

Modelo Lee-Carter Bayesiano para Modelagem da Mortalidade no Rio de Janeiro

Helen Christiny Alves Nascimento e Luana Monteiro Augusto



Universidade Federal do Rio de Janeiro
Instituto de Matemática
Departamento de Métodos Estatísticos

2019

Modelo Lee-Carter Bayesiano para Modelagem da Mortalidade no Rio de Janeiro

Helen Christiny Alves Nascimento e Luana Monteiro Augusto

Projeto final submetido 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 Bacharel em Ciências Atuariais.

Aprovada por:

Prof. João Batista de Moraes Pereira
PhD - IM - UFRJ - Orientador.

Prof. Viviana das Graças Ribeiro Lobo
PhD - IM - UFRJ.

Prof. William Lima Leão
PhD - IM - UFRJ.

Rio de Janeiro, RJ - Brasil

2019

Agradecimentos

Helen Christiny Alves Nascimento

Agradeço a Deus, em primeiro lugar, pela vida que me concedeu. Agradeço por toda força, sabedoria e amor que me deste nessa caminhada. Agradeço por não me deixar sozinha e sempre estar me guiando, protegendo e amparando. Por não desistir de mim, mesmo quando nem eu acreditava. Por me amar de tal forma que entregou seu único filho em meu lugar, dando sentido e motivação a minha vida.

Agradeço à minha família que me apoiou desde o primeiro dia de aula até o último. Compreenderam a dificuldade e me ajudaram para que essa jornada fosse mais leve. Agradeço, especialmente, aos meus pais, Edson e Sonia, que sempre fizeram de tudo para que eu pudesse chegar até aqui. Agradeço à minha irmã Andréa e ao meu namorado Elison que me ouviram e aconselharam nos momentos em que precisei.

Agradeço aos meus amigos que não desistiram da nossa amizade, mesmo muitas vezes eu não podendo curtir como gostariam por estar com afazeres da faculdade. Agradeço por torcerem por mim e se alegrarem comigo. Por direta ou indiretamente cooperaram para essa conquista.

Agradeço à Luana que esteve comigo desde o primeiro dia da graduação até o último. Agradeço por ter aturado todos os meus humores, por trilhar toda essa caminhada comigo e por ser minha amiga na vida. Agradeço por dividir o trabalho de conclusão de curso e por sempre confiar e acreditar em mim.

Agradeço ao meu orientador João, que de pronto aceitou o convite e nos auxiliou para que tudo saísse com excelência. Também pelas aulas maravilhosas que tive a oportunidade de assistir, sem dúvidas o carinho que temos será sempre lembrado.

Agradeço aos professores Viviana e William que aceitaram participar da banca do nosso projeto.

Luana Monteiro Augusto

Agradeço, primeiramente, aos meus pais, por terem me dado uma educação de qualidade, tanto na escola quanto em casa. Sem vocês eu não estaria aqui hoje. Obrigada por sempre confiarem em mim e nas minhas escolhas.

Agradeço a todos os professores que passaram pela minha trajetória e que, de certa forma, foram fundamentais para meu desenvolvimento. Um agradecimento especial ao nosso orientador João Batista por ter nos motivado e ensinado tanto desde o início da faculdade, sempre nos apoiando e acreditando no nosso potencial. Sem dúvidas, marcou as nossas vidas para sempre.

À Helen, minha parceria em todos os períodos dessa graduação. Obrigada pelas risadas, ajudas, paciência e amizade durante todos esses anos, por ser tão compreensiva e ter aceitado a ideia deste projeto final, mesmo sabendo que seria um grande desafio para nós.

À Chebyshev, minha gata tão querida, com seu temperamento difícil. Entretanto, com muito carinho, sendo fundamental ao longo desses últimos anos.

Por fim, obrigada a todos que contribuíram de alguma forma para o encerramento de mais uma etapa na minha vida.

“Educar não é cortar as asas, e sim orientar o voo.”
(Aisha Linda)

Sumário

1	Introdução	1
2	Conjunto de dados	4
2.1	Análise exploratória de dados	5
3	Fundamentos Estatísticos	10
3.1	Modelos dinâmicos	10
3.2	Modelo Lee Carter	11
3.3	Modelo Lee Carter bayesiano	12
3.4	Amostrador de Gibbs	13
3.5	Método FFBS	14
4	Ajuste do Modelo	17
5	Projeção da mortalidade	26
5.1	Distribuição preditiva	26
6	Conclusões	39
A	Distribuições condicionais completas	43
B	Traços das cadeias	44

Lista de Tabelas

- 4.1 Médias a posteriori dos Parâmetros α_x e β_x para diferentes faixas etárias. 21
- 4.2 Médias a posteriori do parâmetro k_t ao longo dos anos por sexo e total. . 23

Lista de Figuras

2.1	Log-mortalidade para diferentes faixas etárias para os anos de 1990, 2000 e 2011.	5
2.2	Log-mortalidade para diferentes faixas etárias ao longo dos anos.	7
2.3	Log-mortalidade para diferentes faixas etárias para os anos 1990, 2000 e 2011. O painel à esquerda apresenta a mortalidade feminina e o painel à direita, a masculina.	8
4.1	Média a posteriori do parâmetro β_x (linha cheia) com respectivos intervalos de 95% de credibilidade a posteriori (área hachurada) para diferentes faixas de idade.	18
4.2	Média a posteriori do parâmetro β_x por sexo e para o total.	18
4.3	Média do parâmetro α_x (linha cheia) com respectivos intervalo de 95% de credibilidade a posteriori (área hachurada) para diferentes faixas de idade.	19
4.4	Média a posteriori do parâmetro α_x estimado por sexo e para o total.	20
4.5	Média do parâmetro k_t (linha cheia) com respectivos intervalo de 95% de credibilidade a posteriori (área hachurada) ao longo dos anos.	22
4.6	Média a posteriori do parâmetro k_t ao longo dos anos por sexo e total.	22
4.7	Histograma dos valores estimados do parâmetro θ por total, feminino e masculino no Amostrador de Gibbs.	24
4.8	Histograma dos valores estimados do parâmetro σ_w^2 por total, feminino e masculino no Amostrador de Gibbs.	25
4.9	Histograma dos valores estimados do parâmetro σ_ϵ^2 por total, feminino e masculino no Amostrador de Gibbs.	25

5.1	Média a posteriori das projeções do parâmetro k_t (linha cheia) com os respectivos intervalos de 95% de credibilidade (área hachurada) ao longo dos anos.	27
5.2	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o total.	29
5.3	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o sexo masculino.	30
5.4	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o sexo feminino.	30
5.5	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 total.	31
5.6	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 para o sexo masculino.	32
5.7	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 para o sexo feminino.	32
5.8	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2021 total.	33
5.9	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos total.	34
5.10	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos para o sexo masculino.	34
5.11	Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos para o sexo feminino.	35
5.12	Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.	36

5.13	Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.	37
5.14	Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.	38
B.1	Traços das cadeias dos parâmetros θ , σ_ϵ^2 , σ_w^2 e β_{5-9} total, feminino e masculino do modelo Lee-Carter bayesiano.	45
B.2	Traços das cadeias dos parâmetros α_{5-9} e k_{2011} total, feminino e masculino do modelo Lee-Carter bayesiano.	46

Capítulo 1

Introdução

Com o passar dos anos, percebe-se o grande declínio na taxa de mortalidade dos países, principalmente daqueles desenvolvidos. Isso deve-se, principalmente, ao avanço da tecnologia em conjunto com a área da saúde. No Brasil, por exemplo, a taxa de mortalidade decaiu de 18/1000 em 1940 para 7/1000 em 1985, com a expectativa de vida crescendo em 20 anos e a mortalidade infantil decaindo de 160/1000 em 1940 para 85/1000 em 1980 de acordo com [Prata \(1992\)](#).

Dentre as mudanças que ocorreram no Brasil, temos como de grande importância a criação do Sistema Único de Saúde (SUS) em 1988, que tem o objetivo de oferecer atendimento a população de maneira gratuita, visto que a saúde é um direito de todo cidadão. O SUS foi e continua sendo essencial para melhorar o perfil sanitário do Brasil. Também foi implementado o programa nacional de vacinação, desde 1973, que fornece vacinas para população contra algumas doenças, resultando na redução das mesmas ([Souza et al., 2018](#)).

Sendo assim, o cenário brasileiro mudou drasticamente nos últimos anos, aumentando a expectativa de vida, diminuindo a mortalidade infantil e também as mortes por doenças infecciosas dado que antigamente a higiene nos ambientes e as vacinas não eram tão acessíveis como hoje em dia. Principalmente no Rio de Janeiro, conhecido nos anos de 1830 por ser uma cidade com ambiente sujo e propício a doenças, levando a um

número bem maior de óbitos do que nascimentos. A cidade enfrentou diversas doenças devastadoras como a febre amarela em 1850 e a cólera em 1855, causando epidemias na sociedade do Rio de Janeiro ([Marcílio, 1993](#)).

O Rio de Janeiro era uma cidade em destaque nos estudos de mortalidade do Brasil antes de 1900, por conta de sua elevada taxa de mortalidade. Somente a partir de 1903, com as medidas de saneamento de Oswaldo Cruz, foi possível começar a reverter o quadro.

Logo, percebe-se como a projeção da mortalidade não é muito previsível visto que vários fatores influenciam na mesma: fatores sócio-econômicos, ambientais, culturais, demográficos, entre outros. Para isso existem diversos métodos de projeção da mortalidade, atualmente um dos mais utilizados é o método Lee-Carter, que se baseia na mortalidade do passado para projetar a mortalidade.

Entender a evolução da mortalidade, particularmente, é importante para o mercado segurador. Dado o cenário atual de crise previdenciária no Brasil, o mercado segurador de previdência vem adquirindo novos clientes. Com isso, é necessário que haja tábuas biométricas que prevejam bem a trajetória da mortalidade nos próximos anos de modo a evitar possíveis cálculos inconscientes com a realidade dos segurados e déficits para a carteira da seguradora, já que a escolha pela utilização de uma determinada tábua biométrica é decisiva para o cálculo de provisões, contribuições e rendas.

A construção de tábuas ainda é um processo incomum, principalmente no Brasil. Temos a tábua disponibilizada anualmente pelo IBGE com a data de referência de primeiro de julho do ano anterior feita utilizando dados de censo. Porém, mudanças têm acontecido na taxa de mortalidade e, conseqüentemente, na expectativa de vida com o passar dos anos, o que pode afetar os resultados de análises de vida e previdência.

Logo, é de suma importância para as seguradoras e até mesmo para uma questão econômica do país ter conhecimento sobre a expectativa de vida dos indivíduos. Quanto

mais tempo vivermos, maior será o tempo de previdência, assim como serão maiores as chances de problemas como massa de desemprego no futuro, entre outros.

O intuito desse trabalho é modelar e projetar a taxa de mortalidade utilizando o modelo Lee-Carter ([Lee e Carter, 1992](#)) para diversas faixas etárias para o Rio de Janeiro do ano de 2012 até o ano de 2021 utilizando um histórico de 21 anos de taxa de mortalidade, ou seja, temos informação de 1990 até 2011. O modelo Lee-Carter é conhecido por bons resultados utilizando dados com muitos anos, em [Lee e Carter \(1992\)](#) os anos utilizados foram de 1900 até 1987, entretanto, os dados do Rio de Janeiro são escassos, além de possuírem uma mudança brusca de mortalidade no passado que acreditamos não se repetir dessa maneira no futuro.

No Capítulo seguinte é apresentada uma descrição do conjunto de dados disponíveis, como foram obtidos, além da análise exploratória dos mesmos. No Capítulo [3](#), é feita uma breve revisão sobre modelos dinâmicos, modelo Lee-Carter, amostrador de Gibbs e método Forward Filtering Backward Sampling (FFBS). O Capítulo [4](#) apresenta a estimação dos parâmetros do modelo proposto. A projeção da log-mortalidade é apresentada no Capítulo [5](#), onde também é apresentada uma comparação da projeção deste trabalho com a projeção do IBGE e com as taxas de mortalidade reais. O Capítulo [6](#) traz as conclusões do trabalho.

Capítulo 2

Conjunto de dados

Os dados utilizados nesse trabalho foram obtidos através do portal Indicadores e Dados Básicos (IDB) do Departamento de Informática do Sistema Único de Saúde (DATASUS) no endereço datasus.saude.gov.br/indicadores-e-dados-basicos-idb.

Extraímos dos indicadores demográficos os itens A.1 - População total e A.8 - Mortalidade proporcional por idade. Apenas a unidade da federação do Rio de Janeiro foi selecionada no período de 1990 até 2011, sendo feito o filtro adicional de sexo “Feminino”, “Masculino” e “Total”. Além disso, foi selecionado no item A.8 o número de óbitos.

Os itens em questão não possuem a mesma quantidade de faixas etárias, sendo assim, foi necessário fazer um ajuste para o item referente à população total, apenas para os anos 1990 e 1992, de forma que a faixa etária “0 a 4 anos” foi proporcionalmente dividida entre “Menor de 1 ano” e “1 a 4 anos” de modo que ficasse comparável ao item de mortalidade. Para o ajuste, foram utilizadas as proporções de população em cada uma das duas faixas do ano subsequente ao ano ajustado. As idades ignoradas, que são os casos em que há não resposta da variável idade, não foram utilizadas para o estudo.

A mortalidade que é a variável de interesse nesse trabalho, foi calculada dividindo o número de óbitos sobre o número total de expostos em cada faixa etária para cada ano

dentre os utilizados. Em resumo, o conjunto de dados corresponde à taxa de mortalidade anual, dividida e agregada por sexo, no Estado do Rio de Janeiro no período de 1990 a 2011 para diferentes faixas de idade. A subseção seguinte apresenta a análise exploratória dos dados em questão.

2.1 Análise exploratória de dados

A partir da análise exploratória dos dados o intuito é observar como a mortalidade se comportou durante os 21 anos (1990-2011) de dados disponíveis para diferentes faixas de idade. Em particular, para fins práticos e de modelagem, optou-se por trabalhar com a variável de interesse na escala logarítmica, isto é, com a log-mortalidade. As análises foram feitas com auxílio do programa [R Core Team \(2018\)](#) e com ajuda do programa Excel para algumas manipulações necessárias.

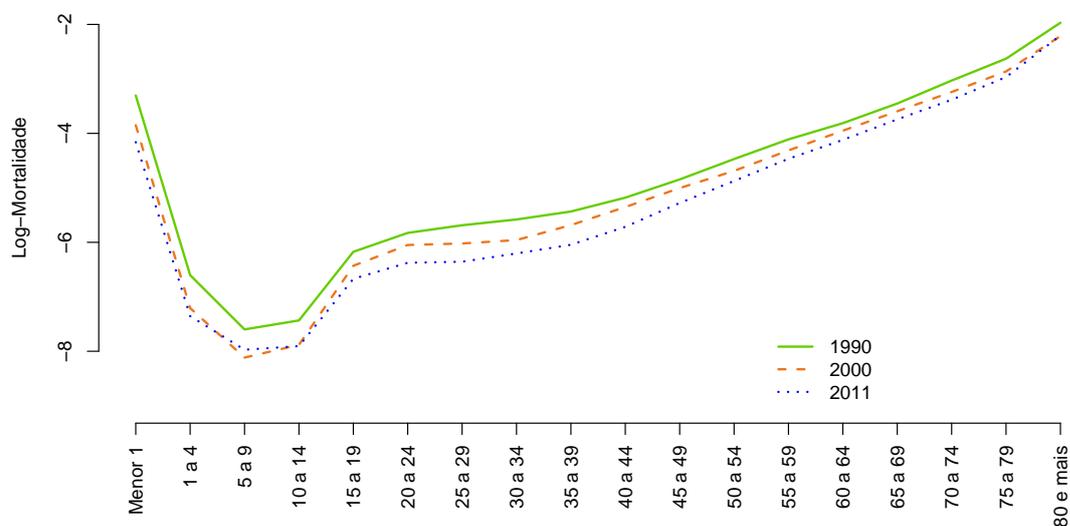


Figura 2.1: Log-mortalidade para diferentes faixas etárias para os anos de 1990, 2000 e 2011.

No gráfico da Figura 2.1 podemos perceber que a curva de log-mortalidade possui algumas características para as faixas etárias. Começando com valores mais elevados para a primeira faixa, a curva após isso sofre uma drástica redução nos valores para as faixas de crianças e uma ascendente mais acentuada para a faixa dos jovens, em seguida a curva se eleva para chegar nos índices de maior mortalidade que são da última faixa.

A evolução na Figura 2.1 nos mostra que ao longo dos anos a curva da função log mortalidade tem a forma parecida, porém houve um deslocamento significativo para algumas idades, ou seja, está diminuindo para algumas faixas. A mortalidade mudou para cada idade ao longo do tempo e, analisando cada idade, pode-se observar as mudanças mais significativas. A Figura 2.1 nos mostra que, para a faixa menor que 1 ano, houve uma diminuição na log-mortalidade: a taxa era de 3,668% em 1990, caiu para 2,127% em 2000 e em 2011 foi de 1,557%.

A mortalidade infantil é um ponto de grande importância para o país, já que a sua redução estava na lista dos Objetivo de Desenvolvimento do Milênio (ODM) para o período de 1990 até 2015 e está nos Objetivos de Desenvolvimento Sustentável (ODS) de 2015 até 2030. O Brasil atingiu em 2015 a meta dos ODM pela redução de mortalidade infantil. Porém, ainda temos um índice elevado, o qual deve ser avaliado para identificar as regiões e causas particulares de determinados lugares. O primeiro ano de vida tem grande concentração de óbitos, por isso existe uma análise separada. Nesses 21 anos observou-se uma diminuição de 57,57% na mortalidade da faixa de crianças com menos de um ano.

Também é possível observar a evolução de cada faixa de idade ao longo desses 21 anos, analisar as variações e observar as mudanças relevantes. A Figura 2.2 abaixo apresenta os valores da log-mortalidade para algumas faixas etárias específicas ao longo dos anos

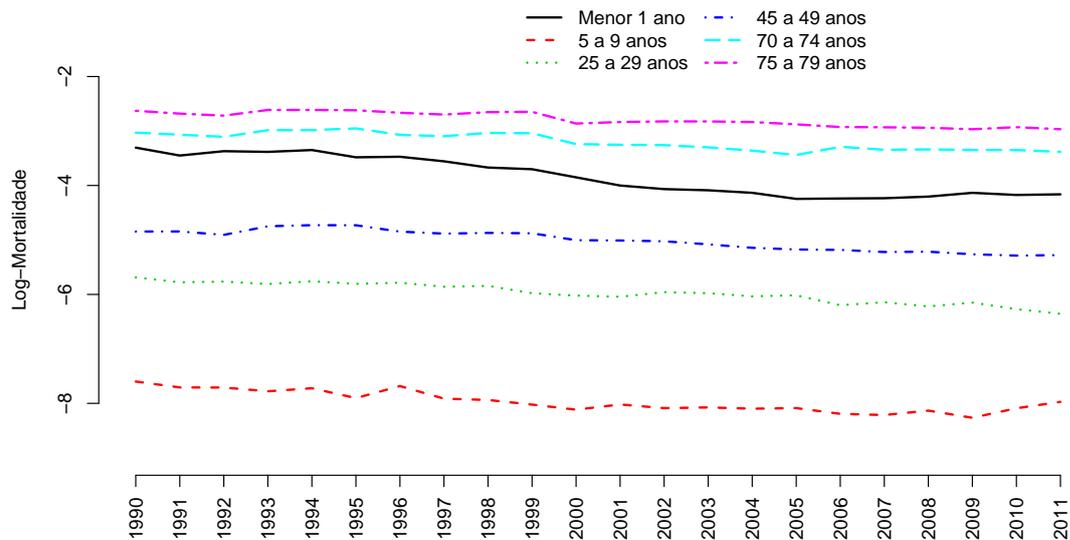


Figura 2.2: Log-mortalidade para diferentes faixas etárias ao longo dos anos.

Observa-se uma redução na mortalidade ao longo dos 21 anos. Isso pode ser um indício de que a expectativa de vida aumentou e os indivíduos passaram a ter uma sobrevivência maior nessas faixas. Pode-se esperar que a curva continue a cair e que haja menor mortalidade com o passar dos anos. Houve uma redução de 29,68% para a faixa de 70 a 74 anos e de 28,67% para a faixa de 75 a 79 anos.

A partir da Figura 2.3, avaliando as log-mortalidades feminina e masculina separadamente, percebe-se uma diferença na forma da curva para todos os anos entre as faixas de 10 a 14 e de 40 a 44 anos. Nota-se que entre as faixas mencionadas, as mulheres têm a log-mortalidade menor, aumentando de maneira linear, enquanto a dos homens tem um aumento rápido para as faixas citadas.

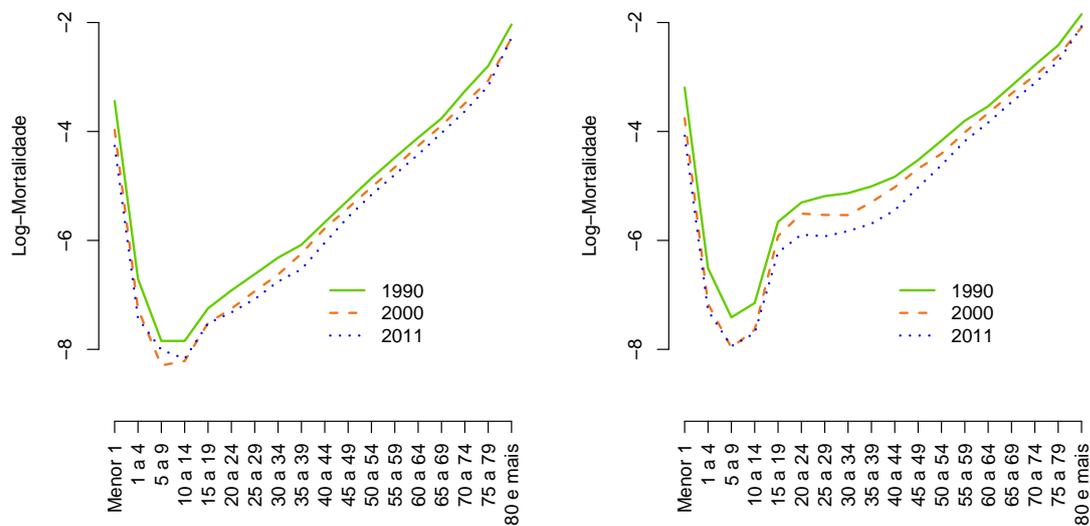


Figura 2.3: Log-mortalidade para diferentes faixas etárias para os anos 1990, 2000 e 2011. O painel à esquerda apresenta a mortalidade feminina e o painel à direita, a masculina.

A disparidade entre a taxa de mortalidade feminina e masculina é analisada mundialmente. Segundo o relatório disponibilizado pela Organização Mundial de Saúde (OMS), as mulheres têm maior expectativa de vida em todas as regiões. O estudo está disponível no endereço <https://apps.who.int/iris/bitstream/handle/10665/311696/WHO-DAD-2019.1-eng.pdf> e aponta alguns fatores que influenciam essa diferenciação na expectativa de vida que são as principais causas de morte de homens e mulheres.

A expectativa de vida ao nascer em 2019, segundo a OMS, é de, em média, 69,8 anos para os meninos e, em média, 74,2 anos para as meninas. As principais causas da menor expectativa de vida ao nascer dos homens em relação às mulheres são doença cardíaca isquêmica, acidentes de trânsito, cânceres de traquéia, brônquios e pulmão, doença pulmonar crônica. Estudos nesse sentido específico para regiões de um país, podem ajudar a constatar melhorias a serem aplicadas e campanhas de conscientização a serem investidas para que homens e mulheres atentem-se ao descobrimento e tratamento

de algumas doenças.

É interessante projetar a mortalidade nos próximos anos visto que a mesma possui mudanças claras ao longo dos anos. É esperado que a expectativa de vida aumente com o passar dos anos por conta do grande aumento da tecnologia e, conseqüentemente, melhorias na área de saúde. O intuito é que esse desenvolvimento seja traduzido e capturado pelo método proposto neste trabalho.

Uma grande dificuldade na projeção da mortalidade no Brasil é que temos dados recentes. Não possuímos um histórico grande de informação. Com isso, a projeção ficará limitada. [Lee e Carter \(1992\)](#), por exemplo, utilizou dados de mortalidade anual de 1900 até 1987.

Capítulo 3

Fundamentos Estatísticos

Esta seção revisa aspectos da modelagem e procedimentos de inferência para a modelagem proposta. Logo abaixo, uma breve apresentação de modelos dinâmicos de forma geral, utilizados para modelagem de séries temporais. Em seguida, será apresentado o modelo Lee-Carter propriamente dito, que pode ser visto como um modelo dinâmico.

3.1 Modelos dinâmicos

Uma série temporal pode ser modelada utilizando diversos componentes como evolução no tempo, tendência e sazonalidade. O modelo que foi desenvolvido para controlar a variação no tempo é o chamado modelo dinâmico, mais detalhes podem ser encontrados em [West e Harrison \(1997\)](#). Também conhecidos como modelos de espaço de estados, podem ser aplicados a séries não estacionárias, sem que haja transformações prévias nos dados e a séries tanto univariadas como multivariadas.

Um caso particular de modelos dinâmicos é o modelo de espaços de estados gaussianos e linear, conhecido como modelo linear dinâmico. O modelo permite que a dependência temporal seja levada em conta por meio de uma estrutura de evolução imposta a parâmetros denominados, neste contexto, parâmetros de estado. A especificação do modelo se completa ao atribuir-se uma distribuição a priori normal, no tempo zero, para o vetor de estados. Formalmente, o modelo é definido como:

$$\theta_0 \sim N(m_0, C_0),$$

$$Y_t = F_t' \theta_t + \nu_t, \quad \nu_t \sim N(0, V_t), \quad (3.1)$$

e

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \quad (3.2)$$

para $t \geq 1$.

Onde,

- Y_t é o vetor de observações ($p \times 1$);
- θ_t é um vetor de parâmetros de estado, p -dimensional, do modelo dinâmico ($n \times 1$);
- F_t é uma matriz de regressão conhecida ($n \times p$);
- G_t é uma matriz de evolução temporal dos parâmetros de estado ($n \times n$);
- V_t é uma matriz conhecida de covariância ligada ao erro observacional ν_t ($p \times p$);
- W_t é uma matriz conhecida de covariância ligada ao erro de evolução dos parâmetros de estado w_t ($n \times n$);

Uma maneira usual de definir o modelo é através da quádrupla $\{F, G, V, W\}_t$

3.2 Modelo Lee Carter

O método de [Lee e Carter \(1992\)](#) é um dos mais famosos métodos de projeção de mortalidade e apresentou-se diferente se comparado com os que já tinham sido propostos anteriormente ([Booth et al., 2006](#)). Dentre as diferenças temos a utilização de um método com dois fatores que são a idade do indivíduo e o tempo e, para obter somente um índice variando no tempo do nível de mortalidade, utiliza uma decomposição de matrizes. O método é um modelo dinâmico em que a extrapolação das tendências passadas é um princípio pressuposto.

O modelo é definido da seguinte maneira, tomando $\ln[m_{(x,t)}] = y_{(x,t)}$:

$$y_{(x,t)} = \alpha_x + \beta_x k_t + \epsilon_{(x,t)} \quad (3.3)$$

onde, $m_{(x,t)}$ é a taxa de mortalidade central (razão entre as mortes para uma dada faixa de idade e a população exposta daquela faixa), para a idade x no ano t , α_x é o padrão médio da mortalidade para a idade específica, β_x é o padrão da mortalidade à idade x quando o nível geral de mortalidade se altera, k_t é o índice de tendência temporal do nível de mortalidade geral e $\epsilon_{(x,t)}$ é o erro aleatório para idade x e tempo t .

O modelo Lee-Carter necessita de algumas restrições obtidas através dos mínimos quadrados da equação principal 3.3 para que seus parâmetros sejam identificáveis. Utiliza-se a normalização, de modo que os β_x somem 1 ($\sum_x \beta_x = 1$) e os k_t somem 0 ($\sum_t k_t = 0$). Consequentemente os α_x serão a média do log-mortalidade no tempo (Lee e Carter, 1992; Wachter, 2014).

3.3 Modelo Lee Carter bayesiano

Pedroza (2006) apresentou uma reformulação do método de Lee-Carter de forma a escrevê-lo como um modelo de espaço de estados. Os modelos de espaço de estados são amplamente utilizados na modelagem de séries temporais. O modelo é dado por:

$$y_{(x,t)} = \alpha_x + \beta_x k_t + \epsilon_{(x,t)}, \quad \epsilon_{(x,t)} \sim N(0, \sigma_\epsilon^2), \quad (3.4)$$

$$k_t = k_{t-1} + \theta + w_t, \quad w_t \sim N(0, \sigma_w^2), \quad (3.5)$$

$$k_0 \sim N(m_0, C_0)$$

onde θ é um parâmetro de *drift* que modela uma tendência linear e w_t é um termo de ruído.

A inferência bayesiana busca quantificar a incerteza acerca da estimação dos parâmetros, o que pode ser visto como uma vantagem na utilização do modelo Lee-Carter bayesiano.

Assumimos as seguintes distribuições a priori para os parâmetros do modelo:

$$\alpha_x \sim N(m_\alpha, C_\alpha),$$

$$\beta_x \sim N(m_\beta, C_\beta),$$

$$\theta \sim N(m_\theta, C_\theta),$$

$$\sigma_\epsilon^2 \sim GI(a, b),$$

$$\sigma_w^2 \sim GI(c, d).$$

Seguindo as mesmas definições e restrições mencionadas na Seção 3.2, temos o modelo Normal multivariado para a log-mortalidade. Com isso, temos uma distribuição conjunta para todas as faixas de idade em qualquer ano. Condicional aos parâmetros, assume-se que os $y_{(x,t)}$ são independentes e identicamente distribuídos (*iid*).

Como a distribuição a posteriori do vetor paramétrico $(\alpha_x, \beta_x, k_t, k_0, \sigma_\epsilon^2, \sigma_w^2 \text{ e } \theta)$ não é conhecida analiticamente, usaremos o Método de Monte Carlo via Cadeias de Markov (MCMC) para obter amostras da mesma. Em particular, utilizaremos o amostrador de Gibbs.

3.4 Amostrador de Gibbs

Os métodos de MCMC (Monte Carlo via Cadeias de Markov) são um conjunto de métodos iterativos que têm a finalidade de aproximar, no enfoque bayesiano, distribuições a posteriori que não podem ser calculadas analiticamente. O amostrador de Gibbs é um algoritmo de MCMC proposto no contexto de processamento de imagens por [Geman e Geman \(1984\)](#) que é utilizado quando tem-se o conhecimento das distribuições condicionais completas dos parâmetros. O método consiste em amostrar das distribuições condicionais completas de forma recursiva.

Suponha que estamos interessados na função $\pi(\mathbf{x})$, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ e que temos as condicionais completas $\pi(\mathbf{x}_i | \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_p)$, $i = 1, \dots, p$. Obter os valores da distribuição conjunta é mais complicado que das distribuições condicionais completas.

O algoritmo é descrito a seguir:

1 - Inicie com $\mathbf{x}^{(0)} = (\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_p^{(0)})$ e um contador $j = 1$ que indica os passos do método.

2 - Amostre iterativamente de cada condicional completa:

$$\mathbf{x}_1^{(j)} \sim \pi(\mathbf{x}_1 | \mathbf{x}_2^{(j-1)}, \dots, \mathbf{x}_p^{(j-1)}),$$

$$\mathbf{x}_2^{(j)} \sim \pi(\mathbf{x}_2 | \mathbf{x}_1^{(j)}, \mathbf{x}_3^{(j-1)}, \dots, \mathbf{x}_p^{(j-1)}),$$

$$\mathbf{x}_3^{(j)} \sim \pi(\mathbf{x}_3 | \mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, \mathbf{x}_4^{(j-1)}, \dots, \mathbf{x}_p^{(j-1)}),$$

e assim até

$$\mathbf{x}_p^{(j)} \sim \pi(\mathbf{x}_p | \mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, \dots, \mathbf{x}_{p-1}^{(j)}).$$

3 - Adicione um passo ao contador se tornando $j + 1$ e retome ao passo 2 até que haja convergência.

3.5 Método FFBS

O método *Forward Filtering Backward Sample* (FFBS) foi proposto por [Frühwirth-Schnatter \(1994\)](#) e [Carter e Kohn \(1994\)](#) e tem o objetivo de auxiliar na amostragem conjunta dos estados para o caso de modelos dinâmicos utilizando filtragem e suavização das distribuições dos parâmetros de interesse.

Para a estimação do parâmetro k_t , utilizamos o referido método para sortear da distribuição condicional completa de $\underline{k} = (k_1, \dots, k_T)'$. Considere $D_t = (\underline{y}_1, \dots, \underline{y}_t)$. Os passos do algoritmo são descritos a seguir:

1 - Passo de filtragem:

Queremos obter $m_t^{(i)}$ e $C_t^{(i)}$, $t = 1, \dots, T$, onde $m_t^{(i)}$ e $C_t^{(i)}$ são o primeiro e segundo momentos da distribuição a posteriori de k_t no tempo t .

i) Posteriori no tempo $t - 1$:

$$k_{t-1}|D_{t-1} \sim N(m_{t-1}, C_{t-1}).$$

ii) Priori no tempo t :

$$k_t|D_{t-1} \sim N(a_t, R_t)$$

onde $a_t = m_{t-1} + \theta$ e $R_t = C_{t-1} + \sigma_w^2$.

iii) Previsão:

$$\underline{y}_t|D_{t-1} \sim N(\underline{f}_t, Q_t),$$

$$\underline{f}_t = \underline{\alpha} + \underline{\beta}a_t,$$

$$Q_t = \underline{\beta}\underline{\beta}'R_t + \sigma_p^2,$$

onde $\underline{\alpha} \sim (\alpha_{x_1}, \alpha_{x_2}, \dots, \alpha_{x_p})'$, $\underline{\beta} \sim (\beta_{x_1}, \beta_{x_2}, \dots, \beta_{x_p})'$ e p representa as faixas etárias utilizadas.

iv) Posteriori no tempo t :

$$\underline{k}_t|D_t \sim N(m_t, C_t)$$

onde $m_t = a_t + R_t\underline{\beta}'Q_t^{-1}(\underline{y}_t - \underline{f}_t)$ e $C_t = R_t - R_t^2\underline{\beta}'Q_t^{-1}\underline{\beta}$.

2 - Passo de suavização:

$$k_t | k_{t+1}, D_t \sim N(h_t, H_t),$$

$$\text{com } h_t = m_t - \frac{C_t(k_{t+1} - a_{t+1})}{R_{t+1}} \text{ e } H_t = C_t - \frac{C_t^2}{R_{t+1}}.$$

Assim podemos sortear sequencialmente de $k_T | D_T \sim N(m_T, C_T)$ e de $k_t | D_t \sim N(h_t, H_t)$, para $t = T - 1, \dots, 1$. Os demais parâmetros do modelo são sorteados de suas respectivas distribuições condicionais completas, as quais estão descritas no Apêndice [A](#).

Capítulo 4

Ajuste do Modelo

O modelo Lee-Carter bayesiano foi ajustado para a log-mortalidade considerando os dados separados por sexo e também para a log-mortalidade total. Para a nossa análise, a avaliação de convergência dos parâmetros estimados será através da visualização do gráfico das cadeias dos respectivos parâmetros excluindo os 200 valores iniciais da cadeia, o chamado período de aquecimento considerando um total de 1000 iterações. Os valores considerados à priori foram: m_0, m_α, m_β e m_θ iguais a 0, C_0, C_α, C_β e C_θ iguais a 100 e a, b, c e d iguais a 0,01.

As Figuras 4.1 e 4.2 apresentam a média a posteriori, assim como os intervalos de 95% de credibilidade a posteriori do parâmetro β_x para as diferentes faixas de idade, por sexo e total (sem considerar a variável sexo). O parâmetro β_x nos traduz a tendência, para a idade x , da mortalidade sendo impactada por k_t , que é o fator que explica o efeito do tempo na log-mortalidade. Pode-se perceber que as estimativas de β_x total se parecem com as estimativas de β_x para o sexo masculino. As estimativas de β_x para o sexo feminino têm diferenças significativas entre as faixas 10 a 14 e 20 a 24 anos onde temos estimativas maiores β_x . Na faixa 25 a 29 anos há um encontro dos valores estimados β_x e entre as faixas 30 a 34 e 45 a 49 anos, as estimativas de β_x para o sexo feminino são menores que as estimativas dos outros dois. Na faixa de 50 a 55 anos há novamente um encontro entre os valores estimados de β_x .

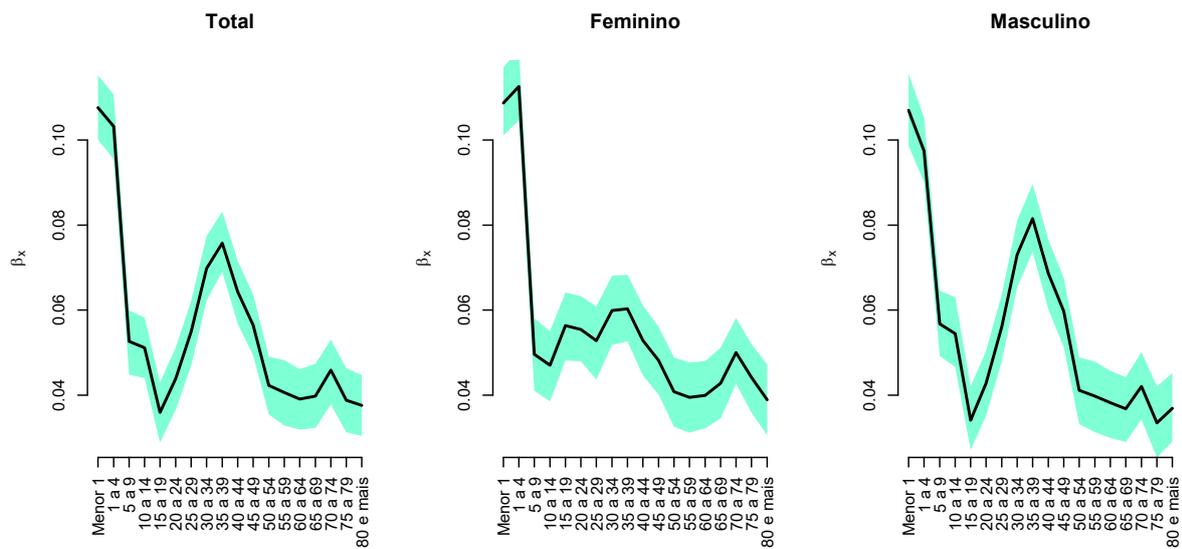


Figura 4.1: Média a posteriori do parâmetro β_x (linha cheia) com respectivos intervalos de 95% de credibilidade a posteriori (área hachurada) para diferentes faixas de idade.

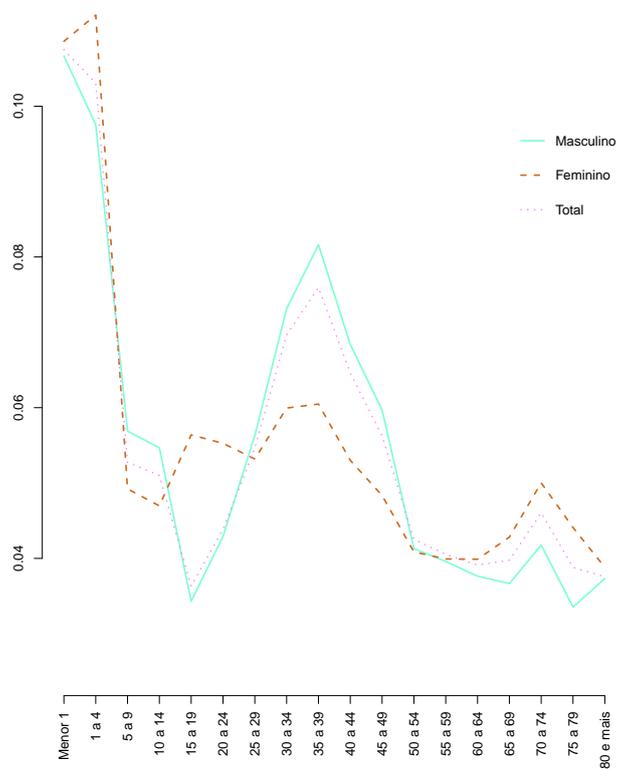


Figura 4.2: Média a posteriori do parâmetro β_x por sexo e para o total.

Na Figura 4.3 que apresenta o sumário a posteriori de α_x , pode-se observar que o intervalo de credibilidade fica quase imperceptível, sabemos que a amplitude do intervalo está relacionada à incerteza que temos a respeito do determinado parâmetro. Portanto, temos que a distribuição de α_x é concentrada.

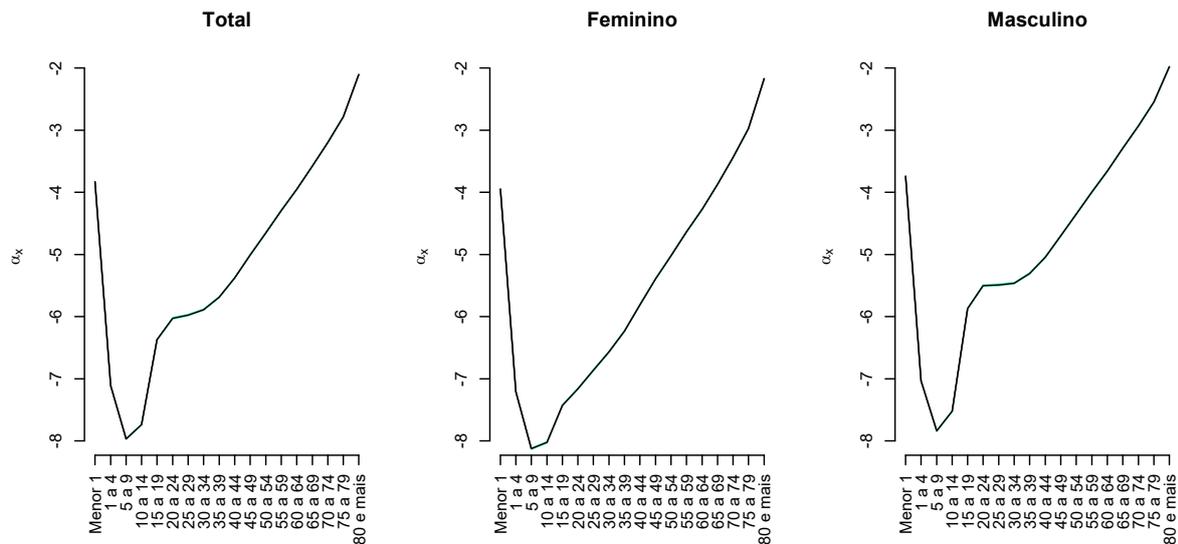


Figura 4.3: Média do parâmetro α_x (linha cheia) com respectivos intervalo de 95% de credibilidade a posteriori (área hachurada) para diferentes faixas de idade.

O impacto diretamente causado pela idade x na taxa de mortalidade é traduzido pelo parâmetro α_x . Como já havíamos observado, há uma diferença significativa na curva de mortalidade para algumas faixas de idade de cada um dos sexos. Tal diferença ficou explícita com a Figura 4.4, temos então que a dissemelhança da curva de log mortalidade entre as faixas de 10 a 14 até 40 a 44 anos é influenciada pelo efeito idade do indivíduo. Portanto, acreditamos que é um fator que será mantido ao longo do tempo.

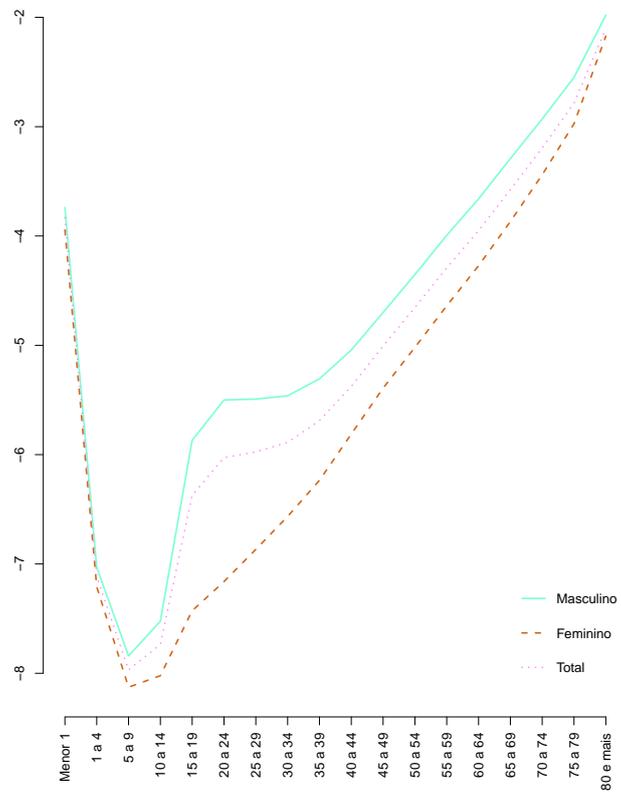


Figura 4.4: Média a posteriori do parâmetro α_x estimado por sexo e para o total.

Tabela 4.1: Médias a posteriori dos Parâmetros α_x e β_x para diferentes faixas etárias.

Faixa Etária	Total		Feminino		Masculino	
	α_x	β_x	α_x	β_x	α_x	β_x
Menor 1 ano	-3,8311	0,1076	-3,9462	0,1086	-3,7395	0,1069
1 a 4 anos	-7,1124	0,1032	-7,2088	0,1125	-7,032	0,0974
5 a 9 anos	-7,9694	0,0526	-8,1271	0,0496	-0,7841	0,0568
10 a 14 anos	-7,7371	0,0512	-8,0219	0,047	-7,5241	0,0545
15 a 19 anos	-6,372	0,036	-7,4287	0,0564	-5,8687	0,0341
20 a 24 anos	-6,0284	0,0439	-7,1611	0,0554	-5,501	0,0428
25 a 29 anos	-5,9753	0,0549	-6,8667	0,0528	-5,4912	0,0561
30 a 34 anos	-5,8883	0,0698	-0,6565	0,0599	-5,4615	0,0731
35 a 39 anos	-5,6877	0,0758	-6,2353	0,0603	-5,3072	0,0815
40 a 44 anos	-5,3755	0,0643	-5,8116	0,0529	-5,0445	0,0686
45 a 49 anos	-5,0076	0,0565	-5,3909	0,0482	-4,7001	0,0597
50 a 54 anos	-4,6535	0,0423	-5,0191	0,0408	-4,3514	0,0412
55 a 59 anos	-4,291	0,0406	-4,6383	0,0395	-3,9937	0,0398
60 a 64 anos	-3,9517	0,0391	-4,2727	0,0399	-3,6591	0,0382
65 a 69 anos	-3,5765	0,0398	-3,8684	0,0428	-3,2907	0,0368
70 a 74 anos	-3,1951	0,0459	-3,4408	0,05	-2,9298	0,0421
75 a 79 anos	-2,7844	0,0388	-2,9741	0,0442	-2,5485	0,0335
80 anos e mais	-2,1028	0,0376	-2,1724	0,0389	-1,9803	0,0369

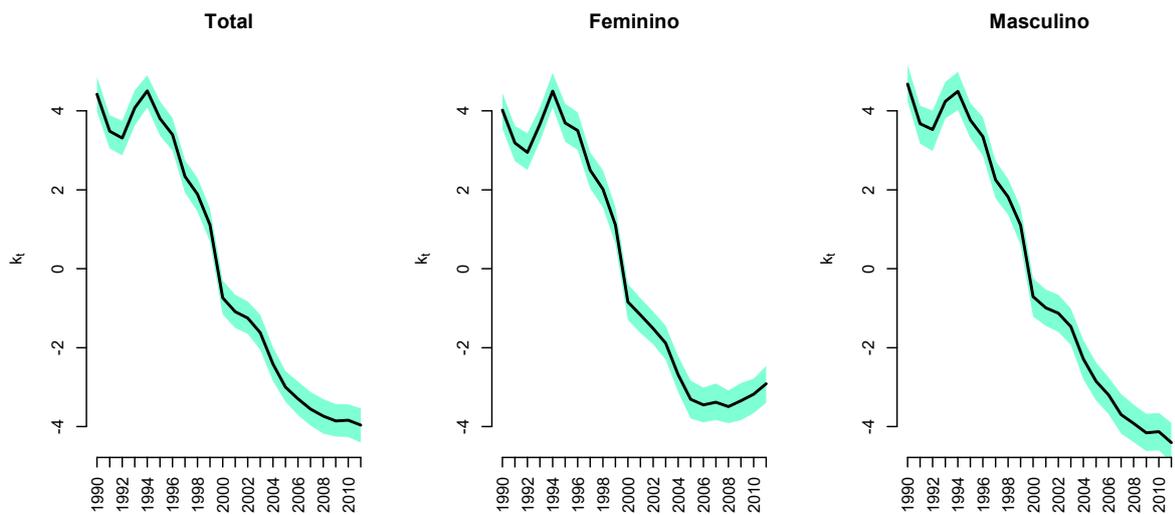


Figura 4.5: Média do parâmetro k_t (linha cheia) com respectivos intervalo de 95% de credibilidade a posteriori (área hachurada) ao longo dos anos.

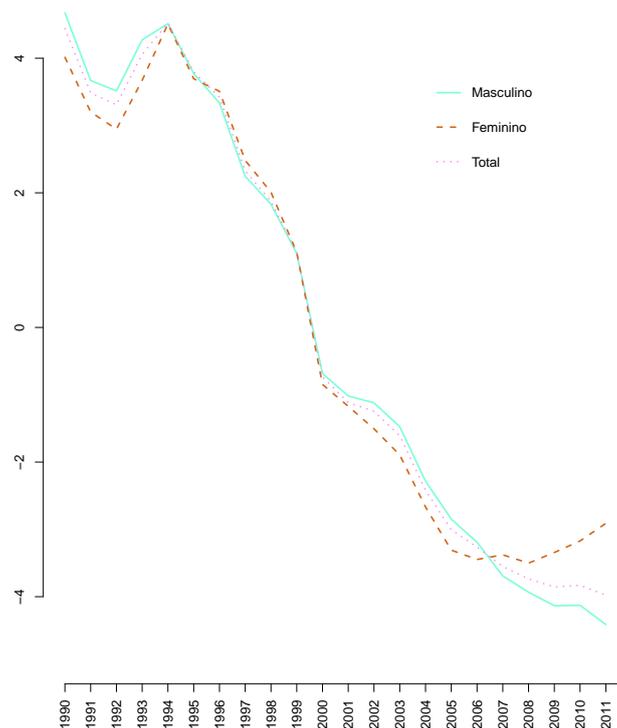


Figura 4.6: Média a posteriori do parâmetro k_t ao longo dos anos por sexo e total.

As Figuras 4.5 e 4.6 apresentam as estimativas dos parâmetros k_t ao longo do tempo. Nota-se que a evolução do parâmetro não é linear ao longo dos anos visto que a mortalidade no Rio de Janeiro vem diminuindo consideravelmente. Através dos gráficos e da Tabela 4.2, percebe-se que a população feminina tem um aumento de k_t a partir do ano de 2006 quando em comparação ao decréscimo do masculino e do total.

Tabela 4.2: Médias a posteriori do parâmetro k_t ao longo dos anos por sexo e total.

Ano	Total	Feminino	Masculino
1990	4,4234	4,0211	4,6769
1991	3,4825	3,1887	3,6753
1992	3,3085	2,9478	3,5289
1993	4,0775	3,6628	4,2424
1994	4,5062	4,5023	4,4933
1995	3,8019	3,6907	3,7676
1996	3,3989	3,5053	3,3506
1997	2,3348	2,4987	2,2526
1998	1,8899	2,0183	1,8220
1999	1,1138	1,1136	1,1069
2000	-0,7348	-0,8407	-0,7048
2001	-1,0883	-1,1695	-0,9877
2002	-1,2511	-1,5140	-1,1227
2003	-1,6172	-1,8830	-1,4666
2004	-2,4162	-2,6871	-2,2868
2005	-2,9967	-3,3041	-2,8525
2006	-3,2957	-3,4476	-3,1953
2007	-3,5523	-3,3781	-3,6964
2008	-3,7341	-3,4902	-3,9187
2009	-3,8541	-3,3436	-4,1558
2010	-3,8344	-3,1781	-4,1247
2011	-3,9623	-2,9132	-4,4045

A convergência dos parâmetros foi analisada visualmente e os gráficos com os traços das cadeias de alguns parâmetros dos modelos ajustados podem ser encontrados no apêndice B. Nas Figuras 4.7, 4.8 e 4.9 é possível visualizar os histogramas das amostras a posteriori θ , σ_w^2 e σ_ϵ^2 . Vale ressaltar que o amostrador de Gibbs foi implementado no programa R [Core Team \(2018\)](#).

Nota-se que o parâmetro θ não parece significativamente diferente de zero. Quanto às variâncias, nota-se que a variância σ_w^2 foi estimada em valores maiores que a variância observacional σ_ϵ^2 , o que também acontece em [Pedroza \(2006\)](#).

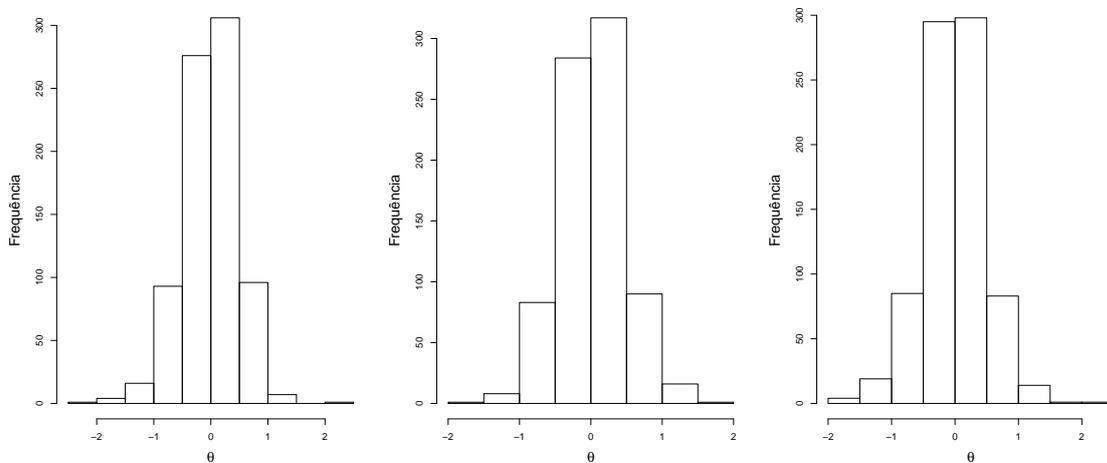


Figura 4.7: Histograma dos valores estimados do parâmetro θ por total, feminino e masculino no Amostrador de Gibbs.

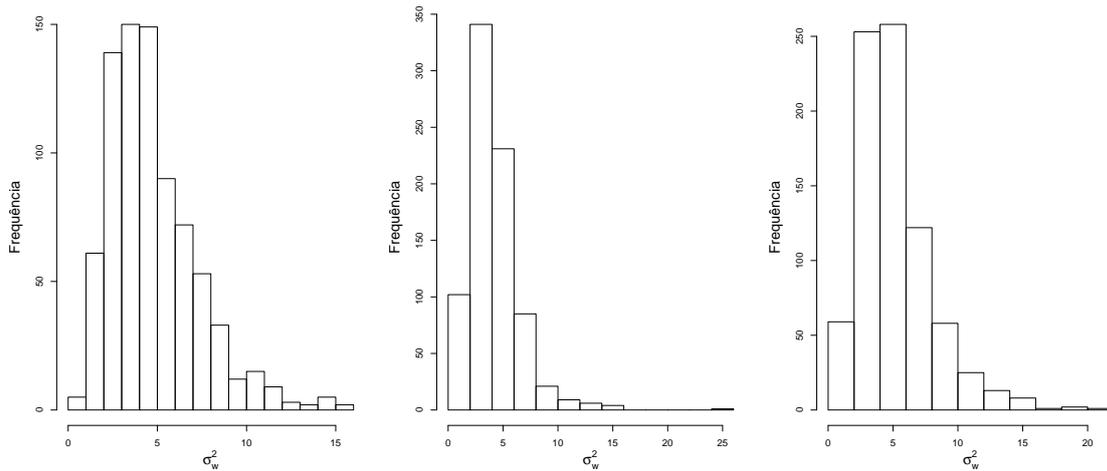


Figura 4.8: Histograma dos valores estimados do parâmetro σ_w^2 por total, feminino e masculino no Amostrador de Gibbs.

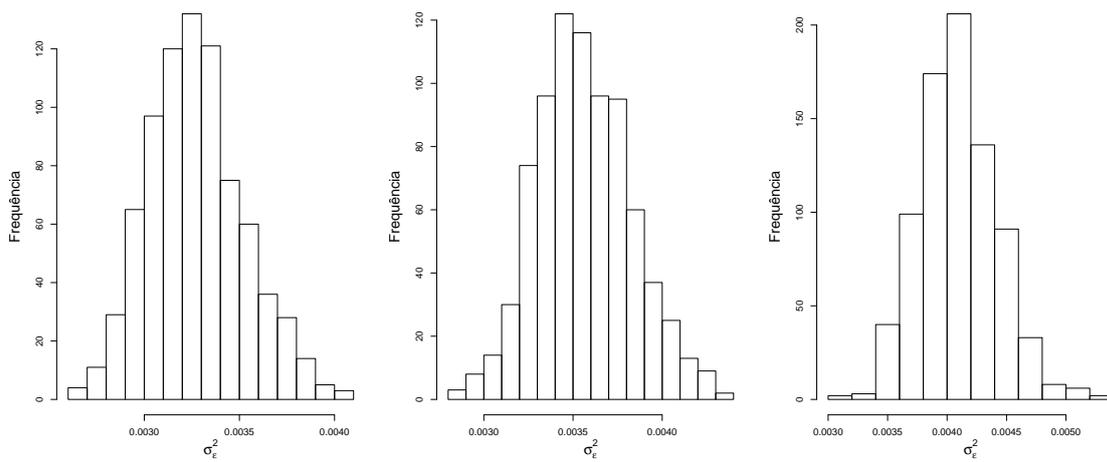


Figura 4.9: Histograma dos valores estimados do parâmetro σ_ϵ^2 por total, feminino e masculino no Amostrador de Gibbs.

Capítulo 5

Projeção da mortalidade

5.1 Distribuição preditiva

Podemos expressar a distribuição preditiva como

$$p(y_{(x,T+1)}|D_T) = \int p(y_{(x,T+1)}|\Theta)p(\Theta|D_T)d\Theta \quad (5.1)$$

onde Θ é o conjunto com todos os parâmetros do modelo. Temos essa expressão, pois temos a hipótese de que $y_{(x,T+1)}$ e D_T são condicionalmente independentes dado os parâmetros. Podemos obter a distribuição preditiva também através de métodos de simulações, assim como utilizamos para a distribuição a posteriori do modelo.

Com a convergência dos parâmetros do modelo garantida, podemos adicionar os passos da predição. Sendo j o contador que indica os passos do Amostrador de Gibbs, temos:

1- Obtenha $k_{(T+1)}^{(j)}$ de

$$k_{(T+1)}^{(j)} \sim N(k_T^{(j)} + \theta^{(j)}, \sigma_w^{2(j)}).$$

2- Obtenha $y_{(x,T+1)}^{(j)}$ de

$$y_{(x,T+1)}^{(j)} \sim N(\alpha_x^{(j)} + \beta_x^{(j)}k_{(T+1)}^{(j)} + \theta^{(j)}, \sigma_\epsilon^{2(j)}).$$

Os passos devem ser repetidos para cada iteração j do amostrador de Gibbs.

O modelo Lee-Carter foi aplicado aos dados de mortalidade do Rio de Janeiro obtidos de 1990 até 2011 e decidiu-se, então, projetar o log-mortalidade para os 10 anos subsequentes, ou seja, até 2021. A projeção foi feita separadamente para homens e mulheres, como já havíamos percebido diferenças significativas de log-mortalidade e também é separada por faixa etária.

Sabe-se que os parâmetros α_x e β_x variam apenas por da faixa etária. Logo apenas o parâmetro k_t será projetado pelos 10 anos seguintes. A Figura 5.1 apresenta a projeção do parâmetros k_t ao longo dos 10 anos a partir de 2012 até 2021. Percebe-se como a incerteza aumenta ao longo dos anos, visualizado pelo aumento do intervalo de credibilidade.

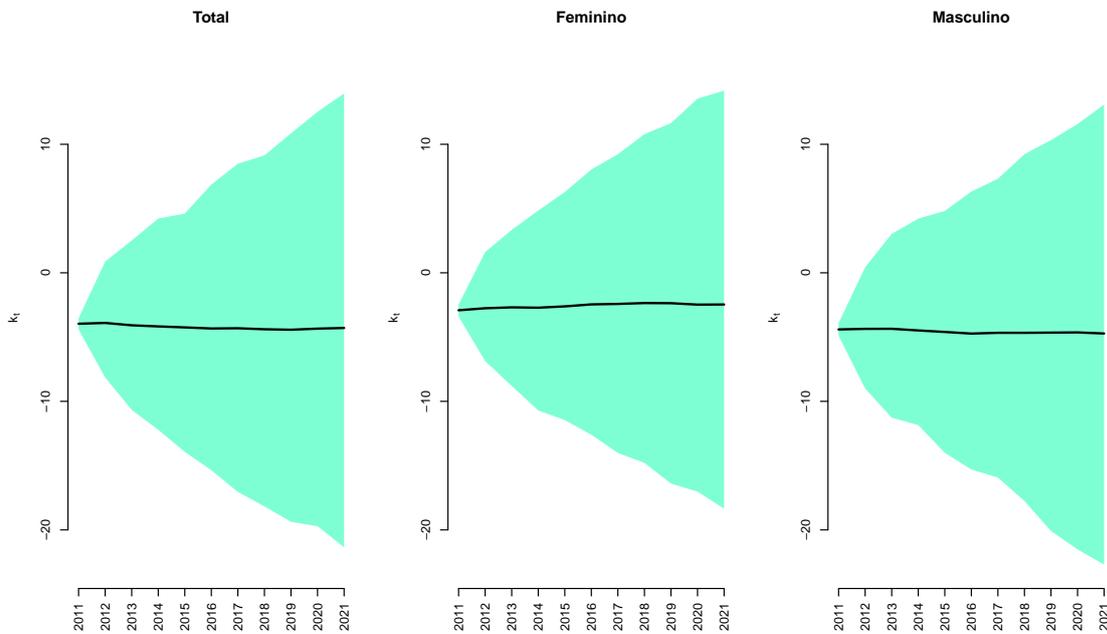


Figura 5.1: Média a posteriori das projeções do parâmetro k_t (linha cheia) com os respectivos intervalos de 95% de credibilidade (área hachurada) ao longo dos anos.

A partir da projeção de k_t , pode-se projetar a mortalidade e este resultado será comparado ao resultado da projeção com as tábuas de mortalidade do IBGE até 2021 encontradas no sítio <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?&t=resultados> e nos dados mais recentes disponibilizados pelo sistema de saúde do rio de janeiro <http://sistemas.saude.rj.gov.br/tabnet/deftohtm.exe?sim/taxasmort.def>. Para isso, utilizaremos a taxa central de mortalidade da faixa etária x no tempo t ($m_{x,t}$), encontrada facilmente avaliando a exponencial do log-mortalidade obtida na projeção.

Os dados que estão sendo comparados possuem as últimas faixas etárias distintas. A projeção do IBGE considera as faixas 80 a 84, 85 a 89 e 90+. É uma atualização interessante já que as pessoas têm vivido mais. Analisar o que tem acontecido nessas últimas 3 faixas seria interessante. Porém, a informação utilizada estava disponível com a última faixa sendo 80+, então foi necessário considerar alguns ajustes para a projeção do IBGE. A informação de população projetada também é disponibilizada separadamente por unidades de federação, então, utilizamos as informações para encontrar o número de óbitos e recalculamos a taxa de mortalidade para a última faixa, 80+.

Além disso, os dados recentes do SUS não possuem a separação das duas faixas iniciais “Menor 1 ano” e “1 a 4 anos”, sendo apenas uma única faixa de “0 a 4 anos”. Como mencionado no Capítulo 2, temos interesse em observar as duas faixas separadamente, pois entendemos que as características de cada uma são muito específicas e a taxa de mortalidade é diferente. Sendo assim, decidiu-se considerar a comparação com os dados recentes a partir da faixa 5 a 9 anos, visto que desagregar a faixa única do SUS nos resultaria em números não tão confiáveis a nível de comparação, pois não temos dados suficientes para fazer a abertura de modo mais verossímil possível.

Podemos comparar a taxa de mortalidade ($m_{x,t}$) encontrada a partir do ajuste do modelo Lee-Carter bayesiano com a curva obtida através da projeção disponibilizada pelo IBGE. Como não há visual diferença entre a projeção do IBGE e o resultado do modelo Lee-Carter para as faixas “Menor que 1” e “1 a 4 anos”, pode-se observar somente

a relação com os dados reais. Na Figura 5.2 é possível observar a taxa de mortalidade sem a distinção de sexo e percebe-se que, para 2012 a diferença entre a curva de mortalidade gerada pelo ajuste do método de Lee-Carter e a gerada pelo resultado da projeção do IBGE é muito pequena, as curvas se distinguem bem pouco e a partir da faixa 50 a 54 anos é possível notar mais claramente tal diferença. A taxa de mortalidade observada está bem mais próxima da taxa obtida através do modelo Lee-Carter bayesiano.

O resultado se repete para a separação entre os sexos feminino e masculino, como apresentado nas Figuras 5.3 e 5.4. Temos a curva do modelo Lee-Carter bayesiano mais próxima da curva com os valores observados do que a curva projetada pelo IBGE e a diferença entre as curvas é mais perceptível nas últimas faixas.

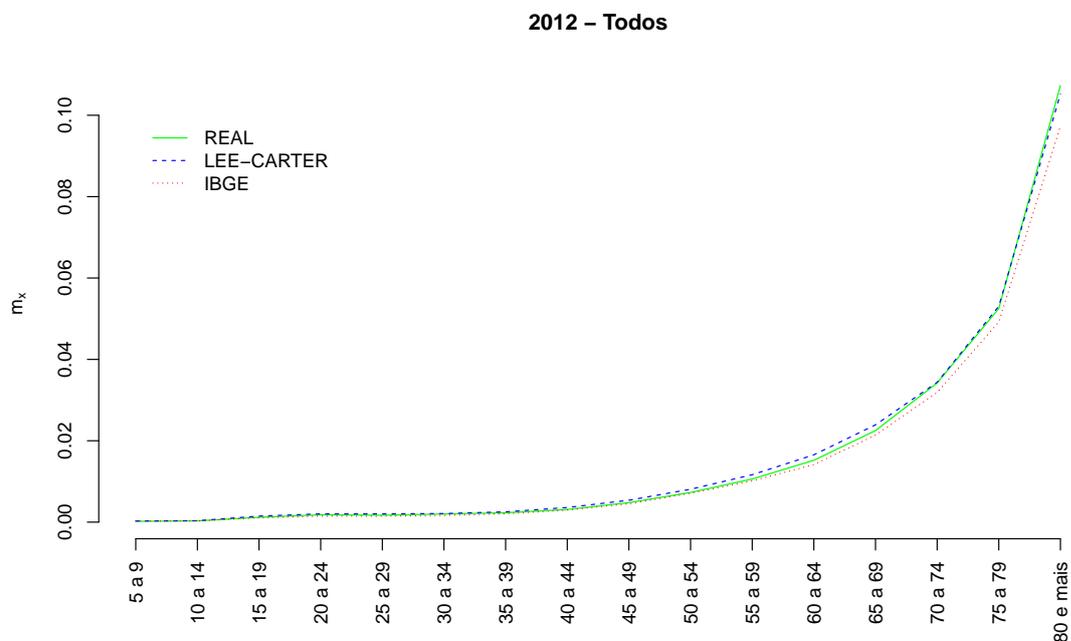


Figura 5.2: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o total.

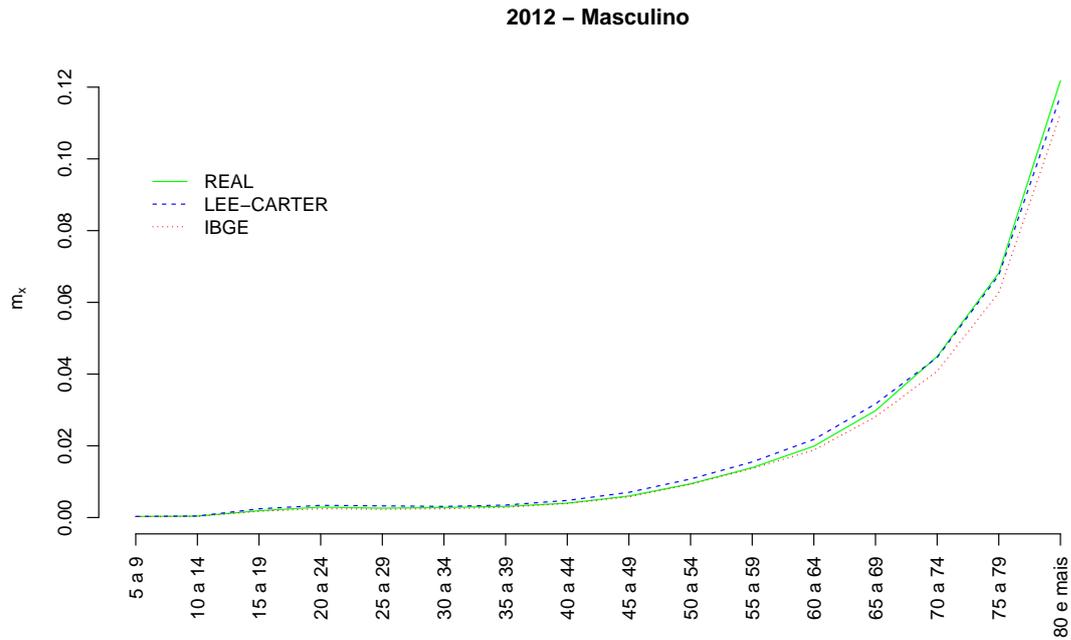


Figura 5.3: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o sexo masculino.

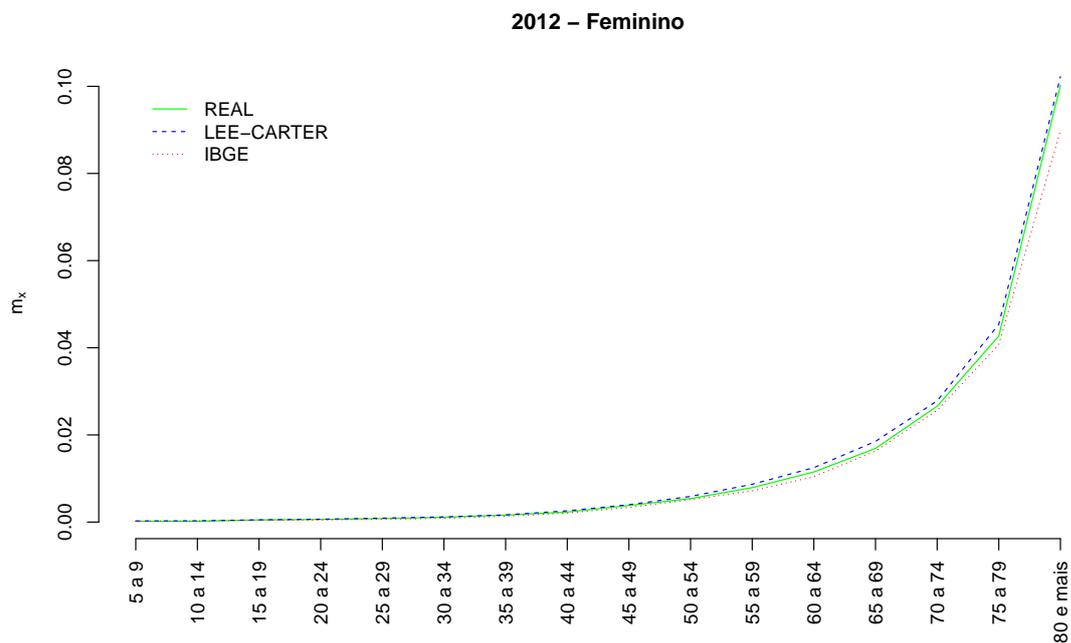


Figura 5.4: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2012 para o sexo feminino.

Separamos também o ano de 2018 para comparação da projeção, pois é o último que temos dados reais observados. A curva com a projeção obtida pelo ajuste do modelo Lee-Carter parece estar representando melhor os dados observados em comparação com a projeção do IBGE. Na Figura 5.7 podemos ver que somente para a faixa “75 a 79 anos” temos a taxa de mortalidade observada mais próxima à projetada pelo IBGE, mas em todo o restante da curva, a projeção obtida do ajuste do modelo proposto se mostra melhor.

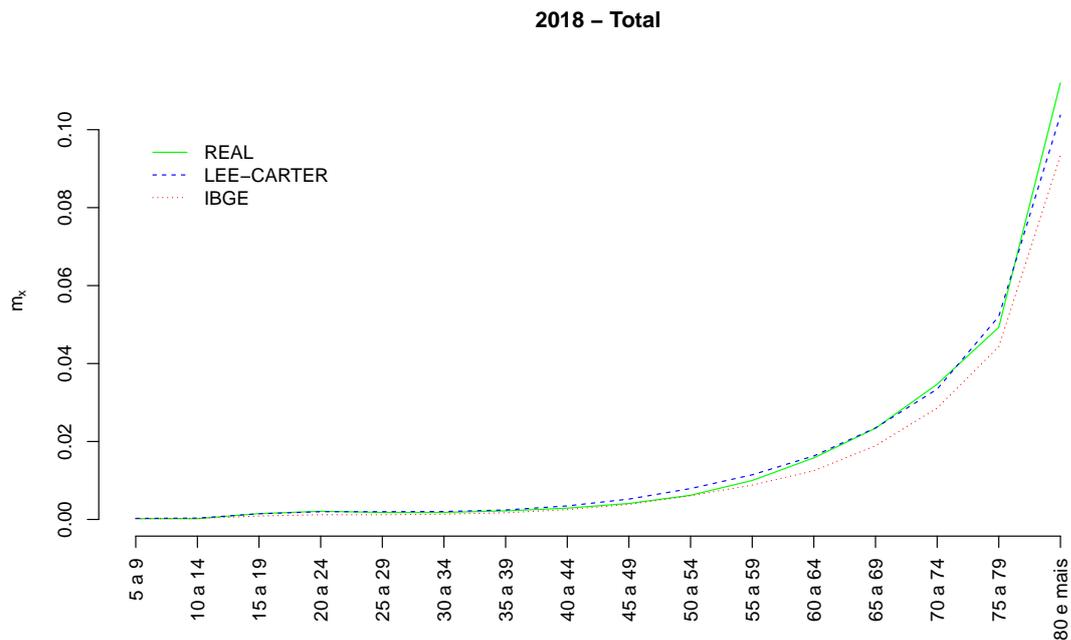


Figura 5.5: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 total.

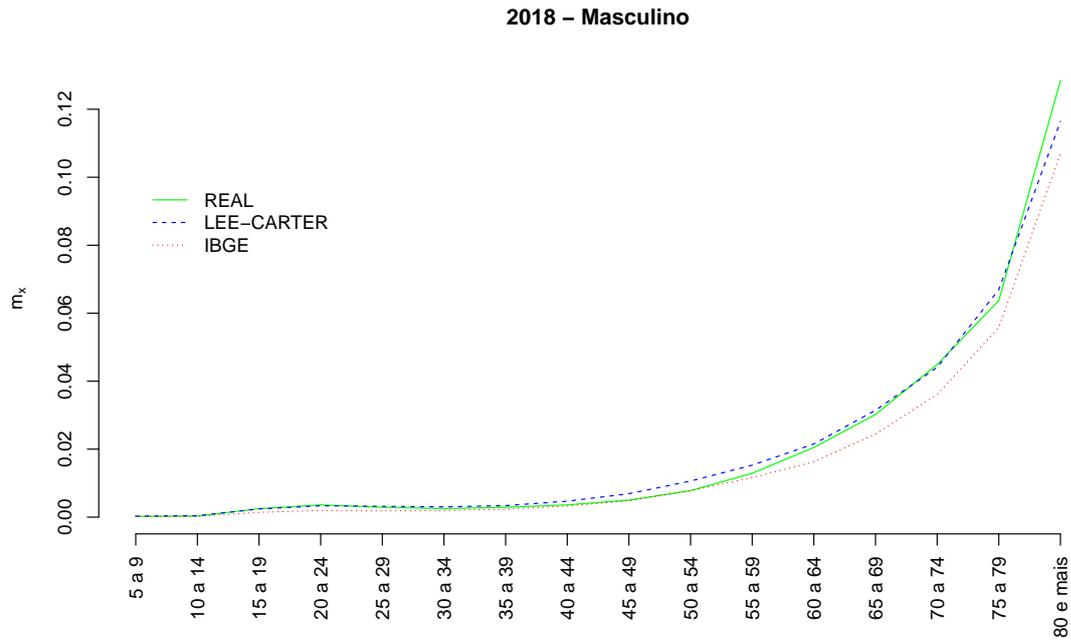


Figura 5.6: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 para o sexo masculino.

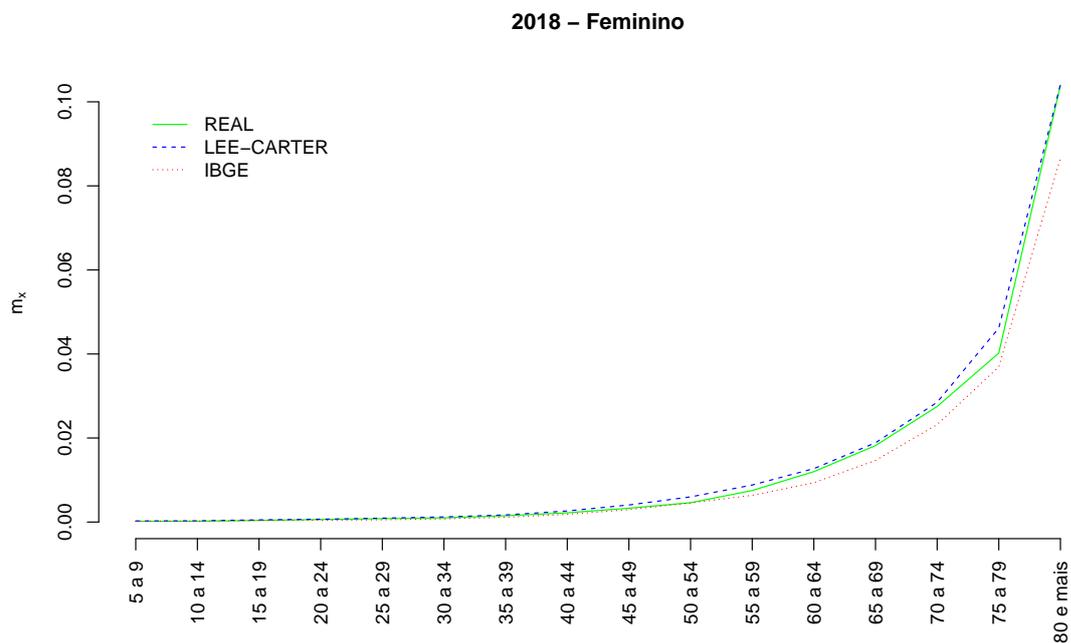


Figura 5.7: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2018 para o sexo feminino.

Já para o ano de 2021 sem considerar o sexo há uma distância maior entre as projeções. A curva do modelo Lee-Carter bayesiano estima uma mortalidade maior que a curva do IBGE principalmente para a faixa “Menor 1” e a partir da faixa “40 a 44” em diante, como apresentado na Figura 5.8.

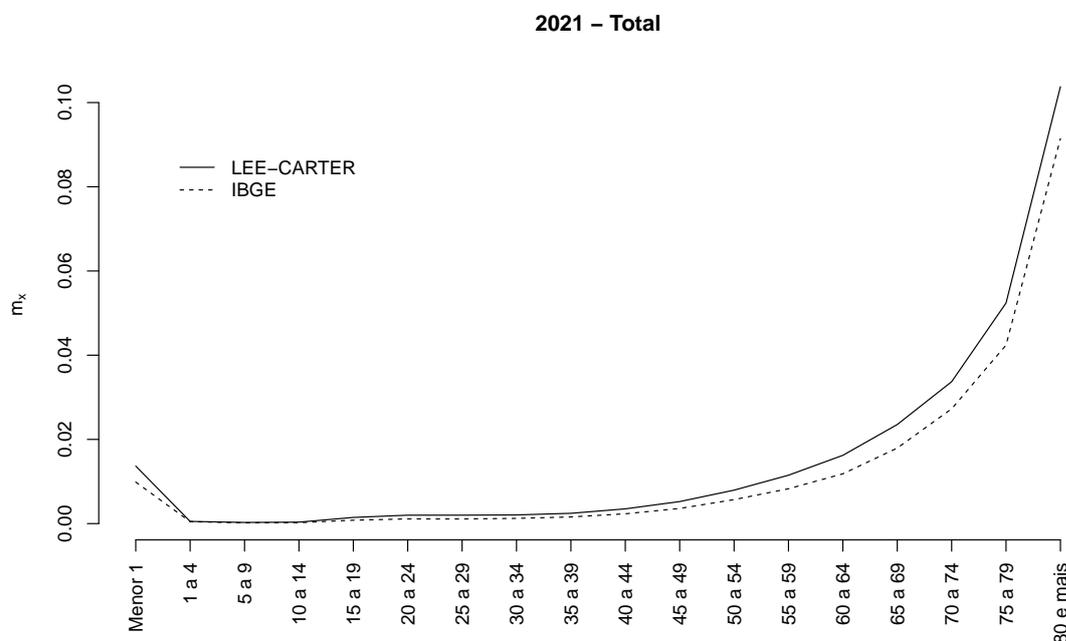


Figura 5.8: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para o ano de 2021 total.

Para que seja melhor avaliada a diferença entre as projeções ao longo do tempo, separamos algumas faixas para comparar com os valores observados. As Figuras 5.9, 5.10 e 5.11 mostram que a projeção do IBGE está considerando uma mortalidade menor e, ainda, uma redução linear da taxa central de mortalidade para a faixa 80+ anos. A projeção pelo modelo Lee-Carter bayesiano se mostra mais próxima à curva com os valores de taxa de mortalidade observados. Nas Figuras 5.9 e 5.10 podemos perceber que a estimação por Lee-Carter bayesiano subestima menos a mortalidade que a projeção feita pelo IBGE. Já em relação à taxa de mortalidade do sexo feminino apresentada na Figura 5.11, o modelo Lee-Carter bayesiano superestima a taxa de mortalidade a menos do ano de 2016 em que houve um pico na taxa, porém, ainda assim se aproxima mais da curva real.

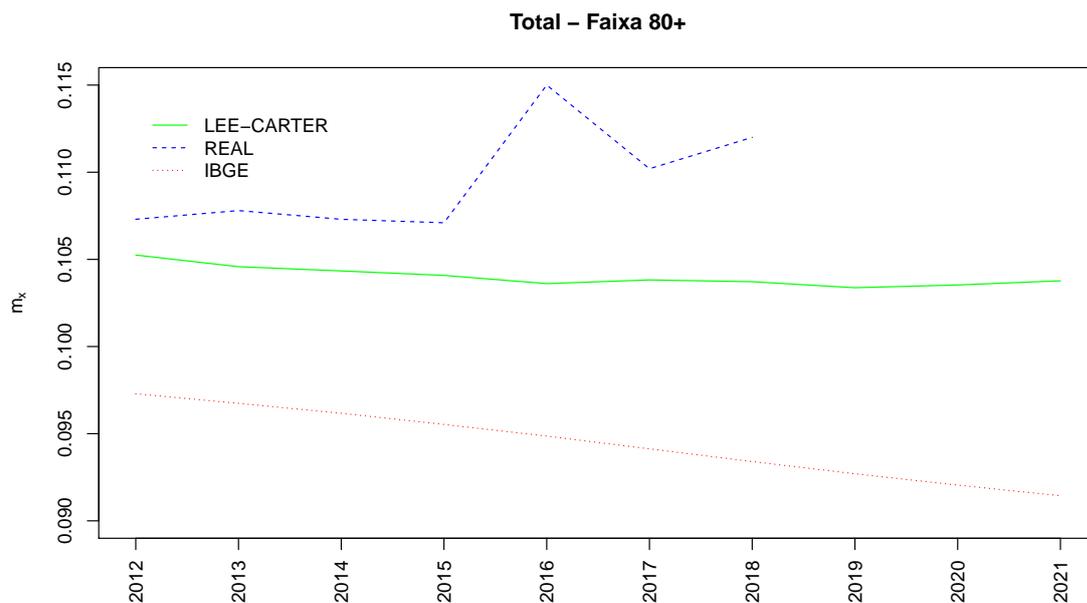


Figura 5.9: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos total.

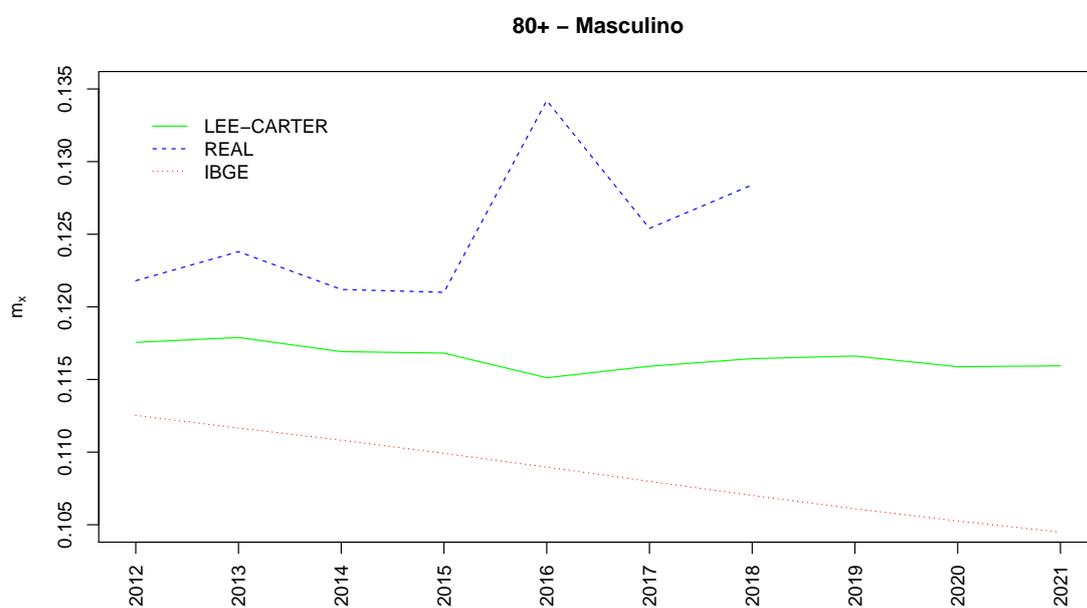


Figura 5.10: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos para o sexo masculino.

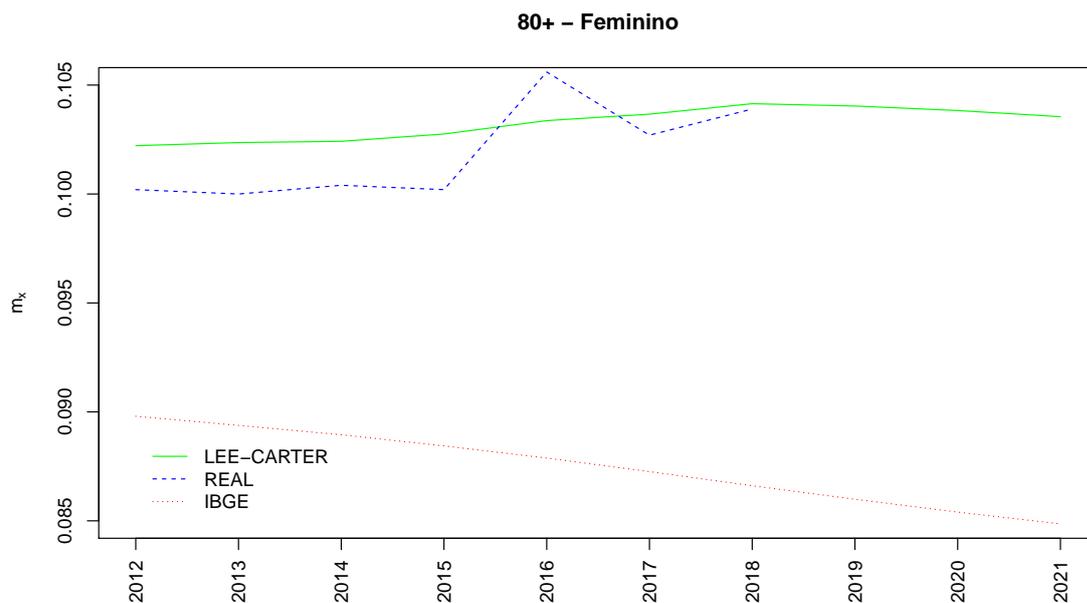


Figura 5.11: Comparação das taxas de mortalidade real, projetadas pelo modelo Lee-Carter e projetadas pelo IBGE para a faixa 80+ anos para o sexo feminino.

Nas Figuras 5.12, 5.13 e 5.14, temos as comparações da evolução de $m_{x,t}$ com o tempo e as respectivas divisões entre sexo feminino e sexo masculino. Há uma grande variação nos valores observamos e, de modo geral, percebemos que para algumas faixas a projeção do IBGE se adequa melhor e para outras a do modelo Lee-Carter.

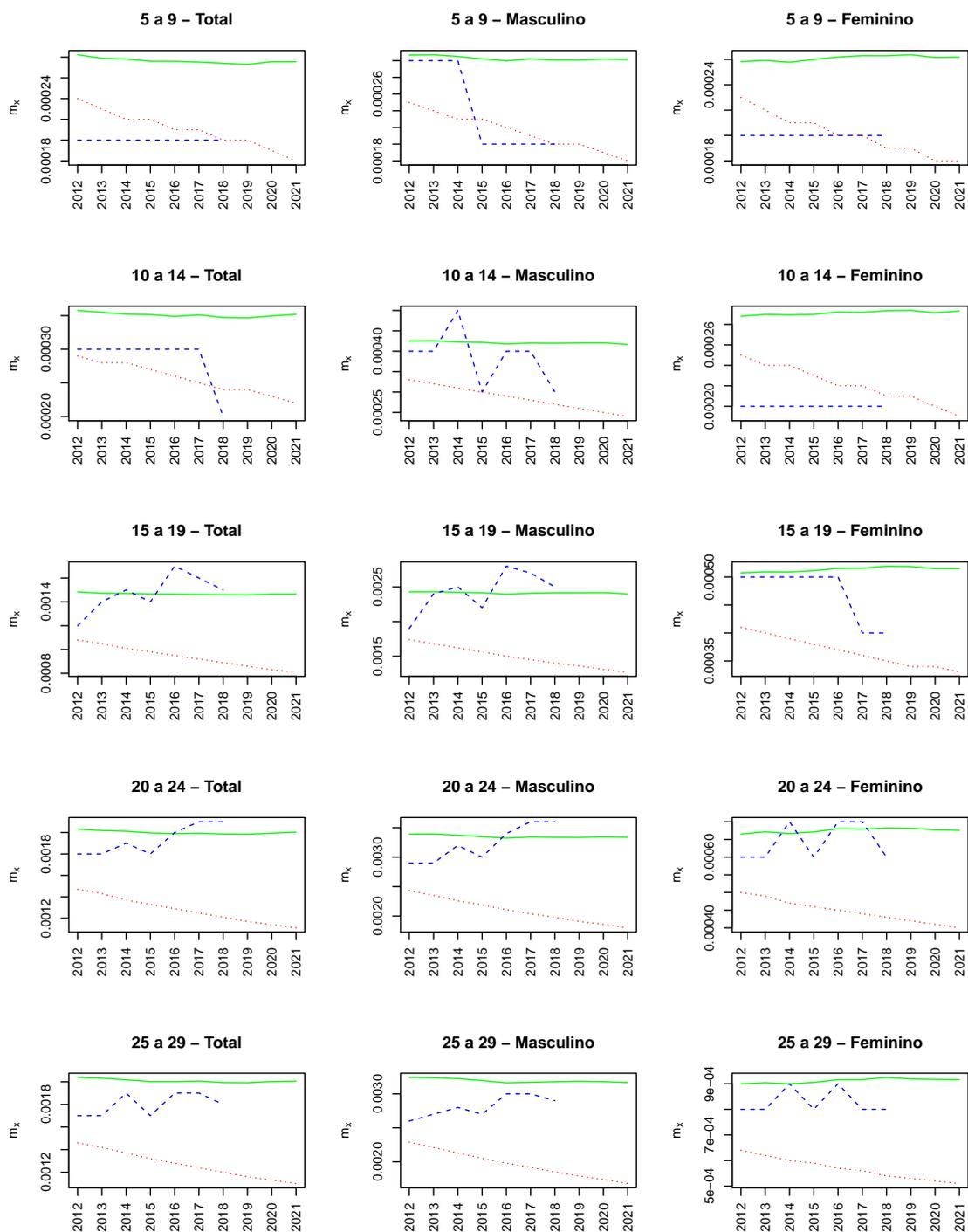


Figura 5.12: Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.

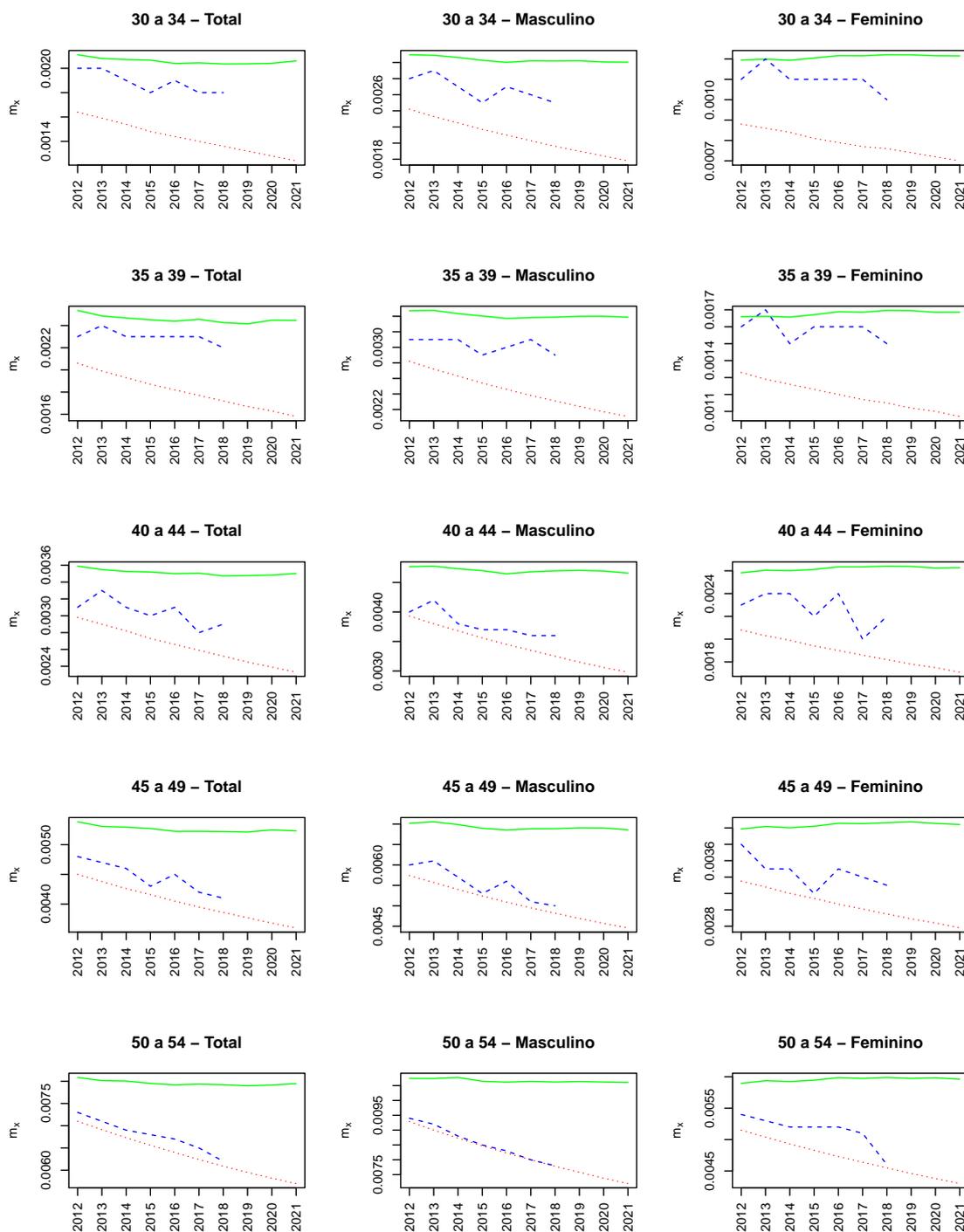


Figura 5.13: Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.

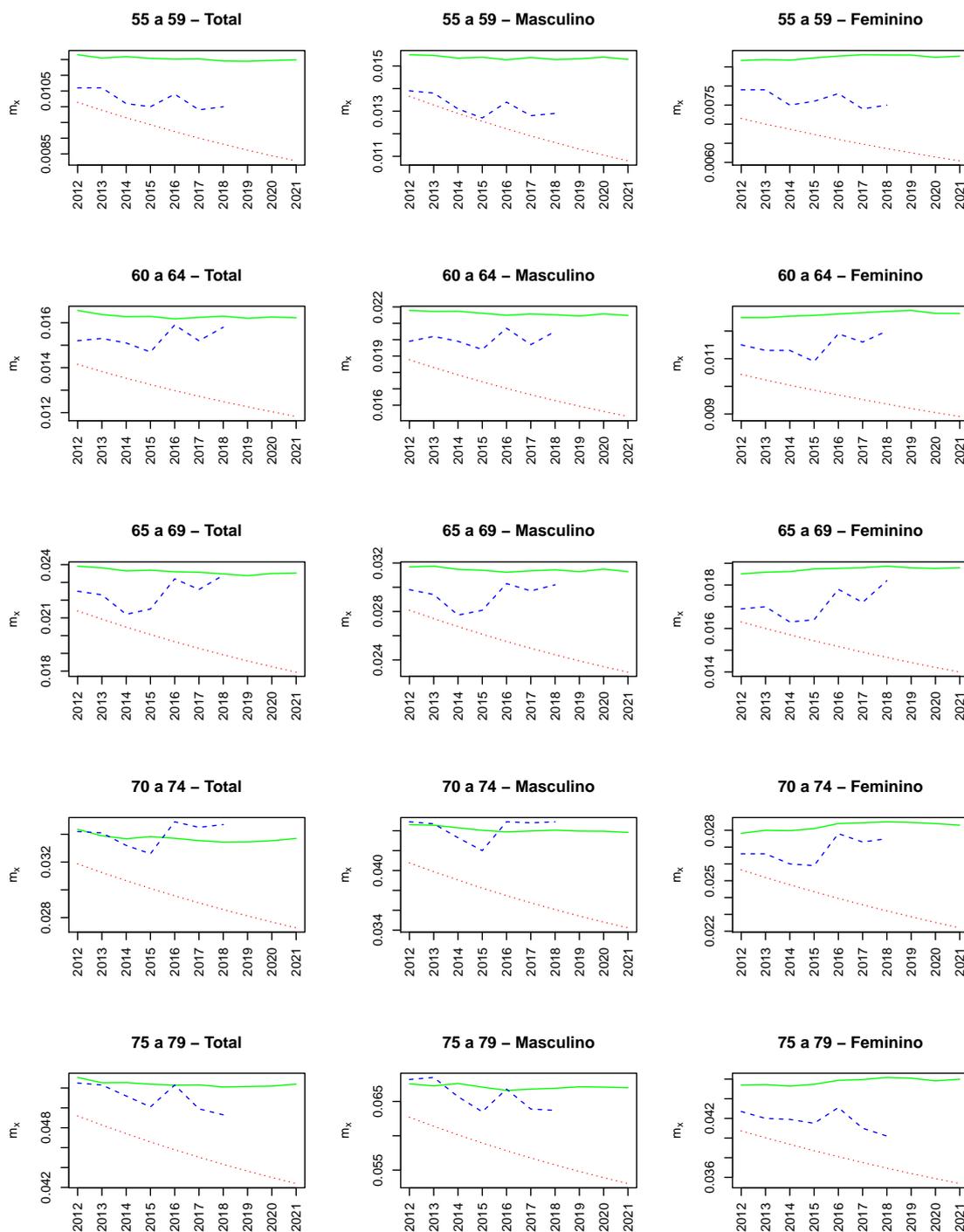


Figura 5.14: Comparação das taxas de mortalidade real (linha tracejada), projetadas pelo modelo Lee-Carter (linha cheia) e projetadas pelo IBGE (linha pontilhada) para diferentes anos e faixas etárias. Paineis à esquerda consideram o total; ao centro, sexo masculino; à direita, sexo feminino.

Capítulo 6

Conclusões

No presente trabalho, o modelo Lee-Carter bayesiano foi utilizado para modelar e projetar as taxas de mortalidade de indivíduos do Estado do Rio de Janeiro. Os dados disponíveis correspondiam a taxas de mortalidade anuais total e divididas por sexo ao longo de 21 anos (de 1990 a 2011) para diversas faixas etárias. O modelo foi ajustado e os parâmetros foram estimados por meio de métodos de simulação estocástica, especificamente o MCMC. Apesar de o conjunto de dados não cobrir tantos anos, após a análise dos resultados, percebe-se que ainda assim a resposta do modelo Lee-Carter bayesiano foi positiva quando comparado ao projetado pelo IBGE e o real observado.

Sobre as possíveis melhorias, [Sartório \(2018\)](#) propõe a utilização de um parâmetro β_x dinâmico com o tempo, pois, segundo ele, poderia capturar as transformações da população americana ao longo do tempo. Analisamos que, para os dados do Brasil, em particular Rio de Janeiro, o modelo bayesiano sem essa modificação foi suficiente para absorver de forma satisfatória as mudanças temporais ocorridas.

A grande desvantagem do método Lee-Carter é não considerar o avanço tecnológico, da área de saúde e impactos da violência, por exemplo. Para estudos futuros é válida a inserção de uma variável que influencie nesse sentido.

Esse trabalho demonstra a facilidade da aplicação do método Lee-Carter bayesiano, proporcionando uma melhor atualização das taxas centrais de mortalidade para os dados estudados, mesmo que o histórico de informações não seja tão extenso. Nota-se que o método resultou em maiores taxas de mortalidade do que o IBGE, mas para um cenário pessimista em um mercador segurador, é uma boa consideração de taxas a ser utilizada.

É preciso ressaltar que os dados da população total do Rio de Janeiro não necessariamente refletirá no cenário de uma seguradora, visto que a base de clientes da mesma pode ser voltada para uma classe específica da população, ocasionando em resultados diferentes. Sendo assim, é necessária a utilização de uma base de dados da própria seguradora, com base no vivenciado pela mesma. Resultando em uma tarifa mais coerente com a carteira da companhia.

Uma possível extensão ao trabalho é a análise espacial dos estados do Brasil, visto que apenas dentro do estado do Rio de Janeiro os dados não seriam tão volumosos e corretamente alocados para alcançar um resultado satisfatório de mudança comportamental na taxa de mortalidade central dependendo da região do estado. Nesse contexto, modelos espaço-temporais poderiam ser úteis na modelagem da correlação espacial entre as regiões, além da temporal já naturalmente tratada.

Referências Bibliográficas

- Booth, H., Hyndman, R. J., Tickle, L. e De Jong, P. (2006) Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, **15**, 289–310.
- Carter, C. K. e Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.
- Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of time series analysis*, **15**, 183–202.
- Geman, S. e Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 721–741.
- Lee, R. D. e Carter, L. R. (1992) Modeling and forecasting US mortality. *Journal of the American statistical association*, **87**, 659–671.
- Marcílio, M. L. (1993) Mortalidade e morbidade da cidade do Rio de Janeiro imperial. *Revista de História*, 53–68.
- Pedroza, C. (2006) A Bayesian forecasting model: predicting US male mortality. *Bio-statistics*, **7**, 530–550.
- Prata, P. R. A. (1992) Transição epidemiológica no Brasil. *Cadernos de Saúde Pública*, **v.8**, p.168–175.

- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Sartório, V. S. (2018) Modelagem da evolução de mortalidade considerando dinâmicas temporais etárias específicas. *Monografia, Escola Nacional de Ciências Estatísticas*.
- Souza, M. d. F. M. d., Malta, D. C., França, E. B. e Barreto, M. L. (2018) Transição da saúde e da doença no Brasil e nas Unidades Federadas durante os 30 anos do Sistema Único de Saúde. *Ciência & Saúde Coletiva*, **23**, 1737–1750.
- Wachter, K. W. (2014) *Essential demographic methods*. Harvard University Press.
- West, M. e Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 2nd edn.

Apêndice A

Distribuições condicionais completas

Abaixo são apresentadas as distribuições condicionais completas de todos os parâmetros envolvidos da especificação do modelo Lee-Carter.

$$1) \theta | \mathbf{y}, \Theta_{-\theta} \sim N \left(\frac{C_\theta \sum_{t=1}^T (k_t - k_{t-1}) + m_\theta w}{TC_\theta + w}, \frac{wC_\theta}{TC_\theta + w} \right)$$

$$2) k_0 | \mathbf{y}, \Theta_{-k_0} \sim N \left(\frac{(k_1 - \theta)C_0 + m_0 w}{C_0 + w}, \frac{wC_0}{C_0 + w} \right)$$

$$3) \sigma_w^2 | \mathbf{y}, \Theta_{-\sigma_w^2} \sim GI \left(\frac{T}{2} + c, \frac{\sum_{t=1}^T (k_t - k_{t-1} - \theta)^2}{2} + d \right)$$

$$4) \sigma_\epsilon^2 | \mathbf{y}, \Theta_{-\sigma_\epsilon^2} \sim GI \left(\frac{NT}{2} + a, \frac{\sum_{x \in X} \sum_{t=1}^T (y_{x,t} - \alpha_x - \beta_x k_t)^2}{2} + b \right)$$

$$5) \beta_x | \mathbf{y}, \Theta_{-\beta_x} \sim N \left(\frac{C_\beta \sum k_t (y_{x,t} - \alpha_x) + \sigma_\epsilon^2 m_\beta}{\sigma_\epsilon^2 + C_\beta \sum k_t^2}, \frac{\sigma_\epsilon^2 C_\beta}{\sigma_\epsilon^2 + C_\beta \sum k_t^2} \right)$$

$$6) \alpha_x | \mathbf{y}, \Theta_{-\alpha_x} \sim N \left(\frac{C_\alpha \sum (y_{x,t} - \beta_x k_t) + \sigma_\epsilon^2 m_\alpha}{\sigma_\epsilon^2 + TC_\alpha}, \frac{\sigma_\epsilon^2 C_\alpha}{\sigma_\epsilon^2 + TC_\alpha} \right)$$

Apêndice B

Traços das cadeias

As Figuras [B.1](#) e [B.2](#) apresentam os traços das cadeias de alguns parâmetros do modelo Lee-Carter. Conforme pode ser observado, o gráficos sugerem a convergência das cadeias.

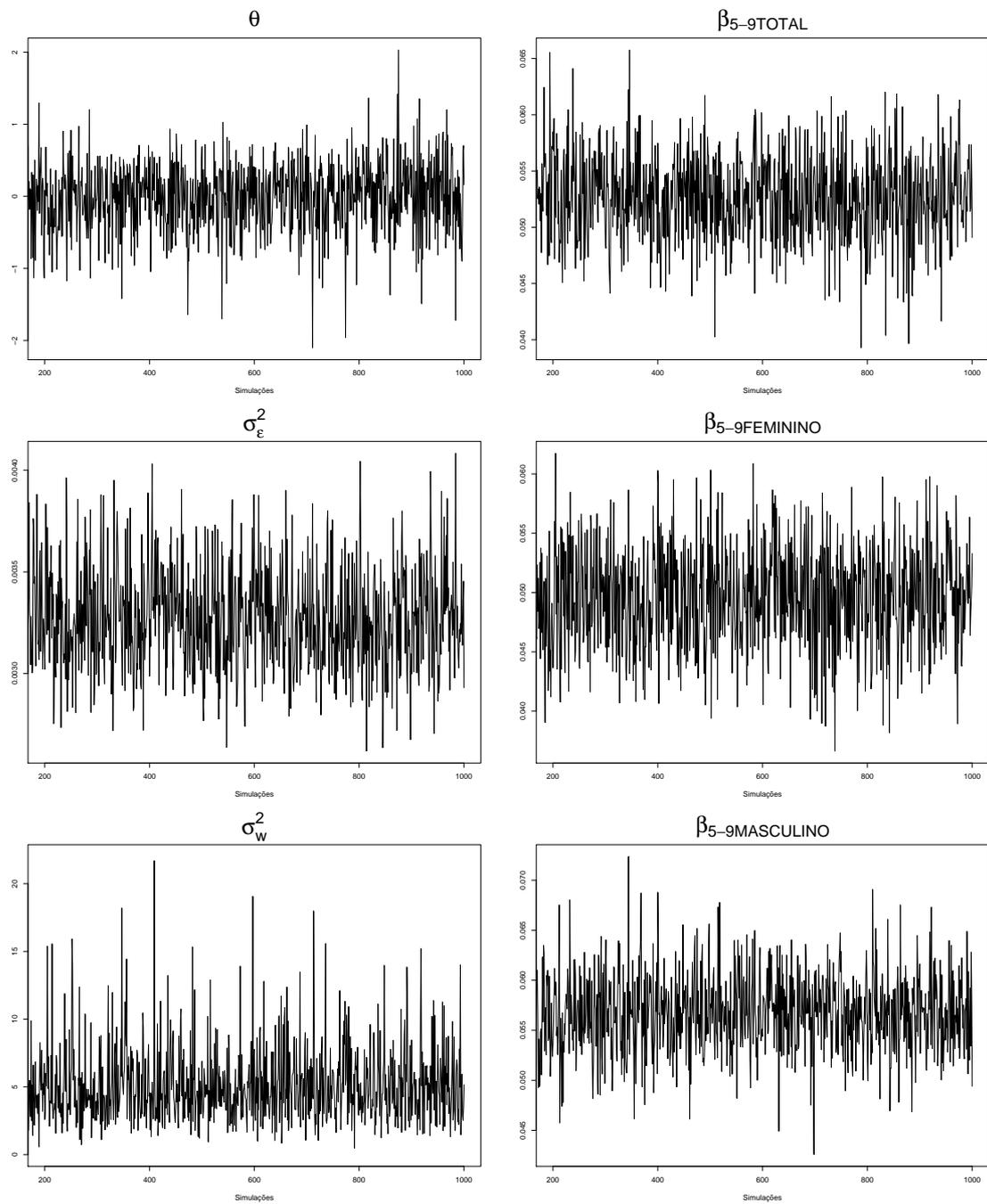


Figura B.1: Traços das cadeias dos parâmetros θ , σ_{ϵ}^2 , σ_w^2 e β_{5-9} total, feminino e masculino do modelo Lee-Carter bayesiano.

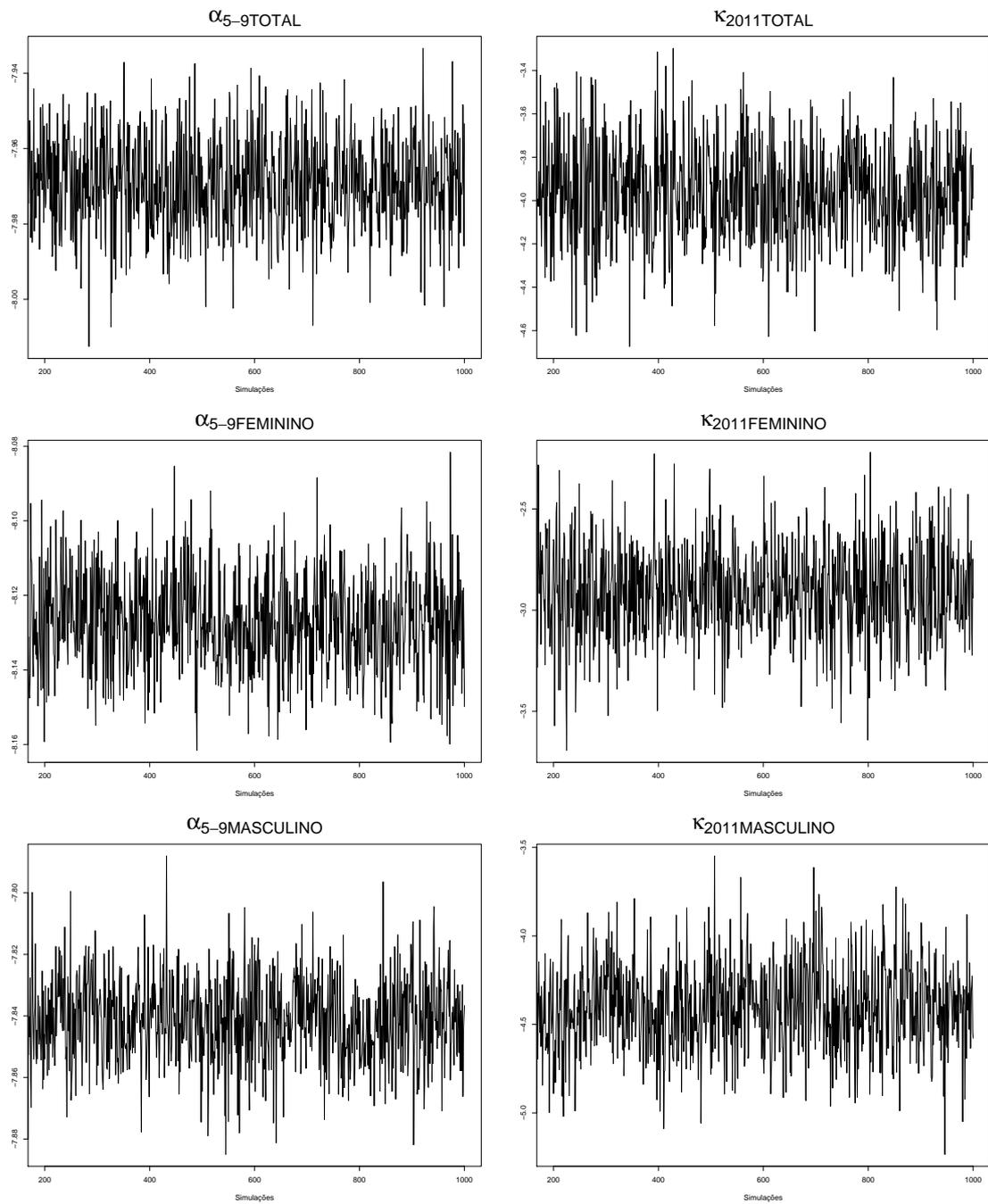


Figura B.2: Traços das cadeias dos parâmetros α_{5-9} e k_{2011} total, feminino e masculino do modelo Lee-Carter bayesiano.