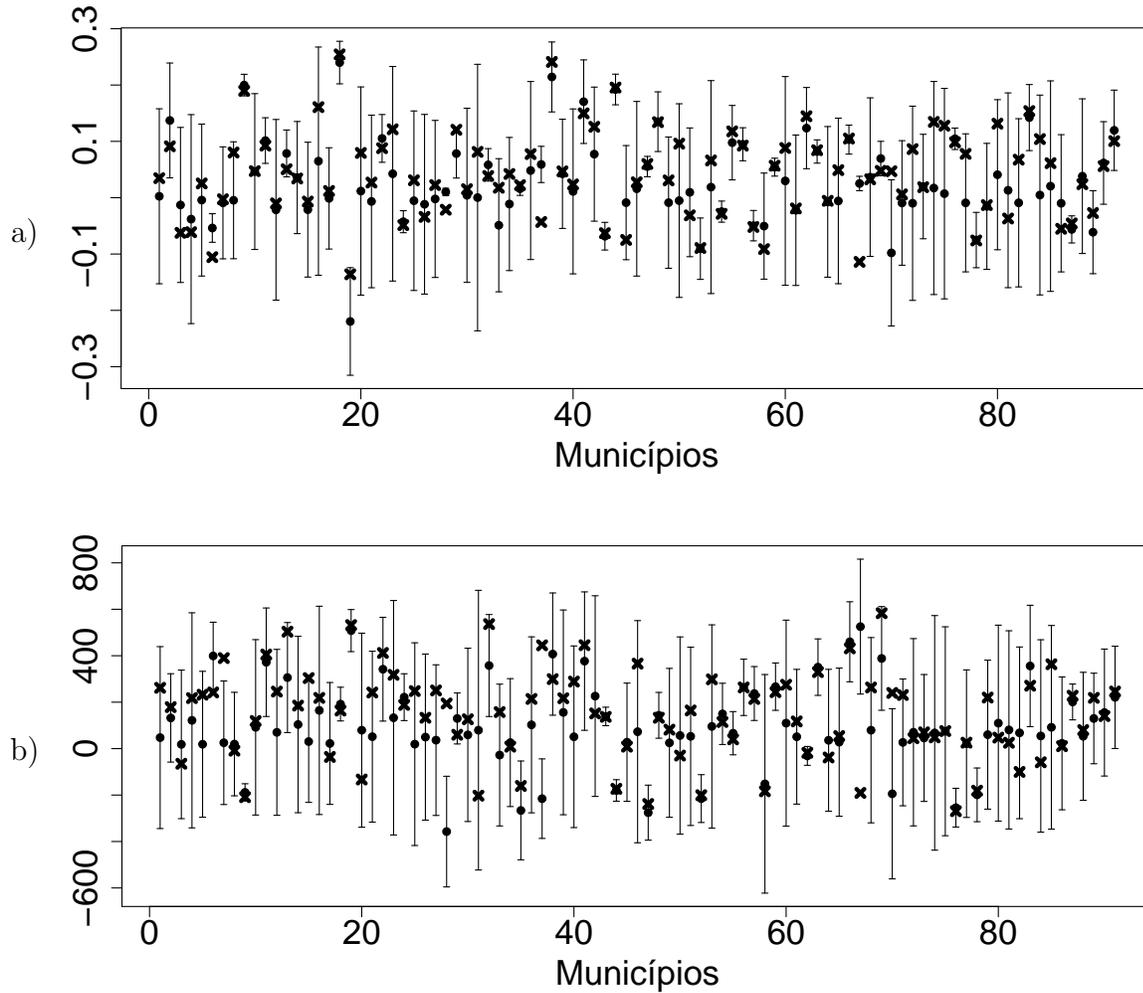


Figura 5.5: Estimação dos parâmetros γ_i (a) e β_i (b) para o Modelo Hierárquico ajustado com conjunto de dados artificiais, onde “•” representa a média a posteriori, “x” o valor usado para simular os dados e a linha informa o intervalo de 95% de credibilidade da média a posteriori.

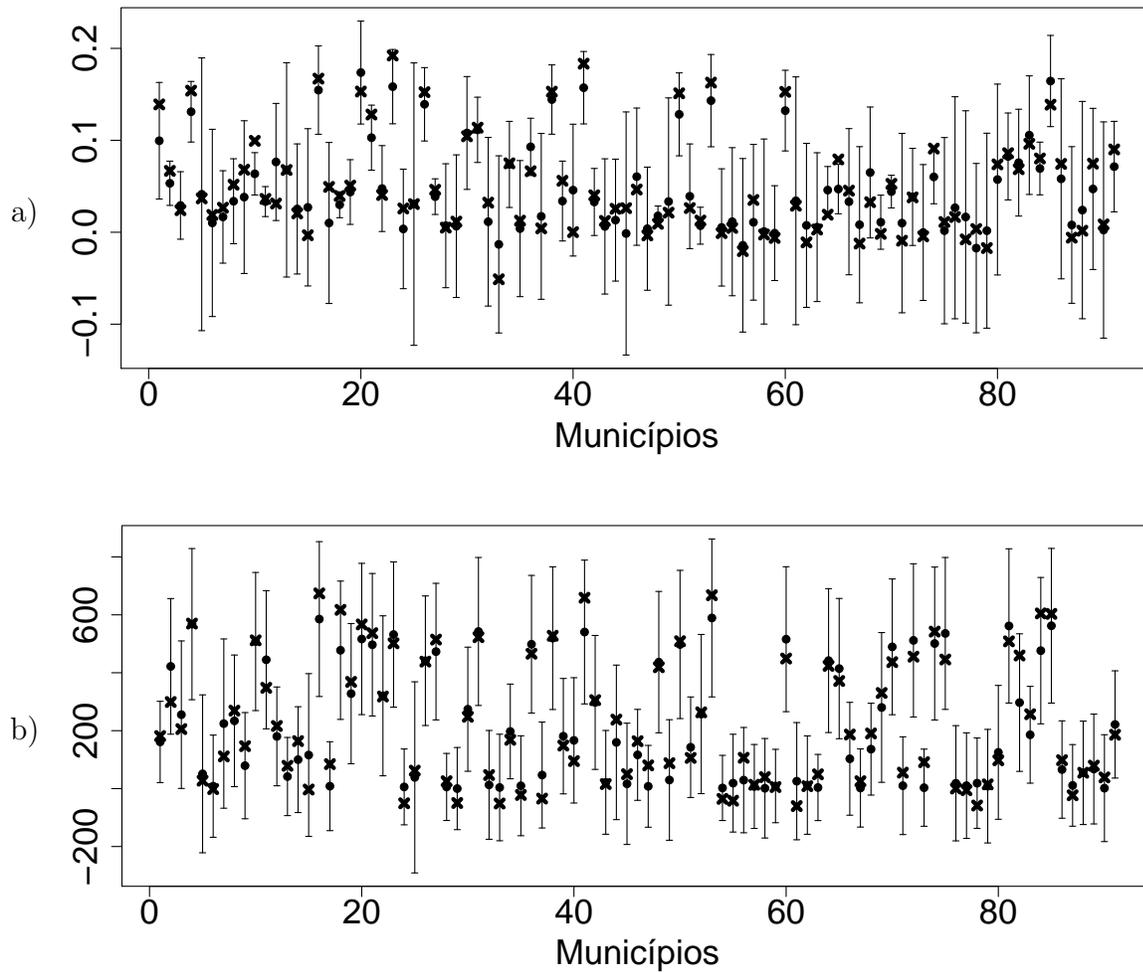


Figura 5.6: Estimação dos parâmetros γ_i (a) e β_i (b) para o Modelo Espacial ajustado com conjunto de dados artificiais, onde “●” representa a média a posteriori destes parâmetros, “×” o valor usado para simular os dados e a linha informa o intervalo de 95% de credibilidade da média a posteriori.

A figura 5.7 exibe, através de histogramas das amostras a posteriori, como o processo de estimação sob o Modelo Espacial foi capaz de localizar os valores usados para gerar os dados, representados pela reta azul. Para o Modelo Hierárquico, o resultado foi similar.

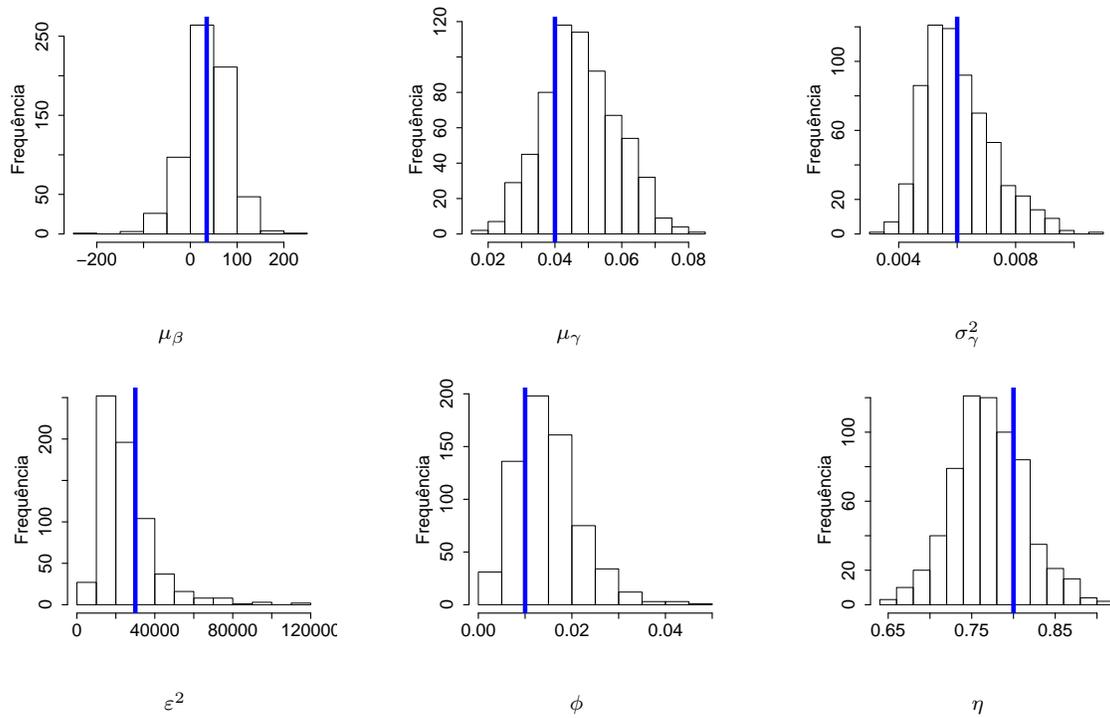


Figura 5.7: Histogramas de amostras a posteriori sob o Modelo Espacial com fator 0,95 e valores usados para gerar os dados (linha azul).

Na figura 5.8 apresentamos um gráfico de dispersão com os 819 (91×9) parâmetros $\mu_{i,t}$ usados para simular os dados do Modelo Hierárquico versus suas estimativas a posteriori. Na figura, a reta azul é a função identidade.

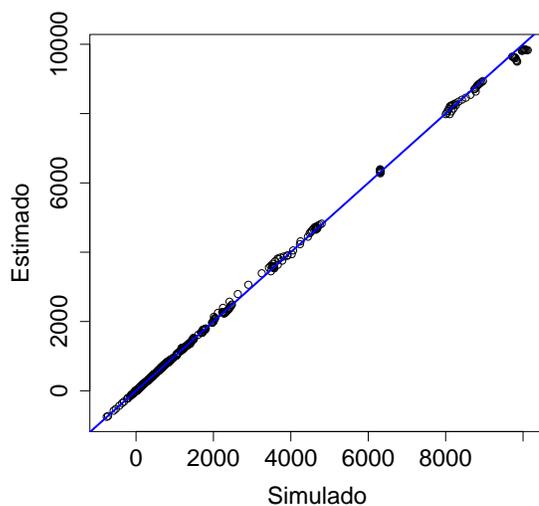


Figura 5.8: $\mu_{i,t}$ usado para simular os dados do modelo Hierárquico versus suas estimativas a posteriori.

Como mostram os resultados desta seção, podemos concluir que a metodologia bayesiana se mostrou eficiente para fazer inferência sobre os modelos considerados.

5.4 Ajuste dos Dados Reais

Analisaremos agora as densidades populacionais estimadas pelos modelos citados, assim como a precisão destas estimativas. Como a análise dos critérios de seleção feita na seção anterior destacou o Modelo Espacial, optamos por apresentar apenas os ajustes deste modelo. As figuras desta seção apresentam em vermelho a média a posteriori $\hat{\mu}_{i,t}$ (linha contínua) e o intervalo de 95% de credibilidade a posteriori (linhas tracejadas) de $\mu_{i,t}$. Os pontos e os triângulos em azul correspondem às observações $y_{i,t}$ para anos de PNAD e de censos, respectivamente. Vale lembrar que os intervalos de credibilidade expostos são para $\mu_{i,t}$ e, portanto, não precisam conter, necessariamente, as observações $y_{i,t}$.

5.4.1 Ajuste com $W_{i,t} = \tau^2$

Analisaremos primeiro os resultados obtidos fixando a variância dos erros de evolução ($W_{i,t} = \tau^2$). Os resultados apresentaram o seguinte padrão: para os municípios menos populosos, o nível $\mu_{i,t}$ estimado oscila muito, tentando se ajustar às estimativas da PNAD; já para os municípios mais populosos, as estimativas da PNAD praticamente não exercem influência sobre o nível.

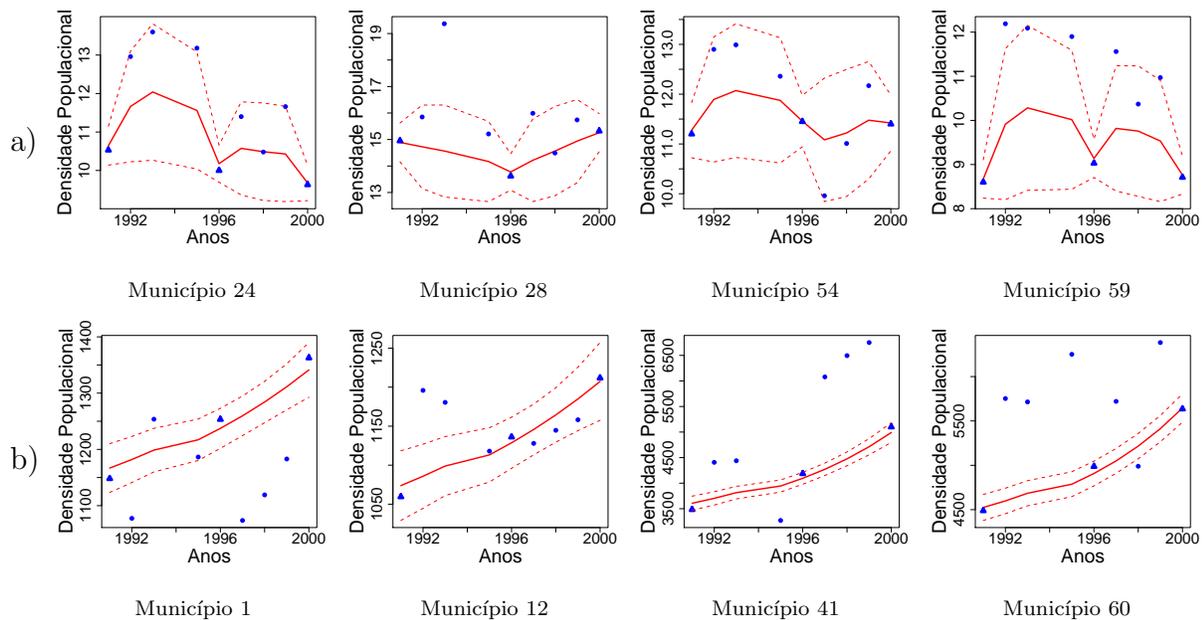


Figura 5.9: Ajustes do Modelo Espacial com a estratégia $W_{i,t} = \tau^2$ para tratar a variância dos erros de evolução.

A figura 5.9 exemplifica esta situação. Nela são apresentados ajustes do Modelo Espacial para 8 municípios, quatro deles com total populacional pequeno (Municípios 24, 28, 54 e 59) e os outros quatro com total populacional elevado (Municípios 1, 12, 41 e 60). A justificativa para este padrão foi atribuída justamente à estratégia utilizada para tratar os erros de evolução. Ao fixar a variância destes erros, estamos assumindo que a perda de informação sobre o nível ao passar do tempo é a mesma para todos os municípios. Assim, é natural que esta perda, “única”, signifique um decaimento de informação relativamente elevado para municípios com erro de cobertura censitário pequeno e um decaimento relativamente pequeno para municípios com erro de cobertura censitário alto. Portanto,

chegamos à conclusão de que usar $W_{i,t} = \tau^2$ não é uma boa estratégia para tratar o conjunto de dados utilizado. O ideal então, seria fazer $W_{i,t} = \tau_i^2$ e tentar estimar os parâmetros τ_i^2 . Porém, julgamos previamente que com esta estratégia teríamos problemas de identificabilidade destes parâmetros. A solução para este problema foi usar fatores de desconto para especificar as variâncias $W_{i,t}$ dos erros de evolução. Na próxima seção apresentaremos os resultados obtidos com a estratégia de descontos.

5.4.2 Ajuste com Fatores de Desconto

Passemos agora à análise dos resultados obtidos usando os fatores de desconto. Com a estratégia de descontos, a perda de informação sobre o nível ao passar para um instante t é proporcional à informação no instante $t - 1$. Deste modo, ao passar de um ano de censo para um ano de PNAD, o decaimento de informação sobre $\mu_{i,t}$ fica vinculado a precisão censitária, evitando o comportamento exibido na figura 5.9, ou seja, evitando que o ajuste seja influenciado indevidamente pelo total populacional do município. Afim de permitir a comparação dos resultados obtidos com $W_{i,t} = \tau^2$ com os resultados obtidos com fatores de desconto, apresentamos na figura 5.10 os mesmos municípios exibidos na figura 5.9, porém com ajuste feito usando o Modelo Espacial com fator 0,90.

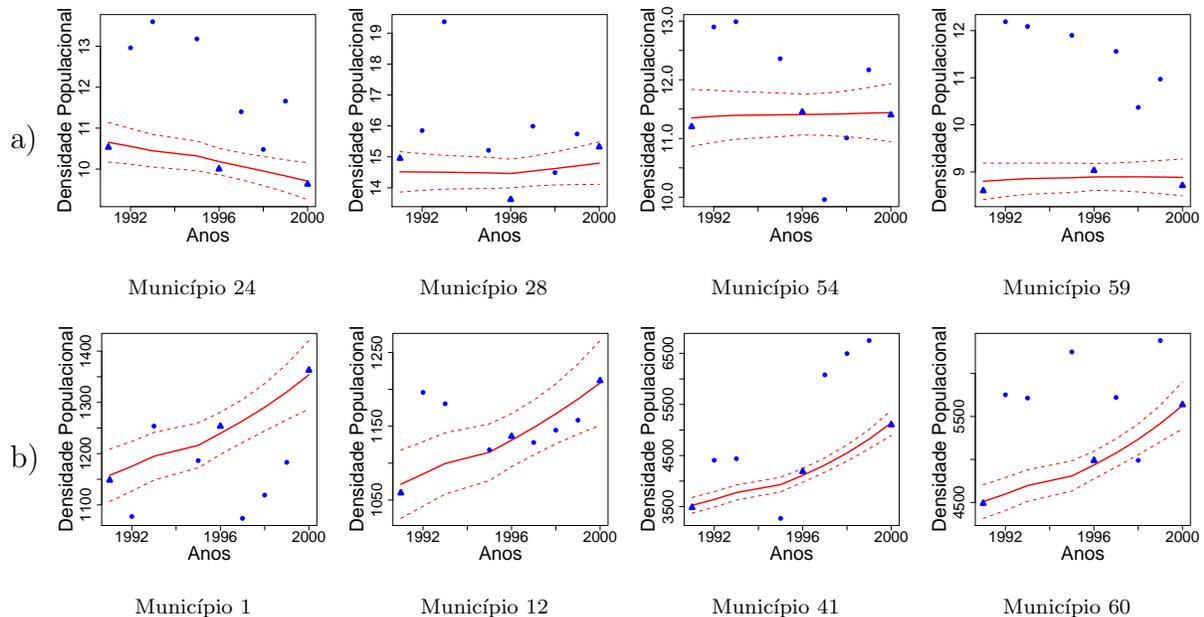


Figura 5.10: Ajuste do Modelo Espacial com fator de desconto 0,90.

Para analisar a influência dos dados na estimação do nível, faremos uso da estatística Q_i , dada por

$$Q_i = \frac{\frac{\sum_{t \in P} s_{i,t}}{6}}{\frac{\sum_{t \in C} s_{i,t}}{3}}, \text{ onde } i \text{ indica o município.}$$

Através desta estatística, podemos ter uma noção grosseira da precisão relativa entre a PNAD e o censo no município i . Quanto menor o valor de Q_i , maior é a indicação de que no município i a PNAD é mais precisa que o censo e, reciprocamente, quanto maior Q_i , maior é a indicação de que o censo é mais preciso.

A figura 5.11 apresenta a estatística Q_i para todos os 91 municípios selecionados. Os pontos “●” marcam alguns municípios com alta precisão relativa e os pontos “×” marcam precisão relativa baixa.

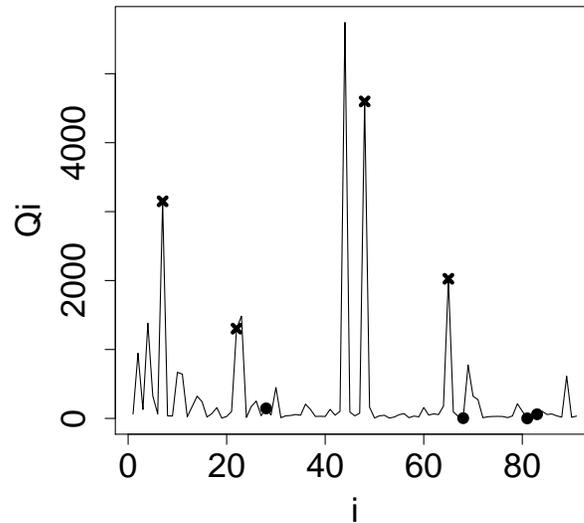


Figura 5.11: Gráfico de dispersão do i (município) versus estatística Q_i (precisão relativa entre PNAD e censo).

Para analisar os resultados obtidos com a estratégia de descontos, optamos por exibir apenas resultados com fatores 0,50 e 0,90, pois os ajustes dos demais fatores utilizados (0,40, 0,80, 0,85, 0,95 e 0,99) foram similares aos ajustes com 0,50 e 0,90.

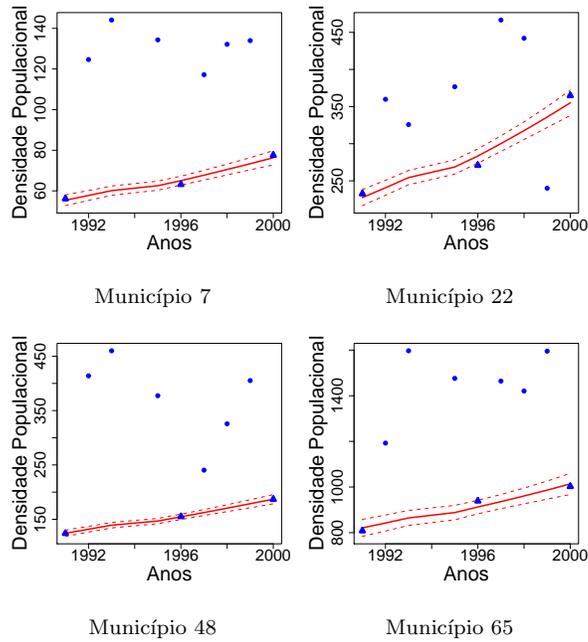


Figura 5.12: Ajustes do Modelo Espacial com fator 0,90 para os municípios marcados por “×” na figura 5.11.

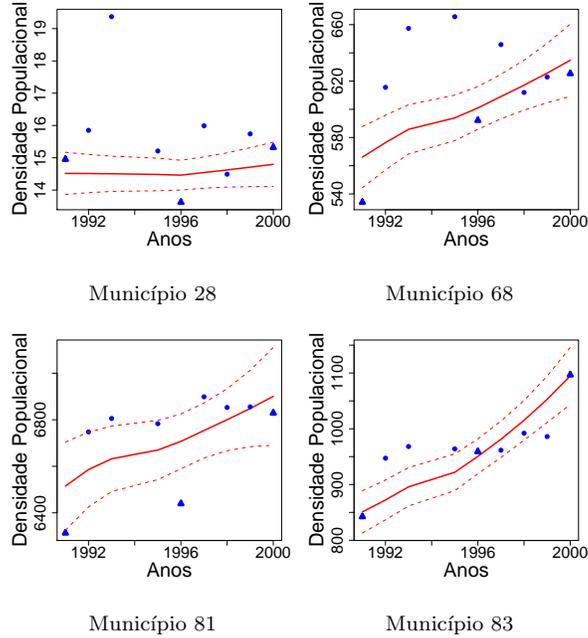


Figura 5.13: Ajustes do Modelo Espacial com fator 0,90 para os municípios marcados por “●” na figura 5.11.

Para os municípios marcados por “×”, a baixa precisão relativa das estimativas da PNAD faz com que o nível $\mu_{i,t}$ sofra pouca influência dos dados não censitários. Para estes municípios, o ajuste é feito basicamente a partir dos censos. A figura 5.12 apresenta o ajuste destes municípios com fator de desconto 0,90.

Já na figura 5.13, apresentamos o ajuste para os municípios marcados por “•”. Nestes municípios, a alta precisão das estimativas faz com que a PNAD exerça mais influência, distorcendo a trajetória do nível $\mu_{i,t}$ de modo significativo.

Portanto, podemos concluir que a modelagem das estimativas das variâncias cumpre a função esperada: informar sobre a precisão das estimativas $y_{i,t}$.

Devido às dificuldades de investigar toda a população, é comum ocorrerem problemas de subcobertura em censos demográficos, ou seja, estimativas menores que o valor verdadeiro da característica de interesse. Os modelos ajustados permitiram identificar indícios de subcobertura do censo em alguns municípios. Na figura 5.14, apresentamos alguns municípios em que a densidade populacional informada nos censos foi sistematicamente menor do que as estimativas obtidas no ajuste dos modelos. Os indícios de subcobertura foram identificados para todos os fatores de desconto testados.

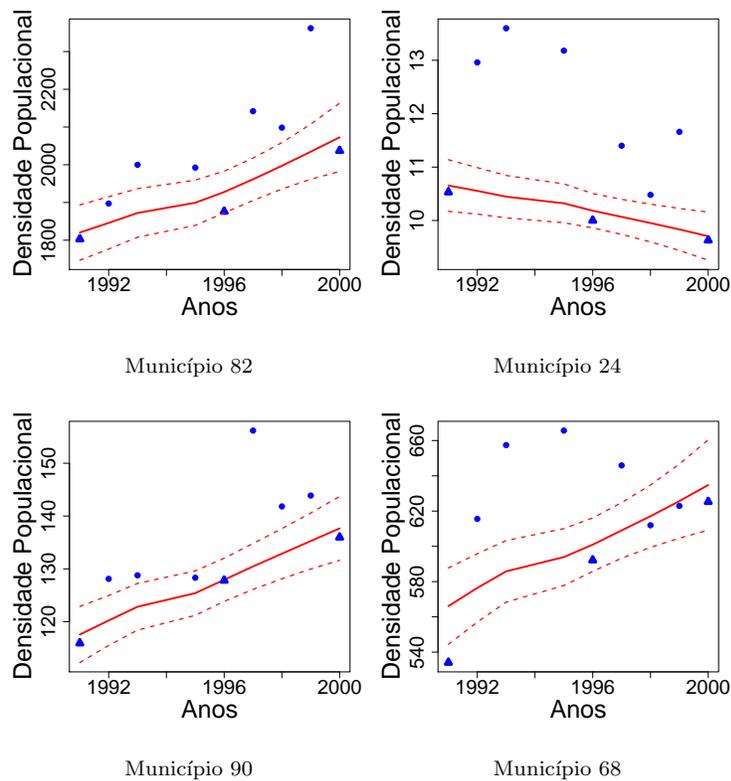


Figura 5.14: Ajustes do Modelo Espacial com fator 0,90 para alguns municípios que apresentaram indícios de subcobertura censitária.

Para comparar o comportamento do ajuste com fatores de desconto distintos, apresentamos na figura 5.15 alguns resultados para o Modelo Espacial com fator de desconto de 0,90 e 0,50. A primeira linha de gráficos exibe os resultados para o fator 0,50 e a segunda para 0,90. A figura exemplifica uma situação já esperada: para fatores de desconto mais baixos, o nível tende a seguir mais as observações da PNAD e, conseqüentemente, o ajuste aos dados é melhor; já para fatores de desconto maiores, a precisão das observações censitárias se propaga ao longo do tempo, diminuindo a influência das observações intercensitárias. Esta análise será confirmada na análise dos critérios DIC e EPD apresentada na próxima seção, onde as medidas \bar{D} e G são menores (melhor ajuste) para fatores de desconto menores.

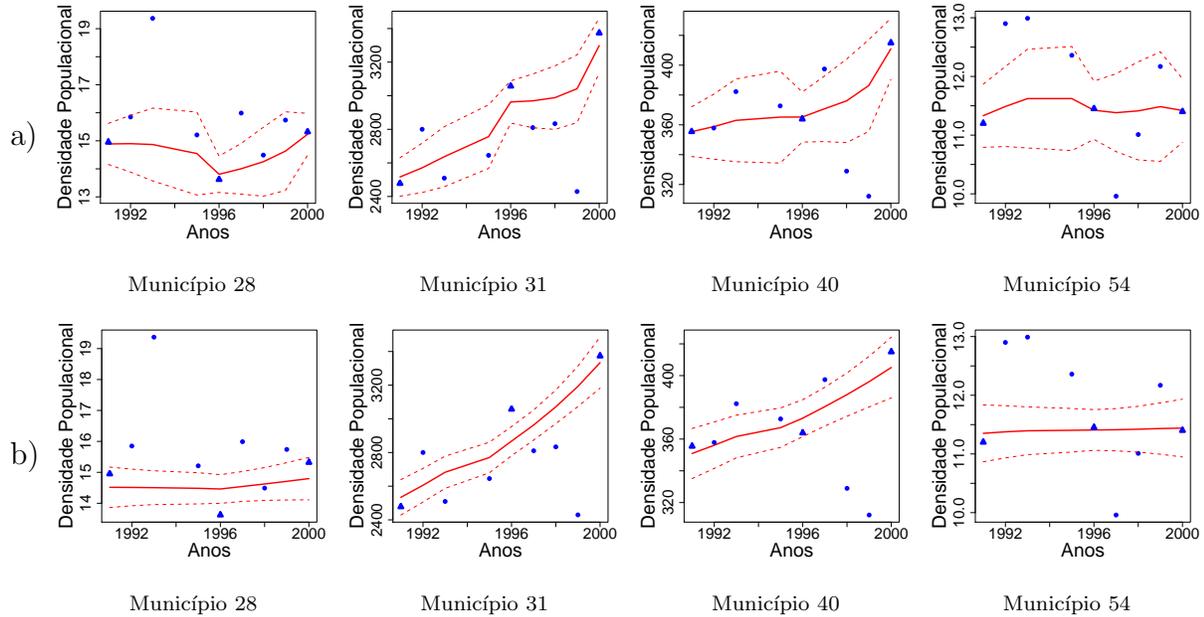


Figura 5.15: Ajustes do Modelo Espacial com fatores de desconto 0,50 (linha a) e 0,90 (linha b).

Os resultados apresentados nesta seção, principalmente os gráficos da figura 5.15, sugerem a seguinte pergunta: “Quanto da informação sobre o nível será perdido com o passar dos anos?”. Como vimos, a resposta para esta questão segundo a nossa proposta está intimamente ligada à especificação do fator de desconto. Porém, tanto a especificação quanto a estimação deste fator são tarefas complicadas. Consideramos, então, a especificação do fator como uma dificuldade inerente ao problema estudado.

5.4.3 Resultados do DIC e do EPD para os Dados Reais

As tabelas 5.1 e 5.2 apresentam os resultados do DIC e do EPD para os modelos propostos.

Comparando o Modelo Hierárquico com o Modelo Espacial para cada fator de desconto utilizado, podemos notar que o número efetivo de parâmetros pD é sistematicamente menor para o Modelo Espacial e que, em compensação, a medida de ajuste \bar{D} é maior neste modelo. No geral, os valores do DIC não apresentaram diferenças significativas entre os modelos, exceto para o fator 0,50, onde o DIC aponta para o Modelo Hierárquico. Assim, como esta diferença foi significativa apenas para o fator 0,50, não

consideramos o DIC como um bom critério para diferenciar os modelos. Ainda pelo critério DIC, podemos notar que a diminuição do fator de desconto de 0,99 à 0,50 é acompanhada pelo aumento do pD e pela diminuição de \bar{D} e do DIC.

		Estatísticas			
Modelo	Fator de Desconto	\bar{D}	\hat{D}	pD	DIC
Espacial	0,40	18015,60	17700,73	314,87	18330,46
	0,50	18019,38	17711,51	307,87	18327,24
	0,80	18127,06	17883,61	243,44	18370,51
	0,85	18144,96	17909,69	235,26	18380,22
	0,90	18163,38	17935,28	228,09	18391,47
	0,95	18178,79	17956,61	222,18	18400,98
	0,99	18188,56	17973,12	215,44	18404,00
Hierárquico	0,40	17991,43	17656,30	335,13	18326,57
	0,50	17995,21	17686,08	309,13	18304,35
	0,80	18119,76	17872,81	246,96	18366,72
	0,85	18138,96	17899,50	239,45	18378,41
	0,90	18156,20	17922,88	233,31	18389,51
	0,95	18174,22	17948,12	226,10	18400,32
	0,99	18183,27	17962,13	223,20	18406,47

Tabela 5.1: Resultados do DIC para o Modelo Hierárquico e para o Modelo Espacial com diferentes fatores de desconto.

Já o EPD aponta significativamente para o Modelo Espacial, com destaque para o fator 0,50, onde a medida G assume o menor valor e o EPD é um dos menores. Para o modelo espacial, podemos notar que a diminuição do fator de desconto é acompanhada por uma tendência de aumento da medida P (maior penalização) e por uma diminuição sistemática da medida G (melhor ajuste). Para o Modelo Espacial, podemos notar que os menores valores do EPD foram obtidos com os fatores 0,50, 0,80 e 0,85. Dentre estes, o fator 0,50 se destaca por apresentar, segundo o EPD, um dos melhores resultados e ainda o melhor ajuste. Logo, consideramos que o EPD permitiu diferenciar bem o Modelo Hierárquico do Espacial, além de ter apontado, assim como o DIC, melhor ajuste e maior medida penalização para fatores menores.

		Estatísticas		
Modelo	Fator de Desconto	G	P	EPD
Espacial	0,40	7,8043E+07	2,5473E+08	2,9375E+08
	0,50	7,8981E+07	2,5207E+08	2,9156E+08
	0,80	8,6014E+07	2,4834E+08	2,9135E+08
	0,85	8,6879E+07	2,4825E+08	2,9169E+08
	0,90	8,7630E+07	2,4922E+08	2,9303E+08
	0,95	8,8394E+07	2,4796E+08	2,9216E+08
	0,99	8,8961E+07	2,4843E+08	2,9291E+08
Hierárquico	0,40	7,9274E+07	2,5580E+08	2,9544E+08
	0,50	7,9289E+07	2,5570E+08	2,9535E+08
	0,80	8,5694E+07	2,5238E+08	2,9523E+08
	0,85	8,6605E+07	2,5212E+08	2,9543E+08
	0,90	8,7938E+07	2,4915E+08	2,9312E+08
	0,95	8,8206E+07	2,5178E+08	2,9588E+08
	0,99	8,8906E+07	2,5078E+08	2,9523E+08

Tabela 5.2: Resultados do EPD para o Modelo Hierárquico e para o Modelo Espacial com diferentes fatores de desconto.

5.5 Previsão para municípios não selecionados pela PNAD

A figura 5.16 exibe as previsões do nível $\mu_{i,t}$ para os 6 municípios não selecionados pela PNAD, obtidas sob o Modelo Espacial com fator 0,50 (em vermelho). Os pontos verdes representam as densidades oficiais publicadas pelo IBGE. Como não há dados da PNAD para estes municípios, não iremos omitir o nome dos mesmos. Como podemos ver, praticamente todas as densidades oficiais ficaram dentro do intervalo de credibilidade. Dentre as previsões, pode-se destacar as do município 92, por aproximarem bem as estimativas oficiais em todos os anos.

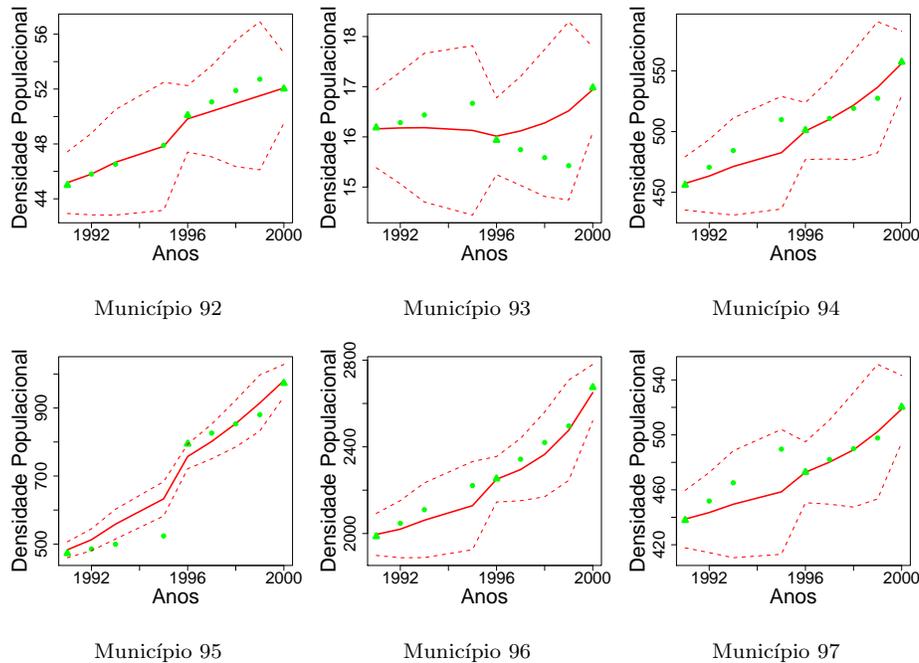


Figura 5.16: Estimativas da densidade populacional publicadas pelo IBGE (pontos verdes), previsões para o nível (linha vermelha) e intervalo de 95% de credibilidade associado (linha tracejada) obtidos sob o Modelo Espacial com fator 0.50.

5.6 Comparação com Dados Externos

Nesta seção, faremos a comparação dos resultados obtidos em nossa análise com resultados obtidos em outras abordagens.

Nas figuras 5.17 e 5.18, apresentamos, respectivamente, o ajuste do Modelo Espacial com fatores 0,50 e 0,80 (em vermelho) e as densidades populacionais oficiais publicadas pelo IBGE (pontos verdes). Como mostram as figuras, a metodologia apresentada aqui gerou estimativas populacionais próximas das publicadas pelo IBGE, com a vantagem de informar o erro associado. Nos apêndices F e G, apresentamos as mesmas informações exibidas nestas figuras, porém para outros municípios.

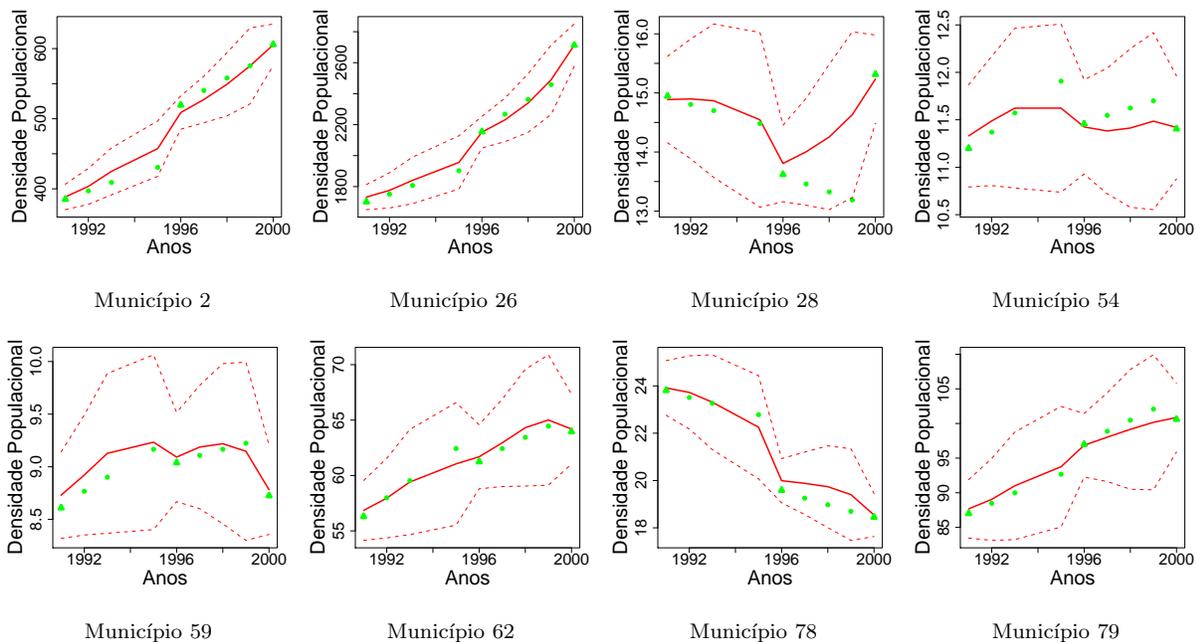


Figura 5.17: Ajuste do Modelo Espacial com fator 0,50 (em vermelho) e estimativas publicadas pelo IBGE (pontos verdes).

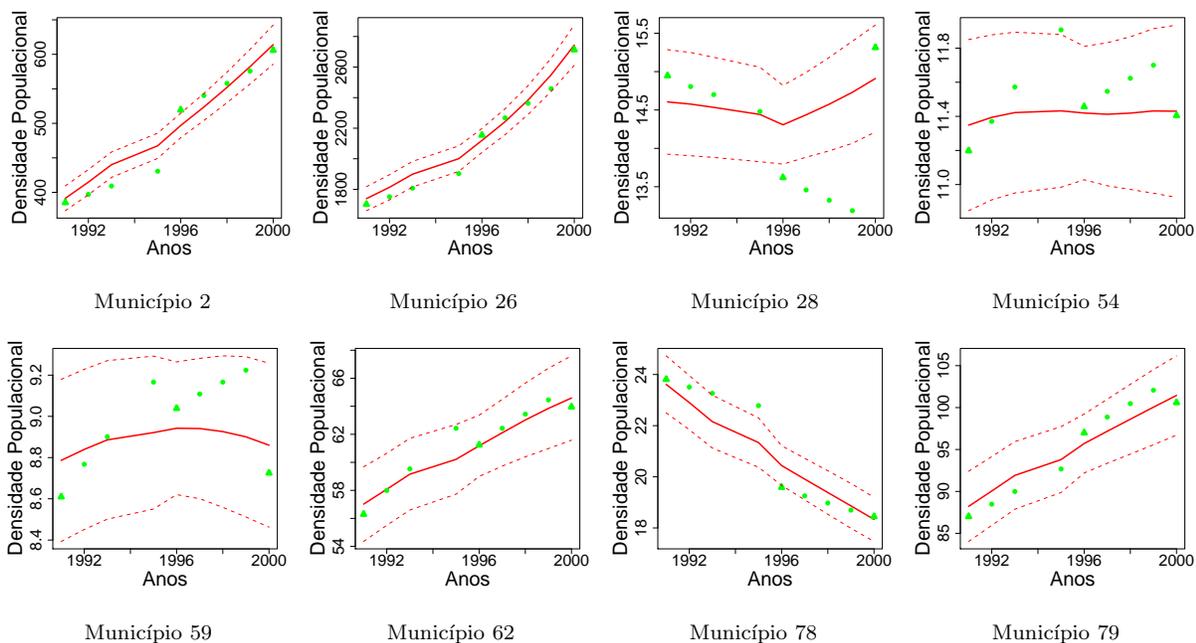


Figura 5.18: Ajuste do Modelo Espacial com fator 0,80 (em vermelho) e estimativas publicadas pelo IBGE (pontos verdes).

Pela análise gráfica apresentada nesta seção, o ajuste com fator de desconto 0,50 foi o que gerou estimativas mais próximas das estimativas oficiais do IBGE (figura 5.17 e apêndice F). Já para fatores de desconto maiores, a trajetória do nível se manteve fiel aos censos.

A figura 5.19 apresenta os resultados obtidos em Souza (2004) para os municípios usados nas figuras 5.17 e 5.18. Comparando as figuras 5.17 e 5.19, podemos notar que em Souza (2004), o ajuste é extremamente influenciado pelos censos e que, por outro lado, nossa abordagem permitiu que o nível se adaptasse melhor aos dados da PNAD.

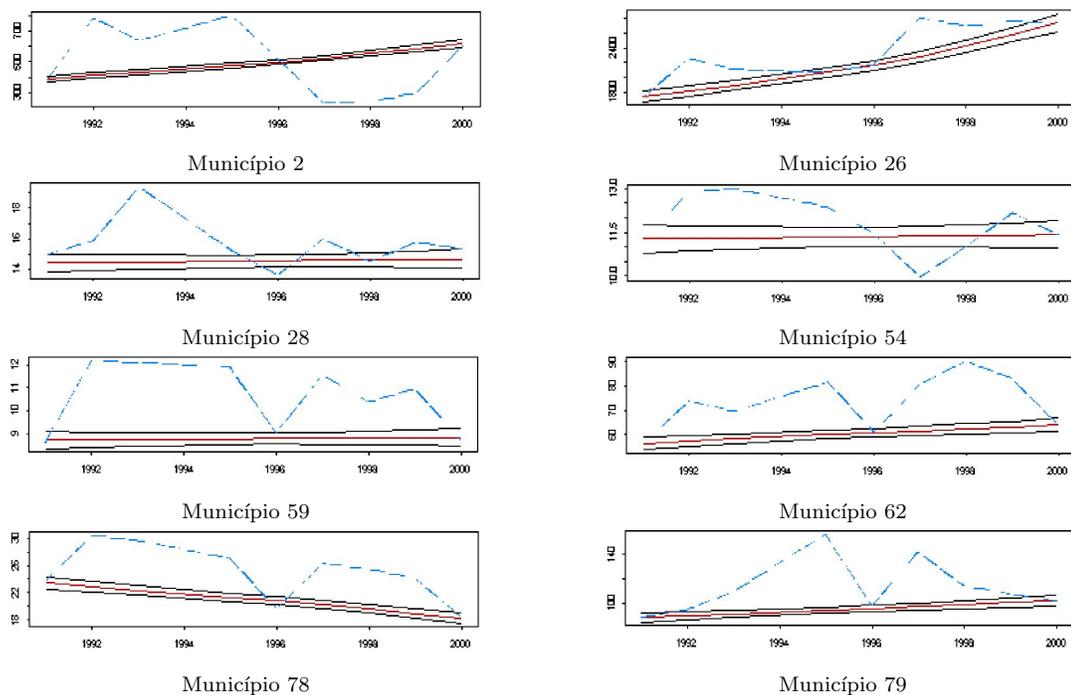


Figura 5.19: Resultados apresentados em Souza (2004), com médias a posteriori para o nível e intervalo de credibilidade em vermelho e em azul as observações $y_{i,t}$.

Portanto, podemos concluir que a abordagem aqui apresentada aproximou bem as estimas oficiais do IBGE em alguns municípios e foi mais flexível que o ajuste em Souza (2004).

Capítulo 6

Conclusões e Trabalhos Futuros

Neste trabalho, realizamos uma análise espaço-temporal da densidade populacional de alguns municípios do estado de São Paulo durante o período de 1991 à 2000. Adotando uma abordagem Bayesiana, utilizamos conjuntamente o Censo Demográfico Brasileiro, a Contagem Populacional e a Pesquisa Nacional por Amostra de Domicílios para obter a distribuição a posteriori das densidades populacionais municipais e, conseqüentemente, produzir estimativas destas densidades apresentando o erro associado. Para tratar a evolução temporal, optamos pelo uso de Modelos Dinâmicos de Crescimento Exponencial e, para considerar de modo diferenciado as precisões das informações, propomos uma modelagem que filtra a quantidade de informação de um estimador a partir da estimativa de sua variância.

Para promover o decaimento de informação ao longo do tempo, foram testadas duas estratégias: $W_{i,t} = \tau^2$ e fatores de desconto. A estratégia $W_{i,t} = \tau^2$ foi invalidada por apresentar um decaimento de informação diferenciado, que depende do tamanho populacional do município. Já a estratégia de descontos se mostrou válida. Com esta estratégia, a perda de informação sobre o nível ficou vinculada à informação no instante anterior e, com isto, o decaimento de informação ao longo do tempo passou a depender adequadamente da precisão dos dados de cada município.

Com o objetivo de verificar se a evolução da densidade populacional de um determinado município sofre influência dos demais, foram testadas duas abordagens: uma hierárquica, que assume uma evolução média comum para todos os municípios, e outra

hierárquica espacial, que assume evolução similar para municípios mais próximos.

Na análise dos ajustes dos modelos foi possível verificar que para fatores de desconto mais baixos, o nível tende a seguir mais as observações da PNAD e, conseqüentemente, o ajuste dos dados foi melhor. Já para fatores de desconto maiores, a precisão das observações censitárias se propaga ao longo do tempo, diminuindo a influência das observações intercensitárias. No geral, os modelos ajustados permitiram concluir que a modelagem das estimativas das variâncias cumpriu a função esperada: informar sobre a precisão das estimativas diretas. Concluimos também que a metodologia apresentada aproximou bem as estimativas oficiais do IBGE e foi mais flexível que o ajuste em Souza (2004).

Na análise de resultados dos critérios DIC e EPD, o Modelo Espacial com fator 0,50 se destacou por apresentar, segundo o EPD, um dos melhores resultados e ainda o melhor ajuste.

Diante de informações muito imprecisas ou de pesquisas com diferentes precisões, é extremamente comum inutilizarmos parte da informação disponível. Porém, com este trabalho, chegamos à conclusão de que o uso conjunto de diferentes fontes de informação é possível e válido. Mostramos que é possível obter estimativas intercensitárias atualizadas a partir de informações para pequenos domínios e ainda identificar erros sistemáticos, como subcobertura ou sobrecobertura.

Vale citar ainda que na pesquisa realizada durante a elaboração deste trabalho, não encontramos nenhuma referência cuja proposta seja modelar conjuntamente estimativas diretas e estimativas da variância. Neste sentido, consideramos este trabalho inovador e esperamos que ele incentive outras pesquisas nesta área. Passemos agora às nossas limitações e perspectivas futuras.

Para pesquisas censitárias onde há forte indícios de subcobertura, pode-se considerar o uso de uma distribuição assimétrica para $y_{i,t}$ ao invés da distribuição simétrica utilizada. Sugerimos ainda que extensões deste trabalho podem ser realizadas pela inclusão de covariáveis, ou variáveis chamadas sintomáticas, que ajudem a explicar a densidade populacional. Essas variáveis podem ser obtidas, por exemplo, a partir de censos, tais como o demográfico e o escolar, registros administrativos, como o Registro Civil, e pes-

quisas por amostragem, como a própria PNAD.

Outras questões a serem investigadas futuramente são: a possibilidade e vantagens de se acrescentar uma estrutura espacial no parâmetro γ_i ; como obter estimativas para os municípios criados dentro do período em estudo; como utilizar informações sobre migração, disponíveis nos questionários das pesquisas, e como compatibilizar as diferentes datas de referência das pesquisas.

Apêndice A

Estimadores da Média e Variância

Para obter estimativas da população dos municípios selecionados pela PNAD nos períodos intercensitários, foi realizada a expansão da amostra da PNAD utilizando um estimador derivado diretamente do plano amostral da pesquisa. O estimador utilizado é apresentado em Klein & Moura (1998), usado em Souza (2004) e será descrito a seguir.

Sejam

- $\hat{Y}_{(d)}$ o estimador do total da população do domínio geográfico d ,
- H_d número de estratos no domínio geográfico d ,
- l_h número de municípios selecionados no h -ésimo estrato,
- P_{hi} probabilidade de seleção do i -ésimo município do h -ésimo estrato,
- m_{hi} número de setores selecionados no i -ésimo município do h -ésimo estrato,
- P_{hij} probabilidade de seleção do j -ésimo setor do i -ésimo município do h -ésimo estrato,
- N_{hij} número de domicílios particulares no j -ésimo setor do i -ésimo município do h -ésimo estrato,
- n_{hij} número de domicílios entrevistados no j -ésimo setor do i -ésimo município do h -ésimo estrato,

- y_{hijk} total de moradores no k -ésimo domicílio entrevistado do j -ésimo setor do i -ésimo município do h -ésimo estrato.

O total populacional do domínio d é estimado por

$$\hat{Y}_{(d)} = \sum_{h=1}^{H(d)} \frac{1}{l_h} \sum_{i=1}^{l_h} \frac{1}{m_{hi} P_{hi}} \sum_{j=1}^{m_{hi}} \frac{N_{hij}}{n_{hij} P_{hij}} \sum_{k=1}^{n_{hij}} y_{hijk}. \quad (1.1)$$

A estimativa $S_{(d)}$ da variância de $\hat{Y}_{(d)}$ pode ser obtida por

$$S_{(d)} = S_1 + S_2, \quad \text{onde}$$

$$S_1 = \sum_{h \in \tilde{A}_d} \frac{1}{l_h(l_h - 1)} \sum_{i=1}^{l_h} \left(\frac{\hat{Y}_{hi}}{P_{hi}} - \hat{Y}_h \right)^2 \quad \text{e}$$

$$S_2 = \sum_{h,i \in A_d} \frac{1}{m_{hi}(m_{hi} - 1)} \sum_{j=1}^{m_{hi}} \left(\frac{\hat{Y}_{hij}}{P_{hij}} - \hat{Y}_{hi} \right)^2$$

com A_d municípios auto-representativos pertencentes ao domínio de interesse d , \tilde{A}_d estratos formados por municípios não auto-representativos pertencentes ao domínio de interesse d ,

$$\hat{Y}_h = \frac{1}{l_h} \sum_{i=1}^{l_h} \frac{\hat{Y}_{hi}}{P_{hi}}, \quad \hat{Y}_{hi} = \frac{1}{m_{hi}} \sum_{i=1}^{m_{hi}} \frac{\hat{Y}_{hij}}{P_{hij}} \quad \text{e} \quad \hat{Y}_{hij} = \frac{N_{hij}}{n_{hij}} \sum_{k=1}^{n_{hij}} y_{hijk}.$$

Como as estimativas foram obtidas para municípios, foram desconsideradas as parcelas e as quantidades do estimador de Klein & Moura (1998), dado em (1.1), relativas a estrato, pois só poderiam ser consideradas se a expansão tivesse sido realizada para um nível geográfico superior ao de município. O mesmo foi feito em relação à variância. Então foram utilizados os seguintes estimadores para obter o tamanho da população e a variância municipais:

$$\hat{Y} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} \frac{N_{hij}}{n_{hij} P_{hij}} \sum_{k=1}^{n_{hij}} y_{hijk} \quad \text{e} \quad (1.2)$$

$$S = \frac{1}{m_{hi}(m_{hi} - 1)} \sum_{j=1}^{m_{hi}} \left(\frac{\hat{Y}_{hij}}{P_{hij}} - \hat{Y}_{hi} \right)^2 \quad (1.3)$$

onde os índices h e i identificam um município.

Quando o município é auto-representativo ou metropolitano, o estimador do tamanho populacional dado em (1.2) é o mesmo que aquele em (1.1), considerando como domínio geográfico o próprio município.

Apêndice B

Taxas de Aceitação

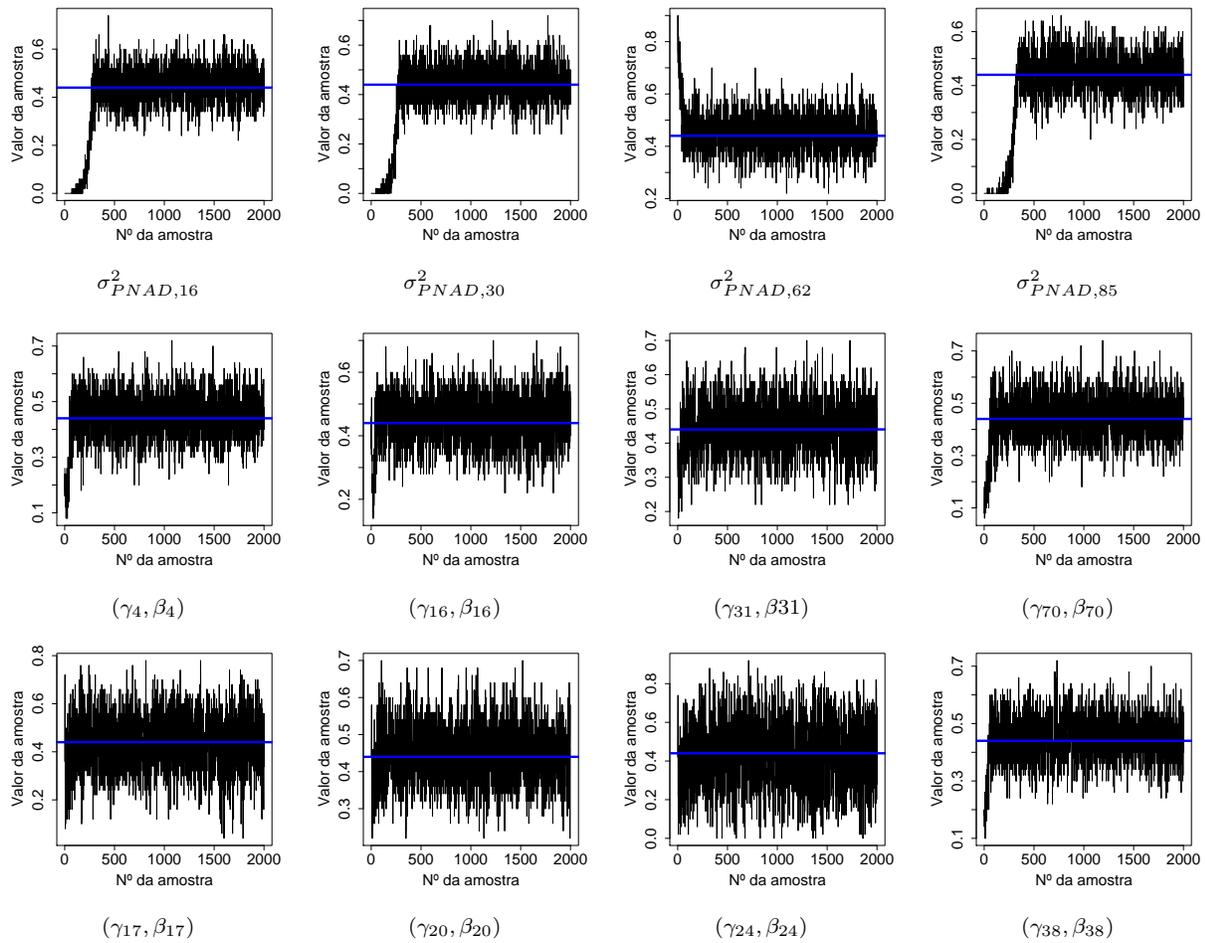


Figura 2.1: Taxas de aceitação obtidas a cada 50 ciclos do Amostrador de Gibbs sob o Modelo Espacial com fator de desconto de 0,90.

Apêndice C

Cálculo das Condicionais Completas

Neste anexo, são apresentados os cálculos das distribuições condicionais completas dos modelos ajustados. Nos cálculos a seguir, I é a matriz identidade 97×97 e $\vec{1}$ é o vetor unitário 97×1 .

Condicionais Completas do Modelo Hierárquico:

- $$\begin{aligned}
 & \bullet p(\mu_\gamma | \gamma, \sigma_\gamma^2) \propto p(\gamma | \mu_\gamma, \sigma_\gamma^2) p(\mu_\gamma) \propto \\
 & \propto \exp \left\{ -\frac{1}{2} (\gamma - \vec{1} \mu_\gamma)' \sigma_\gamma^{-2} I (\gamma - \vec{1} \mu_\gamma) \right\} \\
 & \times \exp \left(-\frac{1}{2} (v_\gamma^2)^{-1} (\mu_\gamma - m_\gamma)^2 \right) \\
 & = \exp \left\{ -\frac{1}{2} (\gamma' \sigma_\gamma^{-2} I \gamma - \gamma' \sigma_\gamma^{-2} I \vec{1} \mu_\gamma - \mu_\gamma \vec{1}' \sigma_\gamma^{-2} I \gamma + \mu_\gamma \vec{1}' \sigma_\gamma^{-2} I \vec{1} \mu_\gamma) \right\} \\
 & \times \exp \left\{ -\frac{1}{2} \left(\mu_\gamma v_\gamma^{-2} \mu_\gamma - \mu_\gamma v_\gamma^{-2} m_\gamma - m_\gamma v_\gamma^{-2} \mu_\gamma + \underbrace{m_\gamma v_\gamma^{-2} m_\gamma}_{\text{vai para a constante de proporcionalidade}} \right) \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} \left(\mu_\gamma \underbrace{(\vec{1}' \sigma_\gamma^{-2} I \vec{1} + v_\gamma^{-2})}_{V^{-1}} \mu_\gamma - \mu_\gamma V^{-1} V \underbrace{(\vec{1}' \sigma_\gamma^{-2} I \gamma + v_\gamma^{-2} m_\gamma)}_M \right) \right\} \\
 & \times \exp \left\{ -\frac{1}{2} \left(-V \underbrace{(\vec{1}' \sigma_\gamma^{-2} I \gamma + v_\gamma^{-2} m_\gamma)}_M V^{-1} \mu_\gamma + \underbrace{MV^{-1}M}_{\text{vem da constante de proporcionalidade}} \right) \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} V^{-1} (\mu_\gamma - M)^2 \right\}
 \end{aligned}$$

Logo $(\mu_\gamma | \gamma, \sigma_\gamma^2) \sim N(M, V)$, com

$$V = \left(\vec{1}' \sigma_\gamma^{-2} I \vec{1} + v_\gamma^{-2} \right)^{-1} \quad \text{e} \quad M = V \left(\vec{1}' \sigma_\gamma^{-2} I \gamma + v_\gamma^{-2} m_\gamma \right).$$

- $$\begin{aligned}
 & p(\sigma_\gamma^2 | \gamma, \mu_\gamma) \propto p(\gamma | \mu_\gamma, \sigma_\gamma^2) p((\sigma_\gamma^2)^{-1}) \propto \\
 & \propto |\sigma_\gamma^2 I|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \underbrace{(\gamma - \vec{1} \mu_\gamma)' I (\gamma - \vec{1} \mu_\gamma)}_K (\sigma_\gamma^2)^{-1} \right\} \left((\sigma_\gamma^2)^{-1} \right)^{a_\gamma - 1} \exp \left\{ -b_\gamma (\sigma_\gamma^2)^{-1} \right\} \\
 & = \left((\sigma_\gamma^2)^{-1} \right)^{\frac{97}{2}} \exp \left\{ -\frac{1}{2} K (\sigma_\gamma^2)^{-1} \right\} \left((\sigma_\gamma^2)^{-1} \right)^{a_\gamma - 1} \exp \left\{ -b_\gamma (\sigma_\gamma^2)^{-1} \right\} \\
 & = \left((\sigma_\gamma^2)^{-1} \right)^{\frac{97}{2} + a_\gamma - 1} \exp \left\{ - \underbrace{\left[b_\gamma + \left(\frac{1}{2} K \right) \right]}_B (\sigma_\gamma^2)^{-1} \right\}.
 \end{aligned}$$

Logo $(\sigma_\gamma^2 | \gamma, \mu_\gamma) \sim GI(A, B)$, com

$$\begin{aligned}
 A &= \frac{97}{2} + a_\gamma, \\
 B &= b_\gamma + \left(\frac{1}{2} K \right) \quad \text{e} \\
 K &= (\gamma - \vec{1} \mu_\gamma)' I (\gamma - \vec{1} \mu_\gamma).
 \end{aligned}$$

- $$\begin{aligned}
 & p(\mu_\beta | \beta, \sigma_\beta^2) \propto p(\beta | \mu_\beta, \sigma_\beta^2) p(\mu_\beta) \propto \\
 & \propto \exp \left\{ -\frac{1}{2} (\beta - \vec{1} \mu_\beta)' \sigma_\beta^{-2} I (\beta - \vec{1} \mu_\beta) \right\} \\
 & \times \exp \left(-\frac{1}{2} (v_\beta^2)^{-1} (\mu_\beta - m_\beta)^2 \right) \\
 & = \exp \left\{ -\frac{1}{2} \left(\beta' \sigma_\beta^{-2} I \beta - \beta' \sigma_\beta^{-2} I \vec{1} \mu_\beta - \mu_\beta \vec{1}' \sigma_\beta^{-2} I \beta + \mu_\beta \vec{1}' \sigma_\beta^{-2} I \vec{1} \mu_\beta \right) \right\} \\
 & \times \exp \left\{ -\frac{1}{2} \left(\begin{array}{l} \mu_\beta v_\beta^{-2} \mu_\beta - \mu_\beta v_\beta^{-2} m_\beta - m_\beta v_\beta^{-2} \mu_\beta + \underbrace{m_\beta v_\beta^{-2} m_\beta}_{\text{vai para a constante de proporcionalidade}} \end{array} \right) \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} \left(\underbrace{\mu_\beta (\vec{1}' \sigma_\beta^{-2} I \vec{1} + v_\beta^{-2})}_{V^{-1}} \mu_\beta - \mu_\beta V^{-1} V \underbrace{(\vec{1}' \sigma_\beta^{-2} I \beta + v_\beta^{-2} m_\beta)}_M \right) \right\} \\
 & \times \exp \left\{ -\frac{1}{2} \left(\underbrace{-V (\vec{1}' \sigma_\beta^{-2} I \beta + v_\beta^{-2} m_\beta)}_M V^{-1} \mu_\beta + \underbrace{MV^{-1} M}_{\text{vem da constante de proporcionalidade}} \right) \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} V^{-1} (\mu_\beta - M)^2 \right\}
 \end{aligned}$$

Logo $(\mu_\beta | \boldsymbol{\beta}, \sigma_\beta^2) \sim N(M, V)$, com

$$V = \left(\vec{1}' \sigma_\beta^{-2} I \vec{1} + v_\beta^{-2} \right)^{-1} \quad \text{e} \quad M = V \left(\vec{1}' \sigma_\beta^{-2} I \boldsymbol{\beta} + v_\beta^{-2} m_\beta \right).$$

- $$\begin{aligned}
 & p(\sigma_\beta^2 | \boldsymbol{\beta}, \mu_\beta) \propto p(\boldsymbol{\beta} | \mu_\beta, \sigma_\beta^2) p\left((\sigma_\beta^2)^{-1}\right) \propto \\
 & \propto |\sigma_\beta^2 I|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \underbrace{(\boldsymbol{\beta} - \vec{1} \mu_\beta)' I (\boldsymbol{\beta} - \vec{1} \mu_\beta)}_K (\sigma_\beta^2)^{-1} \right\} \left((\sigma_\beta^2)^{-1} \right)^{a_\beta - 1} \exp \left\{ -b_\beta (\sigma_\beta^2)^{-1} \right\} \\
 & = \left((\sigma_\beta^2)^{-1} \right)^{\frac{97}{2}} \exp \left\{ -\frac{1}{2} K (\sigma_\beta^2)^{-1} \right\} \left((\sigma_\beta^2)^{-1} \right)^{a_\beta - 1} \exp \left\{ -b_\beta (\sigma_\beta^2)^{-1} \right\} \\
 & = \left((\sigma_\beta^2)^{-1} \right)^{\frac{97}{2} + a_\beta - 1} \exp \left\{ -\underbrace{\left[b_\beta + \left(\frac{1}{2} K \right) \right]}_B (\sigma_\beta^2)^{-1} \right\}.
 \end{aligned}$$

Logo $(\sigma_\beta^2 | \boldsymbol{\beta}, \mu_\beta) \sim GI(A, B)$, com

$$\begin{aligned}
 A &= \frac{97}{2} + a_\beta, \\
 B &= b_\beta + \left(\frac{1}{2} K \right) \quad \text{e} \\
 K &= (\boldsymbol{\beta} - \vec{1} \mu_\beta)' I (\boldsymbol{\beta} - \vec{1} \mu_\beta).
 \end{aligned}$$

- $$\begin{aligned}
 & p(\eta | \mathbf{s}, \boldsymbol{\sigma}_{PNAD}^2, \eta) \propto \left[\prod_{t \in P} \prod_{i=1}^{91} p(s_{i,t} | \sigma_{PNAD,i}^2, \eta) \right] p(\eta) \\
 & \propto \left[\prod_{t \in P} \prod_{i=1}^{91} \frac{b_s^{a_s}}{\Gamma(a_s)} s_{i,t}^{a_s - 1} \exp \{-b_s s_{i,t}\} \right] \eta^{a_\eta - 1} \exp \{-b_\eta \eta\}, \quad \text{com} \\
 & \quad a_s = \eta^{-1} \quad \text{e} \quad b_s = \eta^{-1} (\sigma_{PNAD,i}^2)^{-1}.
 \end{aligned}$$

A condicional completa de η não tem forma fechada.

- $$\begin{aligned}
 & p(\sigma_{PNAD,i}^2 | \mathbf{y}, \mathbf{s}, \boldsymbol{\mu}, \eta) \propto \\
 & \propto \left[\prod_{t \in P} p(y_{i,t} | \mu_{i,t}, \sigma_{PNAD,i}^2) \right] \left[\prod_{t \in P} p(s_{i,t} | \sigma_{PNAD,i}^2, \eta) \right] p(\sigma_{PNAD,i}^2) \\
 & \propto \left[\prod_{t \in P} (\sigma_{PNAD,i}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\sigma_{PNAD,i}^2)^{-1} (y_{i,t} - \mu_{i,t})^2 \right\} \right] \\
 & \times \left[\prod_{t \in P} \frac{b_s^{a_s}}{\Gamma(a_s)} s_{i,t}^{a_s - 1} \exp \{-b_s s_{i,t}\} \right] \cdot (\sigma_{PNAD,i}^2)^{a_\sigma - 1} \exp \{-b_\sigma \sigma_{PNAD,i}^2\}
 \end{aligned}$$

$$= \left[(\sigma_{PNAD,i}^2)^{-\frac{6}{2}} \prod_{t \in P} \exp \left\{ -\frac{1}{2} (\sigma_{PNAD,i}^2)^{-1} (y_{i,t} - \mu_{i,t})^2 \right\} \right] \\ \times \left[\frac{b_s^{6a_s}}{(\Gamma(a_s))^6} \prod_{t \in P} s_{i,t}^{a_s-1} \exp \{ -b_s s_{i,t} \} \right] \cdot (\sigma_{PNAD,i}^2)^{a_s-1} \exp \{ -b_s \sigma_{PNAD,i}^2 \}, \text{ com}$$

$$a_s = \eta^{-1} \text{ e } b_s = \eta^{-1} (\sigma_{PNAD,i}^2)^{-1}.$$

A condicional completa de $\sigma_{PNAD,i}^2$ não tem forma fechada.

$$\bullet p(\gamma_i | \mu, \beta_i, \mu_\gamma, \sigma_\gamma^2) \propto \left(\prod_{t=1}^9 p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \right) p(\gamma_i | \mu_\gamma, \sigma_\gamma^2) \propto \\ \propto \prod_{t=1}^9 \exp \left\{ -\frac{1}{2} \underbrace{W_{i,t}^{-1} (\mu_{i,t} - \mu_{i,t-1} - \beta_i \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1))^2}_{K_t} \right\} \\ \times \exp \left(-\frac{1}{2} \sigma_\gamma^{-2} (\gamma_i - \mu_\gamma)^2 \right) = \exp \left\{ -\frac{1}{2} \left(\sum_{t=1}^9 K_t + \sigma_\gamma^{-2} (\gamma_i - \mu_\gamma)^2 \right) \right\}, \text{ com}$$

$$K_t = W_{i,t}^{-1} (\mu_{i,t} - \mu_{i,t-1} - \beta_i \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1))^2$$

A condicional completa de γ_i não tem forma fechada.

$$\bullet p(\beta_i | \gamma_i, \mu_\beta, \sigma_\beta^2) \propto p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) p(\beta_i | \mu_\beta, \sigma_\beta^2) \\ \propto \prod_{t=1}^9 \exp \left\{ -\frac{1}{2} W_{i,t}^{-1} (\mu_{i,t} - \mu_{i,t-1} - \beta_i \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1))^2 \right\} \\ \times \exp \left(-\frac{1}{2} (\sigma_\beta^2)^{-1} (\beta_i - \mu_\beta)^2 \right) \\ = (\text{fazendo } A_t = \mu_{i,t} - \mu_{i,t-1} \text{ e } B_t = \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1)) = \\ = \left(\prod_{t=1}^9 \exp \left\{ -\frac{1}{2} W_{i,t}^{-1} (A_t - B_t \beta_i)^2 \right\} \right) \exp \left(-\frac{1}{2} (\sigma_\beta^2)^{-1} (\beta_i - \mu_\beta)^2 \right) \\ = \prod_{t=1}^9 \exp \left\{ -\frac{1}{2} (A_t W_{i,t}^{-1} A_t - A_t W_{i,t}^{-1} B_t \beta_i - \beta_i B_t W_{i,t}^{-1} A_t + \beta_i B_t W_{i,t}^{-1} B_t \beta_i) \right\} \\ \times \exp \left\{ -\frac{1}{2} \left(\beta_i (\sigma_\beta^2)^{-1} \beta_i - \beta_i (\sigma_\beta^2)^{-1} \mu_\beta - \mu_\beta (\sigma_\beta^2)^{-1} \beta_i + \underbrace{\mu_\beta (\sigma_\beta^2)^{-1} \mu_\beta}_{\text{vai para a constante de proporcionalidade}} \right) \right\} \\ \propto \exp \left\{ -\frac{1}{2} \left(\beta_i \left(\underbrace{\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right)}_{V^{-1}} + (\sigma_\beta^2)^{-1} \right) \beta_i \right) \right\}.$$

$$\begin{aligned}
& \times \exp \left\{ -\frac{1}{2} \left(-\beta_i V^{-1} V \left(\underbrace{\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right)}_M + (\sigma_\beta^2)^{-1} \mu_\beta \right) \right) \right\} \\
& \times \exp \left\{ -\frac{1}{2} \left(-V \left(\underbrace{\left(\sum_{t=1}^9 A_t W_{i,t}^{-1} B_t \right)}_M + \mu_\beta (\sigma_\beta^2)^{-1} \right) V^{-1} \beta_i + \underbrace{MV^{-1}M}_{\text{vem da constante de proporcionalidade}} \right) \right\} = \\
& \times \exp \left(-\frac{1}{2} (\beta_i V^{-1} \beta_i - \beta_i V^{-1} M - MV^{-1} \beta_i + MV^{-1} M) \right) \\
& = \exp \left\{ -\frac{1}{2} V^{-1} (\beta_i - M)^2 \right\}.
\end{aligned}$$

Logo $(\beta_i | \gamma_i, \mu_\beta, \sigma_\beta^2) \sim N(M, V)$, com

$$\begin{aligned}
V &= \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right) + (\sigma_\beta^2)^{-1} \right)^{-1}, \\
M &= V \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right) + (\sigma_\beta^2)^{-1} \mu_\beta \right), \\
A_t &= \mu_{i,t} - \mu_{i,t-1} \text{ e} \\
B_t &= \exp(\gamma_i (t-1)) (\exp(\gamma_i) - 1).
\end{aligned}$$

Condicionais Completas do Modelo Espacial:

- Sejam $\vec{1}_{96}$ o vetor unitário 96×1 e $\boldsymbol{\beta}_{-i} = [\beta_1 \ \cdots \ \beta_{i-1} \ \beta_{i+1} \ \cdots \ \beta_{97}]'$. Como $(\boldsymbol{\beta} | \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \sim N(\mu_\beta, \Sigma_{\varepsilon, \phi})$, temos que

$$(\beta_i | \boldsymbol{\beta}_{-i}, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \sim N(C, D), \text{ onde}$$

- $C = \mu_\beta + \Sigma_{1,i}, \Sigma_{2,i}^{-1} (\boldsymbol{\beta}_{-i} - \mu_\beta \vec{1}_{96})$,
- $D = \varepsilon^2 - \Sigma_{1,i} \Sigma_{2,i}^{-1} \Sigma'_{1,i}$,
- $\Sigma_{1,i}$ é o vetor de covariâncias entre β_i e $\boldsymbol{\beta}_{-i}$ e
- $\Sigma_{2,i} = \text{Var}(\boldsymbol{\beta}_{-i} | \mu_\beta, \varepsilon^2, \phi, \mathbf{d})$.

Segue abaixo o cálculo da condicional completa de β_i .

$$\begin{aligned}
p(\beta_i | \boldsymbol{\beta}_{-i}, \gamma_i, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) &\propto \left(\prod_{t=1}^9 p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \right) p(\boldsymbol{\beta} | \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \\
&= \left(\prod_{t=1}^9 p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \right) p(\beta_i | \boldsymbol{\beta}_{-i}, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \underbrace{p(\boldsymbol{\beta}_{-i} | \mu_\beta, \varepsilon^2, \phi, \mathbf{d})}_{\text{vai para a constante de proporcionalidade}} \\
&\propto \left(\prod_{t=1}^9 p(\mu_{i,t} | \mu_{i,t-1}, \beta_i, \gamma_i) \right) p(\beta_i | \boldsymbol{\beta}_{-i}, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \\
&\propto \left(\prod_{t=1}^9 \exp \left\{ -\frac{1}{2} W_{i,t}^{-1} (\mu_{i,t} - \mu_{i,t-1} - \beta_i \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1))^2 \right\} \right) \\
&\times \exp \left(-\frac{1}{2} D^{-1} (\beta_i - C)^2 \right) \\
&= (\text{fazendo } A_t = \mu_{i,t} - \mu_{i,t-1} \text{ e } B_t = \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1)) = \\
&= \left(\prod_{t=1}^9 \exp \left\{ -\frac{1}{2} W_{i,t} (A_t - B_t \beta_i)^2 \right\} \right) \exp \left(-\frac{1}{2} D^{-1} (\beta_i - C)^2 \right) \\
&= \left(\prod_{t=1}^9 \exp \left\{ -\frac{1}{2} (A_t W_{i,t}^{-1} A_t - A_t W_{i,t}^{-1} B_t \beta_i - \beta_i B_t W_{i,t}^{-1} A_t + \beta_i B_t W_{i,t}^{-1} B_t \beta_i) \right\} \right) \\
&\times \exp \left\{ -\frac{1}{2} \left(\beta_i D^{-1} \beta_i - \beta_i D^{-1} C - C D^{-1} \beta_i + \underbrace{C D^{-1} C}_{\text{vai para a constante de proporcionalidade}} \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left(\beta_i \left(\underbrace{\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right)}_{V^{-1}} + D^{-1} \right) \beta_i \right) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \left(-\beta_i V^{-1} V \left(\underbrace{\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right)}_M + D^{-1} C \right) \right) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \left(- \left(\underbrace{V \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right) + D^{-1} C \right)}_M \right) V^{-1} \beta_i + \underbrace{M V^{-1} M}_{\text{vem da constante de proporcionalidade}} \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} (\beta_i V^{-1} \beta_i - \beta_i V^{-1} M - M V^{-1} \beta_i + M V^{-1} M) \right\} \\
&= \exp \left\{ -\frac{1}{2} V^{-1} (\beta_i - M)^2 \right\}.
\end{aligned}$$

Logo $(\beta_i | \beta_{-i}, \gamma_i, \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) \sim N(M, V)$, com

$$V = \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} B_t \right) + D^{-1} \right)^{-1},$$

$$M = V \left(\left(\sum_{t=1}^9 B_t W_{i,t}^{-1} A_t \right) + D^{-1} C \right),$$

$$A_t = \mu_{i,t} - \mu_{i,t-1} \text{ e}$$

$$B_t = \exp(\gamma_i(t-1)) (\exp(\gamma_i) - 1).$$

- $$p(\mu_\beta | \beta, \phi, \varepsilon^2, \mathbf{d}) \propto p(\beta | \mu_\beta, \phi, \varepsilon^2, \mathbf{d}) p(\mu_\beta) \propto$$

$$\propto \exp \left\{ -\frac{1}{2} (\beta - \bar{\mathbf{1}} \mu_\beta)' \Sigma_{\varepsilon, \phi}^{-1} (\beta - \bar{\mathbf{1}} \mu_\beta) \right\}$$

$$\times \exp \left(-\frac{1}{2} (v_\beta^2)^{-1} (\mu_\beta - m_\beta)^2 \right)$$

$$= \exp \left\{ -\frac{1}{2} (\beta' \Sigma_{\varepsilon, \phi}^{-1} \beta - \beta' \Sigma_{\varepsilon, \phi}^{-1} \bar{\mathbf{1}} \mu_\beta - \mu_\beta \bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \beta + \mu_\beta \bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \bar{\mathbf{1}} \mu_\beta) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \left(\mu_\beta v_\beta^{-2} \mu_\beta - \mu_\beta v_\beta^{-2} m_\beta - m_\beta v_\beta^{-2} \mu_\beta + \underbrace{m_\beta v_\beta^{-2} m_\beta}_{\text{vai para a constante de proporcionalidade}} \right) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left(\underbrace{\mu_\beta (\bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \bar{\mathbf{1}} + v_\beta^{-2})}_{V^{-1}} \mu_\beta - \mu_\beta V^{-1} V \underbrace{(\bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \beta + v_\beta^{-2} m_\beta)}_M \right) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \left(-V \underbrace{(\bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \beta + v_\beta^{-2} m_\beta)}_M V^{-1} \mu_\beta + \underbrace{MV^{-1}M}_{\text{vem da constante de proporcionalidade}} \right) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} V^{-1} (\mu_\beta - M)^2 \right\}$$

Logo $(\mu_\beta | \beta, \phi, \varepsilon^2, \mathbf{d}) \sim N(M, V)$, com

$$V = \left(\bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \bar{\mathbf{1}} + v_\beta^{-2} \right)^{-1} \text{ e } M = V \left(\bar{\mathbf{1}}' \Sigma_{\varepsilon, \phi}^{-1} \beta + v_\beta^{-2} m_\beta \right).$$

- $$p(\phi | \beta, \mu_\beta, \varepsilon^2, \mathbf{d}) \propto p(\beta | \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) p(\phi)$$

$$\propto |\Sigma_{\varepsilon, \phi}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\beta - \bar{\mathbf{1}} \mu_\beta)' \Sigma_{\varepsilon, \phi}^{-1} (\beta - \bar{\mathbf{1}} \mu_\beta) \right\} (\phi^{-1})^{a_\phi - 1} \exp[-b_\phi (\phi^{-1})]$$

A condicional completa de ϕ não tem forma fechada.

$$\begin{aligned}
& \bullet p(\varepsilon^2 \mid \boldsymbol{\beta}, \mu_\beta, \phi, \mathbf{d}) \propto p(\boldsymbol{\beta} \mid \mu_\beta, \varepsilon^2, \phi, \mathbf{d}) p(\varepsilon^2) \propto \\
& \propto |\Sigma_{\varepsilon, \phi}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \underbrace{(\boldsymbol{\beta} - \bar{\mathbf{1}} \mu_\beta)' \Sigma_\phi^{-1} (\boldsymbol{\beta} - \bar{\mathbf{1}} \mu_\beta)}_C \varepsilon^{-2} \right\} (\varepsilon^{-2})^{a_\varepsilon - 1} \exp[-b_\varepsilon (\varepsilon^{-2})] \\
& \propto (\varepsilon^{-2})^{\frac{m}{2}} \exp \left\{ -\frac{1}{2} C \varepsilon^{-2} \right\} (\varepsilon^{-2})^{a_\varepsilon - 1} \exp \left\{ -b_\varepsilon (\varepsilon^{-2}) \right\} \\
& = (\varepsilon^{-2})^{\underbrace{\frac{m}{2} + a_\varepsilon - 1}_A} \exp \left\{ - \underbrace{\left[b_\varepsilon + \left(\frac{1}{2} C \right) \right]}_B \varepsilon^{-2} \right\}.
\end{aligned}$$

Logo $(\varepsilon^2 \mid \boldsymbol{\beta}, \mu_\beta, \phi, \mathbf{d}) \sim GI(A, B)$, com

$$\begin{aligned}
A &= \frac{97}{2} + a_\varepsilon, \\
B &= b_\varepsilon + \frac{1}{2} C \text{ e} \\
C &= (\boldsymbol{\beta} - \bar{\mathbf{1}} \mu_\beta)' \Sigma_\phi^{-1} (\boldsymbol{\beta} - \bar{\mathbf{1}} \mu_\beta).
\end{aligned}$$

As condicionais completas de γ_i , μ_γ , σ_γ^2 , $\sigma_{PNAD,i}$ e η para o Modelo Espacial são idênticas às do Modelo Hierárquico.

Apêndice D

FFBS - Distribuição de

$$(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$$

$$\begin{aligned} p(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \beta_i, \gamma_i) &= (\text{pelo Teorema de Bayes}) = \\ &= \frac{p(y_{i,t+1}, \dots, y_{i,9} \mid D_t, \mu_{i,t}, \mu_{i,t+1}, \beta_i, \gamma_i) p(\mu_{i,t} \mid D_t, \mu_{i,t+1}, \Phi)}{p(y_{i,t+1}, \dots, y_{i,9} \mid D_t, \mu_{i,t+1}, \Phi)} \\ &= (\text{pois dado } \mu_{i,t+1}, \text{ temos que } y_{i,t+1}, \dots, y_{i,9} \text{ é independente de } \mu_{i,t}) = \\ &= \frac{p(y_{i,t+1}, \dots, y_{i,9} \mid D_t, \mu_{i,t+1}, \Phi) p(\mu_{i,t} \mid D_t, \mu_{i,t+1}, \Phi)}{p(y_{i,t+1}, \dots, y_{i,9} \mid D_t, \mu_{i,t+1}, \Phi)} \\ &= p(\mu_{i,t} \mid D_t, \mu_{i,t+1}, \Phi) \propto (\text{pelo Teorema de Bayes}) \propto \\ &\propto p(\mu_{i,t+1} \mid D_t, \mu_{i,t}, \Phi) p(\mu_{i,t} \mid D_t, \Phi) \\ &= p(\mu_{i,t+1} \mid \mu_{i,t}, \Phi) p(\mu_{i,t} \mid D_t, \Phi) \propto \\ &\propto \exp \left\{ -\frac{1}{2} W_{i,t+1}^{-1} (\mu_{i,t+1} - \mu_{i,t} - \beta_i \exp(\gamma_i t) (\exp(\gamma_i) - 1))^2 \right\} \\ &\times \exp \left\{ -\frac{1}{2} C_{i,t}^{-1} (\mu_{i,t} - m_{i,t})^2 \right\} \\ &= (\text{fazendo } U_{i,t} = \mu_{i,t+1} - \beta_i \exp(\gamma_i t) (\exp(\gamma_i) - 1) \text{ para melhorar a notação}) = \\ &= \exp \left\{ -\frac{1}{2} W_{i,t+1}^{-1} (U_{i,t} - \mu_{i,t})^2 \right\} \exp \left\{ -\frac{1}{2} C_{i,t}^{-1} (\mu_{i,t} - m_{i,t})^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} (U_{i,t} W_{i,t+1}^{-1} U_{i,t} - U_{i,t} W_{i,t+1}^{-1} \mu_{i,t} - \mu_{i,t} W_{i,t+1}^{-1} U_{i,t} + \mu_{i,t} W_{i,t+1}^{-1} \mu_{i,t}) \right\} \end{aligned}$$

$$\begin{aligned}
& \times \exp \left\{ -\frac{1}{2} (\mu_{i,t} C_{i,t}^{-1} \mu_{i,t} - \mu_{i,t} C_{i,t}^{-1} m_{i,t} - m_{i,t} C_{i,t}^{-1} \mu_{i,t} + m_{i,t} C_{i,t}^{-1} m_{i,t}) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left(\underbrace{\mu_{i,t} (W_{i,t+1}^{-1} + C_{i,t}^{-1})}_{V_{i,t}^{-1}} \mu_{i,t} - \mu_{i,t} V_{i,t}^{-1} \underbrace{V_{i,t} (W_{i,t+1}^{-1} U_{i,t} + C_{i,t}^{-1} m_{i,t})}_{M_{i,t}} \right) \right\} \\
& \times \exp \left\{ -\frac{1}{2} \left(- \underbrace{[V_{i,t} (W_{i,t+1}^{-1} U_{i,t} + C_{i,t}^{-1} m_{i,t})]}_{M_{i,t}} V_{i,t}^{-1} \mu_{i,t} + \underbrace{M_{i,t} V_{i,t}^{-1} M_{i,t}}_{\text{vem da constante de proporcionalidade}} \right) \right\} \\
& = \exp \left\{ -\frac{1}{2} V_{i,t}^{-1} (\mu_{i,t} - M_{i,t})^2 \right\}.
\end{aligned}$$

Logo,

$$\begin{aligned}
(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi) & \sim N(M_{i,t}, V_{i,t}), \quad \text{onde} \\
V_{i,t} & = (W_{i,t+1}^{-1} + C_{i,t}^{-1})^{-1}, \\
M_{i,t} & = V_{i,t} (W_{i,t+1}^{-1} U_{i,t} + C_{i,t}^{-1} m_{i,t}) \quad \text{e} \\
U_{i,t} & = \mu_{i,t+1} - \beta_i \exp(\gamma_i t) (\exp(\gamma_i) - 1).
\end{aligned}$$

Como podemos notar, a distribuição $p(\mu_{i,t} \mid D_9, \mu_{i,t+1}, \Phi)$ não depende diretamente das observações $y_{i,t}$, uma vez que estas observações já foram incorporadas por $p(\mu_{i,t} \mid D_t, \Phi)$. Portanto, para os municípios p selecionados para previsão, a geração de $\mu_{p,t}$, para quaisquer tempos, pode ser feita diretamente da distribuição que acabamos de calcular acima.

Apêndice E

Autocorrelações e Histogramas

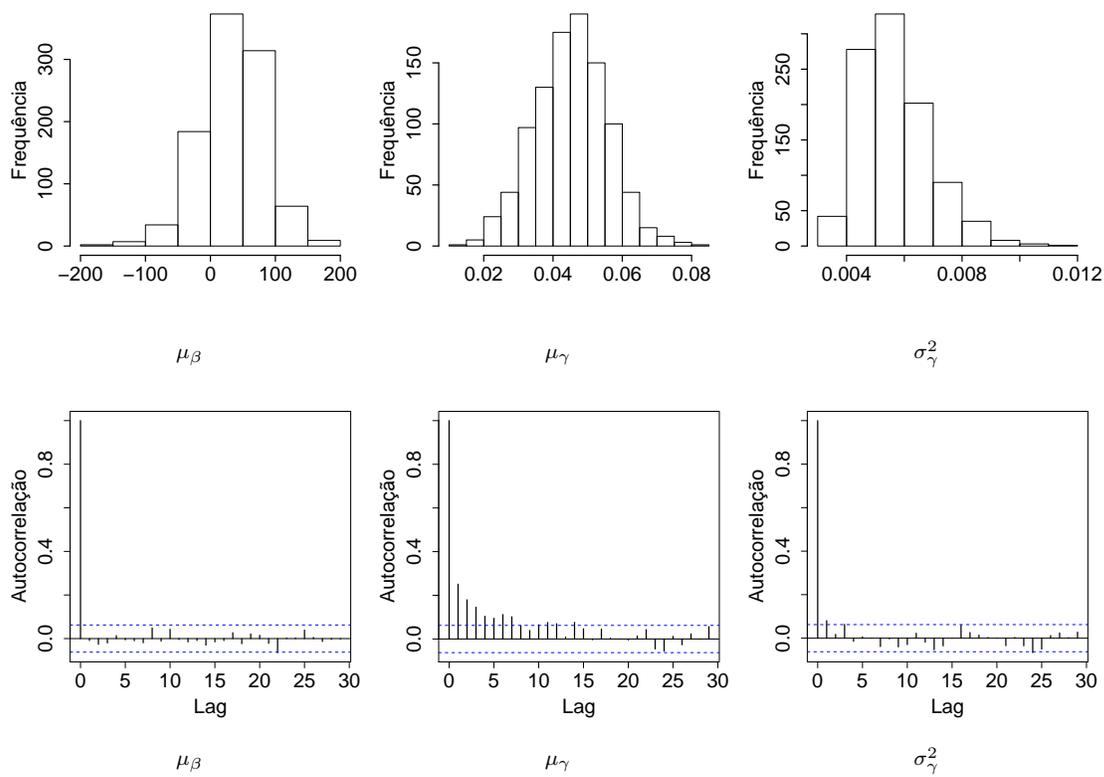


Figura 5.1: Histogramas e funções de autocorrelação de algumas cadeias a posteriori para o Modelo Espacial com fator de desconto 0,90, *burn-in* 2000 e *thin* 150.

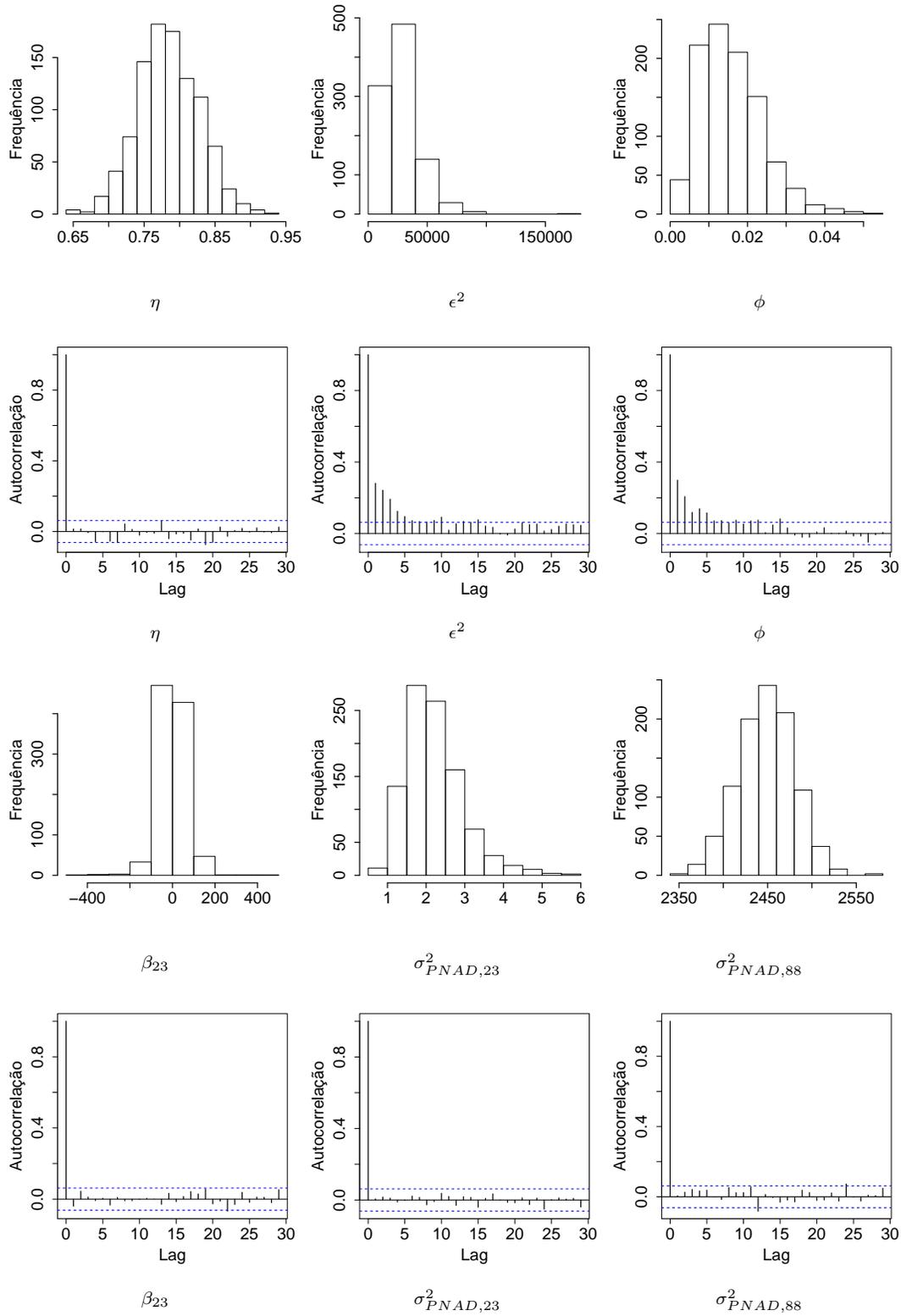


Figura 5.2: Histogramas e funções de autocorrelação de algumas cadeias a posteriori para o Modelo Espacial com fator de desconto 0,90, *burn-in* 2000 e *thin* 150.

Apêndice F

Ajustes com fator 0,50

Ajustes do Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais publicadas pelo IBGE em verde.

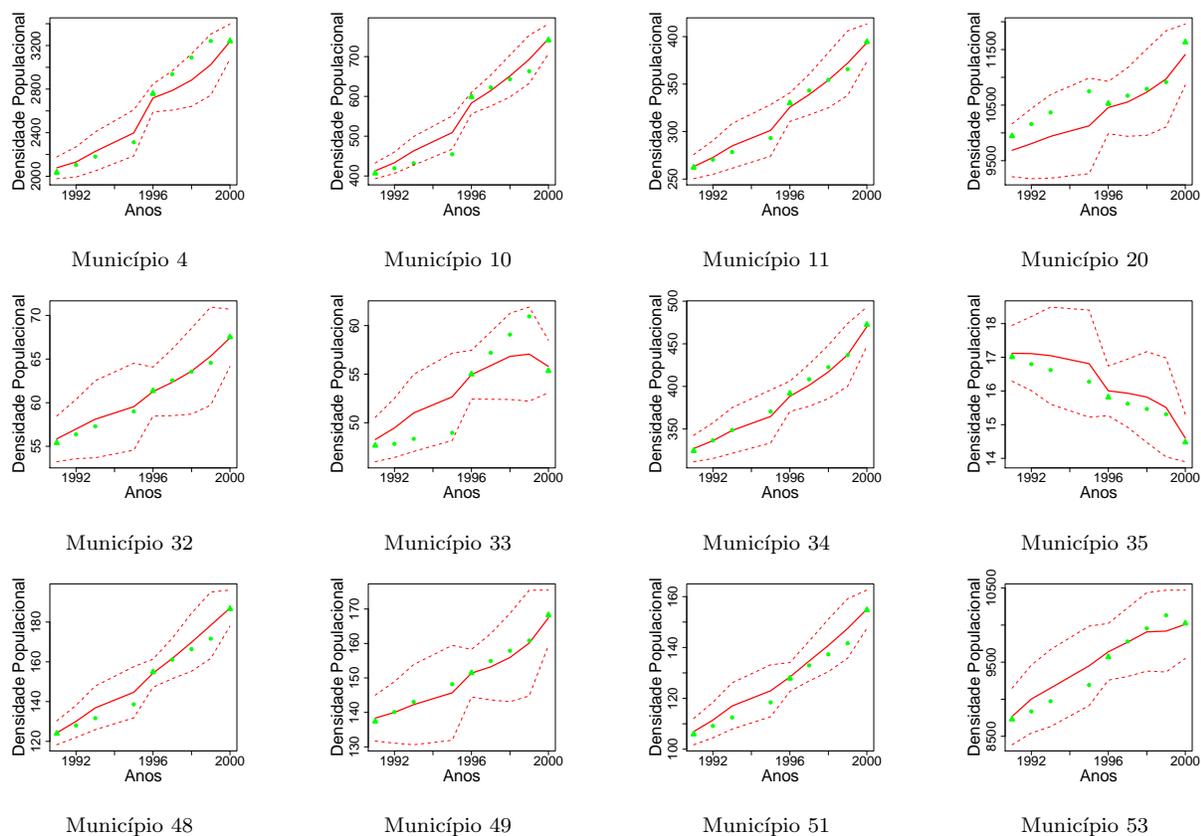


Figura 6.1: Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.

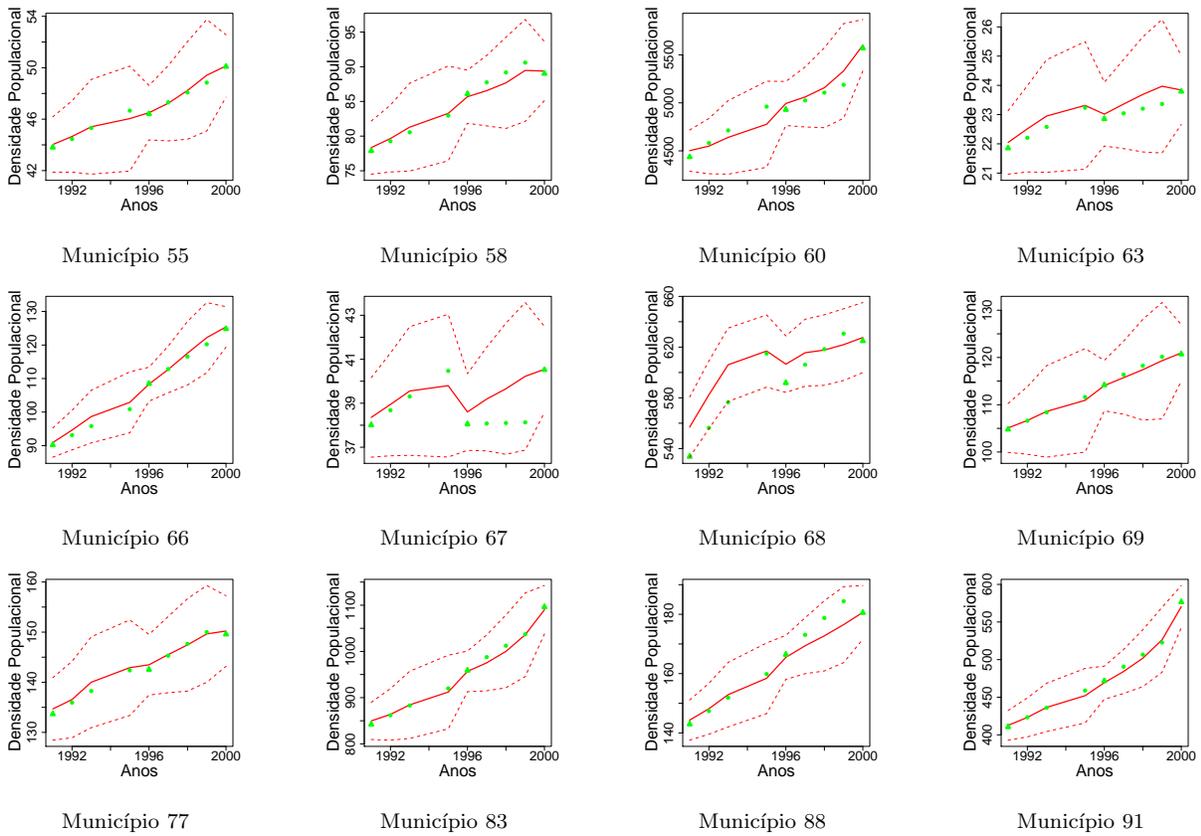


Figura 6.2: Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.

Apêndice G

Ajustes com fator 0,80

Ajustes do Modelo Espacial com fator 0,80 em vermelho e estimativas oficiais publicadas pelo IBGE em verde.

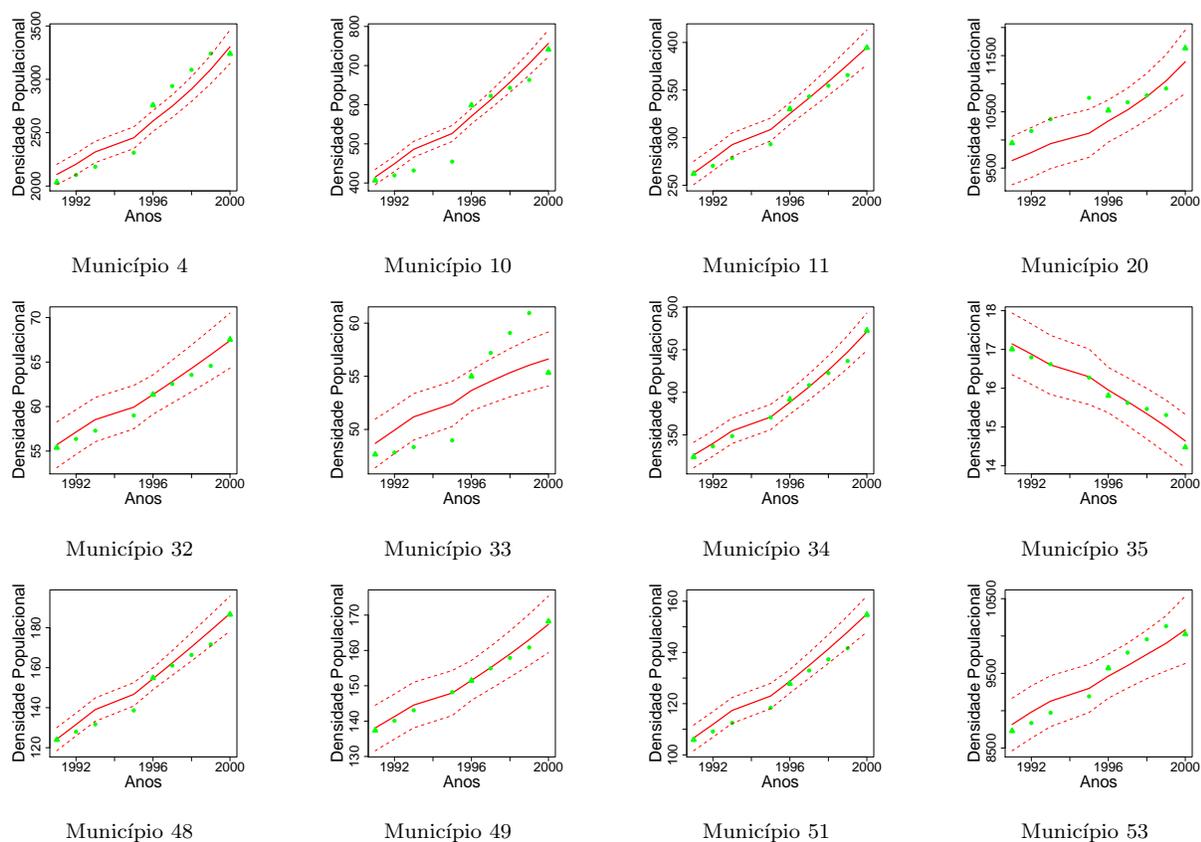


Figura 7.1: Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.

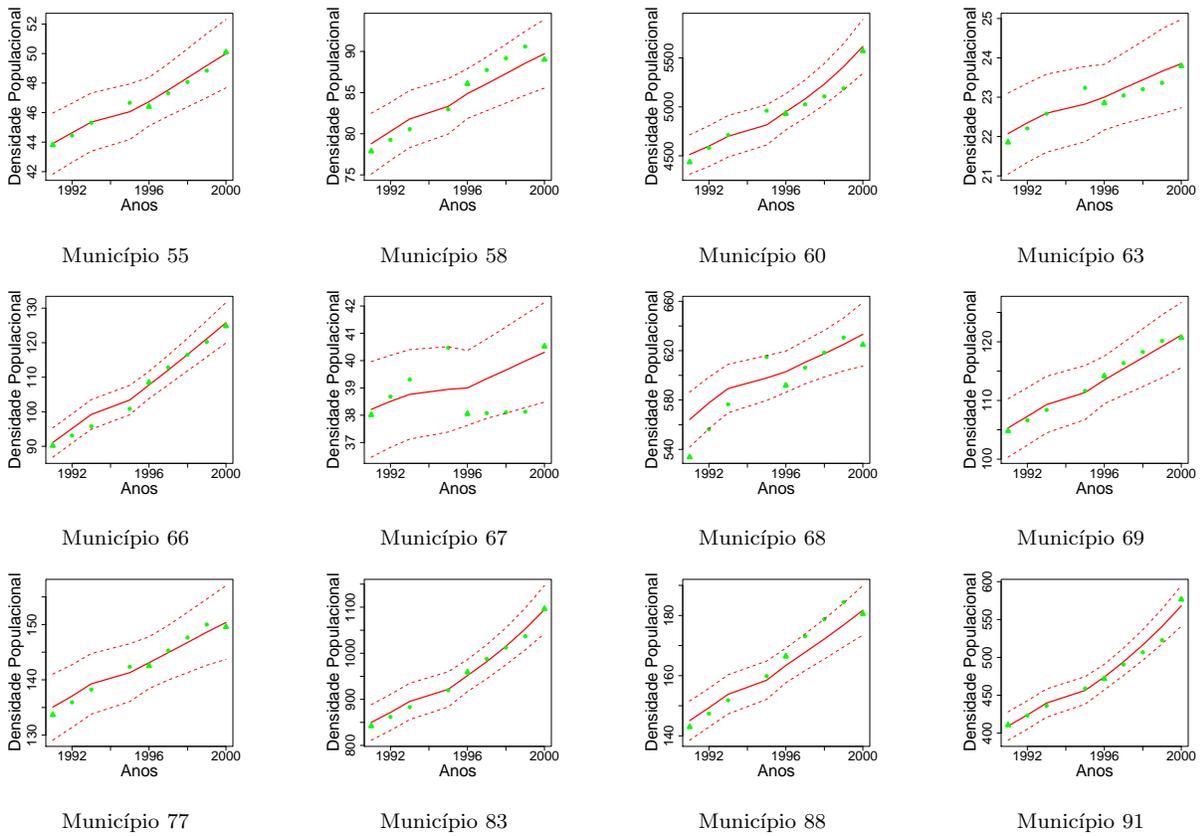


Figura 7.2: Ajustes sob o Modelo Espacial com fator 0,50 em vermelho e estimativas oficiais do IBGE em verde.

Referências Bibliográficas

- Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, CRC Press/Chapman Hall.
- Berger, J. O., de Oliveira, V. & Sanso, B. (2001). Objective bayesian analysis of spatially correlated data, *Journal of the American Statistical Association* **96**(456): 1361–1374.
- Carter, C. K. & Kohn, R. (1994). On gibbs sampling for state space models, *Biometrika* **81**(3): 541–553.
- Cressie, N. A. C. (1993). *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York.
- Diggle, P. & Ribeiro Jr, P. (2000). Model based geostatistics, *Associação Brasileira de Estatística - ABE - 14o SINAPE* .
- Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox*, 3rd edn, Timberlake Consultants Press and Oxford, London.
- Fay, R. E. & Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data, *Journal of the American Statistical Association* **74**(366): 269–277.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *J. Time Ser. Anal.* **15**(2): 183–202.
- Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*, 2nd edn, Chapman & Hall.

- Gamerman, D. & Migon, H. (1999). *Statistical inference - an integrated approach*, Arnold.
- Gelfand, A. E. & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach, *Biometrika* **85**(1): 1–11.
- Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**(410): 398–409.
- Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.
- Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd edn, Edward Arnold, London.
- Harrison, P. J. & Stevens, C. F. (1976). Bayesian forecasting, *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(3): 205–247.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**(1): 97–109.
- Klein, R. & Moura, F. A. S. (1998). Análise dos erros amostrais de indicadores educacionais obtidos através de dados da pesquisa nacional por amostra de domicílios, *Revista Brasileira de Estatística* **59**(211): 47–58.
- Madeira, J. L. & Simões, C. C. S. (1972). Estimativas preliminares da população urbana e rural segundo as unidades da federação, de 1960/1980 por uma nova metodologia, *Revista Brasileira de Estatística* **33**(129): 3–11.
- Metropolis, N., Rosenbulth, A., Rosenbulth, M., Teller, A. & Teller, E. (1953). Equation of state calculations by fast computing machine, *Journal of Chemical Physics* **21**: 1087–91.
- Migon, H. & Gamerman, D. (1993). Generalized exponential growth models: A bayesian approach, *Journal of Forecasting* **12**: 573–584.

- Rao, J. N. K. (2003). *Small area estimation*, Wiley Series in Survey Methodology, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms, *The Annals of Applied Probability* **7**(1): 110–120.
- Roberts, G. O. & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms, *Statistical Science* **16**(4): 351–367.
- Roberts, G. O. & Rosenthal, J. S. (2006). Examples of adaptive MCMC, *Technical report*, Lancaster University.
- Schmidt, A. M., Conceição, M. F. G. & Moreira, G. A. (2006). Investigating the sensitivity of gaussian processes to the choice of their correlation function and prior specifications, *Technical report*, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil.
- Schmidt, A. M., Nobre, A. A. & Ferreira, G. S. (2002). Alguns aspectos da modelagem de dados espacialmente referenciados, *Revista Brasileira de Estatística* **63**(220): 59–88.
- Souza, D. F. (2004). *Estimação de população em nível municipal via modelos hierárquicos e espaciais*, Master's thesis, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Brazil.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(4): 583–639.
- Stroud, J. R., Muller, P. & Sanso, B. (2001). Dynamic models for spatiotemporal data, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63**(4): 673–689.
- West, M. & Harrison, J. (1997). *Bayesian forecasting and dynamic models*, Springer Series in Statistics, Springer-Verlag, New York.

Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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