

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340979229>

# Modelos de Espaço de Estados: Abordagens Clássica e Bayesiana

Book · April 2020

CITATIONS

0

READS

89

1 author:



[Thiago Rezende](#)

Federal University of Minas Gerais

36 PUBLICATIONS 275 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Modelos de espaço de estados para os dados de armadilhas do Aedes aegypti [View project](#)



Um aplicativo com previsões automáticas para o COVID-19 [View project](#)

# MODELOS DE ESPAÇO DE ESTADOS: ABORDAGENS CLÁSSICA E BAYESIANA

Glaura C. Franco  
Departamento de Estatística  
Universidade Federal de Minas Gerais  
<http://www.est.ufmg.br/~glaura>

Dani Gamerman  
Instituto de Matemática  
Universidade Federal do Rio de Janeiro  
<http://dme.ufrj.br/dani>

Thiago Rezende dos Santos  
Departamento de Estatística  
Universidade Federal de Minas Gerais  
[thiagords@ufmg.br](mailto:thiagords@ufmg.br)

*Aos nossos amores*

# Prefácio

Este texto surgiu de uma parceria entre uma estatística clássica e um estatístico Bayesiano na orientação da dissertação de mestrado do terceiro autor deste texto. Poder-se-ia esperar deste trabalho um embate entre estas duas filosofias, com cada um dos autores tentando defender seu ponto de vista. Isso quase certamente levaria a um texto confuso e possivelmente contraditório. Felizmente, os autores decidiram não tomar posições radicais. Vale dizer aqui que o pragmatismo e o não alinhamento do terceiro autor (resistindo a todas as pressões ao longo do trabalho) foi importante para que prevalecesse o clima de harmonia entre os diferentes pontos de vista. Assim, o que o leitor irá encontrar ao longo do texto é um apanhado de parte do que é possível realizar, tanto do ponto de vista clássico, quanto do ponto de vista Bayesiano, no ajuste de séries temporais através de modelos de espaço de estados (MEE).

Este trabalho dedica-se à discussão, em um nível intermediário entre final da Graduação e início de um Programa de Mestrado em Estatística, dos modelos de espaço de estados para a modelagem de séries temporais, utilizando os enfoques clássico e Bayesiano. Estes procedimentos são mais conhecidos, sob o ponto de vista clássico, como modelos estruturais, enquanto que no contexto Bayesiano eles são comumente chamados de modelos dinâmicos. A escolha que fizemos da nomenclatura de modelos de espaço de estados não foi casual; ela é parte do compromisso supracitado de não privilegiar algum dos pontos de vista.

O foco principal está voltado para a comparação dos métodos de estimação clássicos e Bayesianos para se fazer inferência sobre os parâmetros do modelo. Para tanto, utilizamos as técnicas computacionais *bootstrap*, para o enfoque clássico, e *Markov chain Monte Carlo* (MCMC), para o enfoque Bayesiano. Assim, serão apresentados estimadores pontuais assim como a construção de testes de hipóteses e intervalos de confiança assintóticos, *boot-*

*strap* e intervalos de credibilidade para os parâmetros. Existem vários estudos que aplicam o *bootstrap* em MEE sob a abordagem clássica, assim como vários trabalhos que aplicam os métodos MCMC em MEE sob a abordagem Bayesiana, porém não temos conhecimento de muitos trabalhos comparando as duas abordagens. Essa lacuna na literatura foi uma das principais motivações para a elaboração desse texto.

Este trabalho também apresenta uma extensão da metodologia para tratar séries com estrutura não-linear. Exemplos incluem séries que possuem quebras estruturais, utilizando funções de transferência (FT), com parâmetros fixos e variando no tempo para as variáveis de intervenção. A parte final destina-se à apresentação dos modelos de resposta não-Gaussiana e a introdução de uma ampla família de modelos de nível local não-Gaussianos, assim como casos particulares dessa família.

Em todos os tópicos abordados serão apresentados aplicativos, construídos na linguagem Ox, para implementação das metodologias descritas. Serão ilustradas também aplicações a dados reais em cada uma das situações apresentadas.

Este trabalho apresenta apenas um panorama de algumas das várias possibilidades de aplicação de MEE, pois esta é uma área muito vasta e que ainda pode ser bastante explorada em várias outras direções não tratadas aqui por nós. Fornecemos inúmeras referências onde o leitor interessado poderá aprofundar seu conhecimento sobre os temas discutidos. Nosso intuito foi repassar os conhecimentos que adquirimos ao longo do caminho que trilhamos. Esperamos que esse livro estimule outros pesquisadores a encontrar seus próprios caminhos dentro do universo dos MEE.

Desejamos agradecer à Comissão Organizadora da 13ª ESTE pela oportunidade a nós concedida. Gostaríamos também de agradecer à Fundação Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Belo Horizonte (IPEAD) e ao prof. Paulo S. Lucio (Estatística-UFRN) por ceder alguns dados utilizados nas aplicações a séries reais. Finalmente, não podemos nos esquecer da grande ajuda recebida pelas agências de pesquisa CNPq, CAPES, FAPEMIG e FAPERJ no desenvolvimento de nossas pesquisas.

BH e RJ, 31 de maio de 2009.

# Sumário

<b>Prefácio</b>	<b>iii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Modelos de Espaço de Estados</b>	<b>6</b>
2.1 Definição . . . . .	6
2.2 Alguns modelos específicos . . . . .	7
2.3 A forma de espaço de estados . . . . .	11
2.3.1 A forma de espaço de estados do MNL . . . . .	13
2.3.2 A forma de espaço de estados do MTL . . . . .	13
2.3.3 A forma de espaço de estados do MEB . . . . .	13
2.4 Filtro de Kalman . . . . .	14
2.5 Suavização . . . . .	16
2.6 Estimação dos hiperparâmetros . . . . .	17
2.6.1 Estimador de máxima verossimilhança . . . . .	17
2.6.2 Estimador de Bayes . . . . .	18
2.7 Previsão . . . . .	20
2.8 Aplicação a séries reais . . . . .	21
2.8.1 Série do IPCA . . . . .	22
2.8.2 Série do ICV . . . . .	22
2.8.3 Série da precipitação de SO <sub>4</sub> . . . . .	23
<b>3 Implementação Computacional</b>	<b>25</b>
3.1 Método de otimização BFGS . . . . .	25
3.2 Métodos de simulação MCMC . . . . .	31
3.3 <i>Bootstrap</i> . . . . .	40
3.4 Estudos de simulação . . . . .	43

<b>4</b>	<b>Intervalos e Testes de Hipóteses</b>	<b>46</b>
4.1	Intervalos de confiança . . . . .	47
4.1.1	Intervalos de confiança assintóticos . . . . .	47
4.1.2	Intervalos de confiança <i>bootstrap</i> . . . . .	51
4.1.3	Intervalos de credibilidade . . . . .	54
4.1.4	Estudos de simulação . . . . .	54
4.2	Testes de hipóteses . . . . .	56
4.2.1	Teste <i>bootstrap</i> usando a estatística da razão de verossimilhança . . . . .	58
4.2.2	Teste <i>bootstrap</i> usando a estatística escore modificado . . . . .	59
4.2.3	Teste de Nyblom e Mäkeläinen . . . . .	60
4.2.4	Estudos de simulação . . . . .	60
4.2.5	Aplicação a séries reais . . . . .	64
<b>5</b>	<b>Modelos de Espaço de Estados Não-Lineares</b>	<b>67</b>
5.1	Uma estrutura geral . . . . .	68
5.2	Modelos condicionalmente lineares . . . . .	70
5.3	Função de transferência . . . . .	71
5.3.1	Análise de intervenção em modelos estruturais . . . . .	72
5.3.2	Inferência . . . . .	76
5.4	Estudos de simulação . . . . .	80
5.4.1	Modelo 1 . . . . .	80
5.4.2	Modelo 2 . . . . .	82
5.5	Aplicação a séries reais . . . . .	86
5.5.1	Série IPCA . . . . .	86
5.5.2	Série IBOVESPA . . . . .	89
<b>6</b>	<b>Modelos de Espaços de Estados Não-Gaussianos</b>	<b>92</b>
6.1	Modelos Lineares Dinâmicos Generalizados . . . . .	94
6.2	Família Gama de modelos dinâmicos . . . . .	95
6.2.1	Definição . . . . .	96
6.2.2	Procedimento de inferência . . . . .	98
6.3	Casos particulares da FGMD . . . . .	99
6.3.1	Modelo Poisson . . . . .	99
6.3.2	Modelo Gama . . . . .	101
6.3.3	Modelo Weibull . . . . .	102
6.3.4	Modelo Pareto . . . . .	103
6.3.5	Modelo Beta . . . . .	104

6.3.6	Modelo Normal com média conhecida . . . . .	104
6.3.7	Transformações de modelos . . . . .	105
6.4	Estudos de simulação . . . . .	106
6.5	Aplicação a séries reais . . . . .	108



# Capítulo 1

## Introdução

A metodologia de *modelos estruturais*, ou *modelos dinâmicos*, é uma das várias abordagens existentes para a modelagem de séries temporais. A premissa básica destes modelos consiste em admitir a existência de componentes não-observáveis de tendência, sazonalidade, ciclo e ruído aleatório.

A idéia dessa decomposição da série temporal surgiu nos trabalhos de Holt (1957) e Winters (1960), que desenvolveram as técnicas de alisamento exponencial. Aproveitando essa idéia, na década de 60 surgiram alguns trabalhos formalizando a metodologia de modelos estruturais, dentre os quais pode-se citar os de Muth (1960), Theil & Wage (1964) e Nervole & Wage (1964). Já na década de 70 surgiram os primeiros modelos de previsão Bayesianos utilizando a modelagem dinâmica, nos trabalhos de Harrison & Stevens (1971, 1976). Entretanto, mediante a dificuldade computacional da época e o aparecimento dos modelos ARIMA de Box & Jenkins (1976), procedimentos utilizando a idéia de decomposição em componentes não-observáveis só voltaram a ser desenvolvidos no final da década de 80.

Apesar do sucesso obtido com a modelagem de Box & Jenkins, o problema da interpretação real da estrutura ARIMA identificada para uma série, assim como a restrição de parâmetros constantes, motivaram os estatísticos clássicos a procurar alternativas metodológicas que melhor descrevessem o mecanismo do sistema gerador da série. Embora estes problemas não ocorressem nos procedimentos Bayesianos, a implementação prática do modelo requeria do usuário o conhecimento de quantidades de difícil interpretação.

Assim, do ponto de vista clássico, com a criação do *aplicativo* STAMP (*Structural Time Series Analyser, Modeller and Predictor*) no final da década de 80, e com a importante colaboração de Andrew Harvey (Harvey, 1989,

1993), os modelos estruturais tornaram-se mais populares. Do ponto de vista Bayesiano, o *aplicativo* BATS (*Bayesian Analysis of Time Series*) e os trabalhos de West, Harrison e colaboradores (West & Harrison, 1986; West, Harrison & Migon, 1985; West, Harrison & Pole, 1988) foram fundamentais para a disseminação e uso dos modelos dinâmicos. Atualmente existem outros aplicativos que fazem a estimação de modelos estruturais, tais como *SsfPack* versão 2.2<sup>1</sup>, disponível na linguagem *Ox* (Koopman, Shephard & Doornik, 1999) e *dln* (Petris, 2008), disponível na linguagem R (2005). Ambos disponibilizam várias rotinas como as de suavização e de estimação.

A principal vantagem dos modelos propostos a partir da década de 80 é que, ao invés de ajustar funções que descrevem os componentes não observáveis aos dados históricos, estes procedimentos procuram identificar as componentes básicas na série e o modelo resultante é obtido a partir da composição desses elementos. Esses modelos são então escritos na forma de espaço de estados. O modelo assim expresso permite a utilização do filtro de Kalman (Kalman, 1960), tanto na abordagem clássica quanto na Bayesiana, para a estimação sequencial do vetor de estado e das previsões requeridas. A diferença básica entre as duas abordagens é que, no caso clássico, as quantidades desconhecidas são estimadas através da maximização da função de verossimilhança. Já o procedimento Bayesiano permite ao usuário a obtenção, por meios subjetivos, de estimativas sequenciais destas quantidades desconhecidas ou estimação por meio de suas distribuições *a posteriori*.

Levando em conta a decomposição da série em componentes não-observáveis, alguns modelos específicos podem ser obtidos, tais como o modelo de nível local (MNL), o modelo de tendência linear local (MTL) e o modelo estrutural básico (MEB), que são construídos baseados nas suposições de independência, homoscedasticidade e normalidade dos erros. Além disto, várias estruturas para séries temporais podem ser obtidas utilizando-se a forma de espaço de estados.

Os livros de Harvey (1989) e West & Harrison (1997) são textos fundamentais sobre o assunto e fonte principal de referência para análises sob os pontos de vista clássico e Bayesiano, respectivamente. Um outro trabalho que apresenta as versões clássica e Bayesiana de modelos de espaço de estados é o livro de Souza (1989), onde o autor aborda os modelos citados acima, utilizando os *softwares* existentes na época (STAMP e BATS) para estimação

---

<sup>1</sup>Esse pacote está disponível em <http://www.ssfpack.com>

dos parâmetros, não fazendo, entretanto, comparações entre as duas abordagens. Além destes, existem ainda os trabalhos de Pereira e Migon (1985 e 1986) sobre o assunto.

Neste trabalho, estimações clássica e Bayesiana dos parâmetros dos modelos serão abordadas, fazendo-se uma comparação entre as duas metodologias. No contexto clássico, utiliza-se a estimação por máxima verossimilhança, através de procedimentos numéricos. No contexto Bayesiano, os métodos de *Markov chain Monte Carlo* (MCMC) são usados para se obter características da distribuição *a posteriori* dos parâmetros do modelo, como a média e os quantis. A implementação do MCMC segue as recomendações contidas no estudo comparativo que Reis, Salazar & Gamerman (2006) fizeram para modelos de espaço de estados. Todos os programas serão implementados na linguagem *Ox* (Doornik, 1999). Parte dos programas serão construídos utilizando algumas rotinas do pacote *SsfPack* versão 2.2 (Koopman, Shephard & Doornik, 1999).

Intervalos de confiança (clássicos) para os parâmetros serão construídos utilizando-se várias metodologias. A primeira delas é através da teoria assintótica que, para modelos normais, está relativamente bem desenvolvida. O *bootstrap* (Efron, 1979) é outra boa alternativa, principalmente se a distribuição dos ruídos não é conhecida, ou as amostras disponíveis são pequenas. Stoffer & Wall (1991) propuseram um procedimento que possibilita a aplicação da técnica *bootstrap* em modelos estruturais e este será o procedimento usado neste trabalho para a obtenção da série *bootstrap*. Os intervalos *bootstrap* empregados serão os mais utilizados na literatura, a saber, o percentílico, o *bootstrap-t*, o intervalo com correção de vício e o intervalo com correção de vício e aceleração (Efron & Tibshirani, 1993). Estimação intervalar para os parâmetros sob a ótica Bayesiana também será apresentada através do cálculo de intervalos de credibilidade. Comparações entre os vários tipos de intervalos, modelos e procedimentos de estimação serão apresentadas para alguns modelos importantes dessa classe seguindo Franco & Souza (2002), Franco *et al.* (2008) e Santos & Franco (2009).

Os métodos computacionais citados acima, a saber *bootstrap* e MCMC, são apenas algumas das formas de se implementar a inferência sobre os parâmetros dos modelos. Alguns procedimentos, como filtros de partículas (ver Carvalho *et al.* (2009) e artigos lá referenciados) e amostragem por importância (ver Durbin & Koopman (2000) e artigos lá citados), estão entre os métodos mais recentes, sendo computacionalmente mais rápidos. Porém, exigem um maior conhecimento do usuário para sua aplicação e ainda estão

em desenvolvimento. Tanto o *bootstrap* quanto o MCMC têm a seu favor o fato de serem confiáveis por já terem sido exaustivamente testados ao longo das últimas décadas, funcionam bem em variados contextos aplicados e são de fácil utilização.

Um outro tópico importante é a realização de testes de hipóteses para as variâncias dos ruídos, conhecidos como hiperparâmetros. Testes *bootstrap*, baseados nos trabalhos de Franco, Koopman & Souza (1999) e Franco & Souza (2002), serão utilizados.

Várias extensões visando relaxar as hipóteses de linearidade e observações normais foram propostas na literatura porém, na maioria dos casos, a tratabilidade analítica é perdida. Diz-se que um modelo de espaço de estados é não-linear se a dependência nos parâmetros não se dá de forma linear. As abordagens mais comuns são baseadas no filtro de Kalman estendido (West & Harrison, 1997) e suas inúmeras variantes. Essa extensão consiste basicamente em uma linearização baseada na expansão de Taylor e também é apresentada nesse trabalho. Uma variante de modelos não-lineares consiste nos modelos chamados de condicionalmente lineares, onde alguns dos termos dependem de quantidades desconhecidas. A importante família de modelos de função de transferência pode ser colocada dentro dessa classe e uma comparação dos métodos de inferência será exibida, usando resultados de Santos, Franco & Gamerman (2009).

Já para os casos nos quais as suposições de normalidade e/ou homoscedasticidade são violadas, uma solução natural é tentar uma transformação nos dados. Porém, nem sempre uma transformação é suficiente para que a suposição de normalidade seja satisfeita. Logo, outra possibilidade é construir modelos com resposta não-Gaussiana (Smith & Miller, 1986), o que inclusive traz um ganho de interpretação do modelo (West, Harrison & Migon, 1985; West & Harrison, 1997). Uma estrutura mais geral, denominada *modelos lineares dinâmicos generalizados* (MLDG), foi proposta por West, Harrison & Migon (1985), despertando grande interesse devido à aplicabilidade dos mesmos em diversas áreas do conhecimento. Exemplos de modelos não-normais serão apresentados, ilustrando a situação onde ainda é possível obter analiticamente a verossimilhança marginal dos parâmetros. Isso é possível devido à forma de evolução proposta por Smith & Miller (1986). Assim, séries temporais Poisson e exponenciais são apresentadas, incluindo uma estrutura de regressão (Harvey & Fernandes, 1989).

Este texto está organizado da seguinte forma. No Capítulo 2, a metodologia dos modelos de espaço de estados, com exemplos de algumas séries reais

é apresentada. Os métodos computacionais *bootstrap* e MCMC são descritos no Capítulo 3, que também exhibe uma comparação empírica entre os métodos clássicos e Bayesianos para se fazer inferência sobre os parâmetros dos modelos. No Capítulo 4 métodos de obtenção de estimativas intervalares e testes de hipóteses para os parâmetros são apresentados. No Capítulo 5 é abordado o tópico de modelos não-lineares, com a metodologia sendo aplicada a funções de transferência. Finalmente, no Capítulo 6 são apresentados modelos dinâmicos não-Gaussianos e alguns casos particulares de interesse.

## Capítulo 2

# Modelos de Espaço de Estados

Os modelos de espaço de estados (MEE) vêm sendo cada vez mais utilizados para modelagem e previsão em séries temporais. Esse avanço se deu notadamente após o lançamento dos livros de Harvey (1989) e West & Harrison (1997), onde a teoria de MEE foi exaustivamente descrita. Este tipo de modelo supõe que os movimentos característicos de uma série temporal  $\{y_t\}, t = 1, \dots, n$ , podem ser decompostos em componentes não-observáveis, como por exemplo, tendência, sazonalidade, componente cíclica e componente aleatória ou erro. A vantagem deste procedimento é que os componentes têm uma interpretação direta, devido à maneira que o modelo é construído. Além disto, a flexibilidade dessa modelagem, incluindo a capacidade de lidar com dados multivariados e processos não-estacionários, confere-lhes uma significativa vantagem frente às demais metodologias de análise de séries temporais (Spall, 1988).

Os modelos estruturais são geralmente escritos na forma de espaço de estados, para possibilitar o uso do filtro de Kalman como ferramenta básica para a estimação e previsão. Inferências sobre os parâmetros do modelo podem ser feitas tanto utilizando métodos clássicos quanto Bayesianos, e a comparação entre estes procedimentos é o tema principal deste capítulo.

### 2.1 Definição

Uma série temporal  $\{y_t\}, t = 1, \dots, n$  pode ser decomposta em componentes não-observáveis, sendo que os componentes que ocorrem com maior frequência na prática são:

1. Tendência ( $\mu_t$ ): refere-se à direção geral segundo a qual o gráfico da série se desenvolve ao longo de um intervalo de tempo.

2. Componente Sazonal ( $\gamma_t$ ): refere-se a padrões semelhantes que uma série temporal parece obedecer durante sucessivos instantes de tempo, resultantes de eventos periódicos que ocorrem repetidamente.

3. Componente cíclica ( $\delta_t$ ): refere-se às oscilações em longo prazo ou aos desvios em torno da reta ou da curva de tendência, segundo algum padrão.

4. Componente aleatória ou erro ( $\epsilon_t$ ): refere-se aos deslocamentos esporádicos das séries temporais, provocados por eventos casuais.

Assim, uma série temporal univariada  $\{y_t\}$ ,  $t = 1, 2, \dots, n$  pode ser escrita da seguinte forma:

$$y_t = \mu_t + \gamma_t + \delta_t + \epsilon_t,$$

sendo  $\epsilon_t \sim [0, \sigma_\epsilon^2]$ , independentes entre si.

## 2.2 Alguns modelos específicos

Neste estudo, abordam-se os modelos de nível local (MNL), tendência linear local (MTL), de tendência polinomial (MTP) e estrutural básico (MEB). O modelo com o componente de ciclo foge ao escopo deste estudo e não será apresentado. Mais detalhes sobre esses modelos podem ser encontrados em West & Harrison (1997) e Harvey (1989).

### Modelos de Nível Local (MNL)

O modelo de nível local, ou modelo linear dinâmico (MLD) de primeira ordem, é o mais simples, porque possui apenas o componente de nível ( $\mu_t$ ) e do erro ( $\epsilon_t$ ). Ele consiste em um passeio aleatório acrescido de um erro. A série se movimenta ao longo do tempo sem uma trajetória fixa, pois o nível atual é igual ao nível no período anterior acrescido de um ruído branco. A forma do modelo é dada por:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad (2.1)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (2.2)$$

com  $\epsilon_t$  e  $\eta_t$  não-correlacionados e  $t = 1, \dots, n$ .

Um exemplo de série real que parece apresentar o comportamento de um MNL é exibido na Figura 2.1. Os dados referem-se ao Índice de Preço ao

Consumidor Amplo de Belo Horizonte (IPCA), no período de julho de 1997 a junho de 2008. Esse índice mede a evolução da renda em famílias que gastam de 1 a 40 salários mínimos por mês. A menos de um salto em meados de outubro de 2002, série parece oscilar em torno de um nível constante.

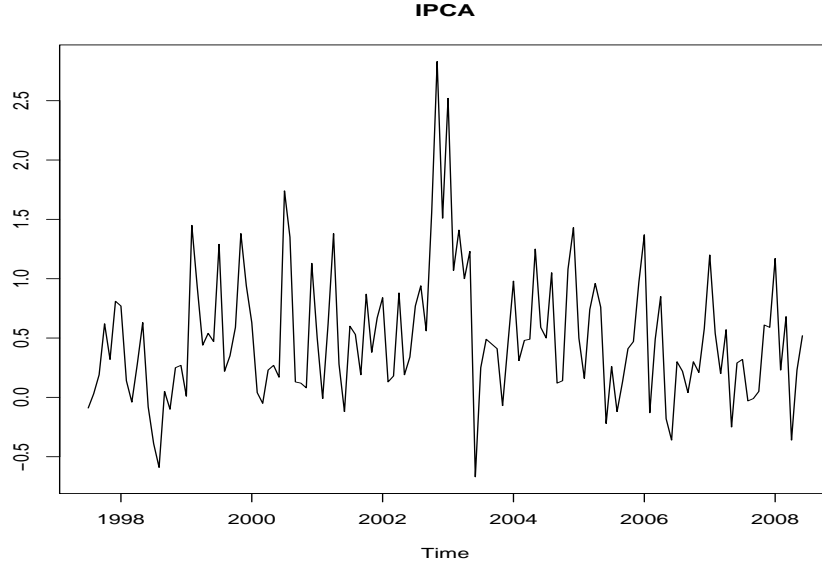


Figura 2.1: Série do IPCA

### Modelo de Tendência Linear Local (MTL)

Este modelo também é conhecido como modelo linear dinâmico (MLD) de segunda ordem. A tendência em uma série pode ser observada quando existe um aumento ou decréscimo gradual do nível das observações ao longo do tempo. Para  $t = 1, \dots, n$ , tem-se:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \xi_t, & \xi_t &\sim N(0, \sigma_\xi^2), \end{aligned}$$

onde  $\epsilon_t$ ,  $\eta_t$  e  $\xi_t$  são erros aleatórios mutuamente não-correlacionados.



### Modelo de Tendência Polinomial (MTP)

O modelo de tendência polinomial é um modelo de tendência que engloba e generaliza os outros dois modelos apresentados anteriormente. O *modelo de tendência polinomial de ordem  $k$*  pode ser definido como:

$$\begin{aligned} y_t &= \alpha_{t,1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \\ \alpha_{t,j} &= \alpha_{t-1,j} + \alpha_{t,j+1} + \eta_{t,j}, \quad (j = 1, \dots, k-1), \\ \alpha_{t,k} &= \alpha_{t-1,k} + \eta_{t,k}, \end{aligned}$$

onde  $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$  e  $\mathbf{Q}_t$  é uma matriz de covariâncias diagonal com entradas  $\sigma_1^2, \dots, \sigma_k^2$ , para  $t = 1, \dots, n$ .

Se  $k = 1$  e  $\alpha_t = \mu_t$ , o MNL é obtido. O MTL também é um caso particular desse modelo quando  $k = 2$  e  $\boldsymbol{\alpha}_t = (\mu_t, \beta_t)'$ . Já o *modelo linear dinâmico de tendências quadráticas* pode ser obtido fazendo  $k = 3$  e  $\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \iota_t)'$ .

Na Figura 2.2 apresenta-se um exemplo de série real com uma tendência crescente ao longo do tempo. Esta série representa o logaritmo do Índice do Custo de Vida (ICV) no município de São Paulo, no período de janeiro de 1970 a junho de 1980.

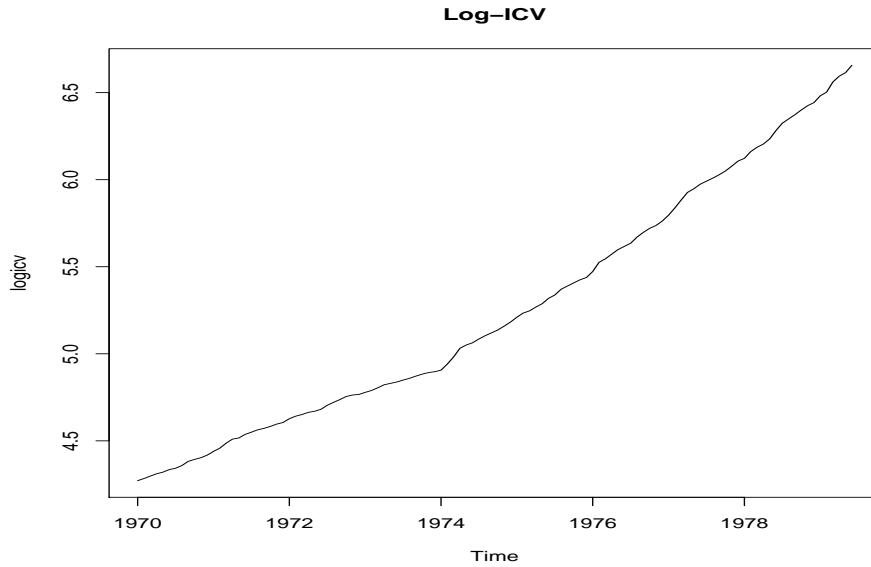


Figura 2.2: Série do logaritmo do ICV

### Modelo Estrutural Básico (MEB)

O componente sazonal deve ser acrescentado ao modelo quando a série apresenta algum tipo de periodicidade que ocorre regularmente, por exemplo de ano em ano. Portanto, as séries sazonais são caracterizadas por apresentarem correlações altas em defasagens sazonais. O período sazonal, que se denotará por  $s$ , pode ser semanal com dados diários ( $s = 7$ ), mensal com dados diários ( $s = 30$ ), trimestral com dados mensais ( $s = 4$ ), ou, mais comumente, anual com dados mensais ( $s = 12$ ).

Há duas maneiras principais de se modelar a sazonalidade. A primeira é modelar a sazonalidade por fatores e, a segunda, por componentes harmônicos ou trigonométricas. A segunda não é discutida neste estudo, entretanto o leitor interessado pode consultar os trabalhos de Harvey (1989) e West & Harrison (1997).

O Modelo Estrutural Básico é o modelo de tendência linear local acrescido do componente sazonal ( $\gamma_t$ ) e é definido na forma:

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \xi_t, & \xi_t &\sim N(0, \sigma_\xi^2), \\ \gamma_t &= -\gamma_{t-1} - \dots - \gamma_{t-s+1} + \omega_t, & \omega_t &\sim N(0, \sigma_\omega^2), \end{aligned}$$

assumindo que  $t = 1, \dots, n$  e que  $s$  refere-se ao número de períodos sazonais.  $\epsilon_t$ ,  $\eta_t$ ,  $\xi_t$  e  $\omega_t$  são distúrbios tipo ruído branco mutuamente não-correlacionados.

A sazonalidade nada mais é que um fator de classificação com  $s$  níveis. Com a presença de outros fatores no modelo, impõe-se a restrição de que a soma dos componentes sazonais seja zero, isto é,  $\sum_{j=0}^{s-1} \gamma_{t-j} = 0$ , para garantir a identificação dos níveis sazonais. A modelagem desta forma assume que o fator é estático, o que pode não ser o mais apropriado para o contexto tipicamente dinâmico de séries temporais. Obtém-se uma modelagem estocástica para o componente sazonal fazendo  $\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$ , sendo  $\omega_t \sim N(0, \sigma_\omega^2)$ .

Um exemplo de série real com sazonalidade é mostrado na Figura 2.3, referente à série do logaritmo da precipitação de sulfato ( $SO_4$ ) em Nova York, EUA, no período de janeiro de 1985 a julho de 2007. Pode-se perceber que a série apresenta um componente de sazonalidade com uma leve tendência decrescente.

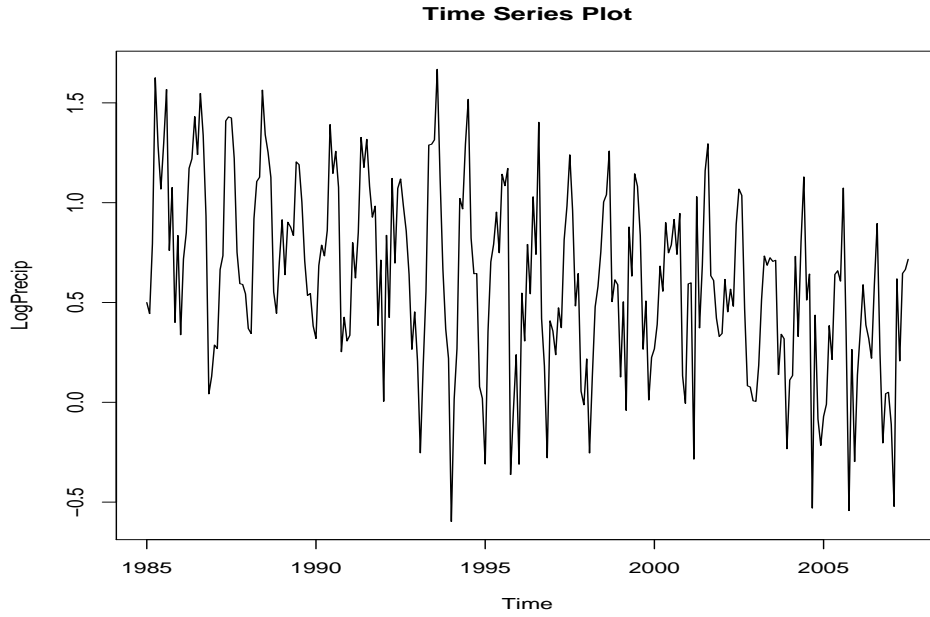


Figura 2.3: Logaritmo da Precipitação de SO4 em Nova York.

## 2.3 A forma de espaço de estados

A representação de espaço de estados de um sistema é um conceito fundamental na teoria de controle moderna (Wei, 1990). Modelos de espaço de estados foram originalmente desenvolvidos por engenheiros de controle, particularmente para aplicação em sistemas de navegação, como, por exemplo, no monitoramento da posição de objetos, como mísseis e foguetes. Mais tarde verificou-se que esses modelos eram também valiosos na análise de muitos outros tipos de séries temporais. De fato, a modelagem em espaço de estados é muito flexível e aplicável em uma grande variedade de problemas.

O modelo na forma de espaço de estados tem apenas duas equações - a equação das observações (ou de medida) e a equação do estado (ou de transição), dadas, respectivamente, por

$$y_t = \mathbf{z}_t' \boldsymbol{\alpha}_t + d_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \quad (2.3)$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t) \quad (2.4)$$

onde  $t = 1, 2, \dots, n$ ,  $\epsilon_t$  são ruídos não correlacionados,  $\boldsymbol{\eta}_t$  é um vetor de ruídos serialmente não-correlacionados, cuja matriz de covariâncias é dada por  $\mathbf{Q}_t$  e  $\boldsymbol{\alpha}_t$  é o vetor de estados. Assume-se também que  $\epsilon_t$  e  $\boldsymbol{\eta}_t$  sejam independentes

entre si para  $t = 1, 2, \dots, n$ . As matrizes  $\mathbf{z}_t$ ,  $\mathbf{T}_t$  e  $\mathbf{R}_t$  são conhecidas como matrizes do sistema e assumidas determinísticas e conhecidas. Os termos  $d_t$  e  $\mathbf{c}_t$  são covariáveis que podem ser inseridas nas equações das observações e de transição, respectivamente. O modelo de espaço de estados é dito invariante no tempo quando as matrizes de sistema não mudam ao longo do tempo. O modelo tem as seguintes suposições  $E(\epsilon_t \boldsymbol{\eta}_j) = \mathbf{0}$  e  $E(\boldsymbol{\eta}_t \boldsymbol{\alpha}_0) = \mathbf{0}$ ,  $\forall j, t = 1, \dots, n$ .

O modelo é completado com a especificação inicial  $\boldsymbol{\alpha}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$ , onde  $\boldsymbol{\alpha}_0$  é o vetor de estados inicial. As quantidades desconhecidas do modelo podem ser divididas em dois grupos: os parâmetros estáticos, denominados hiperparâmetros, como as variâncias dos erros associados a cada componente não-observável, e os parâmetros de estado  $\boldsymbol{\alpha}_t$ . Uma diferença entre as metodologias clássica e Bayesiana aparece aqui. A primeira abordagem considera os parâmetros de estados como variáveis latentes auxiliares cuja inferência não é objeto primordial de estudo. A segunda linha considera os parâmetros de estados como qualquer outro parâmetro e, portanto, são características para as quais deve-se realizar inferências.

Várias estruturas para séries temporais podem ser obtidas através das equações apresentadas acima. Por exemplo, pode-se adicionar variáveis explicativas, variáveis indicadoras para incluir informação sobre quebras estruturais, componentes de tendência e sazonalidade, modelos com erros normais e não normais, estruturas não lineares, dentre outros.

Uma das grandes vantagens dos modelos de espaço de estados decorre de sua natureza Markoviana, que permite que os cálculos sejam feitos recursivamente, de forma particularmente conveniente à computação. Segundo Durbin & Koopman (2001), essa natureza recursiva dos modelos e as técnicas computacionais usadas para analisá-los conferem aos modelos de espaço de estados grande flexibilidade, pois permitem que alterações na estrutura do sistema sejam feitas com relativa facilidade.

Outro ponto a favor da modelagem em espaço de estados é que ela permite trabalhar com observações faltantes e agregação temporal (Harvey, 1989). Além disso, modelos podem ser reformulados para tempo contínuo, habilitando-os, por exemplo, a tratar situações com observações irregularmente espaçadas no tempo.

Aqui é apresentada a forma de espaço de estados para séries temporais univariadas, mas esse modelo contempla também séries multivariadas. O interessado neste tópico pode consultar os livros de Harvey (1989) e West & Harrison (1997).

### 2.3.1 A forma de espaço de estados do MNL

Como foi visto na Seção 2.2, o modelo de nível local é o modelo mais simples, pois não segue nenhuma tendência fixa e é composto apenas por um componente de nível adicionado de um ruído. A forma de espaço de estados pode ser aplicada ao MNL, encontrando-se as seguintes quantidades para as matrizes de sistema e de covariância:

$$z'_t = 1, T_t = 1, R_t = 1, h_t = \sigma_\varepsilon^2 \text{ e } Q_t = \sigma_\eta^2.$$

O vetor de estados é dado por  $\alpha_t = \mu_t$ .

### 2.3.2 A forma de espaço de estados do MTL

Assumindo que a série temporal tenha um movimento crescente ou decrescente e também um comportamento de passeio aleatório, as seguintes componentes são obtidas:

$$\mathbf{z}'_t = \begin{bmatrix} 1 & 0 \end{bmatrix}, \mathbf{R}_t = \mathbf{I}_2, \mathbf{T}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, h_t = \sigma_\varepsilon^2 \text{ e } \mathbf{Q}_t = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix}.$$

O vetor de estados é dado por  $\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$ .

### 2.3.3 A forma de espaço de estados do MEB

Como pode ser claramente visto, o MEB foge à regra de dependência de primeira ordem para os parâmetros de estado, pois sua componente sazonal  $\gamma_t$  exibe dependência de ordem  $s - 1$ . Para poder reescrevê-lo na forma (2.3)-(2.4), é necessário aumentar o parâmetro de estado a cada tempo  $t$  pela inclusão de um ciclo sazonal completo na forma  $\alpha'_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \dots, \gamma_{t-s+1})$ .

Para os modelos com sazonalidade modelados como acima, a forma de espaço de estados é dada por:

$$\mathbf{z}'_t = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 \\ & & & & \sim & \end{bmatrix}_{1 \times (s+1)}, \mathbf{R}_t = \mathbf{I}_{s+1}$$

$$\mathbf{T}_t = \begin{bmatrix} 1 & 1 & | & & & 0 \\ 0 & 1 & | & & & \sim \\ - & - & | & - & - & - & - & - \\ & & | & -1 & -1 & \dots & -1 & -1 \\ 0 & & | & 1 & 0 & \dots & 0 & 0 \\ \sim & & | & 0 & 1 & \dots & 0 & 0 \\ & & | & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & | & 0 & 0 & \dots & 1 & 0 \end{bmatrix}_{(s+1) \times (s+1)},$$

$$\mathbf{Q}_t = \begin{bmatrix} \sigma_\eta^2 & & & & & \\ & \sigma_\xi^2 & & & & 0 \\ & & \sigma_\omega^2 & & & \sim \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & \sim & & & & 0 \end{bmatrix}_{(s+1) \times (s+1)} \quad \text{e } h_t = \sigma_\varepsilon^2.$$

## 2.4 Filtro de Kalman

Como o vetor de estados  $\boldsymbol{\alpha}_t$  é não-observável, é necessário um algoritmo recursivo para estimá-lo. O Filtro de Kalman (FK) (Kalman, 1960) é um algoritmo recursivo que determina a estimativa do vetor de estados no tempo  $t$  dada toda a informação disponível até o instante  $t - 1$ . Outros algoritmos recursivos poderiam ser escolhidos, mas o FK é preferido devido a algumas propriedades interessantes, sendo a principal delas o fato de ele coincidir com médias e variâncias *a posteriori* calculadas sob o ponto de vista Bayesiano.

Definindo que  $\mathbf{Y}_{t-1}$  é o vetor das observações até o instante  $t - 1$ ,  $E(\boldsymbol{\alpha}_0) = \mathbf{a}_0$ ,  $Cov(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_0) = \mathbf{P}_0$ , supondo-se que as observações têm distribuição Normal e utilizando as propriedades da Normal multivariada, tem-se que:

1.  $(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) \sim N(\mathbf{a}_{t|t-1}, \mathbf{P}_{t|t-1})$ , sendo  $\mathbf{a}_{t|t-1} = E(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) = \mathbf{T}_t \mathbf{a}_{t-1} + \mathbf{c}_t$  e  $\mathbf{P}_{t|t-1} = Var(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) = \mathbf{T}_t \mathbf{P}_{t-1} \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$ ,
2.  $(y_t | \mathbf{Y}_{t-1}) \sim N(\tilde{y}_{t|t-1}, F_t)$ , sendo  $\tilde{y}_{t|t-1} = E(y_t | \mathbf{Y}_{t-1}) = \mathbf{z}_t' \mathbf{a}_{t|t-1} + d_t$  e  $F_t = Var(y_t | \mathbf{Y}_{t-1}) = \mathbf{z}_t' \mathbf{P}_{t|t-1} \mathbf{z}_t + h_t$ .

Através de recursividade, encontra-se também que  $(\boldsymbol{\alpha}_t | \mathbf{Y}_t) \sim N(\mathbf{a}_t, \mathbf{P}_t)$

onde

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{z}_t' F_t^{-1} (y_t - \tilde{y}_{t|t-1}), \quad (2.5)$$

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{z}_t' F_t^{-1} \mathbf{z}_t \mathbf{P}_{t|t-1}. \quad (2.6)$$

As equações (2.5)-(2.6) são conhecidas como equações de atualização do filtro de Kalman, pois através delas é possível atualizar os estimadores  $\mathbf{a}_t$  e  $\mathbf{P}_t$  do vetor de estado no instante  $t$ . Deve-se observar que  $\mathbf{a}_t$  é um vetor de estimativas das componentes de estados. Por exemplo, para o MEB, o vetor de estados é dado por  $\boldsymbol{\alpha}_t' = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \dots, \gamma_{t-s+1})$ . Assim, as estimativas destas quantidades serão denotadas neste texto por  $\mathbf{a}_t^{(\mu)}$ ,  $\mathbf{a}_t^{(\beta)}$ ,  $\mathbf{a}_t^{(\gamma_1)}$ ,  $\mathbf{a}_t^{(\gamma_2)}$ , ...,  $\mathbf{a}_t^{(\gamma_s)}$ .

Para facilitar o manuseio das fórmulas, denota-se por  $\nu_t$  o erro de previsão um passo à frente ou inovação

$$\nu_t = y_t - \tilde{y}_{t|t-1} \Rightarrow \nu_t = y_t - \mathbf{z}_t' \mathbf{a}_{t|t-1} - d_t.$$

Assim,  $E(\nu_t | \mathbf{Y}_{t-1}) = 0$  e  $Var(\nu_t | \mathbf{Y}_{t-1}) = F_t$ . Então, substituindo  $\nu_t$  em (2.5)-(2.6) tem-se

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{T}_{t+1}^{-1} \mathbf{K}_t \nu_t$$

e

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{T}_{t+1}^{-1} \mathbf{K}_t \mathbf{z}_t' \mathbf{P}_{t|t-1},$$

onde

$$\mathbf{K}_t = \mathbf{T}_{t+1} \mathbf{P}_{t|t-1} \mathbf{z}_t' F_t^{-1}$$

é conhecida como matriz de ganho de Kalman. Para simplificar as equações do FK, pode-se escrever  $\mathbf{a}_{t+1|t}$  e  $\mathbf{P}_{t+1|t}$  diretamente de  $\mathbf{a}_{t|t-1}$  e  $\mathbf{P}_{t|t-1}$ . Logo, tem-se que

$$\mathbf{a}_{t+1|t} = \mathbf{T}_{t+1} \mathbf{a}_{t|t-1} + \mathbf{K}_t \nu_t + \mathbf{c}_{t+1} \quad (2.7)$$

$$\mathbf{P}_{t+1|t} = \mathbf{T}_{t+1} \mathbf{P}_{t|t-1} \mathbf{T}_{t+1}' - \mathbf{K}_t F_t \mathbf{K}_t' + \mathbf{R}_{t+1} \mathbf{Q}_{t+1} \mathbf{R}_{t+1}'. \quad (2.8)$$

O filtro de Kalman é inicializado com os valores de  $\mathbf{a}_0$  e  $\mathbf{P}_0$ . Escolhas típicas são dadas por  $\mathbf{a}_0 = \mathbf{0}$  e  $\mathbf{P}_0$  como sendo uma matriz com valores suficientemente grandes na diagonal (Harvey, 1989), o que caracteriza ausência de informação *a priori*.

Como exemplo de aplicação do FK, será considerado o modelo de nível local, descrito nas equações (2.1)-(2.2). Tomando-se as equações do MNL na

forma de espaço de estados (2.3)-(2.4) e os valores encontrados das matrizes de sistema, substituindo-os nas fórmulas do filtro de Kalman e atendendo às suposições, encontram-se, para o MNL, equações mais simples e fáceis de serem usadas, porém com a mesma finalidade:

$$\begin{aligned}\nu_t &= y_t - \mathbf{a}_{t|t-1}, \\ \mathbf{a}_{t+1|t} &= \mathbf{a}_{t|t-1} + K_t \nu_t, \\ F_t &= \mathbf{P}_{t|t-1} + \sigma_\varepsilon^2, \\ K_t &= \mathbf{P}_{t|t-1} / F_t, \\ \mathbf{P}_{t+1|t} &= \mathbf{P}_{t|t-1} - K_t^2 F_t + \sigma_\eta^2.\end{aligned}$$

O FK tem outras características além de predição como, por exemplo, a suavização - que será definida a seguir.

## 2.5 Suavização

O FK pode ser utilizado também para estimar o vetor de estado,  $\boldsymbol{\alpha}_t$ , baseado em toda informação da amostra obtida,  $\mathbf{Y}_n$ . Existem vários métodos, que não são apresentados aqui, para calcular os estimadores suavizados do estado  $\mathbf{a}_{t|n} = E(\boldsymbol{\alpha}_t | \mathbf{Y}_n)$  e suas respectivas variâncias  $\mathbf{P}_{t|n} = Var(\boldsymbol{\alpha}_t | \mathbf{Y}_n)$ . De acordo com a definição dada em Koopman (1992), esses estimadores são chamados de suavizadores do estado. Pode se mostrar que  $(\boldsymbol{\alpha}_t | \mathbf{Y}_n) \sim N(\mathbf{a}_{t|n}, \mathbf{P}_{t|n})$ . Esses momentos podem ser encontrados recursivamente através das seguintes equações:

$$\mathbf{r}_{t-1} = \mathbf{z}_t F_t^{-1} + \mathbf{L}_t' \mathbf{r}_t, \quad (2.9)$$

$$\mathbf{N}_{t-1} = \mathbf{z}_t' F_t^{-1} \mathbf{z}_t + \mathbf{L}_t' \mathbf{N}_t \mathbf{L}_t, \quad (2.10)$$

$$\mathbf{a}_{t|n} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1}, \quad (2.11)$$

$$\mathbf{P}_{t|n} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{N}_{t-1} \mathbf{P}_{t|t-1}, \quad (2.12)$$

em que  $\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{z}_t'$ . São necessárias as quantidades  $\mathbf{a}_{t|t-1}$ ,  $\mathbf{P}_{t|t-1}$ ,  $\mathbf{K}_t$ ,  $F_t$  e  $\nu_t$  obtidas através do FK para a execução do algoritmo, o qual é iniciado no instante  $t = n$  e o suavizador do estado é obtido na ordem inversa até  $t = 1$ . Os valores iniciais  $\mathbf{r}_n$  e  $\mathbf{N}_n$  podem ser fixados em zero. Analogamente ao que foi exposto na atualização, aqui também  $\mathbf{a}_{t|n}$  é um vetor de estimativas das componentes de estados. Portanto, utilizando o mesmo exemplo do MEB, as estimativas suavizadas do vetor de estados serão denotadas por  $\mathbf{a}_{t|n}^{(\mu)}$ ,  $\mathbf{a}_{t|n}^{(\beta)}$ ,



$\mathbf{a}_{t|n}^{(\gamma_1)}, \mathbf{a}_{t|n}^{(\gamma_2)}, \dots, \mathbf{a}_{t|n}^{(\gamma_s)}$ . Exposições mais completas desse tópico estão em Harvey (1989) e West & Harrison (1997).

## 2.6 Estimação dos hiperparâmetros

Seja  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)'$  o vetor de hiperparâmetros. Por exemplo, os vetores paramétricos dos modelos MNL, MTL, MTP e MEB são  $\boldsymbol{\psi} = (\sigma_\eta^2, \sigma_\epsilon^2)'$ ,  $\boldsymbol{\psi} = (\sigma_\eta^2, \sigma_\xi^2, \sigma_\epsilon^2)'$ ,  $\boldsymbol{\psi} = (\sigma_1^2, \dots, \sigma_k^2, \sigma_\epsilon^2)'$  e  $\boldsymbol{\psi} = (\sigma_\eta^2, \sigma_\xi^2, \sigma_\omega^2, \sigma_\epsilon^2)'$ , respectivamente. Por se tratar de variâncias, o espaço paramétrico dos vetores é o  $\mathbb{R}_+^p$ .

Todos os cálculos realizados nas seções anteriores pressupõem o conhecimento desses hiperparâmetros. Assim, todas as quantidades obtidas, como médias e variâncias, são funções dos hiperparâmetros. Quando esses são conhecidos, o procedimento de inferência se encerra nas seções anteriores. Entretanto, isso raramente acontece na prática e métodos de inferência para essas quantidades são necessários.

No contexto clássico, o método de estimação por máxima verossimilhança será usado para estimar de forma pontual os hiperparâmetros do modelo através da maximização da função densidade conjunta  $p(y_1, \dots, y_n; \boldsymbol{\psi})$  em relação a  $\boldsymbol{\psi}$ . Nos procedimentos Bayesianos, indicadores de medida central da distribuição *a posteriori*, como média e mediana, serão utilizados como estimadores pontuais dos hiperparâmetros do modelo.

### 2.6.1 Estimador de máxima verossimilhança

A função de verossimilhança pode ser calculada através das quantidades obtidas pelo filtro de Kalman, supondo que  $(y_t | \mathbf{Y}_{t-1}) \sim N(\tilde{y}_{t|t-1}, F_t)$ , onde  $\mathbf{Y}_{t-1} = (y_1, \dots, y_{t-1})$ . Neste caso, a função de densidade preditiva é dada por:

$$p(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\psi}) = (2\pi)^{-1/2} |F_t|^{-1/2} \exp \left\{ -\frac{1}{2} (y_t - \tilde{y}_{t|t-1})' F_t^{-1} (y_t - \tilde{y}_{t|t-1}) \right\}.$$

Substituindo  $\nu_t = y_t - \tilde{y}_{t|t-1}$  na equação anterior tem-se que:

$$p(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\psi}) = (2\pi)^{-1/2} |F_t|^{-1/2} \exp \left\{ -\frac{1}{2} \nu_t' F_t^{-1} \nu_t \right\}.$$

Para encontrar o estimador de máxima verossimilhança (EMV) de  $\boldsymbol{\psi}$ , deve-se maximizar a função de verossimilhança, que neste caso é o produtório das distribuições preditivas, pois  $L(\boldsymbol{\psi}; \mathbf{Y}_n) = p(y_1, \dots, y_n | \boldsymbol{\psi}) = \prod_{t=1}^n p(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\psi})$ .

Aplicando-se o logaritmo natural em  $L(\boldsymbol{\psi}; \mathbf{Y}_n)$  para simplificar os cálculos, a função de log-verossimilhança é obtida:

$$\ln L(\boldsymbol{\psi}; \mathbf{Y}_n) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln |F_t| - \frac{1}{2} \sum_{t=1}^n \boldsymbol{\nu}_t' F_t^{-1} \boldsymbol{\nu}_t. \quad (2.13)$$

O vetor  $\hat{\boldsymbol{\psi}}$  de hiperparâmetros estimados é obtido maximizando a função de log-verossimilhança (2.13). Como esta é uma função não-linear do vetor de hiperparâmetros, os EMV não podem ser obtidos analiticamente. Essa estimação deve ser realizada numericamente através de algoritmos de otimização, descritos no Capítulo 3.

A vantagem de se considerar o método de máxima verossimilhança para obter estimadores dos hiperparâmetros é que se pode valer das propriedades ótimas que esses estimadores gozam sob certas condições de regularidade (Migon & Gamerman, 1999; Casella & Berger, 2002). Por exemplo, os EMV são assintoticamente consistentes e não-viciados. A normalidade assintótica é outra propriedade importante dos estimadores de máxima verossimilhança que será utilizada para construir intervalos de confiança assintóticos para os hiperparâmetros.

### 2.6.2 Estimador de Bayes

A inferência Bayesiana consiste em descrever a incerteza inicial de um pesquisador sobre uma quantidade desconhecida ou parâmetro, através de uma distribuição de probabilidade *a priori*  $\pi(\cdot)$ , e depois combinar com a informação proveniente dos dados através da verossimilhança, resultando na distribuição *a posteriori*. Se uma distribuição *a priori*  $\pi(\boldsymbol{\psi})$  é especificada para  $\boldsymbol{\psi}$ , então a distribuição *a posteriori* de  $\boldsymbol{\psi}$  é dada por:

$$\pi(\boldsymbol{\psi} | \mathbf{Y}_n) = \frac{L(\boldsymbol{\psi}; \mathbf{Y}_n) \pi(\boldsymbol{\psi})}{\int L(\boldsymbol{\lambda}; \mathbf{Y}_n) \pi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}}.$$

Além disso, a distribuição a priori para os parâmetros de estado  $\{\boldsymbol{\alpha}_t\}$  condicional aos hiperparâmetros  $\boldsymbol{\psi}$  é obtida com a especificação (2.4) em conjunto com a especificação inicial para  $\boldsymbol{\alpha}_0$ .

Na maioria das vezes, não é uma tarefa trivial especificar  $\pi(\boldsymbol{\psi})$ . Uma estratégia comum é a adoção de distribuições *a priori* vagas (com valores de variância grandes). No entanto, esta pode não ser a estratégia mais adequada, uma vez que distribuições *a priori* vagas para um parâmetro podem ser informativas para outras transformações desse parâmetro, segundo Migon & Gamerman (1999). Na ausência de informação suficiente *a priori*, tem sido uma prática comum a especificação de distribuições *a priori* não-informativas apropriadas, obtidas através de algum método formal. A mais comum dessas distribuições *a priori* é a de Jeffreys (1961).

Neste trabalho serão utilizadas as seguintes distribuições *a priori* para os hiperparâmetros:

**Uniforme:**  $\pi(\boldsymbol{\psi}) \propto c$ , para todos os valores possíveis de  $\boldsymbol{\psi}$ , e 0, caso contrário, onde  $c \in \mathbb{R}_+$ .

**Jeffreys:**  $\pi(\boldsymbol{\psi}) \propto \det(\mathbf{I}(\boldsymbol{\psi}))^{1/2}$ , onde  $\mathbf{I}(\boldsymbol{\psi})$  é a matriz de informação de Fisher cuja forma para os MEE é dada por:

$$\mathbf{I}_{ij}(\boldsymbol{\psi}) = \frac{1}{2} \sum_t \left[ \text{tr} \left[ F_t^{-1} \frac{\partial F_t}{\partial \psi_i} F_t^{-1} \frac{\partial F_t}{\partial \psi_j} \right] \right] + E \left[ \sum_t \left( \frac{\partial \nu_t}{\partial \psi_i} \right)' F_t^{-1} \frac{\partial \nu_t}{\partial \psi_j} \right], \quad (2.14)$$

onde  $i, j = 1, \dots, p$  e  $t = 1, \dots, n$ .

A distribuição *a posteriori* contém toda a informação disponível sobre  $\boldsymbol{\psi}$  após os dados serem observados. Apesar disto, muitas vezes é preciso resumir-la em um único número. Quando este número é obtido minimizando uma função de perda previamente escolhida, ele é denominado estimador de Bayes. Seja  $G(\boldsymbol{\psi}, \boldsymbol{\delta})$  a função de perda segundo a qual se quer ser penalizado, onde  $\boldsymbol{\psi}$  é o valor real e  $\boldsymbol{\delta}$  são possíveis estimativas de  $\boldsymbol{\psi}$ . O estimador de Bayes para  $\boldsymbol{\psi}$  é o valor  $\boldsymbol{\delta}$  que torna mínima a perda esperada *a posteriori*, isto é:

$$\hat{\boldsymbol{\psi}}_{EB} = \min_{\boldsymbol{\delta}} E(G(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Y}_n) = \min_{\boldsymbol{\delta}} \int G(\boldsymbol{\psi}, \boldsymbol{\delta}) \pi(\boldsymbol{\psi} | \mathbf{Y}_n) d\boldsymbol{\psi}.$$

Se a função de perda quadrática  $G(\boldsymbol{\psi}, \boldsymbol{\delta}) = (\boldsymbol{\psi} - \boldsymbol{\delta})^2$  é considerada, o estimador de Bayes é a média *a posteriori*. Se, por outro lado, escolhe-se as funções de perda 0-1 e perda absoluta, a moda e a mediana *a posteriori* são os estimadores de Bayes  $\hat{\boldsymbol{\psi}}_{EB}$ , respectivamente. Deve-se ressaltar que diferentes funções de perda geram diferentes estimadores de Bayes e deve-se frisar que a escolha da função de perda é completamente subjetiva (ver Migon & Gamerman (1999)).

## 2.7 Previsão

Um dos objetivos principais de se modelar séries temporais é prever valores futuros. Portanto, o desenvolvimento das funções de previsão torna-se um importante tópico na análise de séries temporais. A previsão de um valor futuro  $y_{n+k}$  baseada em toda informação disponível na amostra,  $\mathbf{Y}_n$ , pode ser obtida pela combinação da equação (2.3) no tempo  $n+k$  com a equação do estado  $k$  passos à frente dada por

$$\boldsymbol{\alpha}_{n+k} = \left( \prod_{i=1}^k \mathbf{T}_{n+i} \right) \boldsymbol{\alpha}_n + \sum_{i=1}^k \left( \prod_{j=1}^{k-i} \mathbf{T}_{n+k-j+1} \right) \mathbf{c}_{n+i} + \sum_{i=1}^k \left( \prod_{j=1}^{k-i} \mathbf{T}'_{n+k-j+1} \right) \mathbf{R}'_{n+i} \boldsymbol{\eta}_{n+i}. \quad (2.15)$$

Para obter a previsão  $k$  passos à frente, define-se  $\tilde{y}_{n+k|n}(\boldsymbol{\psi}) = E(y_{n+k} | \mathbf{Y}_n, \boldsymbol{\psi})$ . Utilizando a forma de espaço de estados em (2.3)-(2.4), temos que

$$\tilde{y}_{n+k|n}(\boldsymbol{\psi}) = \mathbf{z}'_{n+k} \left[ \left( \prod_{i=1}^k \mathbf{T}_{n+i} \right) \mathbf{a}_n + \sum_{i=1}^k \left( \prod_{j=1}^{k-i} \mathbf{T}_{n+k-j+1} \right) \mathbf{c}_{n+i} \right] + d_{n+k}. \quad (2.16)$$

A variância de  $(y_{n+k} | \mathbf{Y}_n, \boldsymbol{\psi})$  é dada por

$$\begin{aligned} & \mathbf{z}'_{n+k} \left( \prod_{i=1}^k \mathbf{T}_{n+i} \right) \mathbf{P}_n \left( \prod_{i=1}^k \mathbf{T}'_{n+i} \right) \mathbf{z}_{n+k} + \mathbf{z}'_{n+k} \sum_{i=1}^k \left( \prod_{j=1}^{k-i} \mathbf{T}_{n+k-j+1} \right) \mathbf{R}_{n+i} \mathbf{Q}_{n+i} \mathbf{R}'_{n+i} \\ & \left( \prod_{j=1}^{k-i} \mathbf{T}'_{n+k-j+1} \right) \mathbf{z}_{n+k} + h_{n+k}. \end{aligned} \quad (2.17)$$

Como exemplo, a previsão  $k$  passos à frente para o MNL é dada por  $\tilde{y}_{n+k|n}(\boldsymbol{\psi}) = \mathbf{a}_n$  e a sua variância é igual a  $\mathbf{P}_n + k\sigma_\eta^2 + \sigma_\varepsilon^2$ .

Deve-se observar que a obtenção de  $\tilde{y}_{n+k|n}(\boldsymbol{\psi})$  e sua variância não encerra o processo de inferência pois eles ainda podem depender de  $\boldsymbol{\psi}$ . A seguir será mostrado como essa dependência é eliminada pelas diferentes abordagens aqui tratadas.

### i) Abordagem clássica

Neste caso, a função de previsão é dada por  $\tilde{y}_{n+k|n}^{(c)}(\hat{\boldsymbol{\psi}})$ , ou seja, ela é obtida substituindo-se o vetor de parâmetros  $\boldsymbol{\psi}$  pelo seu estimador de máxima verossimilhança  $\hat{\boldsymbol{\psi}}$  (Brockwell & Davis, 1996). Da mesma forma, a variância da função de previsão é obtida substituindo  $\boldsymbol{\psi}$  por  $\hat{\boldsymbol{\psi}}$  em (2.17).

### ii) Abordagem Bayesiana

Nessa abordagem, o hiperparâmetro  $\psi$  é eliminado através de sua integração. Em assim sendo, a incerteza sobre todos os seus possíveis valores é incorporada. Isso é diferente da simples substituição de  $\psi$  por um único valor, o que inevitavelmente leva à subestimação da incerteza.

A distribuição preditiva  $y_{n+k}|\mathbf{Y}_n$  é calculada integrando fora em  $\psi$  da seguinte maneira:

$$p(y_{n+k}|\mathbf{Y}_n) = \int p(y_{n+k}|\psi, \mathbf{Y}_n)\pi(\psi|\mathbf{Y}_n)d\psi = E_\pi[p(y_{n+k}|\psi, \mathbf{Y}_n)]. \quad (2.18)$$

Essa distribuição carrega toda informação preditiva do modelo, portanto resumos da mesma tais como média e percentis podem ser calculados. Ela pode ser obtida combinando a distribuição de previsão de  $y_{n+k}|\mathbf{Y}_n, \psi$  que é Normal com média e variância dadas por (2.16) e (2.17), respectivamente, com pesos dados pela distribuição *a posteriori* de  $\psi$ . Essa integração raramente pode ser realizada analiticamente e métodos de aproximação precisam ser utilizados.

A função de previsão é então dada pela média da distribuição preditiva de  $y_{n+k}|\mathbf{Y}_n$  e é denotada por  $\tilde{y}_{n+k|n}^{(b)}$ . É obtida analogamente resolvendo-se a integral

$$\tilde{y}_{n+k|n}^{(b)} = \int \tilde{y}_{n+k}(\psi)\pi(\psi|\mathbf{Y}_n)d\psi = E_\pi[\tilde{y}_{n+k}(\psi)|\psi, \mathbf{Y}_n].$$

## 2.8 Aplicação a séries reais

Nesta seção, são vistos exemplos de aplicações do modelo de nível local (MNL), modelo de tendência linear local (MTL) e modelo estrutural básico (MEB) ajustados às séries temporais reais IPCA, ICV e  $SO_4$ , respectivamente. As estimativas pontuais clássica e Bayesiana dos hiperparâmetros dos modelos são obtidas e avaliadas para cada uma das séries. Assume-se a distribuição *a priori* Uniforme para os hiperparâmetros dos modelos dessas séries, porém vale ressaltar que pode-se considerar qualquer outra distribuição *a priori*, como por exemplo a de Jeffreys. Maiores detalhes e descrição dos algoritmos de otimização, bem como os programas utilizados no ajuste dos modelos, encontram-se no Capítulo 3.

### 2.8.1 Série do IPCA

A série analisada do Índice de Preços ao Consumidor Amplo (IPCA) de Belo Horizonte, coletada pelo IPEAD - *Fundação Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Minas Gerais* - é composta de 132 observações mensais no período de julho de 1997 a junho de 2008. Esse índice mede a evolução da renda em famílias que gastam de 1 a 40 salários mínimos por mês. Como observado na Figura 2.1, a série parece apresentar um comportamento de um Modelo de Nível Local (MNL), apesar de possuir observações aberrantes em torno de outubro de 2002.

Na Tabela 2.1 as estimações clássica e Bayesiana para os hiperparâmetros de um MNL ajustado à série do IPCA são avaliadas. Percebe-se que as estimativas estão bem próximas em ambas perspectivas, com o EMV apresentando valores levemente inferiores aos estimadores Bayesianos. Além disto, verifica-se que as estimativas de  $\sigma_\eta^2$  são menores que as de  $\sigma_\epsilon^2$ , o que geralmente ocorre na prática, fornecendo uma razão sinal-ruído,  $\sigma_\eta^2/\sigma_\epsilon^2$ , de aproximadamente 0,1.

Tabela 2.1: Ajuste do MNL à série IPCA.

	Estimação clássica	Estimação Bayesiana	
Hiperparâmetros	EMV	Média	Mediana
$\sigma_\eta^2$	0,024	0,029	0,027
$\sigma_\epsilon^2$	0,207	0,210	0,209

### 2.8.2 Série do ICV

Essa série, mostrada na Figura 2.2, representa o logaritmo do Índice do Custo de Vida (ICV) no município de São Paulo, com 126 observações mensais de janeiro de 1970 a junho de 1980. Observa-se que a série possui uma tendência crescente, portanto o MTL seria o modelo mais indicado a ser ajustado a esta série.

Da Tabela 2.2, nota-se que as estimativas Bayesianas apresentam valores muito próximos e novamente o EMV possui valores mais baixos. A estimativa de  $\sigma_\eta^2$  apresenta o maior valor, seguida de  $\sigma_\xi^2$  e de  $\sigma_\epsilon^2$ . Esta série também foi analisada por Morettin & Toloi (2004) (Cap. 13) e estes autores encontram valores muito similares aos apresentados aqui para os hiperparâmetros.

Tabela 2.2: Ajuste do MTL à série do logaritmo do ICV.

	Estimação clássica	Estimação Bayesiana	
Hiperparâmetros	EMV	Média	Mediana
$\sigma_\eta^2$	$7,73 \times 10^{-5}$	$7,84 \times 10^{-5}$	$7,84 \times 10^{-5}$
$\sigma_\xi^2$	$1,57 \times 10^{-6}$	$1,86 \times 10^{-6}$	$1,77 \times 10^{-6}$
$\sigma_\epsilon^2$	$9,37 \times 10^{-12}$	$9,38 \times 10^{-8}$	$7,83 \times 10^{-8}$

### 2.8.3 Série da precipitação de SO<sub>4</sub>

Na Figura 2.3 foi apresentada a série do logaritmo da precipitação de sulfato ( $SO_4$ ) em Nova York, EUA, com 271 observações mensais no período de janeiro de 1985 a julho de 2007. Como a série parece apresentar sazonalidade e tendência, o MEB será ajustado a esta série.

A Tabela 2.3 mostra os resultados do ajuste com os estimadores EMV, média *a posteriori* e mediana *a posteriori*. Comparando-se as duas abordagens, nota-se que as estimativas das variâncias do sistema são muito baixas, parecendo ser nulas. Como  $\sigma_\epsilon^2$  é o único hiperparâmetro que parece não ser igual a zero, a conclusão é que tanto a sazonalidade quanto a tendência nesta série parecem ser não estocásticas.

Tabela 2.3: Ajuste do MEB à série do logaritmo da precipitação de SO<sub>4</sub>.

	Estimação clássica	Estimação Bayesiana	
Hiperparâmetros	EMV	Média	Mediana
$\sigma_\eta^2$	$1,29 \times 10^{-4}$	$2,13 \times 10^{-4}$	$1,98 \times 10^{-4}$
$\sigma_\xi^2$	$1,42 \times 10^{-12}$	$2,00 \times 10^{-7}$	$1,14 \times 10^{-7}$
$\sigma_\omega^2$	$2,37 \times 10^{-4}$	$4,69 \times 10^{-4}$	$4,14 \times 10^{-4}$
$\sigma_\epsilon^2$	0,068	0,068	0,068

A título de ilustração, na Figura 2.4 são apresentadas as estimativas suavizadas do vetor de estados do MEB, sob o enfoque clássico, ajustado à série do logaritmo da precipitação de  $SO_4$  (linha cheia). As linhas tracejadas nos quatro gráficos da Figura 2.4 representam, respectivamente, estimativas suavizadas da média da série ( $\mathbf{a}_{t|n}^{(\mu)} + \mathbf{a}_{t|n}^{(\gamma)}$ ), do componente de nível ( $\mathbf{a}_{t|n}^{(\mu)}$ ), do componente de tendência ( $\mathbf{a}_{t|n}^{(\beta)}$ ) e do componente de sazonalidade ( $\mathbf{a}_{t|n}^{(\gamma)}$ ). Pode se observar que a média estimada acompanha bem o comportamento da série, que o nível apresenta uma inclinação praticamente inexistente mas bastante estável e a presença de sazonalidade também bastante estável, confirmando os resultados obtidos na Tabela 2.3.

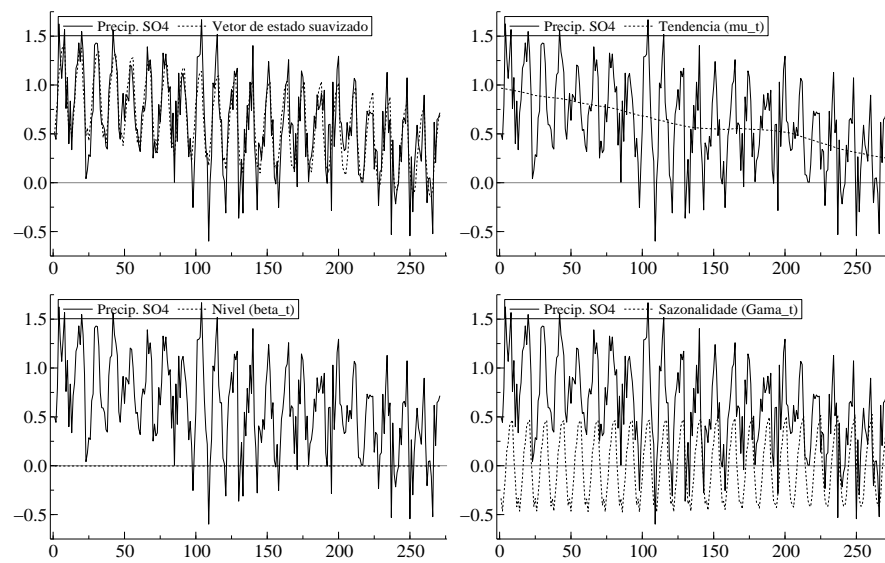


Figura 2.4: Gráficos das estimativas suavizadas do vetor de estado para a série do logaritmo da precipitação de  $SO_4$ .



## Capítulo 3

# Implementação Computacional

Para os MEE, as estimativas dos hiperparâmetros, tanto na abordagem clássica quanto na Bayesiana, só podem ser obtidas através de métodos numéricos. Na parte clássica, um método de maximização numérica da função de verossimilhança deve ser utilizado. Neste trabalho, o procedimento escolhido foi o BFGS, que é um método robusto para obtenção de máximos e mínimos de uma função. No contexto Bayesiano, o cálculo da distribuição *a posteriori* é difícil de ser obtido e assim a estimação dos parâmetros dos modelos será feita através dos métodos de *Markov chain Monte Carlo* (MCMC). Além desses procedimentos, a técnica *bootstrap* também será descrita, pois a mesma será utilizada em capítulos posteriores para o cálculo de intervalos de confiança e testes de hipóteses no contexto clássico.

### 3.1 Método de otimização BFGS

O método de otimização de Broyden-Fletcher-Goldfarb-Shanno (proposto simultaneamente por Broyden (1970), Fletcher (1970), Goldfarb (1970) e Shanno (1970)), usualmente conhecido como BFGS, é um método para resolver problemas de otimização não linear sem restrições. Ele é derivado dos métodos de otimização de Newton, que assumem que a função pode ser aproximada localmente como uma expansão de Taylor quadrática na região em torno do ponto ótimo, e usa as primeira e segunda derivadas para encontrar o ponto estacionário.

O método BFGS vem sendo utilizado pelos autores na maximização da função de verossimilhança em MEE e tem apresentado desempenho satis-

fatório (ver Franco *et al.* (2008)), principalmente pela sua simplicidade e rapidez. Certamente muitos outros métodos de maximização irrestrita são aplicáveis, havendo uma vasta literatura a ser consultada sobre esta e outras técnicas de otimização. Entre outros, veja Fletcher (1987), Gill, Murray & Wright (1981), Cramer (1986) e o clássico livro de Press *et al.* (1988). Deve-se notar que muitos textos sobre otimização abordam problemas de minimização em vez de maximização, mas, claro, isto é apenas o caso de reversão de um sinal algébrico.

Um detalhamento mais aprofundado do procedimento BFGS não será abordado aqui (ver Avriel (2003) para maiores detalhes), mas em linhas gerais pode-se dizer que o método encontra-se no intermediário entre a simplicidade do método do gradiente e a rapidez do método de Newton, razão pela qual é conhecido como um método quasi-Newton. No BFGS, assim como em outros métodos quasi-Newton, ao invés de a matriz hessiana ser calculada exatamente, como é o caso do método de Newton, ela é aproximada por um processo iterativo finito, via derivadas de primeira ordem, procurando um compromisso entre a rapidez de convergência e a dificuldade da avaliação da inversa a cada passo.

O método BFGS segue o algoritmo apresentado de forma bastante simplificada abaixo, mas que define claramente parâmetros importantes, tais como o número máximo de iterações, MaxIt, a tolerância máxima do vetor gradiente,  $\varepsilon_1$ , e a tolerância máxima do passo,  $\varepsilon_2$ . O algoritmo iterativo tem a seguinte estrutura:

---

**Algoritmo 3.1:** Método BFGS

---

(dados  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  )

**escolha**  $\psi_0 \in \mathbb{R}^n$

**faça**  $j = 0$

**repita** os passos abaixo

**calcule**  $\nabla f(\psi^j)$  (gradiente)

**calcule**  $D(\psi^j)$  (aproximação para a inversa da hessiana  $H^{-1}(\psi^j)$ )

**faça**  $\psi^{j+1} \leftarrow \psi^j - \lambda_j D(\psi^j) \nabla f(\psi^j)$

**faça**  $j \leftarrow j + 1$

**até** ( $j \geq \text{MaxIt}$ ) **ou** ( $|\nabla f(\psi^j)| \leq \varepsilon_1$ ) **ou** ( $|\psi^j - \psi^{j-1}| \leq \varepsilon_2$ )

---



```

////////////////////////////////////
main() {
// declaracao de variaveis
  decl vP, vFunc, vYt, ii, MAXIT=50;
  decl Sigma2Eta,Sigma2Eps;
  decl LLM=<
    CMP_LEVEL,1.0,0,0;
    CMP_IRREG,1.0,0,0>; //especificando o MNL
// carregando os dados
  vYt=loadmat("ipca.xls");
  g_STSvYt=vYt;
// inicializando as variaveis
  GetSsfStsm(LLM, &Phi, &Omega, &Sigma);
  vP=zeros(2,1);
  vP[0]=0.5*log(Omega[0][0]);
  vP[1]=0.5*log(Omega[1][1]);
// maximizando a log-verossimilhança
  MaxControl(MAXIT,-1);
  ii=MaxBFGS(likelihood, &vP, &vFunc, 0, TRUE);
  Sigma2Eta=Omega[0][0];
  Sigma2Eps=Omega[1][1];
\\ imprimindo os resultados
  println("\t\n***** Ajuste do MNL*****\t\n");
  println("Parametro\ Estimativa");
  println("Sigma2Eta\t",Sigma2Eta,"\t");
  println("Sigma2Eps\t",Sigma2Eps,"\t");
  println("\t\n*****\t\n");
}

```

Esse programa produz o seguinte resultado, que foi apresentado na Tabela 2.1 para o EMV:

```

***** Ajuste do MNL*****

Parametros\ Estimativas
Sigma2Eta  0.0243642
Sigma2Eps  0.206828

*****

```

**Exemplo 3.2:** Abaixo é apresentado o programa principal para o ajuste do modelo de tendência linear local (MTL) à série do logaritmo do ICV (Figura 2.2).

**Código 3.2:** Estimação do MTL na linguagem *Ox* usando BFGS

---

```
// Este programa estima os hiperparametros Sigma2Eta, Sigma2Csi e Sigma2Eps do
// MTL utilizando o metodo BFGS para maximizacao da funcao de verossimilhanca

#include <oxstd.h>
#import <maximize>
#include</packages/ssfpack/ssfpack.h>
#include "STSTBootPack.ox"

main(){
// declaracao de variaveis
decl vYt, STSmOmega;
decl Sigma2Eta, Sigma2Csi, Sigma2Eps;
decl LTL=<
    CMP_LEVEL,1.0,0,0;
    CMP_SLOPE,1.0,0,0;
    CMP_IRREG,1.0,0,0>; //especificando o MTL
// carregando os dados
vYt=loadmat("ICV.xls");
// estimacao dos hiperparametros
STSmOmega=STSAadjust(vYt,LTL);
Sigma2Eta=STSmOmega[0][0];
Sigma2Csi=STSmOmega[1][1];
Sigma2Eps=STSmOmega[rows(STSmOmega)-1][rows(STSmOmega)-1];
// imprimindo os resultados
println("\t\n***** Ajuste do MTL*****\t\n");
println("Parametro\ Estimativa");
println("Sigma2Eta\t",Sigma2Eta,"\t");
println("Sigma2Csi\t",Sigma2Csi,"\t");
println("Sigma2Eps\t",Sigma2Eps,"\t");
println("\t\n*****\t\n");
}
```

---

Este código produz o resultado apresentado abaixo, que foi mostrado na Tabela 2.2 para o EMV:

```
***** Ajuste do MTL*****

Parametro\ Estimativa
Sigma2Eta    7,73e-05
Sigma2Csi    1,57e-06
Sigma2Eps    9,37e-012

*****
```

**Exemplo 3.3:** Abaixo é apresentado o programa principal para o ajuste do modelo estrutural básico (MEB) à série do logaritmo da precipitação de  $SO_4$  (Figura 2.3).

**Código 3.3:** Estimação do MEB na linguagem *Or* usando BFGS

---

```

// Este programa estima os hiperparametros Sigma2Eta, Sigma2Csi, Sigma2Omega e Sigma2Eps
// do MEB utilizando o metodo BFGS para maximizacao da funcao de verossimilhanca

#include <oxstd.h>
#include <oxdraw.h>
#include <oxfloat.h>
#import <maximize>
#include </packages/ssfpack/ssfpack.h>
#include "STSBootPack.ox"
main(){
// declaracao de variaveis
    decl vYt, STSmOmega, mPhi, mSigma, mOmega, mp, mest, cm;
    decl Sigma2Eta, Sigma2Csi, Sigma2Omg, Sigma2Eps;
    decl SBM=<
        CMP_LEVEL,1.0,0,0;
        CMP_SLOPE,1.0,0,0;
        CMP_SEAS_DUMMY,1.0,12,0;
        CMP_IRREG,1.0,0,0>; //especificando o MEB
// carregando os dados
    vYt=loadmat("Precip.xls");
    vYt=vYt[60:];
// estimacao dos hiperparametros
    STSmOmega=STSAdjust(vYt,SBM);
    Sigma2Eta=STSmOmega[0][0];
    Sigma2Csi=STSmOmega[1][1];
    Sigma2Omg=STSmOmega[2][2];
    Sigma2Eps=STSmOmega[rows(STSmOmega)-1][rows(STSmOmega)-1];
// imprimindo os resultados
    println("\t\n***** Ajuste do MEB*****\t\n");
    println("Parametro\ Estimativa");
    println("Sigma2Eta\t",Sigma2Eta,"\t");
    println("Sigma2Csi\t",Sigma2Csi,"\t");
    println("Sigma2Omg\t",Sigma2Omg,"\t");
    println("Sigma2Eps\t",Sigma2Eps,"\t");
    println("\t\n*****\t\n");
// construindo os graficos do vetor de espacos
    GetSsfStsm(SBM, &mPhi, &mOmega,&mSigma);
    mp=SsfMomentEst(ST_SM0, &mest, vYt, mPhi, STSmOmega, mSigma);
    cm = columns(mPhi);
    DrawTMatrix(0, vYt, {"Precip. So4"}, 0, 1, 1);
    DrawTMatrix(0, mest[cm][], {"Vetor de estado suavizado"}, 0, 1, 1, 0, 3);
    DrawTMatrix(1, vYt, {"Precip. So4"}, 0, 1, 1);
    DrawTMatrix(1, mest[0][], {"Tendencia suavizada"}, 0, 1, 1, 0, 3);
    DrawTMatrix(2, vYt, {"Precip. So4"},0, 1, 1);
    DrawTMatrix(2, mest[1][], {"Nivel suavizada"},0, 1, 1, 0, 3);
    DrawTMatrix(3, vYt, {"Precip. So4"},0, 1, 1);
    DrawTMatrix(3, mest[3][], {"Gama_t suavizado"},0, 1, 1, 0, 3);
    SaveDrawWindow("graficosMEB.eps");
}

```

---

Este código produz as estimativas apresentadas abaixo (também reproduzidas na Tabela 2.3 para o EMV) e os gráficos que foram mostrados na Figura 2.4.

```
***** Ajuste do MEB*****

Parametro\ Estimativa
Sigma2Eta    0.000129299
Sigma2Csi    1.42253e-012
Sigma20mg    0.000236934
Sigma2Eps    0.0682181

*****
```

## 3.2 Métodos de simulação MCMC

Os métodos *Markov chain Monte Carlo* (MCMC) possibilitam a geração de distribuições de maneira indireta, usando algoritmos como o amostrador de Gibbs e Metropolis-Hastings. O MCMC consiste, essencialmente, em simulação Monte Carlo usando cadeias de Markov. A idéia central dos métodos MCMC é construir uma cadeia de Markov da qual seja fácil gerar uma trajetória e que tenha distribuição de equilíbrio igual à distribuição de interesse (ver Gamerman & Lopes (2006) e Gilks, Richardson & Spiegelhalter (1996)). Após um número suficientemente grande de iterações, a cadeia deve convergir para a distribuição de interesse. Porém, em alguns casos a cadeia pode ainda apresentar altas autocorrelações entre seus valores, algo que não é desejado. Qualquer que seja o caso, o teorema ergódico (Gamerman & Lopes, 2006) garante que, para uma cadeia de Markov  $\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \dots\}$  satisfazendo determinadas condições relativamente brandas e com  $E_{\pi}(t(\boldsymbol{\psi})) < \infty$ ,

$$\bar{t}_N = N^{-1} \sum_{k=1}^N t(\boldsymbol{\psi}^{(k)}) \longrightarrow E_{\pi}(t(\boldsymbol{\psi})), \quad (3.1)$$

quando  $N \longrightarrow \infty$  com probabilidade 1, onde  $N$  é o número de valores da cadeia e  $t(\boldsymbol{\psi})$  é alguma função de interesse de  $\boldsymbol{\psi}$ . Por exemplo, no caso da média *a posteriori*,  $t(\boldsymbol{\psi}) = \boldsymbol{\psi}$ . O teorema garante que a média ergódica dos valores da cadeia fornece um estimador consistente de características da distribuição limite  $\pi(\cdot)$ .

Esses métodos são muito utilizados na estatística Bayesiana para obter amostras de uma densidade *a posteriori*  $\pi(\boldsymbol{\psi}|\mathbf{Y}_n)$ , cuja geração direta é custosa ou complicada.

### i) Algoritmo de Metropolis-Hastings

Esse algoritmo foi proposto por Metropolis *et al.* (1953), generalizado por Hastings (1970) e será utilizado neste trabalho para obter as médias, medianas e quantis das diferentes distribuições de interesse. Dada uma distribuição alvo  $\pi(\cdot)$ , o algoritmo de Metropolis-Hastings (M-H) e suas variantes fornecem meios para gerar cadeias de Markov ergódicas que tenham  $\pi(\cdot)$  como distribuição de equilíbrio. Esse algoritmo é utilizado para simular uma distribuição *a posteriori*, utilizando uma densidade de proposição ou de transição de um estado  $x$  para um  $y$ ,  $q(y|x)$ , que é uma distribuição condicional que governa as transições do estado, tendo as seguintes propriedades:

- $\int q(y|x)dy = 1$ ;
- $q(y|x)$  pode ser avaliada para todo  $x$  e  $y$ ;
- para cada  $x$  é possível gerar realizações que tenham distribuição  $q(\cdot|x)$ .

Seja  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_p)'$  o conjunto de parâmetros de interesse. O algoritmo de Metropolis-Hastings é inicializado a partir de um ponto arbitrário  $\boldsymbol{\psi}^{(0)}$ . Ele passa de um ponto  $\boldsymbol{\psi}^{(j-1)}$  a outro  $\boldsymbol{\psi}^{(j)}$  da cadeia de acordo com o algoritmo abaixo:

#### Algoritmo 3.2: Metropolis-Hastings

---

**faça**  $j = 1$   
**repita** os passos abaixo  
     **simule**  $\boldsymbol{\psi}^{(j)} \sim q(\cdot|\boldsymbol{\psi}^{(j-1)})$   
     **calcule** a razão de Hastings  
         
$$R = [\pi(\boldsymbol{\psi}^{(j)})q(\boldsymbol{\psi}^{(j-1)}|\boldsymbol{\psi}^{(j)})] / [\pi(\boldsymbol{\psi}^{(j-1)})q(\boldsymbol{\psi}^{(j)}|\boldsymbol{\psi}^{(j-1)})]$$
  
     **obtenha** o próximo valor da cadeia como sendo  $\boldsymbol{\psi}^{(j)}$  com probabilidade  
         igual a  $\min(1, R)$  ou  $\boldsymbol{\psi}^{(j-1)}$  com probabilidade complementar  
     **faça**  $j = j + 1$   
**até** a convergência ser alcançada.

---



Às vezes, é muito difícil ou não é possível amostrar  $\boldsymbol{\psi} = (\psi_i, \dots, \psi_p)'$  conjuntamente de forma computacionalmente eficiente. Uma solução, nestes casos, é utilizar uma variação do algoritmo de Metropolis conhecida como M-H híbrido ou M-H por blocos.

### ii) Algoritmo de Metropolis-Hastings híbrido

Neste estudo, adota-se uma versão híbrida do algoritmo de M-H (Geman & Lopes, 2006). Nessa versão, os  $p$  hiperparâmetros de  $\boldsymbol{\psi}$  são atualizados separadamente, com diferentes densidades de proposição. Os seguintes passos compõem o algoritmo:

---

#### Algoritmo 3.3: Metropolis-Hastings híbrido

---

**inicie**  $\boldsymbol{\psi}_0$

**faça**  $j = 1$  (contador das iterações)

**repita** os passos abaixo

**faça**  $i = 1$  (contador das componentes)

**repita** os passos abaixo

**gere**  $\psi_i^{(j)} \sim q_i(\cdot | \psi_i^{(j-1)})$

**calcule** a razão de Hastings

$$R_i = \left[ \pi_i(\psi_i^{(j)}) q_i(\psi_i^{(j-1)} | \psi_i^{(j)}) \right] / \left[ \pi_i(\psi_i^{(j-1)}) q_i(\psi_i^{(j)} | \psi_i^{(j-1)}) \right],$$

        onde  $\pi_i(\psi_i)$  é a condicional completa de  $\psi_i$ . Note que  $\frac{\pi_i(\psi_i^{(j)})}{\pi_i(\psi_i^{(j-1)})} =$

$$\frac{\pi(\Psi_i^{(j)})}{\pi(\Psi_i^{(j-1)})}, \text{ em que } \Psi_i^{(j)} = (\psi_1^{(j)}, \dots, \psi_{i-1}^{(j)}, \psi_i^{(j)}, \psi_{i+1}^{(j-1)}, \dots, \psi_p^{(j-1)})'$$

**faça** a probabilidade de aceitação igual a  $\min\{1, R_i\}$

**obtenha** o próximo valor da cadeia como sendo  $\psi_i^{(j)}$  com

            probabilidade igual a  $\min\{1, R_i\}$ , ou  $\psi_i^{(j-1)}$  com probabilidade complementar

**faça**  $i = i + 1$

**até**  $i = p$

**faça**  $j = j + 1$

**até** a convergência ser alcançada.

---

As probabilidades de aceitação  $R_i$  são definidas de forma a garantir a reversibilidade da cadeia e sua convergência para a distribuição de equilíbrio  $\pi(\cdot)$ .

Há várias possibilidades de densidades de proposição na literatura. Neste trabalho, usa-se a densidade de proposição sendo um passeio aleatório  $q_i(\psi_i^{(j)}|\psi_i^{(j-1)}) = N(\psi_i^{(j-1)}, \sigma_i^2)$  (Chib & Greenberg, 1995). Densidades de proposição que geram passos curtos conduzirão a altas taxas de aceitação, porém com baixa mistura, posto que a cadeia caminhará lentamente. Por outro lado, densidades que geram passos muito longos tenderão a ser rejeitadas, mantendo a cadeia estagnada e levando a baixas taxas de aceitação. Por isso, os valores de  $\sigma_i^2$  são definidos de forma que as taxas de aceitação dos hiperparâmetros estejam entre 20% e 50% (Chib & Greenberg, 1995).

Em alguns casos existem restrições no espaço paramétrico, por exemplo a semi-reta positiva para variâncias. Nestes casos, faz-se necessário restringir a densidade proposta a esse intervalo e as distribuições das proposições passam a ser normais truncadas para garantir que sejam preservadas as propriedades definidas anteriormente. No caso do espaço paramétrico ser dado pelo intervalo  $[a, b]$ , a razão de Hastings passaria a ser

$$R_i = \left[ \pi_i(\psi_i^{(j)}) q_i(\psi_i^{(j-1)}|\psi_i^{(j)}) \left( \Phi((a - \psi_i^{(j-1)})/\sigma) - \Phi((b - \psi_i^{(j-1)})/\sigma) \right) \right] /$$

$$\left[ \pi_i(\psi_i^{(j-1)}) q_i(\psi_i^{(j)}|\psi_i^{(j-1)}) \left( \Phi((a - \psi_i^{(j)})/\sigma) - \Phi((b - \psi_i^{(j)})/\sigma) \right) \right],$$

onde  $\Phi(\cdot)$  denota a função de distribuição da Normal-padrão.

Outro aspecto importante é a avaliação da convergência do algoritmo. Isso pode ser verificado de inúmeras formas. Uma das mais utilizadas é a dada pelo método de Gelman (1996), no qual múltiplas cadeias com diferentes valores iniciais são usadas. Esse método está disponível no R (2005) no pacote *Coda* (Plummer *et al.*, 2005). O traço da cadeia gerada é observado, procurando movimentos qualitativamente similares, os quais são um forte indício que a convergência da cadeia para a distribuição estacionária foi atingida. Além disso, a função de autocorrelação das cadeias geradas deve ter um decaimento rápido, característica de distribuições estacionárias. A questão do número de cadeias que se deve utilizar é um tanto quanto controversa na literatura, pois alguns autores sugerem muitas cadeias pequenas, outros pesquisadores sugerem muitas cadeias longas ou, até mesmo, uma

cadeia muita longa (ver Gilks, Richardson & Spiegelhalter (1996)). Os valores iniciais devem ser escolhidos com cuidado, para que o *burn-in* da cadeia não fique muito longo. Mas, se a cadeia é irredutível, os valores iniciais não afetarão a distribuição estacionária.

**Exemplo 3.4:** Novamente, considere a série IPCA (Figura 2.1). Assim como no método de máxima verossimilhança, o MNL será utilizado aqui para a modelagem da série. O algoritmo para estimação Bayesiana dos hiperparâmetros  $\sigma_\epsilon^2$  e  $\sigma_\eta^2$ , usando distribuição *a priori* Uniforme, é apresentado abaixo:

**Código 3.4:** Estimação do MNL na linguagem *Ox* usando MCMC

---

```
// Este programa estima os hiperparametros Sigma2Eta e Sigma2Eps
// do MNL utilizando MCMC para obtencao de uma amostra da distribuicao
// a posteriori da media e da mediana, usando uma priori uniforme

#include <oxstd.h>
#include <oxdraw.h>
#include <oxprob.h>
#include "MCMCMNL.ox"
main(){
// declaracao de variaveis
    decl vYt;
// carregando os dados:
    vYt=loadmat("ipca.xls");
// estimacao dos hiperparametros
    MCMCMNL(vYt,10300,300,1,24466,0.04,0.06);
    //primeiro parametro: serie
    //segundo parametro: numero de iteracoes do algoritmo
    //terceiro parametro: burn-in
    //quarto parametro: lag
    //quinto parametro: semente aleatória
    //sexto parametro: sigma da densidade de prop. de sigma2eta
    //setimo parametro: sigma da densidade de prop. de sigma2eps
}
```

---

Este código produz o resultado abaixo, que foi apresentado na Tabela 2.1 para as estimativas Bayesianas. Pode-se verificar que as taxas de aceitação para ambos os hiperparâmetros foram tomadas dentro dos limites especificados de 20% a 50%.

\*\*\*\*\* Ajuste do MNL\*\*\*\*\*

Estimativas Bayesianas:

Parâmetro	media	mediana
sigma2eta	0.029	0.027
sigma2eps	0.210	0.209

Taxas de Aceitacao:

Tx.Aceit.Sigma2Eta(%)=47.8932

Tx.Aceit.Sigma2Eps(%)=48.5534

\*\*\*\*\*

A Figura 3.1 apresenta os gráficos de avaliação da convergência da cadeia de cada hiperparâmetro do ajuste do MNL sob o enfoque Bayesiano. São exibidos histogramas, gráficos de autocorrelação (ACF), da média ergódica e do traço da cadeia de cada variável. Os gráficos indicam a convergência da cadeia, pois existe um decrescimento rápido na ACF, os gráficos da média ergódica mostram que a mesma tende a se estabilizar quando o número de valores da cadeia aumenta e os gráficos do traço não apresentam nenhum comportamento tendencioso.

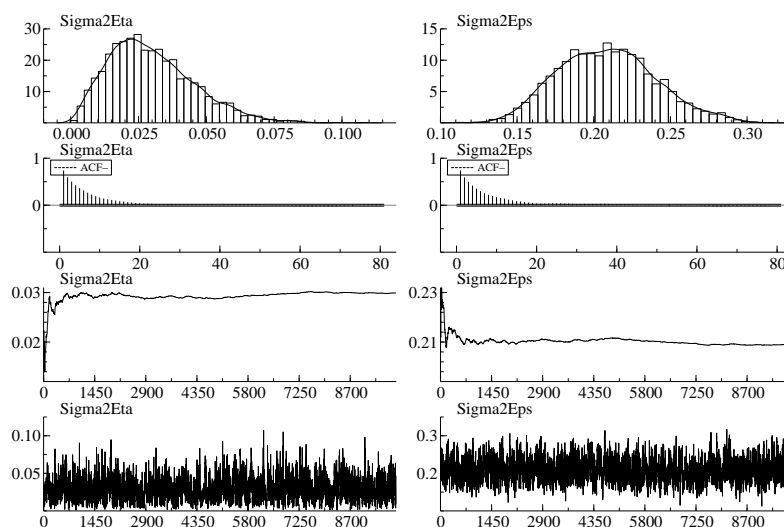


Figura 3.1: Histograma, gráficos de autocorrelação, da média ergódica e do traço da cadeia de cada hiperparâmetro para o ajuste do MNL à série IPCA.

**Exemplo 3.5:** Abaixo é apresentado o programa principal para a estimação Bayesiana do modelo de tendência linear local (MTL), ajustado à série do logaritmo do ICV (Figura 2.2). Usa-se a distribuição *a priori* Uniforme.

---

**Código 3.5:** Estimação do MTL na linguagem *Ox* usando MCMC

---

```
// Este programa estima os hiperparametros Sigma2Eta, Sigma2Csi e Sigma2Eps
// do MTL utilizando MCMC para obtencao de uma amostra da distribuicao
// a posteriori da media e da mediana, usando uma priori uniforme

#include <oxstd.h>
#include <oxdraw.h>
#include <oxprob.h>
#include "MCMCMTL.ox"

main(){
// declaracao de variaveis
  decl vYt;
// carregando os dados:
  vYt=loadmat("ICV.xls");
// estimacao dos hiperparametros
  MCMCMTL(vYt,10300,300,1,24466,3.1,1,0.9);
  //primeiro parametro: serie
  //segundo parametro: numero de iteracoes do algoritmo
  //terceiro parametro: burn-in
  //quarto parametro: lag
  //quinto parametro: semente aleatória
  //sexto parametro: sigma da densidade de prop. de sigma2eta
  //setimo parametro: sigma da densidade de prop. de sigma2csi
  //oitavo parametro: sigma da densidade de prop. de sigma2eps
}
```

---

Este código produz o resultado abaixo, que foi apresentado na Tabela 2.2 para as estimativas Bayesianas.

\*\*\*\*\* Ajuste do MTL\*\*\*\*\*

Estimativas Bayesianas:

Parâmetro	media	mediana
sigma2eta	7,84e-05	7,84e-05
sigma2csi	1,86e-06	1,77e-06
sigma2eps	9,38e-08	7,83e-08

Taxas de Aceitacao:

Tx.Aceit.Sigma2Eta(%)=46.6408

Tx.Aceit.Sigma2Csi(%)=45.5437  
Tx.Aceit.Sigma2Eps(%)=46.1748

\*\*\*\*\*

A Figura 3.2 apresenta os gráficos de avaliação da convergência da cadeia do ajuste do MTL. Os gráficos indicam a convergência da cadeia, pois a função de autocorrelação apresenta um decaimento rápido, a média ergódica se estabiliza em torno de um único valor e o traço da cadeia parece flutuar em torno dos mesmos valores ao longo das iterações.

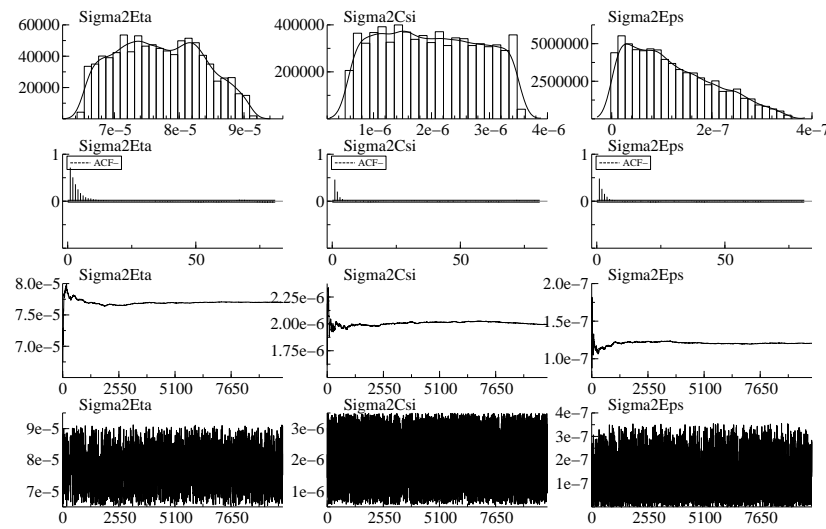


Figura 3.2: Histograma, gráficos de autocorrelação, da média ergódica e do traço da cadeia de cada hiperparâmetro para o ajuste do MTL à série do logaritmo do ICV.

**Exemplo 3.6:** Abaixo é apresentado o programa principal para a estimação Bayesiana do modelo estrutural básico (MEB), ajustado à série do logaritmo da precipitação de  $SO_4$  (Figura 2.3). Usa-se a distribuição *a priori* uniforme.

**Código 3.6:** Estimação do MEB na linguagem *Ox* usando MCMC

---

```
// Este programa estima os hiperparametros Sigma2Eta, Sigma2Csi, Sigma20mega e Sigma2Eps
// do MEB utilizando MCMC para obtencao de uma amostra da distribuicao
// a posteriori da media e da mediana, usando uma priori uniforme
```

```

#include <oxstd.h>
#include <oxdraw.h>
#include <oxprob.h>
#include "MCMCMEB.ox"

main(){
// declaracao de variaveis
    decl vYt;
// carregando os dados:
    vYt=loadmat("Precip.xls");
    vYt=vYt[60:];
// estimacao dos hiperparametros
    MCMCMEB(vYt,10300,300,1,24466,0.0008,0.7e-06,0.0017,0.017);
    //primeiro parametro: serie
    //segundo parametro: numero de iteracoes do algoritmo
    //terceiro parametro: burn-in
    //quarto parametro: lag
    //quinto parametro: semente aleatória
    //sexto parametro: sigma da densidade de prop. de sigma2eta
    //setimo parametro: sigma da densidade de prop. de sigma2csi
    //oitavo parametro: sigma da densidade de prop. de sigma2omega
    //nono parametro: sigma da densidade de prop. de sigma2eps
}

```

Esse programa produz a seguinte saída, cujos dados foram apresentados na Tabela 2.3:

\*\*\*\*\* Ajuste do MEB\*\*\*\*\*

Estimativas Bayesianas:

Parâmetro	media	mediana
sigma2_eta	2,13e-4	1,98e-4
sigma2_csi	2,00e-7	1,14e-7
sigma2_omg	4,69e-4	4,14e-4
sigma2_eps	0,068	0,068

Taxas de Aceitacao:

```

Tx.Aceit.Sigma2Eta(%)=35.2621
Tx.Aceit.Sigma2Csi(%)=33.7379
Tx.Aceit.Sigma2Eps(%)=37.0583
Tx.Aceit.Sigma2Omg(%)=34.932

```

\*\*\*\*\*

A Figura 3.3 apresenta os gráficos de avaliação da convergência da cadeia de cada hiperparâmetro do ajuste do MEB. Os gráficos indicam a con-

vergência da cadeia pois apresentam o mesmo comportamento qualitativo dos gráficos dos exemplos anteriores.

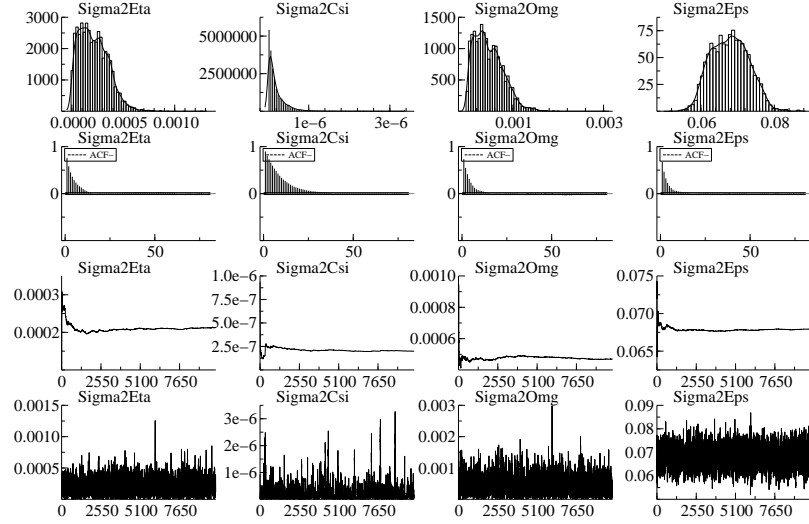


Figura 3.3: Histograma, gráficos de autocorrelação, da média ergódica e do traço da cadeia de cada hiperparâmetro para o ajuste do MEBà série do logaritmo de  $SO_4$ .

### 3.3 Bootstrap

O *bootstrap* (Efron, 1979) é um método computacional desenvolvido para inferência estatística. A idéia central desse método consiste em aproximar a distribuição dos dados ( $F$ ) pela distribuição empírica ( $\hat{F}$ ), baseada nas observações obtidas de  $F$ . Um resultado importante é que a distribuição estimada de  $F$  converge, em probabilidade, para a distribuição dos dados.

Existem basicamente duas maneiras de se realizar o *bootstrap*: não-paramétrica e paramétrica. Na primeira, a reamostragem é feita, com reposição, na própria amostra e cada observação tem a chance de  $1/n$  de ser reamostrada, ou seja, a distribuição estimada empírica  $\hat{F}_{npar}$  é definida como uma distribuição discreta. Na segunda,  $\hat{F}_{par}$  é uma estimativa de  $F$  (conhecida e derivada de algum modelo paramétrico), isto é, as observações são reamostradas da distribuição geradora dos dados cujos parâmetros desconhecidos são



substituídos por suas respectivas estimativas. O *bootstrap* paramétrico é útil em problemas onde se tem algum conhecimento sobre a forma da distribuição da população.

Suponha que a estatística de interesse é dada por  $T(\mathbf{X})$ , onde  $\mathbf{X}$  é uma amostra aleatória da distribuição dos dados. Usando o procedimento *bootstrap*, as  $B$  amostras *bootstrap* são obtidas  $\mathbf{X}^{*1}, \mathbf{X}^{*2}, \dots, \mathbf{X}^{*B}$  e, consequentemente, as réplicas *bootstrap* de qualquer estatística  $T^*(\mathbf{X}^{*b})$  para  $b = 1, 2, \dots, B$ . Pode-se usar  $T^*(\mathbf{X}^*)$  como uma aproximação da estatística  $T(\mathbf{X})$  e, portanto, fazer inferências sobre o parâmetro de interesse. Esse procedimento é denominado como princípio *plug-in* por Efron & Tibshirani (1993). Segundo os mesmos, a função  $\psi = T_F(\mathbf{X})$  da distribuição de  $F$  é estimada pela mesma distribuição empírica  $\hat{F}$ ,  $\hat{\psi} = T_{\hat{F}}(\mathbf{X})$ , logo  $\hat{\psi}$  é a estimativa do parâmetro  $\psi$ .

Efron & Tibshirani (1993) discorrem sobre duas vantagens do método *bootstrap* em relação aos métodos tradicionais. A primeira é que, quando o modo não-paramétrico é usado, o analista ou pesquisador não precisa fazer suposições paramétricas sobre a população latente. A segunda é que, quando o modo paramétrico é utilizado, ele obtém respostas mais acuradas que as fórmulas tradicionais.

Sabe-se que para fazer inferências usando o *bootstrap* é necessário que a suposição de independência das observações seja válida. Levando em consideração que as observações em séries temporais são correlacionadas, uma das possibilidades é aplicar o *bootstrap* nos resíduos do modelo ajustado. Neste trabalho, será utilizada uma proposta de Stoffer & Wall (1991) para a realização do *bootstrap* em modelos estruturais.

#### i) *Bootstrap* não-paramétrico nos resíduos

Seja o modelo na forma de espaço de estados dado em (2.3)-(2.4) com  $d_t = 0$  e  $\mathbf{c}_t = \mathbf{0}$ . Inicialmente, deve-se estimar os hiperparâmetros do modelo que, neste estudo, são as variâncias desconhecidas dos erros  $\varepsilon_t$  e  $\boldsymbol{\eta}_t$  das equações de medida e transição, respectivamente. Para tanto, tem-se que executar o FK para obter os valores das inovações estimadas,  $\hat{\nu}_t$ , e sua variância  $\hat{F}_t$ . Deve-se lembrar que essas quantidades são funções dos hiperparâmetros desconhecidos.

Após a estimação dos hiperparâmetros, por máxima verossimilhança, obtém-se as inovações estimadas,  $\hat{\nu}_t$ , que são reamostradas para a construção da série *bootstrap*. Antes, porém, é necessário centrar e padronizar as inovações para garantir que todos os resíduos possuam, pelo menos, os dois

primeiros momentos iguais (Stoffer & Wall, 1991). Logo,

$$\hat{e}_t = \frac{\hat{\nu}_t - \bar{\hat{\nu}}_t}{\sqrt{\hat{F}_t}},$$

$t = 1, 2, \dots, n$ , onde  $\bar{\hat{\nu}}_t = \frac{1}{n} \sum_{j=1}^n \hat{\nu}_j$ . Se o modelo está bem ajustado, as amostras podem ser tomadas, com reposição, de  $\hat{e}_t$ ,  $t=1, 2, \dots, n$  para obter as inovações *bootstrap*,  $\hat{e}_t^*$ .

A série *bootstrap*,  $y_t^*$ , é construída, recursivamente, utilizando-se as inovações *bootstrap* obtidas acima e os valores  $\hat{F}_t$  e  $\hat{\mathbf{K}}_t$  calculados pelo FK. Inicialmente, as equações do filtro de Kalman para  $\mathbf{a}_{t+1|t}$  e  $y_t$  são escritas em função das inovações. Portanto, tem-se que

$$\mathbf{a}_{t+1|t} = \mathbf{T}_t \mathbf{a}_{t|t-1} + \mathbf{K}_t \nu_t$$

$$y_t = \mathbf{z}_t' \mathbf{a}_{t|t-1} + \nu_t.$$

A seguir, o vetor  $\mathbf{S}_t = \begin{bmatrix} \mathbf{a}_{t+1|t} \\ y_t \end{bmatrix}$  é definido como

$$\mathbf{S}_t = \mathbf{A}_t \mathbf{S}_{t-1} + \mathbf{B}_t e_t, \quad (3.2)$$

$$t = 1, 2, \dots, n, \text{ onde } \mathbf{A}_t = \begin{bmatrix} \mathbf{T}_t & 0 \\ \mathbf{z}_t & 0 \end{bmatrix} \text{ e } \mathbf{B}_t = \begin{bmatrix} \mathbf{K}_t \sqrt{F_t} \\ \sqrt{F_t} \end{bmatrix}.$$

Portanto, resolvendo a equação (3.2) com  $\hat{e}_t^*$  no lugar de  $e_t$  e utilizando os valores estimados  $F_t$  e  $\mathbf{K}_t$ , obtém-se a série *bootstrap*  $y_t^*$ ,  $t = 1, 2, \dots, n$ .

## ii ) *Bootstrap* paramétrico nos resíduos

Diferentemente do *bootstrap* não-paramétrico, no qual os resíduos do modelo ajustado são reamostrados com reposição, o *bootstrap* paramétrico utiliza apenas as estimativas dos parâmetros da série original. O procedimento é realizado da seguinte forma: primeiro, os parâmetros do modelo são estimados baseados na série original. A seguir, as variâncias dos erros nas equações das observações e de transição em (2.3)-(2.4) são substituídas pelas suas respectivas estimativas. Realizando reamostragens, os termos de erro *bootstrap*  $\hat{e}_t^*$  e  $\hat{\boldsymbol{\eta}}_t^*$  são obtidos e a série temporal *bootstrap*,  $y_t^*$ , pode ser construída, para  $t = 1, 2, \dots, n$ , através das equações:

$$y_t^* = \mathbf{z}_t' \boldsymbol{\alpha}_t^* + d_t + \hat{e}_t^*,$$

$$\boldsymbol{\alpha}_t^* = \mathbf{T}_t \boldsymbol{\alpha}_{t-1}^* + \mathbf{c}_t + \mathbf{R}_t \hat{\boldsymbol{\eta}}_t^*.$$

### 3.4 Estudos de simulação

Nesta seção, o desempenho dos estimadores de máxima verossimilhança (EMV) e dos estimadores de Bayes (EB-Mediana e EB-Media), com *prioris* uniforme e Jeffreys, foram avaliados para o MNL, MTL e MEB através de experimentos Monte Carlo (MC), implementados na linguagem Ox. Os resultados são apresentados, respectivamente, nas Tabelas 3.1 a 3.3, com as estimativas comparadas através do vício e do erro quadrático médio (EQM). Os resultados das replicações *bootstrap* também são mostrados nas tabelas, com o intuito de verificar se esta técnica produz resultados similares aos obtidos pelo EMV. Várias combinações dos valores dos hiperparâmetros foram utilizadas e os resultados foram muito similares a estes apresentados aqui. Séries de tamanho  $n = 60$  e  $200$  foram geradas. Para os estimadores Bayesianos, 10.300 amostras foram geradas das quais as 300 primeiras foram excluídas. Para o estimador de máxima verossimilhança foi considerado o número de iterações do algoritmo de maximização numérica BFGS igual a 50. O número de repetições MC e *bootstrap* foram fixados em 500. Todas as tabelas desta seção foram retiradas de Santos & Franco (2009).

Da Tabela 3.1, pode ser visto que os estimadores Bayesianos apresentam, para o componente de nível ( $\sigma_\eta^2$ ), um vício maior que o EMV. As estimativas Bayesianas tendem a superestimar o valor real deste hiperparâmetro, mesmo para séries de tamanho  $n = 200$ . O EQM, entretanto, tem a mesma magnitude para todos os estimadores. Com respeito ao componente de erro,  $\sigma_\epsilon^2$ , não há muita diferença entre os métodos. Pode-se notar também que o *bootstrap* imita o comportamento do EMV, o que permite a essa técnica ser usada como ferramenta para construir intervalos de confiança neste caso. Comparando os estimadores Bayesianos, a distribuição *a priori* de Jeffreys parece levar a estimativas mais próximas do valor real dos hiperparâmetros do que a distribuição *a priori* uniforme. Como esperado, o EQM decresce quando o tamanho da amostra aumenta.

Os resultados para o MTL são apresentados na Tabela 3.2. As mesmas conclusões obtidas para os hiperparâmetros  $\sigma_\eta^2$  e  $\sigma_\epsilon^2$  no modelo MNL com respeito ao vício e ao EQM podem ser verificadas aqui. Quanto ao componente de tendência,  $\sigma_\xi^2$ , parece não existir muita diferença entre os procedimentos, sendo todas as estimativas muito próximas do valor real do hiperparâmetro.

Simulações para o MEB são apresentadas na Tabela 3.3. Parece que a introdução do componente sazonal no modelo traz algumas mudanças nos métodos de estimação com respeito às conclusões retiradas para os outros

Tabela 3.1: Resultados da estimação no MNL segundo os métodos de máxima verossimilhança, *bootstrap* e estimação Bayesiana.

n	$\psi$	EMV	BOOT	EB-Mediana	EB-Media	EB-Mediana	EB-Media
		<i>Priori</i> uniforme		<i>Priori</i> Jeffreys			
60	$\sigma_\eta^2 = 0,5$	0,483	0,517	0,654	0,725	0,513	0,580
		(-0,017)	(0,017)	(0,154)	(0,225)	(0,013)	(0,080)
		<b>0,063</b>	<b>0,072</b>	<b>0,093</b>	<b>0,124</b>	<b>0,057</b>	<b>0,066</b>
	$\sigma_\epsilon^2 = 1,0$	1,017	1,002	1,020	1,055	1,055	1,088
		(0,017)	(0,002)	(0,020)	(0,055)	(0,055)	(0,087)
		<b>0,081</b>	<b>0,086</b>	<b>0,084</b>	<b>0,089</b>	<b>0,101</b>	<b>0,108</b>
200	$\sigma_\eta^2 = 0,5$	0,500	0,503	0,540	0,560	0,510	{0,529}
		(0,000)	(0,003)	(0,040)	(0,060)	(0,010)	(0,029)
		<b>0,017</b>	<b>0,017</b>	<b>0,026</b>	<b>0,026</b>	<b>0,016</b>	<b>0,017</b>
	$\sigma_\epsilon^2 = 1,0$	1,000	1,001	1,020	1,004	1,020	1,013
		(0,000)	(0,001)	(0,020)	(0,004)	(0,020)	(0,013)
		<b>0,016</b>	<b>0,017</b>	<b>0,024</b>	<b>0,024</b>	<b>0,025</b>	<b>0,025</b>

Obs.: Os números entre parênteses são o vício e, em negrito, o erro quadrático médio.

Tabela 3.2: Resultados da estimação no MTL segundo os métodos de máxima verossimilhança, *bootstrap* e estimação Bayesiana.

n	$\psi$	EMV	BOOT	EB-Mediana	EB-Media	EB-Mediana	EB-Media
		<i>Priori</i> uniforme		<i>Priori</i> Jeffreys			
60	$\sigma_\eta^2 = 0,5$	0,562	0,670	0,681	0,758	0,688	0,740
		(0,062)	(0,170)	(0,181)	(0,258)	(0,188)	(0,240)
		<b>0,364</b>	<b>0,287</b>	<b>0,160</b>	<b>0,201</b>	<b>0,199</b>	<b>0,215</b>
	$\sigma_\xi^2 = 0,1$	0,097	0,100	0,199	0,233	0,070	0,076
		(-0,003)	(0,000)	(0,099)	(0,133)	(-0,030)	(-0,024)
		<b>0,005</b>	<b>0,004</b>	<b>0,236</b>	<b>0,252</b>	<b>0,004</b>	<b>0,004</b>
	$\sigma_\epsilon^2 = 1,0$	0,963	0,912	0,943	0,952	0,907	0,916
		(-0,037)	(-0,088)	(-0,057)	(-0,048)	(-0,093)	(-0,084)
		<b>0,153</b>	<b>0,132</b>	<b>0,098</b>	<b>0,095</b>	<b>0,085</b>	<b>0,082</b>
	$\sigma_\eta^2 = 0,5$	0,522	0,530	0,553	0,554	0,575	0,573
		(0,022)	(0,030)	(0,053)	(0,054)	(0,075)	(0,073)
		<b>0,120</b>	<b>0,112</b>	<b>0,091</b>	<b>0,090</b>	<b>0,098</b>	<b>0,097</b>
200	$\sigma_\xi^2 = 0,1$	0,099	0,099	0,107	0,092	0,110	0,097
		(-0,001)	(-0,001)	(0,007)	(-0,008)	(0,010)	(-0,003)
		<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>
	$\sigma_\epsilon^2 = 1,0$	0,982	0,979	0,991	0,955	0,993	0,956
		(-0,018)	(-0,022)	(-0,009)	(-0,045)	(-0,007)	(-0,044)
		<b>0,045</b>	<b>0,044</b>	<b>0,038</b>	<b>0,038</b>	<b>0,043</b>	<b>0,042</b>

Obs.: Os números entre parênteses são o vício e, em negrito, o erro quadrático médio.

dois modelos. Por exemplo, o vício ainda é maior para os estimadores Bayesianos no caso do componente de nível,  $\sigma_\eta^2$ , mas agora os EB com a distribuição *a priori* uniforme apresentam o menor EQM. Para a tendência,  $\sigma_\xi^2$ , os EB com a distribuição *a priori* uniforme possuem maior vício que os outros procedimentos, entretanto o EQM é muito similar para todos eles. O componente de erro,  $\sigma_\epsilon^2$ , tem menor vício para os EB com distribuição *a priori* de Jeffreys, mas menor EQM para os EB com distribuição *a priori* uniforme, entretanto os resultados são muito parecidos para tamanho de amostra  $n = 200$ . Com respeito ao componente sazonal, o EMV é muito similar aos EB com distribuição *a priori* de Jeffreys, apresentando vício e EQM menores. Novamente, o *bootstrap* mostrou um desempenho próximo dos resultados do EMV. Comparando os estimadores Bayesianos, a mediana apresenta menores valores do vício e do EQM, independente da distribuição *a priori* utilizada, exceto para  $\sigma_\epsilon^2$ , onde a média parece ter um desempenho melhor.

Tabela 3.3: Resultados da estimação no MEB segundo os métodos de máxima verossimilhança, *bootstrap* e estimação Bayesiana.

n	$\psi$	EMV	BOOT	EB-Mediana	EB-Media	EB-Mediana	EB-Media
		Priori uniforme				Priori Jeffreys	
60	$\sigma_\eta^2 = 0,5$	0,550	0,507	0,614	0,663	0,616	0,677
		(0,050)	(0,007)	(0,114)	(0,163)	(0,116)	(0,177)
		<b>0,223</b>	<b>0,143</b>	<b>0,110</b>	<b>0,124</b>	<b>0,172</b>	<b>0,188</b>
	$\sigma_\xi^2 = 0,03$	0,026	0,036	0,049	0,056	0,023	0,038
		(-0,004)	(0,006)	(0,019)	(0,026)	(-0,007)	(0,008)
		<b>0,001</b>	<b>0,002</b>	<b>0,001</b>	<b>0,002</b>	<b>0,001</b>	<b>0,001</b>
	$\sigma_\omega^2 = 0,1$	0,105	0,097	0,152	0,166	0,103	0,124
		(0,005)	(0,004)	(0,052)	(0,066)	(0,003)	(0,024)
		<b>0,007</b>	<b>0,006</b>	<b>0,011</b>	<b>0,013</b>	<b>0,007</b>	<b>0,008</b>
	$\sigma_\epsilon^2 = 1,0$	0,973	1,109	0,914	0,932	0,991	1,009
		(-0,027)	(0,109)	(-0,086)	(-0,068)	(0,009)	(0,009)
		<b>0,197</b>	<b>0,218</b>	<b>0,140</b>	<b>0,132</b>	<b>0,175</b>	<b>0,167</b>
200	$\sigma_\eta^2 = 0,5$	0,523	0,500	0,539	0,551	0,521	0,536
		(0,023)	(0,000)	(0,039)	(0,051)	(0,021)	(0,036)
		<b>0,072</b>	<b>0,080</b>	<b>0,055</b>	<b>0,058</b>	<b>0,075</b>	<b>0,073</b>
	$\sigma_\xi^2 = 0,03$	0,029	0,030	0,034	0,036	0,029	0,033
		(-0,001)	(0,000)	(0,004)	(0,006)	(0,001)	(0,003)
		<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
	$\sigma_\omega^2 = 0,1$	0,100	0,100	0,112	0,115	0,096	0,102
		(0,000)	(0,000)	(0,012)	(0,015)	(0,004)	(0,002)
		<b>0,001</b>	<b>0,001</b>	<b>0,002</b>	<b>0,002</b>	<b>0,001</b>	<b>0,001</b>
	$\sigma_\epsilon^2 = 1,0$	0,996	1,053	0,989	0,994	1,004	1,008
		(-0,004)	(0,053)	(-0,011)	(0,006)	(0,004)	(0,008)
		<b>0,054</b>	<b>0,071</b>	<b>0,057</b>	<b>0,054</b>	<b>0,058</b>	<b>0,057</b>

Obs.: Os números entre parênteses são o vício e, em negrito, o erro quadrático médio.

## Capítulo 4

# Intervalos e Testes de Hipóteses

Neste capítulo são apresentados alguns métodos de construção de intervalos de confiança, intervalos de credibilidade e testes de hipóteses para os hiperparâmetros dos modelos descritos no Capítulo 2.

Utilizando propriedades assintóticas dos estimadores de máxima verossimilhança, intervalos de confiança assintóticos para os hiperparâmetros, baseados na distribuição Normal, podem ser obtidos. Porém, o cálculo destes intervalos se torna complicado devido à dificuldade em se obter a variância assintótica dos hiperparâmetros (Harvey, 1989). Além disto, nem sempre as séries reais são grandes o suficiente para se assumir uma distribuição Normal assintótica para os dados. Assim, intervalos de confiança *bootstrap* e intervalos de credibilidade são boas opções para evitar problemas como os citados anteriormente.

Nos trabalhos de Efron & Tibshirani (1993) e Shao & Tu (1995) são apresentados os principais métodos de construção de intervalos de confiança *bootstrap*, dentre eles o *bootstrap-t*, o *bootstrap* percentílico e os métodos de correção para vício (BC) e de correção para vício e aceleração (BCa). Estes autores fazem um estudo minucioso do modo de cálculo dos intervalos, assim como de suas propriedades assintóticas. Além disto, comparações entre os procedimentos definem as vantagens e desvantagens de cada um deles (ver Franco & Souza (2002)).

Os intervalos de credibilidade Bayesianos, ao contrário dos intervalos de confiança, são calculados de forma natural através da distribuição *a posteriori* do parâmetro de interesse. Idealmente, os intervalos devem ser definidos com o menor tamanho e maior nível de credibilidade possíveis.

Testes assintóticos para testar hipóteses com respeito aos hiperparâmetros

em modelos estruturais foram desenvolvidos por alguns autores, tais como LaMotte & McWhorter (1978), Nyblom & Makelainen (1983) e Harvey & Streibel (1998). Neste capítulo, testes baseados na técnica *bootstrap* são propostos e se configuram como uma boa alternativa em situações onde a distribuição da estatística de teste sob a hipótese nula é desconhecida. Recentemente, muitos autores têm estudado o comportamento de testes *bootstrap* em séries temporais. Entre eles podemos citar os trabalhos de Dettea & Weißbach (2009), Pérez-Alonso (2007), Cavaliere & Taylor (2008), Paparoditis & Politis (2005), Fuchun & Tkacz (2006) e Palm, Smeekes & Urbain (2008).

A Seção 4.1 descreve os principais métodos de construção de intervalos de confiança e intervalos de credibilidade existentes para se fazer inferência sobre os hiperparâmetros de modelos estruturais, além de apresentar um estudo de simulação comparando estes intervalos. A Seção 4.2 introduz testes *bootstrap* para se realizar testes de hipóteses para os hiperparâmetros dos modelos, com um estudo de simulação para o MNL e o MTL e uma aplicação a dados reais de energia elétrica.

## 4.1 Intervalos de confiança

### 4.1.1 Intervalos de confiança assintóticos

Sob certas condições de regularidade, o estimador de máxima verossimilhança,  $\hat{\psi}$ , é assintoticamente Normal com média  $\psi$  e matriz de covariância assintótica  $aVar(\hat{\psi}) = n^{-1}IA^{-1}(\psi)$ , na qual  $IA(\psi) = \lim_{n \rightarrow \infty} n^{-1}I(\psi)$  e  $I(\psi)$  é a matriz de informação de Fisher. Além disso, em amostras finitas,  $aVar(\hat{\psi})$  pode ser considerada como o inverso da matriz de informação (Harvey, 1989). Usando a distribuição assintótica dos estimadores de máxima verossimilhança, intervalos de confiança assintóticos podem ser construídos para os hiperparâmetros (ver Casella & Berger (2002) e Migon & Gamerman (1999)).

Seja  $\psi_i$ ,  $i = 1, \dots, p$ , um hiperparâmetro qualquer no vetor  $\psi$ . Então, um intervalo de confiança assintótico de  $100(1 - \kappa)\%$  para  $\psi_i$  é dado por:

$$\hat{\psi}_i \pm z_{\kappa/2} \sqrt{aVar(\hat{\psi}_i)},$$

onde  $z_{\kappa/2}$  é o  $\kappa/2$  percentil da distribuição Normal-padrão e  $aVar(\hat{\psi}_i)$  é obtido

dos elementos da diagonal da matriz  $\mathbf{aVar}(\hat{\boldsymbol{\psi}})$ . Para que as condições de regularidade sejam satisfeitas, em modelos mais comportados como MNL, MTL e MEB, basta que os hiperparâmetros sejam pontos interiores ao espaço paramétrico (ver Harvey (1989)).

O cálculo das derivadas do erro de previsão e de sua variância com respeito aos hiperparâmetros da matriz de informação não é uma tarefa trivial. Harvey (1989) propõe uma forma de aproximação numérica no cálculo das derivadas de  $F_t$  e  $\nu_t$ , que é extremamente útil no cálculo da matriz informação de Fisher cuja forma, em uma expressão simplificada, é dada por:

$$\mathbf{I}_{ij}(\boldsymbol{\psi}) = \frac{1}{2} \sum_t \left[ tr \left[ F_t^{-1} \frac{\partial F_t}{\partial \psi_i} F_t^{-1} \frac{\partial F_t}{\partial \psi_j} \right] \right] + E \left[ \sum_t \left( \frac{\partial \nu_t}{\partial \psi_i} \right)' F_t^{-1} \frac{\partial \nu_t}{\partial \psi_j} \right], \quad (4.1)$$

onde  $i, j = 1, \dots, p$  e  $t = 1, \dots, n$ . As derivadas de  $F_t$  e  $\nu_t$  podem ser calculadas pelo seguinte procedimento:

**Algoritmo 4.1:** Cálculo das derivadas de  $F_t$  e  $\nu_t$

---

**faça**  $i = 1$

**repita** os seguintes passos

**acrescente** uma pequena quantidade,  $\delta_i$ , ao hiperparâmetro  $\psi_i$

**execute** o Filtro de Kalman com o novo valor do hiperparâmetro ( $\psi_i$ ) e mantenha constante os valores dos outros hiperparâmetros

**obtenha** o novo vetor do erro de previsão,  $\nu_t^{(i)}$ , e sua variância,  $F_t^{(i)}$

**calcule** as expressões  $\delta_i^{-1}[\nu_t^{(i)} - \nu_t]$  e  $\delta_i^{-1}[F_t^{(i)} - F_t]$  que são aproximações numéricas para as requeridas derivadas

**faça**  $i = i + 1$

**até**  $i = p$ .

---

Em Franco *et al.* (2008) foi realizado um estudo Monte Carlo para a determinação da quantidade  $\delta_i$ . As Figuras (4.1)-(4.3) mostram que, para os modelos analisados neste trabalho,  $\delta_i = 0,0001$  é uma boa escolha, para todo  $i = 1, \dots, p$ .

**Exemplo 4.1:** Voltando à série do logaritmo da precipitação de  $SO_4$ , deseje-se agora estimar intervalos de confiança assintóticos para os hiperparâmetros



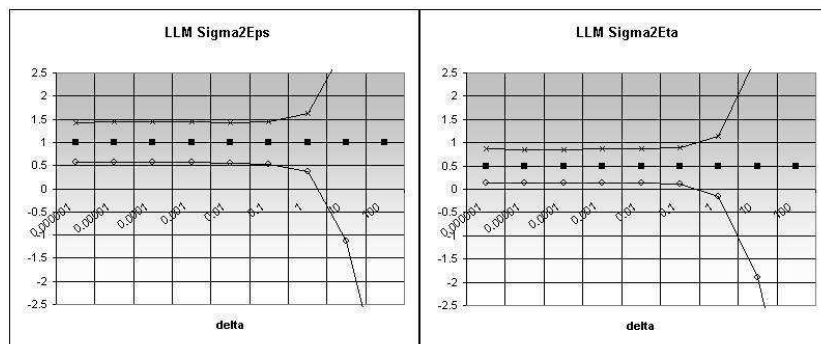


Figura 4.1: Estudo Monte Carlo para determinar  $\delta$  no MNL.

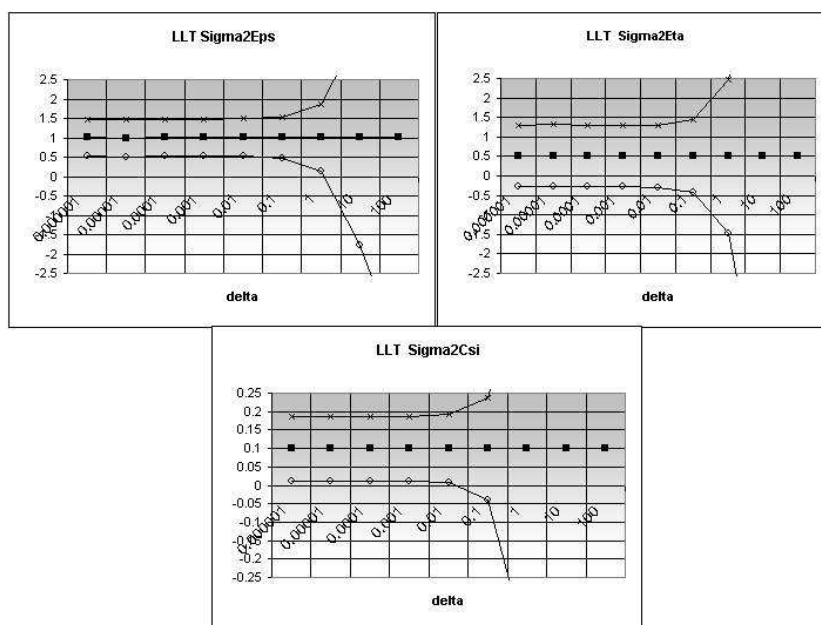


Figura 4.2: Estudo Monte Carlo para determinar  $\delta$  no MTL.

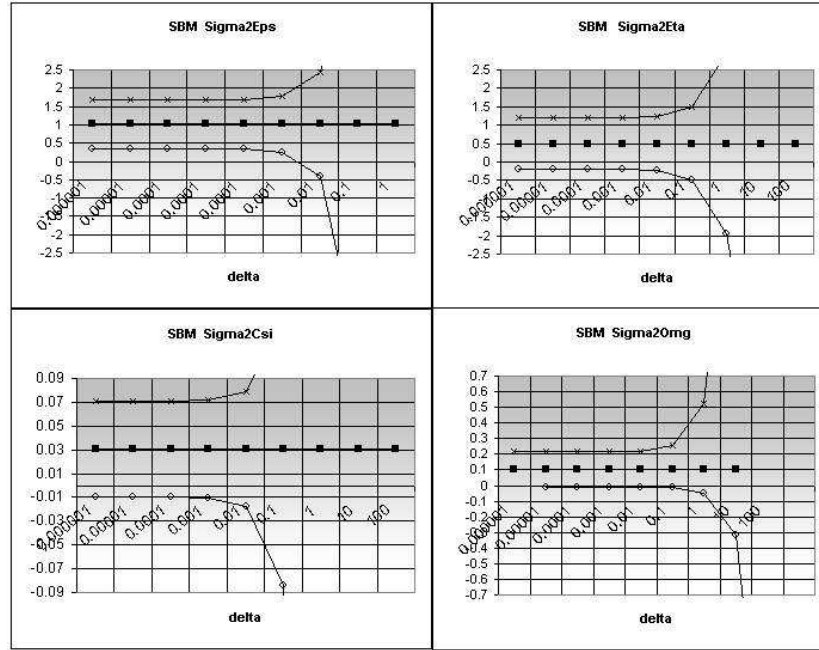


Figura 4.3: Estudo Monte Carlo para determinar  $\delta$  no MEB.

do MEB ajustado à série. A Tabela 4.1 apresenta os intervalos para um nível de confiança de 95%.

Pelos resultados obtidos acima, verifica-se que os intervalos de confiança dos hiperparâmetros  $\sigma_\eta^2$ ,  $\sigma_\xi^2$  e  $\sigma_\omega^2$  estão bem concentrados em torno de zero, o que poderia levar à conclusão de que o único hiperparâmetro realmente importante no modelo seja  $\sigma_\epsilon^2$ . Observa-se também que os intervalos assintóticos podem levar a valores fora do espaço paramétrico, já que os limites inferiores nos três primeiros casos foram negativos.

Tabela 4.1: Intervalos de confiança assintóticos de 95% para os hiperparâmetros do MEB ajustado à série do logaritmo de  $SO_4$ .

Hiperparâmetro	Limite inferior	Limite superior
$\sigma_\eta^2$	$-4,1 \times 10^{-4}$	$6,7 \times 10^{-4}$
$\sigma_\xi^2$	$-5,0 \times 10^{-5}$	$5,0 \times 10^{-5}$
$\sigma_\omega^2$	$-5,6 \times 10^{-4}$	$10,4 \times 10^{-4}$
$\sigma_\epsilon^2$	0,054	0,092

### 4.1.2 Intervalos de confiança *bootstrap*

Os intervalos de confiança *bootstrap*, propostos por Efron & Tibshirani (1993), são utilizados aqui para se fazer inferências sobre os parâmetros dos modelos MNL, MTL e MEB.

- **Intervalo de confiança *bootstrap* percentílico**

O intervalo de confiança *bootstrap* percentílico consiste no seguinte. Após a estimação dos valores dos hiperparâmetros para cada uma das  $B$  séries *bootstrap*, ordena-se esses valores e toma-se o percentil  $100(\kappa/2)^{\circ}$  como o limite inferior do intervalo e o percentil  $100(1 - \kappa/2)^{\circ}$  como o limite superior do intervalo. Pode-se definir o intervalo  $100(1 - \kappa)\%$  para  $\psi_i$  como

$$\left[ \hat{\psi}_i^{*(\kappa/2)}; \hat{\psi}_i^{*(1-\kappa/2)} \right],$$

onde  $\hat{\psi}_i^*$ ,  $i = 1, \dots, p$ , é o vetor das estimativas ordenadas das  $B$  séries *bootstrap* de um hiperparâmetro qualquer em  $\psi$ .

- **Intervalo de confiança *bootstrap* BC**

Esse método também utiliza os percentis da distribuição *bootstrap*, mas não exatamente o  $100(\kappa/2)^{\circ}$  e o  $100(1 - \kappa/2)^{\circ}$ . Ao invés disso, ele corrige esses valores para possíveis vícios na estimação de  $\psi$ , através de uma quantidade  $\hat{z}_0$  que mede o vício mediano de  $\hat{\psi}$ .

O intervalo BC (*Bias-corrected*)  $100(1 - \kappa)\%$  é definido por

$$\left[ \hat{\psi}_i^{*(\kappa_1)}; \hat{\psi}_i^{*(\kappa_2)} \right],$$

onde  $i = 1, \dots, p$ ,  $\kappa_1 = \Phi(2\hat{z}_0 + z^{(\kappa/2)})$  e  $\kappa_2 = \Phi(2\hat{z}_0 + z^{(1-\kappa/2)})$ . A função  $\Phi(\cdot)$  é a função de distribuição acumulada da Normal-padrão e  $z^{(k/2)}$  seu  $100(k/2)^{\circ}$  percentil.

O valor de  $\hat{z}_0$  é calculado usando a proporção de amostras *bootstrap* que são menores que os hiperparâmetros estimados usando a série original:

$$\hat{z}_0 = \Phi^{-1} \left( \frac{\#\hat{\psi}_i^*(b) < \hat{\psi}_i}{B} \right).$$

• **Intervalo de confiança *bootstrap* BCa**

No intervalo BCa (*Bias-corrected and accelerated*), além do cálculo da quantidade  $\hat{z}_0$ , também é necessário calcular a aceleração, dada por

$$\hat{a} = \frac{\sum_{i=1}^n \left( \hat{\psi}_i^{(\cdot)} - \hat{\psi}_i^{(j)} \right)^3}{\left[ \sum_{i=1}^n \left( \hat{\psi}_i^{(\cdot)} - \hat{\psi}_i^{(j)} \right)^2 \right]^{3/2}},$$

onde  $\hat{\psi}_i^{(j)}$  é uma estimativa do hiperparâmetro com a  $j$ -ésima observação deletada e  $\hat{\psi}_i^{(\cdot)} = \sum_{j=1}^n \hat{\psi}_i^{(j)} / n$ . Então, da mesma forma que se calcula o intervalo BC, calcula-se o intervalo BCa de nível  $100(1 - \kappa)\%$

$$\left[ \hat{\psi}_i^{*(\kappa_1)}; \hat{\psi}_i^{*(\kappa_2)} \right]$$

mas com  $\kappa_1 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 + z(\kappa/2)}{1 - \hat{a}(\hat{z}_0 + z(\kappa/2))} \right)$  e  $\kappa_2 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 + z(1 - \kappa/2)}{1 - \hat{a}(\hat{z}_0 + z(1 - \kappa/2))} \right)$ .

• **Comparações entre os intervalos de confiança *bootstrap***

Uma forma de comparar os intervalos descritos acima consiste em avaliar se a probabilidade de cobertura dos mesmos converge para o nível nominal  $(1 - \kappa)$  quando  $n \rightarrow \infty$  (consistência) e quão rápida é essa convergência (precisão).

**Definição 2.** Um intervalo de confiança  $I_n$  para  $\psi$  com nível nominal  $1 - \kappa$  é consistente se  $P(\psi \in I_n) \rightarrow 1 - \kappa$  quando  $n \rightarrow \infty$ .

Os procedimentos *bootstrap* percentílico, BC e BCa são consistentes, segundo Shao & Tu (1995). Para intervalos bilaterais, os procedimentos *bootstrap* percentílico, BC e BCa são precisos de segunda ordem e também invariantes a transformações (isto é, quando os pontos limites do intervalo são transformados corretamente quando se muda de  $\psi$  para uma função de  $\psi$ ). Diferentemente da noção de acurácia, a definição de correção refere-se a quão próximo o ponto limite do intervalo candidato está do ponto limite do intervalo exato ou ideal. Em situações onde os limites podem ser definidos, a Normal-padrão e a t-Student produzem apenas pontos limites de primeira ordem corretos, enquanto que alguns métodos *bootstrap* produzem pontos limites de confiança corretos de segunda ordem (Efron & Tibshirani, 1993).

O número de replicações *bootstrap* pode variar de um problema para outro. Por exemplo, na estimação do erro padrão,  $B = 50$  é frequentemente suficiente (Efron & Tibshirani, 1993, p.52). Já na construção intervalos de confiança *bootstrap*, Efron & Tibshirani (1993) sugerem um valor de  $B \geq 500$ , dependendo da complexidade do problema e também do custo computacional. Hall (1986) aconselha tomar a probabilidade de cobertura nominal como um múltiplo de  $(B + 1)^{-1}$ . Para intervalos de confiança, na literatura, os valores  $B = 500, 1000$  e  $2000$  são os mais recorrentes. Davidson & Hinkley (1997) propõem utilizar  $B \geq 100$ , para obter uma boa aproximação, se o objetivo for calcular o nível de significância em testes de hipóteses. Às vezes, a melhora nas estimativas de um determinado parâmetro é muito pequena, passando de um determinado valor de  $B$  para um valor maior. Então, o número de réplicas *bootstrap* deve ser escolhido com cautela.

**Exemplo 4.2:** Utilizando-se novamente a série do logaritmo da precipitação de  $SO_4$ , intervalos de confiança *bootstrap* de 95% de confiança para os hiperparâmetros do MEB ajustado à série foram calculados. Os resultados são apresentados na Tabela 4.2. Novamente observa-se que o único componente cujos limites do intervalo não estão concentrados em torno de zero é  $\sigma_\epsilon^2$ . Neste caso, a série parece realmente não apresentar componentes estocásticos para tendência e sazonalidade. Comparados aos intervalos assintóticos mostrados na Tabela 4.1, pode-se verificar que os intervalos *bootstrap* possuem menor amplitude, além de não apresentarem valores negativos para os limites inferiores.

Tabela 4.2: Intervalos de confiança *bootstrap* de 95% para os hiperparâmetros do MEB ajustado à série do logaritmo de  $SO_4$ .

Hiperparâmetro	Limite inferior	Limite superior
$\sigma_\eta^2$	$1,95 \times 10^{-9}$	$4,99 \times 10^{-4}$
$\sigma_\xi^2$	$5,18 \times 10^{-22}$	$1,43 \times 10^{-7}$
$\sigma_\omega^2$	$1,05 \times 10^{-13}$	$1,20 \times 10^{-3}$
$\sigma_\epsilon^2$	0,056	0,083

### 4.1.3 Intervalos de credibilidade

Seja  $\psi_i$ ,  $i = 1, \dots, p$ , um hiperparâmetro qualquer no vetor  $\boldsymbol{\psi}$ . Fixado  $0 < \kappa < 1$ , qualquer intervalo  $(t_1, t_2)'$  satisfazendo

$$\int_{t_1}^{t_2} \pi(\psi_i | \mathbf{Y}_n) d\psi_i = 1 - \kappa$$

é um intervalo de credibilidade para  $\psi_i$  de nível  $100(1 - \kappa)\%$ . Esses intervalos são aproximados pelas correspondentes estatísticas de ordem da amostra gerada via MCMC de  $\psi_i$ , tomando os percentis simétricos. Para maiores detalhes ver Migon & Gamerman (1999).

**Exemplo 4.3:** Na Tabela 4.3 é apresentado o resultado do cálculo do intervalo de credibilidade de 0,95 para os hiperparâmetros do MEB ajustado à série do logaritmo da precipitação de  $SO_4$ . Utilizou-se a distribuição *a priori* uniforme para os hiperparâmetros.

Tabela 4.3: Intervalos de credibilidade de 0,95 para os hiperparâmetros do MEB ajustado à série do logaritmo de  $SO_4$ .

Hiperparâmetro	Limite inferior	Limite superior
$\sigma_\eta^2$	$1,18 \times 10^{-5}$	$5,12 \times 10^{-4}$
$\sigma_\xi^2$	$2,76 \times 10^{-9}$	$8,68 \times 10^{-7}$
$\sigma_\omega^2$	$2,90 \times 10^{-5}$	$1,12 \times 10^{-3}$
$\sigma_\epsilon^2$	0,059	0,077

As mesmas conclusões anteriores podem ser tiradas aqui. Além disto, os intervalos de credibilidade apresentam comportamento bem semelhante aos intervalos *bootstrap*.

### 4.1.4 Estudos de simulação

É bem conhecido que a probabilidade de credibilidade e o nível de confiança têm diferentes significados e interpretações. Citando Casella & Berger (2002), "A primeira reflete a crença subjetiva do pesquisador, enquanto o último reflete a incerteza no procedimento amostral". Assim, neste trabalho, uma tentativa de comparar os resultados obtidos usando esses dois procedimentos será feita. A porcentagem de vezes que os intervalos construídos contêm

o valor verdadeiro dos hiperparâmetros é obtida baseada em um número grande de simulações Monte Carlo. Os resultados, junto com a amplitude dos intervalos, são mostrados nas Tabelas 4.4 a 4.6, para o MNL, MTL e MEB, respectivamente.

Da Tabela 4.4 pode ser visto que, para o MNL, todos os intervalos apresentam, em geral, taxas de cobertura próximas do nível nominal assumido de 0,95, a única exceção sendo o intervalo assintótico para  $n = 60$  e  $\sigma_\eta^2 = 0,5$ . No entanto, os intervalos de confiança *bootstrap* e assintótico devem ser preferidos, porque possuem menores amplitudes.

Tabela 4.4: Intervalos de confiança e de credibilidade de 95% para o MNL.

n	$\psi$	Assint.	Cred.Unif	Cred.Jeff	Perc	BC	BCa
60	$\sigma_\eta^2 = 0,5$	[0,03; 0,94]	[0,25;1,61]	[0,19;1,36]	[0,11;1,13]	[0,16;1,26]	[0,16;1,26]
		(0,91)	(1,36)	(1,17)	(1,02)	(1,10)	(1,10)
		<b>0,88</b>	<b>0,95</b>	<b>0,98</b>	<b>0,90</b>	<b>0,92</b>	<b>0,92</b>
	$\sigma_\epsilon^2 = 1,0$	[0,45;1,58]	[0,22;1,83]	[0,51;1,85]	[0,44;1,62]	[0,47;1,65]	[0,43;1,61]
		(1,13)	(1,61)	(1,34)	(1,18)	(1,18)	(1,18)
		<b>0,95</b>	<b>0,97</b>	<b>0,94</b>	<b>0,94</b>	<b>0,94</b>	<b>0,95</b>
200	$\sigma_\eta^2 = 0,5$	[0,24;0,76]	[0,32;0,92]	[0,30;0,87]	[0,27;0,79]	[0,28;0,81]	[0,28; 0,82]
		(0,52)	(0,60)	(0,57)	(0,52)	(0,53)	(0,54)
		<b>0,93</b>	<b>0,94</b>	<b>0,97</b>	<b>0,94</b>	<b>0,94</b>	<b>0,94</b>
	$\sigma_\epsilon^2 = 1,0$	[0,69;1,31]	[0,72;1,38]	[0,71; 1,36]	[0,70;1,33]	[0,71;1,34]	[0,70; 1,33]
		(0,62)	(0,66)	(0,65)	(0,63)	(0,63)	(0,63)
		<b>0,97</b>	<b>0,96</b>	<b>0,95</b>	<b>0,93</b>	<b>0,97</b>	<b>0,97</b>

Obs.: Os números em colchetes são os limites médios dos intervalos, em parênteses a amplitude e, em negrito, a taxa de cobertura dos intervalos.

Na Tabela 4.5 estão os resultados para o MTL. Para o componente de nível,  $\sigma_\eta^2$ , os intervalos de credibilidade, em geral, fornecem a melhor combinação de taxa de cobertura mais próxima do nível nominal de 0,95 e menores amplitudes. Para o componente de tendência,  $\sigma_\xi^2$ , os intervalos *bootstrap*, especialmente o BC e o BCa, mostram um melhor desempenho. Finalmente, para o componente do erro, os intervalos *bootstrap* para  $\sigma_\epsilon^2$  têm a taxa de cobertura mais próxima de 0,95, entretanto a amplitude desses intervalos é ligeiramente maior do que dos intervalos de credibilidade. Pode ser notado também que, muitas vezes, os intervalos assintóticos apresentam valores negativos para o limite inferior dos intervalos, e isso não é desejável, uma vez que os hiperparâmetros são variâncias.

Para o MEB, mostrado na Tabela 4.6, os intervalos de credibilidade apresentam, em geral, a melhor combinação de menor amplitude para os interva-

Tabela 4.5: Intervalos de confiança e de credibilidade de 95% para o MTL.

n	$\psi$	Assint.	Cred.Unif	Cred.Jeff	Perc	BC	BCa
60	$\sigma_\eta^2 = 0,5$	[-0,67; 1,79]	[0,07;1,89]	[ 0,14;1,65]	[0,01;2,15]	[0,01;2,62]	[0,01;2,55]
		(2,46)	(1,82)	(1,51)	(2,15)	(2,61)	(2,54)
		<b>0,97</b>	<b>0,99</b>	<b>0,93</b>	<b>0,99</b>	<b>0,99</b>	<b>0,99</b>
	$\sigma_\xi^2 = 0,1$	[-0,04;0,23]	[0,04; 0,84]	[0,02;0,17]	[0,01; 0,25]	[0,01;0,33]	[0,01;0,32]
		(0,27)	(0,80)	(0,15)	(0,24)	(0,32)	(0,31)
		<b>0,88</b>	<b>0,89</b>	<b>0,71</b>	<b>0,86</b>	<b>0,97</b>	<b>0,95</b>
	$\sigma_\epsilon^2 = 1,0$	[0,22;1,71]	[0,41;1,52]	[0,44;1,43]	[0,19;1,64]	[0,15;1,64]	[0,16;1,65]
		(1,49)	(1,11)	(0,99)	(1,45)	(1,49)	(1,49)
		<b>0,92</b>	<b>0,93</b>	<b>0,90</b>	<b>0,94</b>	<b>0,96</b>	<b>0,96</b>
200	$\sigma_\eta^2 = 0,5$	[-0,16; 1,20]	[0,14;1,14]	[ 0,16;1,09]	[0,05;1,30]	[0,05;1,46]	[0,05;1,43]
		(1,36)	(1,00)	(0,93)	(1,25)	(1,41)	(1,38)
		<b>0,97</b>	<b>0,93</b>	<b>0,90</b>	<b>0,98</b>	<b>0,99</b>	<b>0,99</b>
	$\sigma_\xi^2 = 0,1$	[ 0,02;0,18]	[0,06; 0,18]	[0,05; 0,15]	[0,04; 0,19]	[0,04;0,19]	[0,04;0,19]
		(0,16)	(0,13)	(0,10)	(0,15)	(0,15)	(0,15)
		<b>0,90</b>	<b>0,90</b>	<b>0,87</b>	<b>0,94</b>	<b>0,95</b>	<b>0,95</b>
	$\sigma_\epsilon^2 = 1,0$	[0,57;1,40]	[0,68;1,31]	[0,66;1,26]	[0,55;1,39]	[0,54;1,38]	[0,54;1,40]
		(0,83)	(0,63)	(0,60)	(0,84)	(0,84)	(0,86)
		<b>0,96</b>	<b>0,92</b>	<b>0,88</b>	<b>0,96</b>	<b>0,95</b>	<b>0,95</b>

Obs.: Os números em colchetes são os limites médios dos intervalos, em parênteses a amplitude e em negrito a taxa de cobertura dos intervalos.

los e taxa de cobertura mais próxima de 0,95. Em alguns casos os intervalos BC e BCa estão mais próximos do ponto 0,95, com amplitude aceitável, como por exemplo em  $n = 200$  com  $\sigma_\omega^2$  e  $\sigma_\epsilon^2$ . Novamente, para o limite inferior do intervalo assintótico, todos os limites são negativos, exceto para  $\sigma_\epsilon^2$ .

## 4.2 Testes de hipóteses

Nesta seção é feita uma revisão geral para a construção de testes *bootstrap* para os hiperparâmetros dos modelos estruturais. Uma atenção especial será dada aos modelos de séries temporais não-estacionárias consistindo de componentes não-observáveis como nível e erro. Certas hipóteses sobre os parâmetros desses modelos permitem aos econométristas testar se a série temporal é estacionária ao invés de não-estacionária.

Em muitas áreas de econometria e estatística o teste da razão de verossimilhança (RV) é usado para testar restrições lineares. Em situações normais, o teste RV tem propriedades assintóticas atrativas, embora a distribuição assintótica do teste RV não seja a distribuição  $\chi^2$  usual quando os parâmetros



Tabela 4.6: Intervalos de confiança e de credibilidade de 95% para o MEB.

n	$\psi$	Assint.	Cred.Unif	Cred.Jeff	Perc	BC	BCa
60	$\sigma_\eta^2 = 0,50$	[-0,34;1,43]	[0,10;1,51]	[0,13;1,58]	[0,00;1,71]	[0,00;2,25]	[0,00; 2,19]
		(1,77)	(1,41)	(1,45)	(1,71)	(2,25)	(2,19)
		<b>0,96</b>	<b>0,97</b>	<b>0,92</b>	<b>0,96</b>	<b>0,99</b>	<b>0,99</b>
	$\sigma_\xi^2 = 0,03$	[-0,02;0,08]	[0,01;0,15]	[0,01;0,21]	[0,00;0,14]	[0,00;0,40]	[0,00;0,38]
		(0,10)	(0,14)	(0,20)	(0,14)	(0,40)	(0,38)
		<b>0,75</b>	<b>0,91</b>	<b>0,96</b>	<b>0,91</b>	<b>0,96</b>	<b>0,96</b>
	$\sigma_\omega^2 = 0,10$	[-0,05;0,27]	[0,04;0,38]	[0,03;0,36]	[0,01;0,28]	[0,01;0,43]	[0,00; 0,44]
		(0,32)	(0,34)	(0,33)	(0,27)	(0,43)	(0,44)
		<b>0,82</b>	<b>0,88</b>	<b>0,92</b>	<b>0,86</b>	<b>0,99</b>	<b>0,99</b>
	$\sigma_\epsilon^2 = 1,00$	[0,12;1,78]	[0,31;1,64]	[0,40;1,71]	[0,14; 2,11]	[0,11;2,11]	[0,11; 2,09]
		(1,66)	(1,33)	(1,31)	(1,97)	(2,00)	(1,98)
		<b>0,90</b>	<b>0,91</b>	<b>0,88</b>	<b>0,99</b>	<b>0,98</b>	<b>0,98</b>
200	$\sigma_\eta^2 = 0,50$	[-0,17;1,20]	[0,13;1,22]	[0,15;1,28]	[0,02;1,45]	[0,04;1,77]	[0,04; 1,75]
		(1,37)	(1,09)	(1,13)	(1,43)	(1,73)	(1,71)
		<b>0,93</b>	<b>0,95</b>	<b>0,89</b>	<b>0,96</b>	<b>0,99</b>	<b>0,99</b>
	$\sigma_\xi^2 = 0,03$	[-0,01;0,07]	[0,01;0,09]	[0,01;0,12]	[0,00;0,08]	[0,00;0,10]	[0,00;0,10]
		(0,08)	(0,08)	(0,11)	(0,08)	(0,09)	(0,10)
		<b>0,82</b>	<b>0,91</b>	<b>0,96</b>	<b>0,89</b>	<b>0,98</b>	<b>0,98</b>
	$\sigma_\omega^2 = 0,10$	[-0,01;0,21]	[0,05;0,26]	[0,03;0,25]	[0,01;0,23]	[0,02;0,26]	[0,02; 0,27]
		(0,22)	(0,21)	(0,22)	(0,22)	(0,24)	(0,25)
		<b>0,87</b>	<b>0,87</b>	<b>0,89</b>	<b>0,89</b>	<b>0,95</b>	<b>0,95</b>
	$\sigma_\epsilon^2 = 1,00$	[0,32;1,60]	[0,46;1,51]	[0,53;1,61]	[0,28; 1,82]	[0,24;1,81]	[0,23; 1,77]
		(1,28)	(1,05)	(1,08)	(1,54)	(1,57)	(1,54)
		<b>0,91</b>	<b>0,92</b>	<b>0,88</b>	<b>0,96</b>	<b>0,96</b>	<b>0,96</b>

Obs.: Os números em colchetes são os limites médios dos intervalos, em parênteses, a amplitude e, em negrito, a taxa de cobertura dos intervalos.

sob a hipótese nula estão no limiar do espaço paramétrico (ver Harvey (1989) para uma discussão mais geral). Assim, um dos objetivos desta seção é utilizar o *bootstrap* no teste RV para calcular a distribuição empírica desta estatística

Uma das desvantagens do método RV é a necessidade de estimar o modelo sob as hipóteses nula e alternativa. Em uma tentativa de superar esse problema, o *bootstrap* também será utilizado em uma versão modificada do teste escore (SM), já que este teste exige a estimação do modelo somente sob a hipótese nula.

Para finalizar a seção, um estudo Monte Carlo é apresentado para investigar o tamanho e poder dos testes *bootstrap* RV e escore, e uma aplicação é feita a uma série real.

### 4.2.1 Teste *bootstrap* usando a estatística da razão de verossimilhança

O teste RV para a hipótese nula

$$H_0 : \psi_i = \psi_{0i}, \quad i = 1, \dots, p$$

é dado por

$$RV = -2[\ln L(\hat{\psi}_0; \mathbf{Y}_n) - \ln L(\hat{\psi}; \mathbf{Y}_n)], \quad (4.2)$$

onde  $\hat{\psi}$  é a estimativa de máxima verossimilhança de  $\psi$  do modelo irrestrito e  $\hat{\psi}_0$  é a estimativa para o modelo sob a hipótese nula. A estatística RV tem, sob a hipótese nula, distribuição assintótica  $\chi^2$  com  $m$  graus de liberdade, denotada por  $\chi_m^2$ . Neste caso,  $m$  é o número de restrições. Deve-se notar que, quando os elementos de  $\psi_0$  estão na fronteira do espaço paramétrico de  $\psi$ , a distribuição assintótica é uma soma de distribuições  $\chi^2$  sob as hipóteses nula e alternativa.

É mostrado por Chernoff (1954) que se um elemento simples é restrito à fronteira do espaço paramétrico, a estatística RV é assintoticamente distribuída como  $(1/2)\chi_0^2 + (1/2)\chi_1^2$ , onde  $\chi_0^2$  é uma distribuição degenerada com toda sua massa na origem. Então, o EMV  $\hat{\psi}$  não terá uma distribuição conjunta assintótica Normal sob a hipótese nula, uma vez que algumas condições de regularidade não são válidas (ver (Harvey, 1989, Seção 5.1.2)). Para tal caso a estatística RV não terá distribuição assintótica  $\chi^2$ .

O *bootstrap* é apropriado e útil em casos como este, nos quais a distribuição assintótica da estatística de teste não pode ser obtida. Assim, o *bootstrap* será aplicado aqui para determinar a distribuição empírica da estatística de teste RV sob a hipótese nula.

Feng & McCulloch (1996) estabelecem condições para a convergência do EMV irrestrito para o verdadeiro valor do parâmetro, quando este pertence ao subconjunto  $\psi_0$ . Franco, Koopman & Souza (1999) verificam a validade destas suposições no caso dos modelos MNL, MTL e MEB, tal que os tamanhos do teste estejam de acordo com os níveis nominais fixados.

Para construir um teste *bootstrap*, deve-se estimar tanto o  $\psi$  irrestrito,  $\hat{\psi}^*$ , quanto  $\psi$  sob  $H_0$ ,  $\psi_0$ , usando as séries *bootstrap*  $y_t^*$ . A estatística assim obtida será denotada por  $RV^*$ . Neste trabalho, ao invés de calcular os valores críticos da distribuição empírica *bootstrap* das estatísticas de teste, segue-se a sugestão de Davidson & MacKinnon (1997) e calcula-se os níveis

de significância diretamente. Esse método também conduz a uma redução considerável no tempo computacional.

A probabilidade de significância, também conhecida como valor-P, para a estatística de teste RV é obtido da amostra *bootstrap* da estatística de teste como

$$P_{RV}^* = \frac{1}{B+1} \left[ 1 + \sum_{i=1}^B I(\hat{RV} \leq \hat{RV}_i^*) \right], \quad (4.3)$$

onde  $\hat{RV}$  é a estatística de teste calculada para a amostra original,  $\hat{RV}_i^*$  é a estatística de teste calculada usando a i-ésima amostra *bootstrap*,  $I(\cdot)$  é a função indicadora e  $B$  é o número de replicações *bootstrap*.

#### 4.2.2 Teste *bootstrap* usando a estatística escore modificado

Uma das desvantagens do teste RV é que o modelo deve ser estimado sob as hipóteses nula e alternativa. Esse fato pode fazer com que os cálculos do teste  $RV^*$  fiquem muito lentos, dependendo do número de amostras *bootstrap* necessárias para atingir o nível desejado de precisão.

Uma estatística de teste alternativa pode ser obtida baseada no vetor escore, a primeira derivada da função de verossimilhança com respeito ao vetor paramétrico  $\psi$ . O vetor escore pode ser avaliado em diferentes pontos, mas é claro que quando é avaliado na estimativa de máxima verossimilhança  $\hat{\psi}$ , o vetor escore é zero. Então, o teste escore modificado (SM) é definido como

$$SM = \frac{\partial \ln L(\psi; \mathbf{Y}_n)}{\partial \psi} \Big|_{\psi=\hat{\psi}_0}. \quad (4.4)$$

Da mesma forma, quando a hipótese nula é verdadeira deve-se ter  $SM(\hat{\psi}_0) = 0$ , mas esse valor tende a aumentar em módulo quando a hipótese nula não é verdadeira. A distribuição assintótica da estatística SM não é conhecida quando a hipótese nula é verdadeira pelas mesmas razões da estatística RV.

A estatística SM tem uma vantagem computacional considerável por necessitar apenas da estimação do parâmetro sob a hipótese nula. A avaliação do teste escore pode ser feita numericamente. Koopman & Shephard (1992) fornecem um procedimento para se calcular o escore exatamente usando

um algoritmo simples de suavização que pode ser aplicado após o filtro de Kalman.

A probabilidade de significância para a estatística de teste SM é obtido da amostra *bootstrap* da estatística de teste por

$$P_{SM}^* = \frac{1}{B+1} \left[ 1 + \sum_{i=1}^B I(\hat{SM} \leq \hat{SM}_i^*) \right], \quad (4.5)$$

onde  $\hat{SM}$  é a estatística de teste calculada para a amostra original,  $\hat{SM}_i^*$  é a estatística de teste calculada usando a  $i$ -ésima amostra *bootstrap*,  $I(\cdot)$  é a função indicadora e  $B$  é o número de replicações *bootstrap*.

### 4.2.3 Teste de Nyblom e Mäkeläinen

Um teste localmente invariante mais poderoso foi introduzido por Nyblom & Makelainen (1983), denotado aqui por NM, para testar a hipótese  $H_0 : \sigma_\eta^2 = 0$  para o modelo de nível local. Nota-se que o teste NM só é válido para o MNL e sua estatística de teste é dada por

$$NM = \frac{\sum_{i=1}^n [\sum_{s=i}^n (y_s - \bar{y})]^2}{(n-1) \sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4.6)$$

Os valores críticos para o teste são dados em Nyblom & Makelainen (1983).

### 4.2.4 Estudos de simulação

O desempenho dos testes descritos acima (RV, SM e NM) pode ser avaliado através de simulações Monte Carlo, onde o tamanho e o poder dos testes são calculados. O procedimento para obter o tamanho e o poder dos testes *bootstrap* é simples, mas pode demandar um tempo computacional elevado. O algoritmo completo para o cálculo dos testes *bootstrap* pode ser encontrado em Franco (1998).

O número de repetições Monte Carlo utilizado será de  $R = 1000$ . Quanto ao número de replicações *bootstrap*, de acordo com Hall (1986), se o nível de cobertura é um múltiplo de  $(1+B)^{-1}$ , o erro do nível de cobertura pode ser reduzido de  $O(n^{-1/2})$  para  $O(n^{-1})$ . Portanto, o número de replicações

*bootstrap* foi fixado em 399, a fim de que se obtenha um valor crítico de nível de cobertura de 0,95. Serão empregados também *bootstraps* paramétrico ( $p$ ) e não-paramétrico ( $np$ ) e as estatísticas de teste resultantes serão denotadas por  $RVn$ ,  $RVnp$ ,  $SMn$  e  $SMnp$ .

Assim, para calcular o tamanho dos testes RV, SM e NM, o seguinte procedimento é repetido  $R = 1000$  vezes.

**Algoritmo 4.2:** Cálculo do tamanho dos testes RV, SM e NM

---

**faça**  $r=1$

**repita** os seguintes passos

**gere** a série por amostragem aleatória do modelo sob a hipótese nula

**calcule** as estatísticas  $RV_{obs}$ ,  $SM_{obs}$  e  $NM_{obs}$

(Apenas para os testes RV e SM):

**simule** séries bootstrap  $y_1^{*b}, y_1^{*b}, \dots, y_n^{*b}$ ,  $b = 1, \dots, B$  sob  $H_0$ , usando os parâmetros estimados na série original

**calcule** as estatísticas de teste  $RV_b^*$  e  $SM_b^*$  para cada série

**obtenha** a aproximação para os valores-P,  $P_{RV}^*$  e  $P_{SM}^*$ , utilizando as equações (4.3) e (4.5), respectivamente

**até**  $r = R$

**calcule** os tamanhos dos testes *bootstrap* como

$$SZ_h = (1/R) \sum_{r=1}^R I(P_h^*(r) < \alpha),$$

onde  $h = RV$  ou  $SM$ ,  $\alpha$  é o nível nominal fixado e  $I(\cdot)$  é a função indicadora.

**calcule** o tamanho do teste NM como

$$SZ_{NM} = (1/R) \sum_{r=1}^R I(NM_{obs}^{(r)} > nm_\alpha),$$

onde  $nm_\alpha$  é o ponto crítico da estatística NM associada com o nível nominal  $\alpha$  dado em Nyblom e Mäkeläinen (1978).

---

Para calcular o poder, o mesmo algoritmo descrito para o cálculo do tamanho dos testes é utilizado, com a modificação de que a série original é gerada do modelo sob a hipótese alternativa.

#### a) Resultados para o Modelo de Nível Local

No modelo de nível local (MNL) as hipóteses de interesse são

$$H_0 : \sigma_\eta^2 = 0 \quad vs \quad H_1 : \sigma_\eta^2 > 0.$$

As análises são feitas considerando duas configurações para a distribuição do erro  $\epsilon_t$ . As escolhas são (a) Normal com média zero e variância  $\sigma_\epsilon^2$ ; (b) t-Student com 5 graus de liberdade. São geradas séries de tamanhos  $n = 30$ ,  $n = 50$  e  $n = 100$ . Testes com níveis nominais de 0,05 e 0,10 foram conduzidos para cada um dos casos descritos acima.

Os resultados são apresentados na Tabela 4.7, onde o subscrito  $p$  nos testes *bootstrap* significa *bootstrap* paramétrico e o subscrito  $np$  significa *bootstrap* não-paramétrico.

Tabela 4.7: Tamanho empírico dos testes *bootstrap* e *NM* para o MNL

$\alpha$	n	Normal					t(5)				
		<i>RVp</i>	<i>RVnp</i>	<i>SMp</i>	<i>SMnp</i>	<i>NM</i>	<i>RVp</i>	<i>RVnp</i>	<i>SMp</i>	<i>SMnp</i>	<i>NM</i>
0,05	30	0,055	0,048	0,064	0,066	0,067	0,045	0,034	0,044	0,042	0,042
	50	0,049	0,042	0,035	0,054	0,050	0,047	0,042	0,055	0,053	0,052
	100	0,056	0,051	0,044	0,052	0,056	0,050	0,040	0,051	0,048	0,046
0,10	30	0,114	0,105	0,124	0,117	0,126	0,089	0,075	0,090	0,093	0,086
	50	0,094	0,097	0,101	0,107	0,104	0,099	0,091	0,115	0,105	0,104
	100	0,109	0,104	0,097	0,103	0,106	0,090	0,093	0,099	0,092	0,103

Como pode-se observar, mesmo quando o tamanho da amostra é pequeno ( $n = 30$ ), as probabilidades de corretamente aceitar  $H_0$  são próximas do tamanho nominal  $\alpha$ , para todos os procedimentos. Além disso, o comportamento dos testes *bootstrap* são equivalentes aos do teste *NM*, com *RVp* ligeiramente superior, principalmente quando o tamanho da amostra é pequeno. É importante ressaltar que o teste *SM*, sendo mais rápido que o teste *RV*, apresenta níveis empíricos tão bons quanto este último. Enfatiza-se que os resultados obtidos com os testes *bootstrap* não-paramétrico, sem fazer qualquer suposição sobre a distribuição dos erros, atingem níveis empíricos muito

próximos aos níveis nominais assumidos. A diferença entre escolher uma distribuição Normal ou t-Student para o componente de erro  $\epsilon_t$  não parece afetar os resultados em modelos de nível local.

Tabela 4.8: Poder empírico dos testes *bootstrap* e *NM* para o MNL ( $\alpha = 0,05$ ).

n	$\sigma_\eta^2$	Normal					t(5)				
		RVp	RVnp	SMp	SMnp	NM	RVp	RVnp	SMp	SMnp	NM
30	0,001	0,151	0,130	0,147	0,143	0,141	0,122	0,108	0,121	0,122	0,123
	0,01	0,588	0,555	0,588	0,489	0,522	0,462	0,427	0,460	0,426	0,428
	0,1	0,919	0,908	0,945	0,755	0,796	0,899	0,886	0,899	0,731	0,775
50	0,001	0,286	0,278	0,296	0,286	0,290	0,205	0,193	0,191	0,193	0,189
	0,01	0,803	0,794	0,773	0,708	0,726	0,721	0,708	0,704	0,653	0,665
	0,1	0,995	0,993	0,995	0,921	0,936	0,976	0,970	0,967	0,876	0,897
100	0,001	0,621	0,619	0,591	0,570	0,568	0,516	0,512	0,489	0,479	0,483
	0,01	0,982	0,981	0,961	0,925	0,926	0,944	0,944	0,926	0,889	0,892
	0,1	1,000	1,000	1,000	0,983	0,987	1,000	1,000	1,000	0,982	0,985

Para o cálculo do poder, avaliam-se três valores de  $\sigma_\eta^2$ : 0,001, 0,01 e 0,1. Os resultados estão resumidos na Tabela 4.8. No estudo do poder, nota-se claramente a superioridade dos testes *bootstrap* em relação ao teste *NM*, a única exceção sendo o *SMnp*. Para séries de tamanho  $n = 100$ , mesmo para valores de  $\sigma_\eta^2$  tão pequenos quanto 0,001, rejeita-se a hipótese nula em aproximadamente 60% das vezes. Para séries de tamanho pequeno, é necessário que  $\sigma_\eta^2$  seja, no mínimo, igual a 0,01 para se obter o mesmo poder. Novamente, o *RVp* tem um desempenho melhor que os outros métodos, mas o *RVnp* e *SMp* estão muito próximos do primeiro. Para o poder pode-se observar algumas diferenças entre os resultados das distribuições Normal e t-Student, com a primeira mostrando um desempenho melhor.

#### b) Resultados para o Modelo de Tendência Linear local

No modelo de tendência linear local (MTL) as hipóteses a serem testadas são

$$H_0 : \sigma_\xi^2 = 0 \quad vs \quad H_1 : \sigma_\xi^2 > 0.$$

Neste caso, a maximização da função de log-verossimilhança é feita com respeito a três parâmetros, então um algoritmo numérico consome relativa-

mente mais tempo. Devido a essa limitação, o número de replicações *bootstrap* foi reduzido para  $B = 99$  (Franco, Koopman & Souza, 1999). A análise é feita apenas considerando a distribuição do erro ( $\epsilon_t$ ) Normal com média zero e variância 0,10. Além disso, o número de repetições do experimento foi mantido em  $R = 100$ . No entanto, relativamente, esses números pequenos de  $B$  e  $R$  foram suficientes para se ter uma idéia do comportamento dos testes *bootstrap* em modelos de tendência linear local.

Da Tabela 4.9 verifica-se o bom desempenho dos testes *bootstrap*, onde o poder é baseado apenas em testes de tamanho 0,05. O teste *RV* apresenta valores para o poder e tamanho mais próximos dos níveis nominais, exceto para as séries de tamanho  $n = 30$ , onde o tamanho está relativamente distante dos níveis assumidos. O *SMp* apresenta resultados aceitáveis para o poder, apesar de ter bons tamanhos somente para séries com  $n = 100$ . O único teste cujos resultados não são aceitáveis é o *SMnp*, com tamanho e poder muito distantes dos valores estabelecidos.

Tabela 4.9: Tamanho e poder empírico dos testes *bootstrap* para o MTL.

		n	<i>RVp</i>	<i>RVnp</i>	<i>SMp</i>	<i>SMnp</i>
tamanho	$\alpha = 0,05$	30	0,22	0,18	0,30	0,23
		50	0,05	0,08	0,18	0,13
		100	0,03	0,04	0,07	0,07
	$\alpha = 0,10$	30	0,25	0,20	0,36	0,31
		50	0,12	0,13	0,27	0,21
		100	0,08	0,08	0,09	0,11
<i>Poder</i> ( $\alpha = 0,05$ )	$\sigma_\xi^2 = 0,10$	30	0,97	0,94	0,98	0,38
		50	1,00	0,98	1,00	0,42
		100	1,00	1,00	0,99	0,26

#### 4.2.5 Aplicação a séries reais

O teste *bootstrap* foi aplicado a dados do setor de energia elétrica na região Nordeste do Brasil. Esses dados foram obtidos de um grande estudo com respeito à quantidade de energia necessária para responder à demanda máxima no horário de pico de 18h a 21h. Deve-se ressaltar que, neste capítulo, o objetivo não concerne particularmente ao estudo detalhado destas séries com o propósito de estimação de modelos e previsões. Assim, buscou-se uma série



que apresentasse apenas o componente de tendência, para que fosse possível verificar o desempenho dos testes *bootstrap*.

A série escolhida foi a de valores mensais do consumo de eletricidade da Companhia Hidrelétrica São Francisco (CHESF), no período de janeiro de 1983 a fevereiro de 1997. Os dados são apresentados na Figura 4.4 e a série parece apresentar um aumento não significativo no seu nível, permitindo que a mesma seja considerada neste trabalho para testar um modelo de nível local, isto é,  $\sigma_\xi^2 = 0$ .

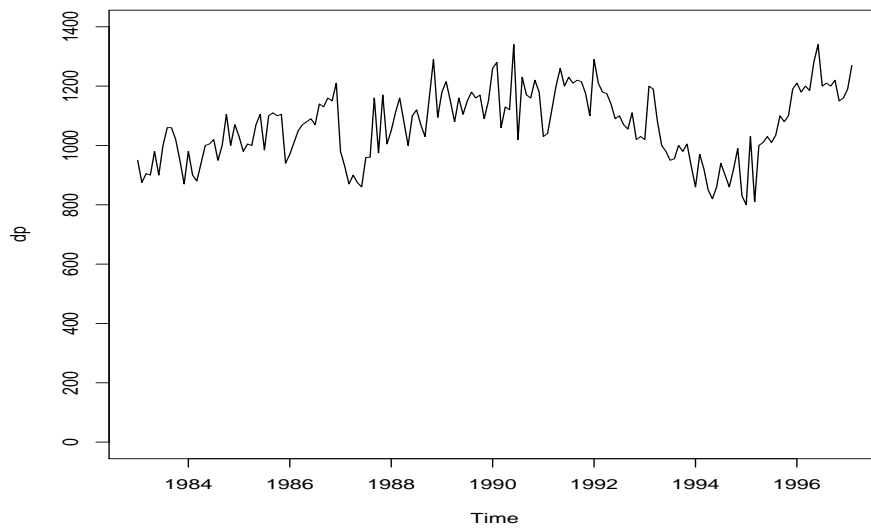


Figura 4.4: Série da demanda de eletricidade na ponta da CHESF

As estimativas de máxima verossimilhança para o MTL ajustado à série são dadas por  $\hat{\sigma}_\epsilon^2 = 3269,6$ ,  $\hat{\sigma}_\eta^2 = 1101,2$  e  $\hat{\sigma}_\xi^2 = 0,0$ . Uma análise dos resíduos foi feita e não houve evidências de correlação nos mesmos.

Empregando o *bootstrap* paramétrico e não-paramétrico para calcular o valor-P das estatísticas de *RV* e *SM* para testar as hipóteses  $H_0 : \sigma_\xi^2 = 0$  contra  $H_0 : \sigma_\xi^2 > 0$ , conclui-se que a hipótese nula não é rejeitada. Os valores-P dos testes paramétricos (1,00 para o *RVp* e 0,98 para o *SMp*) foram maiores que os dos testes não-paramétricos (0,75 para o *RVnp* e 0,52 para o *SMnp*). Então, pode-se concluir que o modelo de tendência linear local não é indicado para essa série. Neste caso, pode-se tentar ajustar um modelo de nível local

para esses dados. Os hiperparâmetros estimados foram  $\hat{\sigma}_\epsilon^2 = 3281,9$  e  $\hat{\sigma}_\eta^2 = 1087,3$  e a análise dos resíduos é muito similar à que foi feita no modelo anterior. Para verificar se o hiperparâmetro  $\sigma_\eta^2$  era significativamente maior que zero, os testes *bootstrap* e *NM* foram utilizados. Todos os testes indicam que  $\sigma_\eta^2 > 0$ , com valores-P iguais a 0,0025 para os testes *bootstrap* e menor que 0,01 para o teste *NM*. Assim, o modelo de nível local parece adequado para modelar essa série.

## Capítulo 5

# Modelos de Espaço de Estados Não-Lineares

Todos os MEE apresentados nesse texto até agora gozaram de algumas propriedades simplificadoras, notadamente linearidade, aditividade dos efeitos e normalidade dos erros. Pudemos ver que essas restrições não impediram que os MEE proporcionassem modelos razoáveis para modelagem de uma quantidade bastante razoável de fenômenos comumente encontrados em séries temporais reais.

Entretanto, como qualquer restrição, elas impõem uma redução no escopo de aplicabilidade dos MEE. A partir deste capítulo trataremos de apresentar classes mais gerais de MEE que relaxam uma ou mais das restrições impostas até aqui. Neste capítulo em particular, iremos nos concentrar nas duas primeiras restrições, a saber, linearidade e aditividade dos efeitos. MEE não-Gaussianos serão tratados no capítulo seguinte.

Esse capítulo será iniciado com uma consideração genérica sobre MEE em sua forma mais geral para depois tratar de MEE não-lineares. Dentro dessa classe de modelos, será dada especial atenção aos modelos condicionalmente lineares. Essa classe de modelos será estudada em detalhes neste capítulo no contexto de um de seus principais exemplos: modelos de função de transferência.

## 5.1 Uma estrutura geral

Para entender melhor os conceitos apresentados na introdução desse capítulo, precisamos recorrer à formulação básica do modelo dada pelas equações (2.3)-(2.4). Nessas equações pode-se perceber claramente que:

- as componentes latentes  $\alpha_t$  apenas interferem no modelo de forma linear em ambas as equações;
- os diferentes efeitos presentes no modelo, a saber, efeito latente, efeito de covariáveis e efeitos de erros não-estruturados são sempre combinados aditivamente;
- os erros de observação e de evolução do sistema são Gaussianos.

Essas três condições correspondem respectivamente ao que foi denominado acima como linearidade, aditividade dos efeitos e normalidade dos erros. Pode até parecer que essas restrições são as características definidoras dos MEE. Esse raciocínio está equivocado. Como mostraremos a seguir, MEE que satisfazem a apenas algumas ou mesmo nenhuma dessas condições podem ser definidos.

No contexto mais geral, as equações (2.3)-(2.4) são substituídas por

$$y_t = \mathbf{z}_t(\alpha_t, d_t, \epsilon_t) \quad (5.1)$$

$$\alpha_t = \mathbf{T}_t(\alpha_{t-1}, \mathbf{c}_t, \boldsymbol{\eta}_t), \quad (5.2)$$

onde  $\mathbf{z}_t$  e  $\mathbf{T}_t$  são funções quaisquer (não necessariamente lineares nem aditivas) de seus argumentos e  $\epsilon_t$  e  $\boldsymbol{\eta}_t$  são perturbações com distribuições arbitrárias (não necessariamente Gaussianas).

Essa classe assim especificada é muito geral para qualquer abordagem exata. Sua inferência sob qualquer ponto de vista só pode ser realizada usando aproximações. Várias simplificações podem ser operadas a partir de imposição de diferentes combinações das restrições acima mencionadas.

Uma formulação menos geral de (5.1)-(5.2) que satisfaça aditividade e Gaussianidade mas ainda bem mais geral que (2.3)-(2.4) é chamada de MEE não-lineares e é dada por

$$y_t = \mathbf{z}_t(\alpha_t) + d_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \quad (5.3)$$

$$\alpha_t = \mathbf{T}_t(\alpha_{t-1}) + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t) \quad (5.4)$$

onde  $\mathbf{z}_t$  e  $\mathbf{T}_t$  são funções quaisquer (não necessariamente lineares) em seus argumentos. Essa classe de modelos é bastante conhecida na literatura, principalmente pelo fato de ter sido desenvolvido para ela um instrumento de aproximação conhecido como filtro de Kalman estendido (FKE) (West & Harrison, 1997).

A base do FKE é bastante simples e intuitiva. O único impedimento para aplicação do filtro de Kalman se deve à não-linearidade das funções  $\mathbf{z}_t$  e  $\mathbf{T}_t$ . Se essas funções pudessem ser (ainda que aproximadamente) linearizadas, esse impedimento deixaria de existir. Esse procedimento de linearização aproximada de funções diferenciáveis é conhecido como expansão de Taylor. Baseado nessa expansão até o termo linear, pode-se propor as aproximações

$$\mathbf{z}_t(\boldsymbol{\alpha}_t) \doteq \mathbf{z}_t(\hat{\boldsymbol{\alpha}}_t) + \left[ \frac{d\mathbf{z}_t}{d\boldsymbol{\alpha}_t}(\hat{\boldsymbol{\alpha}}_t) \right]' (\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t) \quad (5.5)$$

$$\mathbf{T}_t(\boldsymbol{\alpha}_{t-1}) \doteq \mathbf{T}_t(\hat{\boldsymbol{\alpha}}_{t-1}) + \left[ \frac{d\mathbf{T}_t}{d\boldsymbol{\alpha}_{t-1}}(\hat{\boldsymbol{\alpha}}_{t-1}) \right]' (\boldsymbol{\alpha}_{t-1} - \hat{\boldsymbol{\alpha}}_{t-1}), \quad (5.6)$$

onde  $d\mathbf{z}_t/d\boldsymbol{\alpha}_t$  é a matriz de derivadas parciais de  $\mathbf{z}_t$ ,  $d\mathbf{T}_t/d\boldsymbol{\alpha}_{t-1}$  é a matriz de derivadas parciais de  $\mathbf{T}_t$  e  $\hat{\boldsymbol{\alpha}}_t$  e  $\hat{\boldsymbol{\alpha}}_{t-1}$  são valores em torno dos quais é feita a expansão. A escolha desses valores está intimamente ligada ao contexto onde é usada a aproximação.

A expansão de  $\mathbf{T}_t(\boldsymbol{\alpha}_{t-1})$  é usada para obter aproximações para  $\mathbf{a}_{t|t-1}$  e  $\mathbf{P}_{t|t-1}$  e essa operação é feita baseada em  $\mathbf{Y}_{t-1}$ . Logo, o ponto mais apropriado para centrar a aproximação de qualquer função de  $\boldsymbol{\alpha}_{t-1}$  é sua melhor estimativa  $\mathbf{a}_t$ . Analogamente, o ponto mais apropriado para centrar a aproximação de qualquer função de  $\boldsymbol{\alpha}_t$  é sua melhor estimativa  $\mathbf{a}_{t|t-1}$ .

De qualquer forma, uma vez feitas as escolhas dos pontos de aproximação, os mesmos são utilizados em (5.5)-(5.6) e essas expressões podem ser usadas em (5.3)-(5.4). Com isso, o modelo recupera a forma linear de (2.3)-(2.4) e o filtro de Kalman pode ser operado. Na prática, as únicas modificações nas equações (2.3)-(2.4) do FK estão nas expressões de  $\mathbf{a}_{t|t-1}$ ,  $\mathbf{P}_{t|t-1}$ ,  $\tilde{y}_{t|t-1}$  e  $F_t$ .

Através do FKE, suas expressões serão modificadas para

$$\begin{aligned} \mathbf{a}_{t|t-1} &= \mathbf{T}_t(\mathbf{a}_{t-1}) + \mathbf{c}_t \\ \mathbf{P}_{t|t-1} &= \left[ \frac{d\mathbf{T}_t}{d\boldsymbol{\alpha}_{t-1}}(\mathbf{a}_{t-1}) \right] \mathbf{P}_{t-1} \left[ \frac{d\mathbf{T}_t}{d\boldsymbol{\alpha}_{t-1}}(\mathbf{a}_{t-1}) \right]' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \\ \tilde{\mathbf{y}}_{t|t-1} &= \mathbf{z}_t(\mathbf{a}_{t|t-1}) + d_t \\ F_t &= \left[ \frac{d\mathbf{z}_t}{d\boldsymbol{\alpha}_t}(\mathbf{a}_{t|t-1}) \right]' \mathbf{P}_{t|t-1} \left[ \frac{d\mathbf{z}_t}{d\boldsymbol{\alpha}_t}(\mathbf{a}_{t|t-1}) \right] + h_t. \end{aligned}$$

Como era de se esperar, essas aproximações funcionam bem em modelos onde a não-linearidade não é muito acentuada. Para casos mais gerais, o FKE não é recomendado. O motivo de sua apresentação aqui nesse texto é que ele fornece um ponto de partida para pensar outras formas de aproximação em MEE.

## 5.2 Modelos condicionalmente lineares

Ainda dentro da classe de modelos não-lineares definidos em (5.3)-(5.4), pode-se definir uma classe de modelos cuja não-linearidade advém única e exclusivamente pela dependência que componentes do modelo linear (2.3)-(2.4) possam ter em quantidades desconhecidas, denotadas aqui por  $\psi$ . Mais explicitamente, o modelo é dado por

$$y_t = [\mathbf{z}_t(\psi)]' \boldsymbol{\alpha}_t + d_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \quad (5.7)$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t(\psi) \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t). \quad (5.8)$$

Esses modelos são chamados de modelos condicionalmente lineares (MCL) pois condicionalmente a  $\psi$  os modelos são lineares.

Diferente do que foi apresentado nos capítulos anteriores, as matrizes  $\mathbf{z}_t$  e  $\mathbf{T}_t$  não são conhecidas. Elas agora dependem de uma quantidade desconhecida  $\psi$ . Se o valor de  $\psi$  puder ser conhecido, o modelo recai em sua forma anterior e todo o procedimento de inferência via FK e obtenção da verossimilhança marginal segue como antes. Assim, basta incluir os elementos desconhecidos de  $\mathbf{z}_t$  e  $\mathbf{T}_t$  na coleção  $\psi$  de quantidades desconhecidas que o procedimento de inferência segue como anteriormente.

Um exemplo simples de MCL inclui a extensão do MNL dada por

$$y_t = \rho_1 \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad (5.9)$$

$$\mu_t = \rho_2 \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \quad (5.10)$$

O parâmetro  $\rho_1$  desempenha o papel de atenuador do efeito do nível sobre as observações e o parâmetro  $\rho_2$  introduz a possibilidade de estacionariedade no processo latente  $\mu$ . O modelo (5.9)-(5.10) pode ser colocado na formulação geral de MCL fazendo  $\rho_1$  e  $\rho_2$  incluídos em  $\psi$ ,  $\mathbf{z}_t(\psi) = \mathbf{z}_t(\rho_1) = \rho_1$  e  $\mathbf{T}_t(\psi) = \mathbf{T}_t(\rho_2) = \rho_2$ . Note que se  $\rho_1$  e  $\rho_2$  são conhecidos,  $\mathbf{z}_t$  e  $\mathbf{T}_t$  também serão e o modelo recai na estrutura básica (2.3)-(2.4).

Um exemplo mais elaborado de MCL será objeto de um extensivo estudo e será apresentado na próxima seção.

### 5.3 Função de transferência

Séries temporais são frequentemente afetadas por eventos externos, conhecidos como intervenções, que podem mudar a estrutura da série (tais como quebras estruturais, mudanças políticas, etc.). Há evidências na literatura de que as primeiras propostas de análise de intervenção surgiram nas Ciências Sociais com o trabalho de Campbell & Stanley (1968). Embora o termo intervenção tenha sido introduzido pela primeira vez por Glass (1972), foi apenas em 1975 que Box & Tiao desenvolveram a teoria de análise de intervenção para estudar mudanças estruturais em séries temporais (Box & Tiao, 1975).

Ao mesmo tempo, Box & Jenkins (1976) introduziram os modelos de função de transferência (FT) no contexto dos processos Autoregressivos Integrados de Médias Móveis (ARIMA). Os modelos de função de transferência foram criados para mensurar a relação entre uma série resposta e uma ou mais séries explicativas. Por exemplo, no caso de uma série resposta  $y_t$  e uma série  $x_t$  explicativa, a função de transferência relaciona as variáveis através de um filtro linear da forma

$$y_t = \vartheta(B)x_t + \epsilon_t$$

onde  $\vartheta(B) = \sum_{j=-\infty}^{\infty} \vartheta_j B^j$  e  $B$  é o operador de retardo,  $B^k y_t = y_{t-k}$ .

Se a série explicativa  $x_t$  é uma função determinística que descreve as mudanças estruturais em  $y_t$ , a FT pode ser usada para modelar quebras estruturais seguindo a mesma idéia de Box & Tiao (1975).

Como os modelos ARIMA podem ser escritos na forma de espaço de estados, a introdução da idéia de funções de transferência nos modelos estruturais é quase natural (Harvey, 1989; West & Harrison, 1997). A flexibilidade da representação de espaço de estados permite a inserção de covariáveis nas equações das observações e de transição. Alguns trabalhos sobre esse assunto incluem Penzer (2007), de Jong & Penzer (1998), Harvey & Koopman

(1992), Salvador & Gargallo (2004), Ravines, Schmidt & Migon (2008) e Alves, Gamerman & Ferreira (2009).

Neste trabalho, dois tipos de procedimentos de intervenção são descritos: as funções pulso e degrau. Além disso, o parâmetro do modelo associado com a intervenção assume duas formas: na primeira, ele é fixo no tempo e, na segunda, ele varia no tempo. Inferências sobre os parâmetros da FT e variâncias dos erros nas equações da observação e transição são feitas usando as abordagens clássica e Bayesiana. Embora grande parte desses modelos já tenha sido estudada na literatura, não são de conhecimento dos autores trabalhos utilizando inferência clássica para modelar os parâmetros associados à uma intervenção variando no tempo. Intervalos de confiança sob o paradigma clássico são construídos usando o *bootstrap* sob a abordagem paramétrica. Na abordagem Bayesiana, os métodos de *Markov chain Monte Carlo* (MCMC) são utilizados para obter as estimativas pontuais e os intervalos de credibilidade.

### 5.3.1 Análise de intervenção em modelos estruturais

Se uma mudança de nível é observada na série, seu efeito pode ser modelado inserindo um componente  $E_t$  na equação do estado. Esse componente  $E_t$  é escrito como o filtro da função de transferência de Box, Jenkins & Reinsel (1994),

$$E_t = \frac{\beta(B)B^b}{\rho(B)}x_t = \vartheta(B)x_t = \sum_{j=0}^{\infty} \vartheta_j B^j x_t \quad (5.11)$$

onde  $x_t$  é uma variável exógena,  $\beta(B) = \beta_0 - \beta_1 B - \dots - \beta_s B^s$ ,  $\rho(B) = \rho_0 - \rho_1 B - \dots - \rho_r B^r$  e  $b$  é o parâmetro de atraso. Os coeficientes  $\vartheta_j$  no modelo de função de transferência, denotado aqui por  $FT(r, b, s)$ , são denominados função-resposta impulso. Mais detalhes podem ser vistos em Wei (1990).

Neste capítulo, a variável exógena  $x_t$  em (5.11) representará a intervenção. Na prática, há muitas possibilidades para a ocorrência de intervenção. Por exemplo, o impacto de um evento externo pode ser sentido  $b$  períodos após a intervenção, com um efeito apenas no momento da ocorrência. Neste caso, o modelo de função de transferência é um  $FT(0,0,b)$ :

$$E_t = \beta_0 B^b x_t = \beta_0 x_{t-b}.$$

Também pode acontecer de o impacto ser sentido no momento da intervenção,



mas a resposta ser gradual. Então, um modelo apropriado é um FT(1,0,0),

$$E_t = \frac{\beta_0 B^0}{\rho_0 - \rho_1 B} x_t \Rightarrow (\rho_0 - \rho_1 B) E_t = \beta_0 x_t \Rightarrow E_t = \rho E_{t-1} + \beta x_t, \quad (5.12)$$

onde  $\rho = \rho_1/\rho_0$  e  $\beta = \beta_0/\rho_0$ .

Box & Tiao (1975) definem dois tipos comuns de variáveis de intervenção, função pulso e degrau, que podem ser representadas pelas seguintes variáveis indicadoras:

(1) Função degrau: Se a intervenção ocorre em algum tempo  $T$  fixado e o efeito persiste após a intervenção:

$$S_t^T = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} ;$$

(2) Função pulso: Se a intervenção ocorre em algum tempo  $T$  fixado e tem efeito apenas neste instante:

$$I_t^T = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} .$$

A Figura 5.1 apresenta o comportamento do bloco estrutural  $E_t$  para séries de tamanho  $n = 100$  com uma mudança de nível no tempo  $T = 50$ , sob o Modelo FT(1,0,0), para funções pulso e degrau. Para a função pulso nota-se que, quando aumenta-se o valor de  $\rho$ , o retorno da série para o seu nível médio anterior à intervenção torna-se mais lento. Para a função degrau, se  $\rho$  aumenta, a série leva mais tempo para atingir um novo nível médio.

Neste trabalho, FT(1,0,0) será considerada com dois tipos de função-resposta impulso. Na primeira (Modelo 1) o fator de ganho  $\beta$  na equação (5.12) é fixo no tempo, enquanto que o Modelo 2 permite  $\beta$  variar no tempo, ou seja, o fator de ganho  $\beta_t$  depende de  $t$ . Esse último modelo é conhecido como um modelo de função de transferência com fator de ganho dinâmico (ver Alves, Gamerman & Ferreira (2009)).

O modelo estrutural mais simples - o modelo de nível local (MNL), descrito no Capítulo 2 - é considerado com um componente de intervenção. As estruturas dos Modelos 1 e 2 sob a abordagem de nível local são descritas abaixo, com exemplos de séries geradas desses modelos.

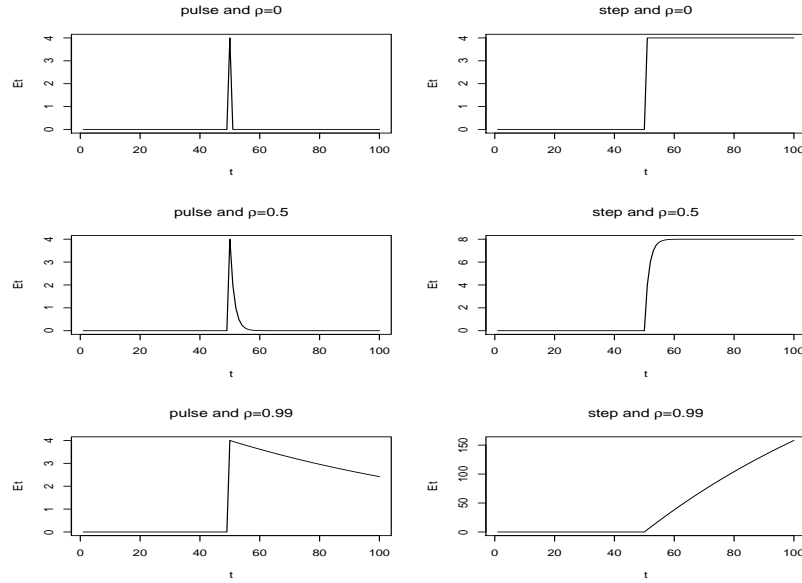


Figura 5.1: Efeito da intervenção das funções pulso e degrau.

### Modelo 1

O MNL com um componente de intervenção, assumindo o coeficiente  $\beta$  fixo no tempo, tem a seguinte estrutura:

$$\begin{cases} y_t = \mu_t + E_t + \epsilon_t, & \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \mu_t = \mu_{t-1} + \eta_t, & \eta_t \sim N(0, \sigma_\eta^2) \\ E_t = \rho E_{t-1} + \beta x_t, \end{cases} \quad (5.13)$$

onde  $t = 1, 2, \dots, n$ ,  $\epsilon_t$  e  $\eta_t$  são independentes e  $0 \leq \rho \leq 1$ .  $\rho$  é definido entre 0 e 1 para que o comportamento da função-resposta impulso tenha um decaimento geométrico. Por essa razão,  $\rho$  é normalmente chamado de persistência ou memória do efeito. A resposta média da série é dada por  $\alpha_t = \mu_t + E_t$ .

O modelo em (5.13) pode ser escrito na forma de espaço de estados com matrizes dadas por

$$\mathbf{Z}'_t = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \mathbf{R}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{T}_t = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}, \quad \boldsymbol{\eta}_t = \begin{bmatrix} \eta_t \\ 0 \end{bmatrix}, \quad h_t = [\sigma_\epsilon^2],$$

$$\mathbf{Q}_t = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad d_t = 0, \quad \mathbf{c}_t = \begin{bmatrix} 0 \\ \beta x_t \end{bmatrix} \text{ e } \boldsymbol{\alpha}_t = \begin{bmatrix} \mu_t \\ E_t \end{bmatrix}.$$

Note que esse modelo é linear condicionalmente ao conhecimento de  $\rho$  e  $\beta$ .

A Figura 5.2 mostra algumas séries simuladas sob o Modelo 1. Como esperado, o comportamento da série é muito similar ao comportamento do bloco  $E_t$  mostrado na Figura 5.1. Para a função pulso, pode-se notar que pequenos valores de  $\rho$  causam na série um efeito similar à presença de um *outlier* no ponto da intervenção. Para grandes valores de  $\rho$ , a série apresenta um salto no ponto da intervenção, mas com um retorno gradual ao valor médio. Para a função degrau, quando  $\rho = 0,00$  ou  $0,50$ , há um salto repentino no tempo da intervenção, com uma mudança de nível, e o restante da série permanece no novo nível. Quando  $\rho = 0,99$ , há uma gradual mudança de nível, e a série leva mais tempo para atingir um novo nível.

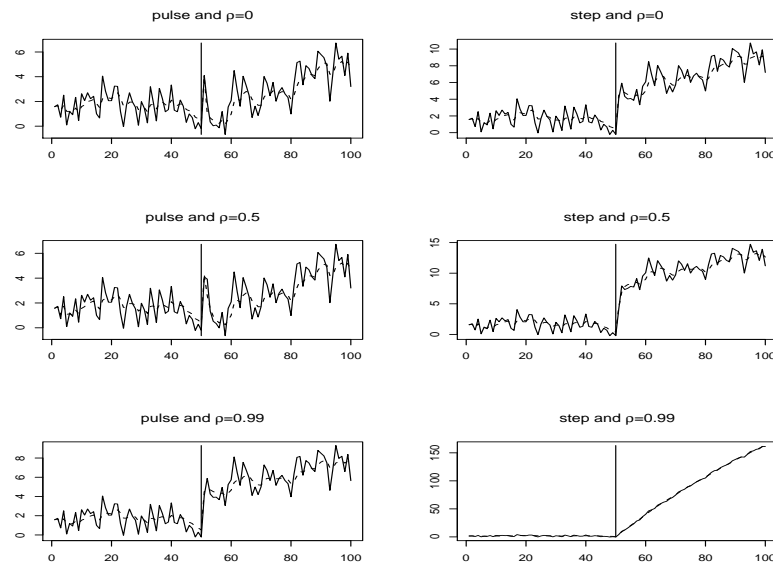


Figura 5.2: Exemplos das funções pulso e degrau para o Modelo 1, com  $n = 100$ ,  $\beta = 4$  e  $\rho = 0,00, 0,50$  e  $0,99$ . A linha vertical indica o instante da intervenção,  $T = 50$ , a linha contínua indica a série temporal,  $y_t$ , e a linha tracejada indica a resposta média  $\alpha_t$ .

**Modelo 2**

O modelo de nível local com um fator de ganho dinâmico (Alves, Gamerman & Ferreira, 2009) é dado por

$$\begin{cases} y_t = \mu_t + E_t + \epsilon_t, & \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \mu_t = \mu_{t-1} + \eta_t, & \eta_t \sim N(0, \sigma_\eta^2) \\ E_t = \rho E_{t-1} + \beta_{t-1} x_t, \\ \beta_t = \beta_{t-1} + \xi_t, & \xi_t \sim N(0, \sigma_\xi^2) \end{cases}$$

onde  $t = 1, 2, \dots, n$ ,  $\epsilon_t$ ,  $\eta_t$  e  $\xi_t$  são conjuntamente independentes.

Ao contrário do Modelo 1, no Modelo 2 o parâmetro  $\beta_t$  é estocástico e é obtido dinamicamente através do tempo por alguma lei estocástica (aqui um passeio aleatório).

O Modelo 2 pode ser escrito na forma de espaço de estados com matrizes dadas por

$$\mathbf{Z}'_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \mathbf{R}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{T}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & x_t \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\eta}_t = \begin{bmatrix} \eta_t \\ 0 \\ \xi_t \end{bmatrix},$$

$$h_t = [\sigma_\epsilon^2], \mathbf{Q}_t = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix}, d_t = 0, \mathbf{c}_t = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ e } \boldsymbol{\alpha}_t = \begin{bmatrix} \mu_t \\ E_t \\ \beta_t \end{bmatrix}.$$

Novamente temos aqui um modelo condicionalmente linear pois se  $\rho$  for conhecido, trata-se de um MEE conforme descrito em (2.3)-(2.4).

A Figura 5.3 apresenta algumas séries simuladas sob o Modelo 2. O comportamento é muito similar ao do Modelo 1, mas neste caso, o  $\beta$  dinâmico causa mais ruído na série. Se uma função pulso é usada, o Modelo 2 não é adequado pois a série não fornece informação para estimar a variação temporal do ganho  $\beta_t$ . Por essa razão, apenas a função degrau será usada no Modelo 2.

**5.3.2 Inferência**

Esta subseção apresenta os métodos de estimação clássico e Bayesiano dos parâmetros dos modelos de FT, assim como previsão de valores futuros.

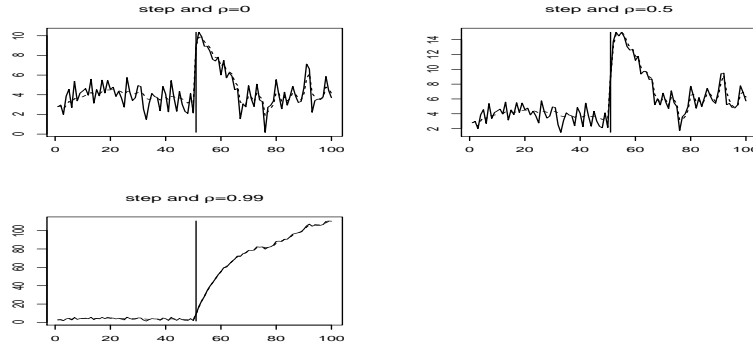


Figura 5.3: Exemplos da função degrau para o Modelo 2, com  $n = 100$  e  $\rho = 0,00, 0,50$  e  $0,99$ . A linha vertical indica o instante da intervenção,  $T = 50$ , a linha contínua indica a série,  $y_t$ , e a linha tracejada indica a resposta média  $\alpha_t$ .

### Estimação

A estimação dos parâmetros dos Modelos 1 e 2 pode ser feita usando métodos clássicos e/ou Bayesianos. Em ambos os casos, o algoritmo do filtro de Kalman (Kalman, 1960) será utilizado para estimar os componentes do vetor de espaço de estados,  $\alpha_t$ , dadas as observações  $\mathbf{Y}_t = \{y_1, \dots, y_t\}$ . Na Seção 2.4 foi visto que o estimador linear de  $\alpha_t$  e sua matriz de variância são dados, respectivamente, por

$$\mathbf{a}_t = E(\alpha_t | \mathbf{Y}_t) \quad \text{e} \quad \mathbf{P}_t = \text{Var}(\alpha_t | \mathbf{Y}_t).$$

O filtro de Kalman pode ser usado para calcular o erro de previsão,  $\nu_t$ , e sua variância

$$\nu_t = y_t - E(y_t | \mathbf{Y}_{t-1}) = y_t - \tilde{y}_{t|t-1} \quad \text{e} \quad F_t = \mathbf{z}_t' \mathbf{P}_{t|t-1} \mathbf{z}_t + h_t,$$

onde  $\mathbf{P}_{t|t-1} = \text{Var}(\alpha_t | \mathbf{Y}_{t-1})$ .

Assumindo que os erros  $\epsilon_t$  e  $\eta_t$  são normalmente distribuídos e o vetor paramétrico é dado por  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_p)'$  - que inclui os hiperparâmetros (ou seja, as variâncias dos termos de erros) e também parâmetros associados com variáveis exógenas ou componentes não-observáveis tais como  $\rho$  e  $\beta$  - a função de verossimilhança  $L(\boldsymbol{\varphi}; \mathbf{Y}_n)$  pode ser calculada depois de integrar fora as variáveis do vetor de estados. Para uma série univariada de tamanho

$n$ , o logaritmo da função de verossimilhança (como definido em 2.13) é dado por

$$\ln L(\boldsymbol{\varphi}; \mathbf{Y}_n) = \ln \prod_{t=1}^n p(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\varphi}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln |F_t| - \frac{1}{2} \sum_{t=1}^n \nu_t' F_t^{-1} \nu_t \quad (5.14)$$

onde  $p = 4$  e  $\boldsymbol{\varphi} = (\sigma_\epsilon^2, \sigma_\eta^2, \rho, \beta)'$  para o Modelo 1 ou  $\boldsymbol{\varphi} = (\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\xi^2, \rho)'$  para o Modelo 2.

Para obter estimadores para  $\boldsymbol{\varphi}$  usando inferência clássica, deve-se maximizar o logaritmo da função de verossimilhança, dado na equação (5.14), com respeito a  $\boldsymbol{\varphi}$ . Como visto na Seção 2.6.1, a função de verossimilhança é uma função não-linear de  $\boldsymbol{\varphi}$ , portanto o algoritmo de otimização BFGS será empregado. Inferências para os parâmetros sob a abordagem clássica serão feitas usando o *bootstrap* paramétrico (Efron, 1979), descrito na Seção 3.3(ii).

Neste capítulo, a distribuição *a priori* uniforme, dada por  $\pi(\boldsymbol{\varphi}) \propto c$ , onde  $c \in \Re$ , é usada para realizar a inferência Bayesiana. Espera-se com isso permitir uma comparação justa com a abordagem clássica. Como a distribuição *a posteriori* não tem uma forma fechada, métodos *Markov chain Monte Carlo* (MCMC) são usados para obter estimativas pontuais e intervalares para os parâmetros (ver Gamerman & Lopes (2006) para detalhes). Neste capítulo, uma versão híbrida do algoritmo de Metropolis-Hastings é adotada, na qual os componentes de  $\boldsymbol{\varphi}$  são atualizados separadamente, com diferentes densidades de proposição. Reis, Salazar & Gamerman (2006) obtêm evidências empíricas em favor deste esquema de amostragem frente a outros esquemas MCMC.

## Previsão

A previsão  $k$  passos à frente,  $\tilde{y}_{n+k|n}(\boldsymbol{\varphi})$ , é obtida pela esperança condicional  $E(y_{n+k} | \mathbf{Y}_n, \boldsymbol{\varphi})$ , como visto na Seção 2.7.

Na abordagem clássica, as funções de previsão para os Modelos 1 e 2 são dadas, respectivamente, por

$$\tilde{y}_{n+k|n}^{(c)} = \mathbf{a}_n^{(\mu)} + \hat{\rho}^k \mathbf{a}_n^{(E)} + \hat{\beta} \sum_{i=1}^k \hat{\rho}^{k-i} x_{n+i}$$

e

$$\tilde{y}_{n+k|n}^{(c)} = \mathbf{a}_n^{(\mu)} + \hat{\rho}^k \mathbf{a}_n^{(E)} + \sum_{i=1}^k \hat{\rho}^{k-i} \mathbf{a}_n^{(\beta)} x_{n+i}$$

onde  $\mathbf{a}_n^{(\mu)}$  é a estimativa do componente  $\mu_t$ ,  $\mathbf{a}_n^E$  é a estimativa do componente  $E_t$  e  $\mathbf{a}_n^{(\beta)}$  é a estimativa do componente  $\beta_n$ , usando o vetor de máxima verossimilhança  $\hat{\boldsymbol{\varphi}} = (\hat{\sigma}_\epsilon^2, \hat{\sigma}_\eta^2, \hat{\rho}, \hat{\beta})$  para o Modelo 1 e  $\hat{\boldsymbol{\varphi}} = (\hat{\sigma}_\epsilon^2, \hat{\sigma}_\eta^2, \hat{\sigma}_\xi^2, \hat{\rho})$  para o Modelo 2.

Um intervalo de confiança *bootstrap* percentílico de nível  $100(1 - \kappa)\%$  para  $y_{n+k}$  é dado por

$$\left[ \tilde{y}_{n+k}^{*(\kappa/2)}, \tilde{y}_{n+k}^{*(1-\kappa/2)} \right],$$

onde  $\tilde{y}_{n+k}^{*(\kappa)}$  é o  $B(\kappa)$ -ésimo valor ordenado da replicação *bootstrap* para a previsão. Neste trabalho, a previsão *bootstrap* é calculada baseada no trabalho de Thombs & Schuncany (1990). Alguns trabalhos que realizam intervalos de previsão *bootstrap* em modelos de espaço de estados são os de Rodriguez & Ruiz (2009) e Wall & Stoffer (2002).

Na abordagem Bayesiana, a função de previsão é dada pela média da distribuição preditiva de  $y_{n+k} | \mathbf{Y}_n$  e é denotada por  $\tilde{y}_{n+k}^{(b)}$ , como definido na Seção 2.7.

Um intervalo de credibilidade de  $100(1 - \kappa)\%$  para  $y_{n+k}$  é dado por

$$\int_{li}^{ls} p(y_{n+k} | \mathbf{Y}_n) d\mathbf{Y}_n = 1 - \kappa.$$

Os limites  $li$  e  $ls$  podem ser obtidos aproximadamente por MCMC e os passos do algoritmo são descritos abaixo, uma vez que uma amostra de  $\boldsymbol{\psi}^{(1)}, \dots, \boldsymbol{\psi}^{(m)}$  é obtida de  $\pi(\boldsymbol{\psi} | \mathbf{Y}_n)$  para cada  $j$  com  $j = 1, \dots, m$ .

**Algoritmo 5.1:** Cálculo do intervalo de credibilidade para  $y_{n+k}$

---

**gere**  $\boldsymbol{\alpha}_n^{(j)}$  da distribuição  $p(\boldsymbol{\alpha}_n | \boldsymbol{\psi}^{(j)}, \mathbf{Y}_n)$ , obtido através do filtro de Kalman

**gere**  $\boldsymbol{\alpha}_{n+k}^{(j)}$  da distribuição de  $p(\boldsymbol{\alpha}_{n+k} | \boldsymbol{\alpha}_n^{(j)}, \boldsymbol{\psi}^{(j)}, \mathbf{Y}_n)$ , obtido pela equação

$$(2.15)$$

**calcule**  $y_{n+k}^{(j)} = \mathbf{z}_{n+k}' \boldsymbol{\alpha}_{n+k}^{(j)} + \epsilon_{n+k}^{(j)}$ , onde  $\epsilon_{n+k}^{(j)}$  é gerado de uma distribuição

Normal com média zero e variância  $\sigma_\epsilon^{2(j)}$

**ordene** os valores de  $y_{n+k}^{(1)}, \dots, y_{n+k}^{(m)}$

**calcule** os percentis  $100(\kappa/2)$  e  $100(1 - \kappa/2)$  como os limites inferior e superior do intervalo, respectivamente.

---

## 5.4 Estudos de simulação

Os procedimentos descritos na Seção 5.3 são agora investigados através de experimentos Monte Carlo (MC), implementados na linguagem Ox (Doornik, 1999). Séries com intervenções pulso e degrau para o modelo de nível local foram simuladas, baseadas no Modelo 1 e Modelo 2 (neste, só intervenção do tipo degrau). O desempenho do estimador de máxima verossimilhança (MLE) e dos estimadores de Bayes - média (Mean), mediana (Med) e moda (Mode), assim como da técnica *bootstrap* (Boot), foi avaliado para séries de tamanho  $n = 100$  e  $\rho = 0,00, 0,50$  e  $0,99$ . Em todos os casos,  $\sigma_\epsilon^2 = 1,00$ ,  $\sigma_\eta^2 = 0,10$ ,  $\sigma_\xi^2 = 0,50$  (para o Modelo 2) e  $\beta = 4$  (para o Modelo 1). Para os estimadores de Bayes, duas cadeias com 2000 amostras foram geradas das quais as 1000 primeiras foram excluídas. Os números de MC e *bootstrap* foram ambos fixados em 500. O nível e a probabilidade dos intervalos de confiança e de credibilidade, respectivamente, foram fixados em 0,95.

### 5.4.1 Modelo 1

As Figuras 5.4 a 5.9 apresentam o vício e o EQM das 500 replicações MC para o Modelo 1 com funções pulso e degrau.

Para a função pulso (Figuras 5.4, 5.5 e 5.6), a primeira conclusão que pode ser obtida é que  $\rho$  é sempre subestimado, exceto no caso  $\rho = 0,00$ . Se  $\rho$  é grande ( $\rho = 0,99$ ), o melhor estimador é a moda, com vício e EQM pequenos, mas para  $\rho = 0,50$  os melhores estimadores são a mediana e média e para  $\rho = 0,00$  o melhor é o EMV. Os outros parâmetros apresentam desempenho satisfatório para todos os métodos. Os melhores estimadores para  $\sigma_\eta^2$  são a média e a moda, independente do valor de  $\rho$ . Para  $\sigma_\epsilon^2$ , todos os procedimentos apresentam o mesmo comportamento. No caso do parâmetro  $\beta$ , se  $\rho$  é grande, todos os procedimentos mostram aproximadamente o mesmo desempenho, mas a média e mediana são ligeiramente melhores para  $\rho = 0,50$  e  $0,00$ . Com respeito ao *bootstrap*, pode-se ver que ele reproduz bem o comportamento do EMV, o que permite a essa técnica ser usada na abordagem clássica para construir os intervalos de confiança para os parâmetros.

Para a função degrau (Figuras 5.7, 5.8 e 5.9),  $\rho$  é subestimado nos casos em que  $\rho = 0,99$  e  $\rho = 0,50$  pelos estimadores Bayesianos. Todos os procedimentos mostram um excelente desempenho para estimar  $\rho$  quando esse parâmetro assume um valor grande, mas para  $\rho = 0,50$  e  $0,00$ , o EMV e a moda *a posteriori* são um pouco melhores. Os desempenhos dos métodos



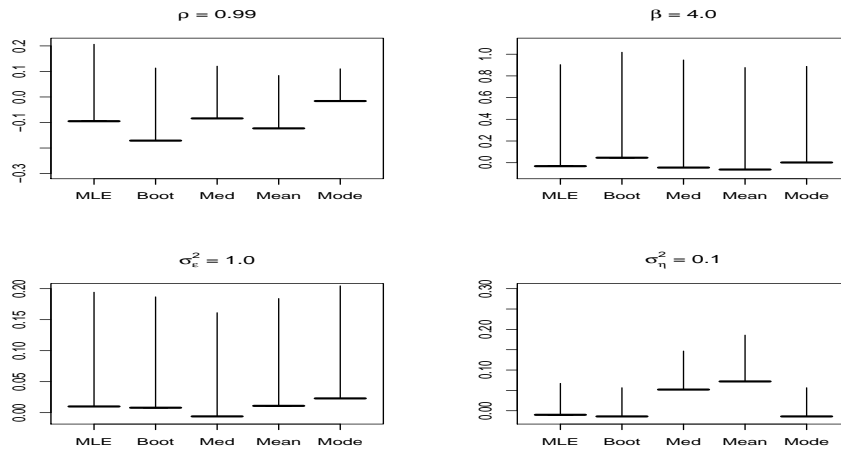


Figura 5.4: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função pulso e  $\rho = 0,99$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

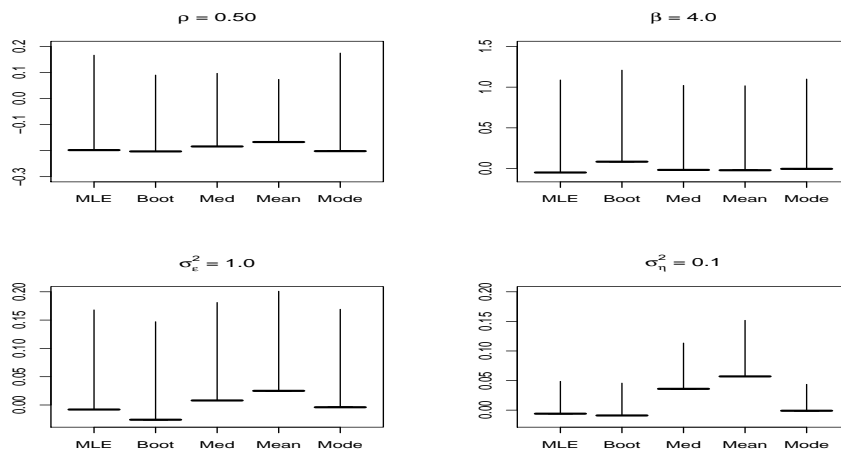


Figura 5.5: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função pulso e  $\rho = 0,50$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

para estimar os outros parâmetros com a função degrau são muito similares à função pulso, exceto no caso  $\rho = 0,99$ . Parece que, neste caso, o EMV não é capaz de estimar  $\sigma_\eta^2$ ,  $\sigma_\epsilon^2$  e  $\beta$  muito precisamente, apresentando vício e EQM maiores que os estimadores Bayesianos. Novamente, o comportamento do *bootstrap* é muito similar ao do EMV.

As Tabelas 5.1 e 5.2 apresentam os intervalos de confiança *bootstrap* e de credibilidade para o Modelo 1 com funções pulso e degrau, respectivamente. Nessas tabelas, além dos estimadores, a previsão um passo à frente é também incluída. Quando  $\rho = 0,99$ , para todos os parâmetros considerados, ambos intervalos têm taxas de cobertura bem próximas do nível nominal 0,95, exceto o intervalo de confiança *bootstrap* para  $\rho$ . Por outro lado, o desempenho do intervalo de credibilidade para a função degrau é bem melhor que o intervalo *bootstrap*, exceto para previsão um passo à frente, onde ambos estão longe do nível de 0,95 assumido. Se  $\rho = 0,50$  ou 0,00, para funções pulso e degrau, ambos intervalos mostram aproximadamente o mesmo comportamento para todos os parâmetros, com taxa de cobertura próxima de 0,95. As únicas exceções são o intervalo *bootstrap* para  $\sigma_\eta^2$ , o qual é ligeiramente pior do que o intervalo de credibilidade, e os intervalos para  $\rho$  na função degrau, onde ambos os métodos apresentam taxas de cobertura abaixo de 0,90. Deve-se ressaltar que os intervalos para  $\rho = 0,00$  não são construídos, pois os intervalos não cobrem o valor verdadeiro deste parâmetro, estando no limite do espaço paramétrico.

### 5.4.2 Modelo 2

As Figuras 5.10, 5.11 e 5.12 apresentam o vício e o EQM sobre as 500 replicações MC para o Modelo 2. Como foi explicado na Seção 5.3, neste caso o modelo não é capaz de estimar a influência da função pulso. Portanto, apenas os resultados para função degrau são apresentados abaixo.

Pode-se notar que, em geral, o EMV e a moda *a posteriori* são ligeiramente melhores que os outros procedimentos para estimar todos os parâmetros, com vício e EQM menores. Mais uma vez,  $\rho$  é subestimado, exceto para  $\rho = 0,00$ .

Na Tabela 5.3 estão os resultados para os intervalos *bootstrap* e de credibilidade com função degrau. Como no Modelo 1 com função degrau, o desempenho do intervalo de credibilidade é novamente melhor do que o do intervalo *bootstrap*, no caso  $\rho = 0,99$ . Se  $\rho = 0,50$  ou 0,00, o intervalo de credibilidade mostra, em geral, taxas de cobertura mais próximas de 0,95,

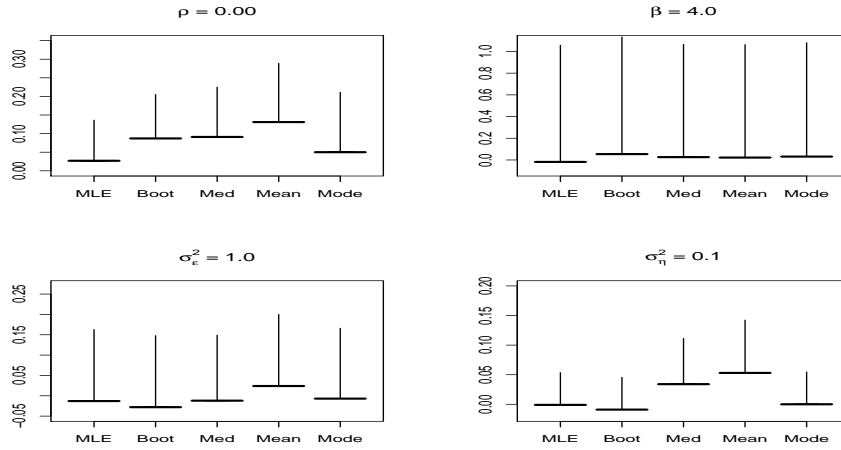


Figura 5.6: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função pulso e  $\rho = 0,00$ . As linhas horizontal e vertical indicam o vício e a raiz quadrada do EQM, respectivamente.

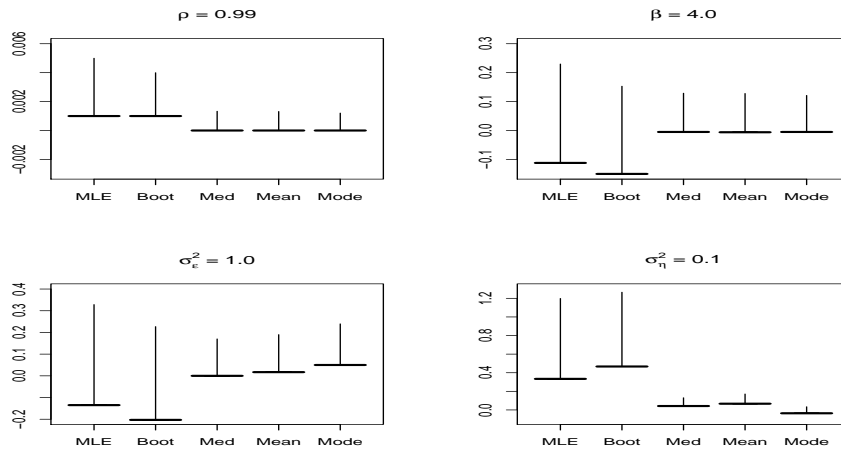


Figura 5.7: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função degrau e  $\rho = 0,99$ . As linhas horizontal e vertical indicam o vício e a raiz quadrada do EQM, respectivamente.

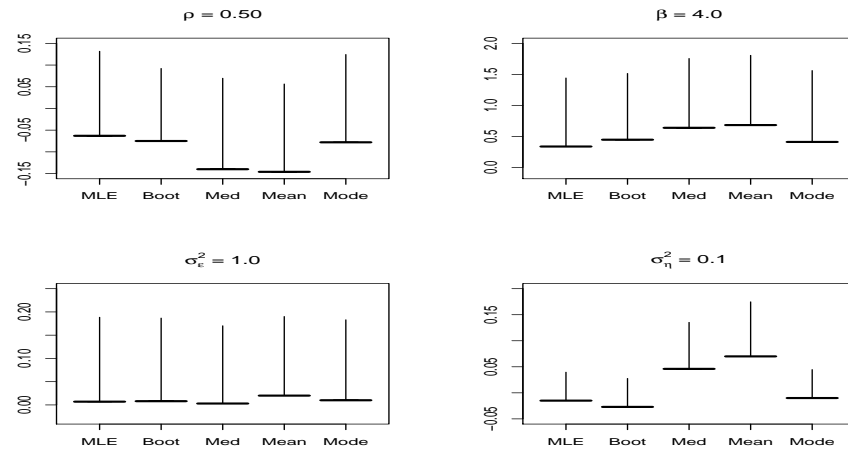


Figura 5.8: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função degrau e  $\rho = 0,50$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

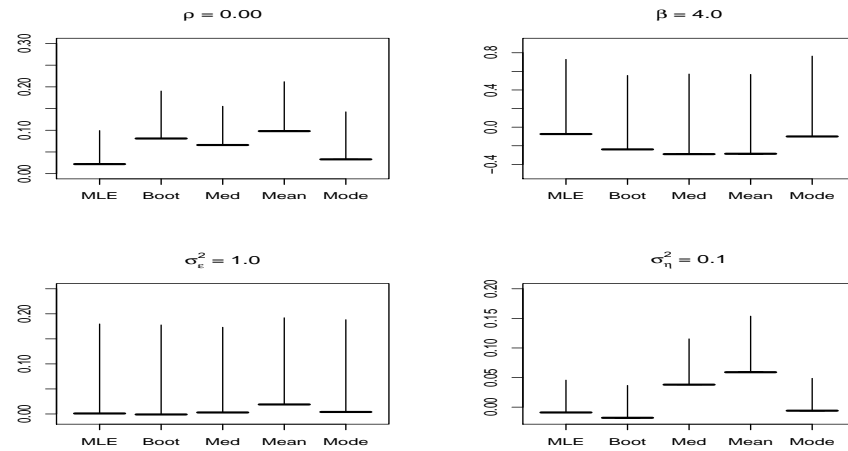


Figura 5.9: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 1 com função degrau e  $\rho = 0,00$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

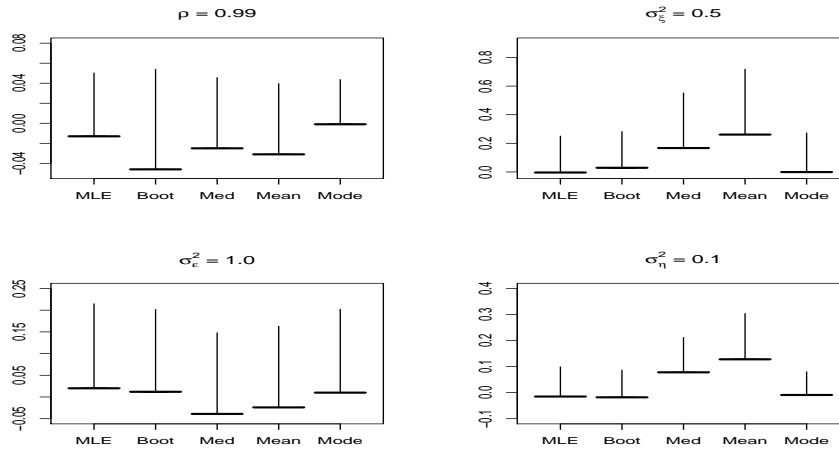


Figura 5.10: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 2 com função degrau e  $\rho = 0,99$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

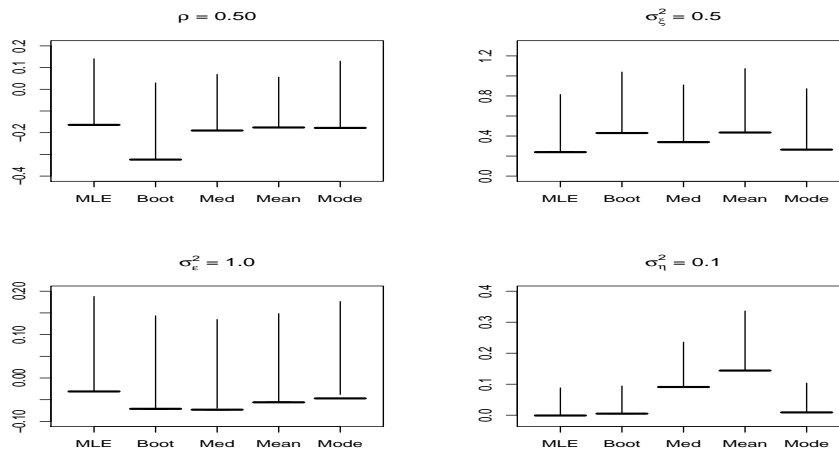


Figura 5.11: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 2 com função degrau e  $\rho = 0,50$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

Tabela 5.1: Intervalos de confiança e de credibilidade para o Modelo 1 com função pulso.

	valor verdadeiro	Intervalo de Confiança			Intervalo de Credibilidade		
		limites médios	amp	cob	limites médios	amp	cob
$\rho$	0,99	[0,177; 0,983]	0,806	0,87	[0,536; 0,994]	0,458	0,95
$\beta$	4,00	[2,167; 5,681]	3,514	0,91	[1,921; 5,845]	3,924	0,97
$\sigma_\varepsilon^2$	1,00	[0,687; 1,371]	0,684	0,94	[0,683; 1,427]	0,744	0,95
$\sigma_\eta^2$	0,10	[0,007; 0,256]	0,249	0,93	[0,043; 0,449]	0,406	0,93
$y_{n+1}$		[0,221; 4,590]	4,369	0,93	[0,013; 4,275]	4,262	0,93
$\rho$	0,50	[0,011; 0,757]	0,746	0,91	[0,040; 0,724]	0,684	0,89
$\beta$	4,00	[1,985; 6,144]	4,159	0,92	[1,724; 6,197]	4,473	0,96
$\sigma_\varepsilon^2$	1,00	[0,668; 1,328]	0,660	0,93	[0,700; 1,445]	0,745	0,95
$\sigma_\eta^2$	0,10	[0,011; 0,213]	0,202	0,87	[0,046; 0,390]	0,344	0,94
$y_{n+1}$		[-2,179; 1,960]	4,139	0,92	[-2,427; 2,024]	4,451	0,95
$\rho$	0,00	-	-	-	-	-	-
$\beta$	4,00	[1,944; 6,148]	4,204	0,94	[1,802; 6,231]	4,429	0,96
$\sigma_\varepsilon^2$	1,00	[0,663; 1,326]	0,663	0,93	[0,702; 1,441]	0,739	0,95
$\sigma_\eta^2$	0,10	[0,015; 0,221]	0,206	0,90	[0,046; 0,377]	0,331	0,94
$y_{n+1}$		[-2,149; 1,944]	4,093	0,92	[-2,424; 2,025]	4,449	0,95

Obs.: amp é a amplitude média e cob é a cobertura média. Não há um único valor verdadeiro para  $y_{n+1}$ , já que cada série tem o seu próprio valor  $y_{n+1}$ .

exceto para  $\sigma_\xi^2$ , onde o intervalo *bootstrap* é melhor. Novamente, os intervalos para  $\rho = 0,00$  não são construídos.

## 5.5 Aplicação a séries reais

Nesta seção, a metodologia anteriormente descrita é aplicada a duas séries temporais reais. A primeira é o Índice de Preço ao Consumidor Amplo (IPCA) na cidade de Belo Horizonte, no período de julho de 1997 a junho de 2008. Essa série parece ter uma intervenção pulso em torno de outubro de 2002. A segunda série é o retorno da BOVESPA (mercado de ações de São Paulo), no período de janeiro de 1991 a agosto de 2008. Para essa série, a intervenção tem a forma de função degrau, apresentando um salto em torno da metade de 1994. Para as estimativas Bayesianas são consideradas duas cadeias com 2000 valores das quais os 1000 primeiros valores são excluídos. As cadeias são inicializadas de pontos diferentes.

### 5.5.1 Série IPCA

A série IPCA, já descrita na Seção 2.2, parece apresentar um comportamento de um Modelo de Nível Local (MNL), com uma intervenção em torno de

Tabela 5.2: Intervalos de confiança e de credibilidade para o Modelo 1 com função degrau.

	valor verdadeiro	Intervalo de Confiança			Intervalo de Credibilidade		
		limites médios	amp	cob	limites médios	amp	cob
$\rho$	0,99	[0,988; 0,998]	0,010	0,75	[0,987; 0,992]	0,005	0,88
$\beta$	4,00	[3,165; 4,169]	1,004	0,78	[3,726; 4,257]	0,531	0,90
$\sigma_\varepsilon^2$	1,00	[0,027; 1,255]	1,228	0,79	[0,688; 1,434]	0,746	0,94
$\sigma_\eta^2$	0,10	[0,201; 2,238]	2,037	0,81	[0,035; 0,443]	0,407	0,95
$y_{n+1}$		[155,64; 160,35]	4,71	0,51	[155,45; 159,83]	4,38	0,38
$\rho$	0,50	[0,139; 0,655]	0,516	0,89	[0,090; 0,601]	0,511	0,81
$\beta$	4,00	[2,839; 6,330]	3,491	0,92	[2,964; 6,634]	3,670	0,91
$\sigma_\varepsilon^2$	1,00	[0,695; 1,365]	0,670	0,94	[0,692; 1,441]	0,749	0,96
$\sigma_\eta^2$	0,10	[0,005; 0,187]	0,182	0,80	[0,046; 0,436]	0,390	0,94
$y_{n+1}$		[5,761; 10,052]	4,291	0,93	[5,536; 9,819]	4,283	0,94
$\rho$	0,00	-	-	-	-	-	-
$\beta$	4,00	[2,135; 5,329]	3,194	0,94	[0,729; 5,465]	4,736	0,95
$\sigma_\varepsilon^2$	1,00	[0,686; 1,360]	0,674	0,93	[0,697; 1,434]	0,737	0,95
$\sigma_\eta^2$	0,10	[0,007; 0,201]	0,194	0,84	[0,046; 0,395]	0,349	0,93
$y_{n+1}$		[1,792; 5,977]	4,185	0,92	[1,543; 5,816]	4,273	0,94

Obs.: amp é a amplitude média e cob é a cobertura média. Não há um único valor verdadeiro para  $y_{n+1}$ , já que cada série tem o seu próprio valor  $y_{n+1}$ .

outubro de 2002. É importante ressaltar que, no final do ano de 2002, Luís Inácio Lula da Silva foi eleito presidente do Brasil, o que pode ter afetado o índice de preços.

Como a intervenção é da forma pulso, o Modelo 1 será utilizado. Da Tabela 5.4 verifica-se que o Modelo 1, em ambas estimações clássica e Bayesiana, apresenta menor AIC do que o MNL puro, justificando a inclusão da intervenção no modelo. A Figura 5.13 apresenta a série IPCA juntamente com a resposta média estimada pelo Modelo 1 utilizando as abordagens clássica e Bayesiana. O gráfico mostra que este modelo acompanha bem o comportamento da série, conseguindo captar a quebra estrutural presente na mesma. A previsão um passo à frente foi calculada para julho de 2008. O valor verdadeiro é 0,35, e pode ser visto que a predição Bayesiana é mais próxima do valor verdadeiro que a clássica, com menor amplitude do intervalo. Na estimacão Bayesiana, os valores referentes ao critério de convergência de Gelman e Rubin (Gelman, 1996) são próximos da unidade, indicando a convergência das cadeias.

Estimativas para os parâmetros são mostradas na Tabela 5.5. Parece que o termo de nível  $\mu_t$  é constante no tempo, já que os intervalos para  $\sigma_\eta^2$  estão bem concentrados em torno de zero. O valor mais provável de  $\rho$ , que mede o efeito da persistência da intervenção, é 0,80. Já a estimativa de  $\beta$  está em torno de 1,97, que corresponde à magnitude do salto da intervenção,

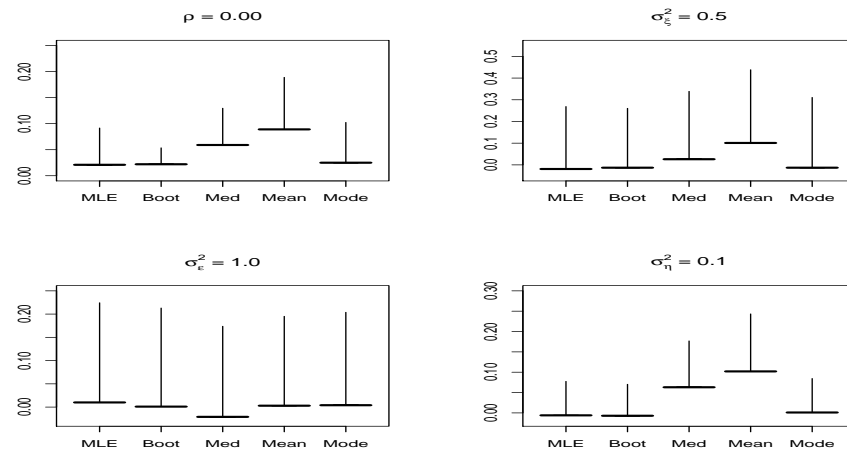


Figura 5.12: Máxima verossimilhança, *bootstrap* e estimação Bayesiana para o Modelo 2 com função degrau e  $\rho = 0,00$ . As linhas horizontal e vertical (o comprimento da linha) indicam o vício e a raiz quadrada do EQM, respectivamente.

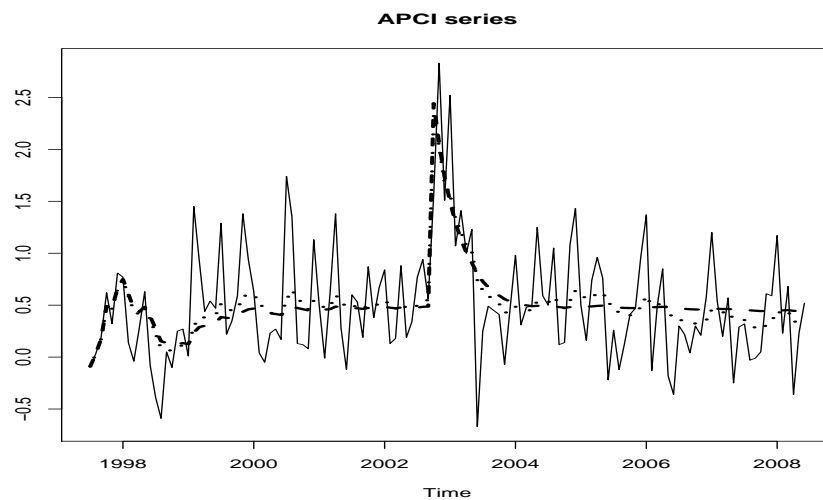


Figura 5.13: A linha contínua indica a série do IPCA,  $y_t$ , e as linhas pontilhada e tracejada indicam as estimativas da resposta média do Modelo 1, segundo as abordagens Bayesiana e clássica, respectivamente.



Tabela 5.3: Intervalos de confiança e de credibilidade para o Modelo 2

	valor verdadeiro	Intervalos de Confiança		Intervalos de Credibilidade	
		limites médios	amp cob	limites médios	amp cob
$\rho$	0,99	[0,671; 0,998]	0,327 0,98	[0,887; 0,997]	0,110 0,96
$\sigma_{\xi}^2$	0,50	[0,128; 1,243]	1,115 0,99	[0,255; 1,815]	1,125 0,93
$\sigma_{\varepsilon}^2$	1,00	[0,656; 1,407]	0,751 0,96	[0,604; 1,436]	0,832 0,97
$\sigma_{\eta}^2$	0,10	[0,001; 0,280]	0,279 0,74	[0,032; 0,722]	0,690 0,95
$y_{n+1}$		[-27,19; -22,71]	4,48 0,27	[-29,49; -20,53]	6,956 0,81
$\rho$	0,50	[0,001; 0,608]	0,607 0,67	[0,103; 0,613]	0,510 0,81
$\sigma_{\xi}^2$	0,50	[0,169; 2,058]	1,889 0,93	[0,240; 2,174]	1,934 0,90
$\sigma_{\varepsilon}^2$	1,00	[0,562; 1,343]	0,781 0,90	[0,559; 1,420]	0,861 0,95
$\sigma_{\eta}^2$	0,10	[0,003; 0,309]	0,306 0,85	[0,041; 0,762]	0,721 0,92
$y_{n+1}$		[-3,434; 0,855]	4,289 0,77	[-3,258; 2,660]	5,920 0,84
$\rho$	0,00	-	- -	-	- -
$\sigma_{\xi}^2$	0,50	[0,052; 1,141]	1,089 0,92	[0,102; 1,526]	1,424 0,97
$\sigma_{\varepsilon}^2$	1,00	[0,636; 1,417]	0,781 0,93	[0,623; 1,471]	0,848 0,97
$\sigma_{\eta}^2$	0,10	[0,003; 0,270]	0,267 0,85	[0,037; 0,588]	0,551 0,93
$y_{n+1}$		[-2,759; 1,368]	4,127 0,79	[-2,662; 3,329]	5,991 0,87

Obs.: amp é a amplitude média e cob é a cobertura média. Não há um único valor verdadeiro para  $y_{n+1}$ , já que cada série tem o seu próprio valor  $y_{n+1}$ .

compatível com uma inspeção visual.

Tabela 5.4: AIC e previsão um passo à frente para a série IPCA.

Modelos	AIC	Previsão (valor verdadeiro=0,350)	Int. Previsão
MNL Clássico	212,650	0,320	[-0,508; 1,456]
MNL Bayesiano	212,600	0,330	[-0,302; 0,962]
Modelo 1 Clássico	198,050	0,419	[-0,407; 1,442]
Modelo 1 Bayesiano	198,101	0,349	[-0,443; 1,430]

### 5.5.2 Série IBOVESPA

A série IBOVESPA é uma série mensal do retorno do mercado de ações de São Paulo, no período de janeiro de 1991 a agosto de 2008. Essa série é composta de 212 observações mensais e parece seguir um MNL com um componente de intervenção da forma degrau, como observado na série formada pela linha contínua na Figura 5.14. A série possui valores acima de zero até a metade de 1994 e próximos de zero após esse período. Uma razão para tal comportamento é a introdução do Plano Real que mudou a moeda no Brasil, implementado em julho de 1994. A fim de comparar os procedimentos, o MNL e os Modelos 1 e 2 com uma intervenção da forma degrau foram ajustados à série IBOVESPA.

Tabela 5.5: Ajuste do Modelo 1 à série IPCA.

	Inferência Clássica		Inferência Bayesiana	
	EMV	Intervalo	Moda	Intervalo
$\rho$	0,797	[0,030; 0,883]	0,801	[0,558; 0,884]
$\beta$	1,975	[1,411; 2,888]	1,968	[0,851; 2,685]
$\sigma_\varepsilon^2$	0,225	[0,167; 0,277]	0,222	[0,168; 0,298]
$\sigma_\eta^2$	$5,88 \times 10^{-12}$	$[2,13 \times 10^{-18}; 1,63 \times 10^{-3}]$	$3,52 \times 10^{-5}$	$[2,88 \times 10^{-5}; 2,52 \times 10^{-2}]$

Tabela 5.6: AIC e previsão um passo à frente para a série IBOVESPA.

Modelos	AIC	previsão (valor verdadeiro=-0,064)	Intervalo previsão
MNL Clássico	-174,389	-0,020	[-0,291; 0,229]
MNL Bayesiano	-174,042	-0,009	[-0,314; 0,294]
Modelo 1 Clássico	-212,488	0,020	[-0,255; 0,296]
Modelo 1 Bayesiano	-212,050	0,003	[-0,229; 0,348]
Modelo 2 Clássico	-255,019	0,010	[-0,291; 0,306]
Modelo 2 Bayesiano	-255,009	-0,016	[-0,342; 0,287]

A Tabela 5.6 mostra o AIC e a previsão um passo à frente usando os três modelos mencionados acima. Pode-se ver que o Modelo 2, em ambas abordagens clássica e Bayesiana, apresenta um menor AIC do que o MNL e o Modelo 1. Novamente, os valores referentes ao critério de convergência de Gelman e Rubin (Gelman, 1996) para as 2 cadeias de cada parâmetro estão próximos da unidade, indicando a convergência das mesmas. Na Figura 5.14 pode-se observar que a resposta média do Modelo 2, tanto clássico quanto Bayesiano, consegue acompanhar o comportamento da série, captando bem o instante e a forma da quebra. O procedimento Bayesiano possui a vantagem de fornecer, para esta série, uma previsão um passo à frente mais próxima do valor real do que o procedimento clássico, porém a amplitude do intervalo é maior. As estimativas dos parâmetros são mostradas na Tabela 5.7. Mais uma vez, parece que o termo de nível  $\mu_t$  é constante no tempo, como pode ser verificado pelos intervalos para  $\sigma_\eta^2$ . O valor de  $\rho$  é muito pequeno neste caso, o que significa que o modelo não guarda uma memória dos instantes anteriores.

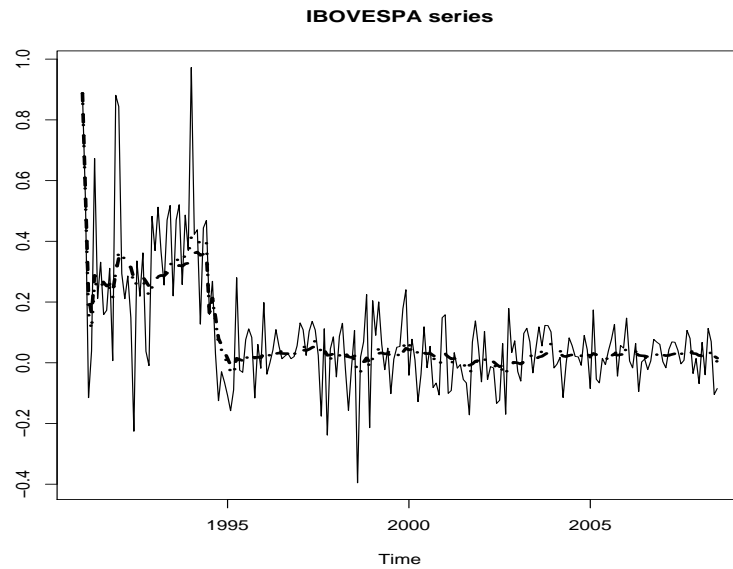


Figura 5.14: A linha contínua indica a série do retorno do IBOVESPA,  $y_t$ , e as linhas pontilhada e tracejada indicam as respostas médias do Modelo 2 Clássico e do Modelo 2 Bayesiano, respectivamente.

Tabela 5.7: Ajuste do Modelo 2 à série IBOVESPA.

	Inferência Clássica		Inferência Bayesiana	
	EMV	Intervalo	Moda	Intervalo
$\rho$	0	[0; 0,022]	0,002	[0,001; 0,184]
$\sigma_{\xi}^2$	0,079	[0,042; 0,123]	0,090	[0,046; 0,164]
$\sigma_{\varepsilon}^2$	0,010	[0,010; 0,012]	0,010	[0,008; 0,013]
$\sigma_{\eta}^2$	$2,30 \times 10^{-13}$	$[3,69 \times 10^{-21}; 3,23 \times 10^{-5}]$	$1,086 \times 10^{-6}$	$[1,815 \times 10^{-7}; 2,427 \times 10^{-4}]$

## Capítulo 6

# Modelos de Espaços de Estados Não-Gaussianos

Na literatura estatística, há vários modelos que são construídos baseados nas suposições de normalidade, homoscedasticidade e independência dos erros, porém nem sempre é possível satisfazer essas suposições. Como foi visto ao longo desse texto, a hipótese de independência dos erros é raramente satisfeita em análise de séries temporais. Entretanto, a suposição de normalidade tem sido central às principais abordagens existentes para tratamento de séries temporais.

A modelagem via espaço de estados, que é o tema abordado neste trabalho, também segue esse padrão. Assim sendo, tudo o que foi apresentado até agora nesse texto foi feito sob a suposição de normalidade. Nesse capítulo, serão apresentadas algumas possibilidades para o tratamento de séries temporais que extrapolam essa restrição.

O ponto de partida para essa extensão é o artigo de Nelder & Wedderburn (1972), que propôs a família por eles denominada de *modelos lineares generalizados* (MLG), unificando vários modelos então existentes de forma isolada em uma classe. A idéia básica desses modelos consiste em abrir o leque de opções para a distribuição da variável-resposta, permitindo que a mesma pertença à família exponencial de distribuições, o que inclusive traz um ganho na questão de interpretação do modelo. A função de ligação dos dados faz o papel de relacionar a média dos dados ao preditor linear, segundo Nelder & Wedderburn (1972) e Dobson (2002).

Há vários trabalhos na literatura propondo modelos que levam em consideração a correlação das observações e o fato da série temporal ser não-

Gaussiana. Trabalhos abordando esse tópico incluem Cox (1981), Kaufmann (1987), Kitagawa (1987), Shephard & Pitt (1997), Smith & Miller (1986), Harvey & Fernandes (1989) e Durbin & Koopman (2000), entre outros. Uma estrutura mais geral, denominada por *Modelos Lineares Dinâmicos Generalizados* (MLDG), foi proposta por West, Harrison & Migon (1985), despertando imenso interesse devido à grande aplicabilidade dos mesmos em diversas áreas do conhecimento. Prova disso é o grande número de trabalhos publicados sobre esses modelos, como por exemplo Gamerman & West (1987), Grunwald, Raftery & Guttorp (1993), Fahrmeir (1987), Fruhwirth-Schnatter (1994), Lindsey & Lambert (1995), Gamerman (1991, 1998), Chiogna & Gaetan (2002), Hemming & Shaw (2002) e Godolphin & Triantafyllopoulos (2006).

Esse capítulo será iniciado por uma seção fazendo uma breve descrição dos *Modelos Lineares Dinâmicos Generalizados* e algumas das abordagens propostas para sua inferência. Esses modelos descrevem a série temporal de acordo com um MEE com a diferença que a equação das observações prescinde da normalidade, sendo substituída por uma especificação na família exponencial e uma função de ligação. Com isso, todas as componentes são assumidas dinâmicas, incluindo as componentes de tendência e de efeito de variáveis explicativas.

O problema com essa classe de modelos é que a tratabilidade analítica é facilmente perdida, mesmo para componentes muito simples. Assim, a verossimilhança preditiva, que é fundamental para o processo de inferência, pode apenas ser obtida de forma aproximada. Uma particularização desses modelos assume que apenas a tendência está flutuando e os efeitos das co-variáveis são fixos ao longo do tempo. Para esses casos, existe uma classe bastante ampla de modelos que permite o cálculo exato da verossimilhança preditiva.

Assim, o objetivo principal deste capítulo é propor na sua segunda seção a *família Gama de modelos dinâmicos*, onde essa tratabilidade analítica é obtida. Essa família é obtida a partir de uma generalização de um resultado de Smith & Miller (1986). Esses autores propuseram um modelo dinâmico para séries temporais com dados exponenciais e transformações um-a-um dessas séries. Eles propuseram uma equação de evolução exata para qualquer série temporal com distribuição pertencente a essa família, possibilitando assim a integração analítica dos estados e a obtenção da verossimilhança preditiva.

A *família Gama de modelos dinâmicos* será apresentada na Seção 6.2.

Seus principais resultados teóricos serão fornecidos bem como será descrita a forma de fazer inferência. Novamente, a abordagem clássica, utilizando o estimador de máxima verossimilhança (EMV), e a abordagem Bayesiana, utilizando métodos MCMC para obter os estimadores Bayesianos (EB), são consideradas no processo de inferência. Além disso, os intervalos de confiança assintótico e de credibilidade são construídos para os parâmetros. A seguir, na Seção 6.3 serão apresentados casos particulares de interesse dentro dessa família. A Seção 6.4 trata da comparação dos pontos de vista clássico e Bayesiano através de exercícios de simulação. O capítulo é concluído com uma aplicação a dados reais.

## 6.1 Modelos Lineares Dinâmicos Generalizados

Essa classe de modelos foi introduzida por West, Harrison & Migon (1985) e consiste numa extensão direta do MEE dado em (2.3)-(2.4) ao permitir especificações mais gerais para a distribuição das observações da série, assumida pertencer à família exponencial de distribuições. A outra novidade aqui com relação à especificação do Capítulo 2 é a presença de uma função de ligação para estabelecer a conexão entre as equações das observações e dos estados. Mais especificamente, diz-se que uma série temporal  $y_t$  é descrita por um Modelo Linear Dinâmico Generalizado (MDLG) se segue

$$p(y_t|\theta_t, \phi) = a(y_t, \phi) \exp\{\phi[\theta_t y_t - b(\theta_t)]\} \quad \text{onde } \mu_t = E(y_t|\theta_t), \quad (6.1)$$

$$g(\mu_t) = \mathbf{z}_t' \boldsymbol{\alpha}_t + d_t, \quad (6.2)$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t), \quad (6.3)$$

onde o suporte de variação de  $y_t$  não depende de  $\theta_t$ ,  $\boldsymbol{\eta}_t$  é um vetor de ruídos serialmente não-correlacionados, cuja matriz de covariâncias é dada por  $\mathbf{Q}_t$  e  $\boldsymbol{\alpha}_t$  é o vetor de estados, para  $t = 1, 2, \dots, n$ . Assume-se também que  $\boldsymbol{\eta}_t$  sejam independentes entre si com média 0, para  $t = 1, 2, \dots, n$ . Novamente, as matrizes  $\mathbf{z}_t$ ,  $\mathbf{T}_t$  e  $\mathbf{R}_t$  são conhecidas como matrizes do sistema e assumidas determinísticas e conhecidas e os termos  $d_t$  e  $\mathbf{c}_t$  são covariáveis que podem ser inseridas nas equações das observações e de transição, respectivamente. O modelo é completado com a especificação inicial  $\boldsymbol{\alpha}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$ .

Note que a equação de observação depende do parâmetro canônico  $\theta_t$  e do parâmetro de dispersão  $\phi$ , aqui assumido constante ao longo do tempo.

Os hiperparâmetros do modelo (6.1)-(6.3) são dados pela dispersão  $\phi$  e pelas componentes desconhecidas em  $\mathbf{Q}_t$ .

As principais distribuições usadas na Estatística pertencem à família exponencial de distribuições. É um simples exercício de acomodação de valores das funções  $a$  e  $b$  mostrar que a Poisson, binomial, Gama e Normal pertencem a essa família. Importantes distribuições não incluídas na família exponencial são a uniforme e a  $t$ -Student.

Resultados sobre conjugação dos parâmetros canônicos foram então utilizados por West, Harrison & Migon (1985) para propor formas aproximadas para um procedimento sequencial de inferência supondo conhecidos os hiperparâmetros  $\psi$ . Com isso, as distribuições preditivas podem ser obtidas e seu produto fornece a verossimilhança preditiva para os hiperparâmetros. Inferência tanto sob a ótica clássica quanto Bayesiana pode ser realizada. Entretanto, não é do conhecimento dos autores nenhum exercício desse tipo na literatura. Vale ressaltar que essa verossimilhança só pode ser obtida a partir de uma aproximação do ciclo de inferência. Esse fato talvez ajude a explicar essa lacuna na literatura. Na seção seguinte, será apresentada uma classe relativamente ampla de modelos dinâmicos onde também não se assume normalidade dos dados, mas a verossimilhança preditiva pode ser obtida de forma exata.

A maioria dos trabalhos na área seguiu preferencialmente dois caminhos. No primeiro deles, optou-se por assumir hiperparâmetros conhecidos ou estimados por algum procedimento *ad-hoc*, como uso de fatores de desconto. Essa foi a abordagem original de West, Harrison & Migon (1985) mas também foi utilizada por outros autores como Singh & Roberts (1992) e Fahrmeir & Wagenpfeil (1997). O segundo caminho foi tratar todos os parâmetros de estados e hiperparâmetros conjuntamente e aproximar a distribuição *a posteriori* conjunta através de amostras obtidas via algum algoritmo de simulação. Existem inúmeras formas de fazer isso: Shephard & Pitt (1997), Gamerman (1998) e Knorr-Held (1999) usaram MCMC com diferentes estratégias de formulação de propostas ao passo que Durbin & Koopman (2000) usaram simulação Monte Carlo via reamostragem por importância.

## 6.2 Família Gama de modelos dinâmicos

Smith & Miller (1986) e Harvey & Fernandes (1989) apresentaram casos particulares de modelos dinâmicos de resposta não-Gaussiana. Neste trabalho,

a partir desses casos é feita uma generalização, isto é, uma família mais ampla - que os engloba - é introduzida, denominada *família Gama de modelos dinâmicos* (FGMD). Uma vantagem desses modelos frente aos MLDG é que a equação de evolução é exata e não aproximada. Por outro lado, uma desvantagem é que é complicado inserir outros componentes estocásticos no mesmo, como por exemplo, os componentes de tendência e sazonalidade. Entretanto, o efeito de componentes determinísticos podem ser captados no modelo através de covariáveis.

### 6.2.1 Definição

De uma maneira geral, define-se que uma variável aleatória  $Y$  possui uma distribuição na família Gama (FG) de distribuições, se a sua densidade/função de probabilidade é escrita na forma:

$$p(y|\mu, \varphi) = a(y, \varphi)\mu^{b(y, \varphi)} \exp(-\mu c(y, \varphi)), \quad (6.4)$$

onde  $y \in H(\varphi) \subset \Re$  e  $0$ , caso contrário. As funções  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$ ,  $c(\cdot, \cdot)$  e  $H(\cdot)$  são tais que  $p(y|\mu, \varphi) \geq 0$  e que a integral de Lebesgue  $\int p(y|\mu, \varphi) dy = 1$ . Caso  $b(y, \varphi) = b(\varphi)$  ou  $c(y, \varphi) = c(\varphi)$  e  $H(\varphi)$  seja uma função constante (que não dependa de  $\varphi$ ), a família gama de distribuições torna-se um caso especial da família exponencial de distribuições.

A FGMD será definida da seguinte forma:

1. a função de ligação  $g$  relaciona o preditor com o parâmetro  $\mu_t$  através da relação  $\mu_t = \lambda_t g(x_t, \beta)$ , onde  $x_t$  é um vetor de covariáveis,  $\beta$  contém coeficientes de regressão (um dos componentes de  $\varphi$ ) e  $\lambda_t$  é o parâmetro responsável pela descrição do nível dinâmico. Se o preditor for linear então,  $g(x_t, \beta) = g(x'_t, \beta)$ .
2. A dinâmica do nível  $\lambda_t$  é dada pela equação de evolução  $\lambda_t = w^{-1}\lambda_{t-1}\varsigma_t$ , onde  $\varsigma_t \sim \text{Beta}(wa_{t-1}, (1-w)a_{t-1})$ , isto é,

$$w \frac{\lambda_t}{\lambda_{t-1}} \mid \lambda_{t-1} \sim \text{Beta}(wa_{t-1}, (1-w)a_{t-1});$$

3. A dinâmica do nível é inicializada com a especificação inicial *a priori*  $\lambda_0|Y_0 \sim \text{Gama}(a_0, b_0)$ . Logo, usando a propriedade de escala da distribuição Gama,  $\mu_0|Y_0 \sim \text{Gama}(a_0, b_0[g(x_t, \beta)]^{-1})$ .



Uma das especificações mais comuns para a ligação  $g$  é a função logarítmica. É interessante destacar que, nesse caso, é obtida a equação de evolução  $\ln(\lambda_t) = \ln(\lambda_{t-1}) + \varsigma_t^*$ , onde  $\varsigma_t^* = \ln(\varsigma_t/w) \in \mathfrak{R}$ . Essa equação é similar à usual equação de evolução dada por um passeio aleatório, como no MNL.  $w$  varia entre 0 e 1 e também compõe  $\boldsymbol{\varphi}$ . Como será visto a seguir,  $w$  cumpre a função de aumentar multiplicativamente a variância devido à passagem do tempo. Assim, ele desempenha um papel similar ao das variâncias do sistema e idêntico ao dos fatores de desconto, usados na abordagem Bayesiana para substituir essas variâncias.

Em geral, deseja-se modelar a variância ou a média dessa distribuição, que será função de alguns parâmetros invariantes no tempo e do parâmetro de escala  $\mu_t$ , isto é,  $E(y_t|\mu_t, \boldsymbol{\varphi}) = f(\mu_t, \boldsymbol{\varphi})$ .

**Teorema 1.** Se o modelo está definido na forma descrita na equação (6.4), os seguintes resultados podem ser obtidos, para  $t = 1, \dots, n$

1. a distribuição *a priori*  $\lambda_t|\mathbf{Y}_{t-1}$  segue uma distribuição Gama( $a_{t|t-1}, b_{t|t-1}$ ), onde

$$a_{t|t-1} = wa_{t-1}, \quad (6.5)$$

$$b_{t|t-1} = wb_{t-1}, \quad (6.6)$$

com  $0 < w \leq 1$ .

2. a distribuição *a priori*  $\mu_t|\mathbf{Y}_{t-1}$  é Gama( $a_{t|t-1}^*, b_{t|t-1}^*$ ), onde

$$a_{t|t-1}^* = wa_{t-1}, \quad (6.7)$$

$$b_{t|t-1}^* = wb_{t-1}[g(x_t, \boldsymbol{\beta})]^{-1}. \quad (6.8)$$

3. A distribuição *a posteriori* de  $\mu_t|\mathbf{Y}_t$  é Gama ( $a_t^*, b_t^*$ ), onde

$$a_t^* = a_{t|t-1}^* + b(y_t, \boldsymbol{\varphi}), \quad (6.9)$$

$$b_t^* = b_{t|t-1}^* + c(y_t, \boldsymbol{\varphi}). \quad (6.10)$$

4.  $\lambda_t = \mu_t[g(x_t, \boldsymbol{\beta})]^{-1}|\mathbf{Y}_t$  tem também distribuição Gama( $a_t, b_t$ ), onde

$$a_t = a_{t|t-1} + b(y_t, \boldsymbol{\varphi}), \quad (6.11)$$

$$b_t = b_{t|t-1} + c(y_t, \boldsymbol{\varphi})g(x_t, \boldsymbol{\beta}). \quad (6.12)$$

5. A função de densidade preditiva é dada por

$$p(y_t|\mathbf{Y}_{t-1}, \boldsymbol{\varphi}) = \frac{\Gamma(b(y_t, \boldsymbol{\varphi}) + a_{t|t-1}^*)a(y_t, \boldsymbol{\varphi})(b_{t|t-1}^*)^{a_{t|t-1}^*}}{\Gamma(a_{t|t-1}^*)[c(y_t, \boldsymbol{\varphi}) + b_{t|t-1}^*]^{b(y_t, \boldsymbol{\varphi}) + a_{t|t-1}^*}}, \quad y_t \in H(\boldsymbol{\varphi}). \quad (6.13)$$

A prova do Teorema 1 pode ser vista em Santos (2009). Fica fácil ver a partir de (6.9)- (6.10) que  $Var(\lambda_t|\mathbf{Y}_{t-1}) = w^{-1}a_{t-1}/b_{t-1}^2 = w^{-1}Var(\lambda_{t-1}|\mathbf{Y}_{t-1})$ . Pensando-se em termos da precisão (inverso da variância) como medida de informação, tem-se que a passagem de  $t-1 \rightarrow t$  implica que apenas 100w% da informação é preservada. É exatamente esse o uso de fatores de desconto em West & Harrison (1997). Esses fatores medem a quantidade de informação (medida pela precisão do sistema) preservada com o passar do tempo.

### 6.2.2 Procedimento de inferência

A função de log-verossimilhança é dada por

$$\ln L(\boldsymbol{\varphi}; \mathbf{Y}_n) = \ln \prod_{t=1}^n p(y_t|\mathbf{Y}_{t-1}, \boldsymbol{\varphi}) = \sum_{t=1}^n \ln \Gamma(a_{t|t-1}^* + b(y_t, \boldsymbol{\varphi})) - \ln \Gamma(a_{t|t-1}^*) + a_{t|t-1}^* \ln b_{t|t-1}^* + \ln(a(y_t, \boldsymbol{\varphi})) - (b(y_t, \boldsymbol{\varphi}) + a_{t|t-1}^*) \ln(c(y_t, \boldsymbol{\varphi}) + b_{t|t-1}^*), \quad (6.14)$$

onde  $\boldsymbol{\varphi}$  é composto por  $w$ ,  $\beta$  e por parâmetros específicos do modelo e  $\mathbf{Y}_n = (y_1, \dots, y_n)'$ . Na equação (6.14), às vezes é considerada a soma apenas a partir do tempo  $\tau$ , onde  $\tau$  é o instante da primeira observação diferente de zero, se a distribuição *a priori* Gama for não-informativa, isto é,  $a_0, b_0 \rightarrow 0$ . Observa-se que

$$\begin{aligned} p(\mu_1|\mathbf{Y}_1) &\propto p(y_1|\mu_1)p(\mu_1|Y_0) \\ &\propto \exp\{-\mu_1[c(y_1, \boldsymbol{\varphi}) + wb_0/g(x_1, \boldsymbol{\beta})]\} \mu_1^{b(y_1, \boldsymbol{\varphi}) + wa_0}. \end{aligned}$$

Se  $a_0, b_0 \rightarrow 0$ , então  $p(\mu_1|\mathbf{Y}_1) \rightarrow Gama((b(y_1, \boldsymbol{\varphi}), c(y_1, \boldsymbol{\varphi}))$ . Portanto, se  $b(y_1, \boldsymbol{\varphi}) = 0$  a distribuição *a posteriori* pode ser imprópria. Logo, as funções de densidade preditivas não estão definidas e, consequentemente, a função de log-verossimilhança também pode não estar definida até a primeira observação tal que  $b(y_t, \boldsymbol{\varphi}) \neq 0$ . Uma distribuição própria para  $\mu_t$  pode ser obtida no tempo  $t = \tau$ , em que  $\tau$  é o índice da primeira observação tal que

$b(y_t, \varphi)$  seja diferente de zero. Ressalta-se que se  $a_0 > 0$  e  $b_0 > 0$ , não é necessária a utilização de  $\tau$ .

O intervalo de confiança assintótico para  $\varphi$  é construído baseado em uma aproximação numérica para a matriz de informação de Fisher obtida do próprio método de maximização numérica BFGS, utilizado para maximizar a função de log-verossimilhança com respeito a  $\varphi$ .

Já para se fazer inferência Bayesiana sobre os parâmetros do modelo, é usado o método MCMC do algoritmo de Metropolis-Hastings (Gamerman & Lopes, 2006) a fim de se obter uma amostra da distribuição *a posteriori* do vetor de parâmetros  $\varphi$ , cuja forma é dada por:

$$\pi(\varphi | \mathbf{Y}_n) \propto L(\varphi; \mathbf{Y}_n) \pi(\varphi), \quad (6.15)$$

onde  $L(\varphi; \mathbf{Y}_n)$  é a função de verossimilhança obtida em (6.14) e  $\pi(\varphi)$  é a distribuição *a priori* para  $\varphi$ . Neste trabalho, uma distribuição *a priori* uniforme, dada por  $\pi(\varphi) \propto c$  para todos os valores possíveis de  $\varphi$  e 0, caso contrário, é usada.

Intervalos de credibilidade para  $\varphi_i$ ,  $i = 1, \dots, p$  são construídos como se segue. Dado um valor  $0 < \kappa < 1$ , todo intervalo  $(t_1, t_2)'$  satisfazendo

$$\int_{t_1}^{t_2} \pi(\varphi_i | \mathbf{Y}_n) d\varphi_i = 1 - \kappa$$

é um intervalo de credibilidade para  $\varphi_i$  com nível  $100(1 - \kappa)\%$ .

## 6.3 Casos particulares da FGMD

Nas próximas subseções, serão discutidos alguns casos particulares da família Gama de modelos dinâmicos tais como os modelos Poisson, Gama, Weibull, Beta e Normal com média conhecida.

### 6.3.1 Modelo Poisson

Suponha que uma observação no tempo  $t$  é retirada de uma distribuição de Poisson com média  $\mu_t$  e função de probabilidade

$$p(y_t | \mu_t, \varphi) = \mu_t^{y_t} \exp(-\mu_t) / y_t! \quad (6.16)$$

onde  $y_t = 0, 1, \dots$  e  $\mu_t = \lambda_t g(x_t, \beta)$ . Esse modelo pertence à família Gama de modelos dinâmicos em que  $a(y_t, \varphi) = (y_t!)^{-1}$ ,  $b(y_t, \varphi) = y_t$  e  $c(y_t, \varphi) = 1$ . Logo,  $\varphi = (w, \beta)'$ .

A distribuição *a priori* é a mesma do Teorema 1. Com as funções  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$  identificadas, utilizando o Teorema 1, a distribuição *a posteriori* de  $\mu_t | \mathbf{Y}_t$  é dada pela distribuição Gama com parâmetros

$$a_t^* = a_{t|t-1}^* + y_t,$$

$$b_t^* = b_{t|t-1}^* + 1.$$

Logo, segue-se que  $\lambda_t = \mu_t [g(x_t, \beta)]^{-1} | \mathbf{Y}_t$  tem também distribuição Gama com parâmetros

$$a_t = w a_{t-1} + y_t,$$

$$b_t = w b_{t-1} + g(x_t, \beta).$$

Substituindo as funções  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$ ,  $c(\cdot, \cdot)$  e usando o Teorema 1, obtém-se a distribuição preditiva, que é Binomial negativa, cuja função de probabilidade é dada por

$$p(y_t | \mathbf{Y}_{t-1}, \varphi) = \binom{a_{t|t-1}^* + y_t + 1}{y_t} (b_{t|t-1}^*)^{a_{t|t-1}^*} (1 + b_{t|t-1}^*)^{-(a_{t|t-1}^* + y_t)},$$

para  $y_t = 0, 1, 2, \dots$  e

$$\binom{a_{t|t-1}^* + y_t + 1}{y_t} = \frac{\Gamma(a_{t|t-1}^* + y_t + 1)}{\Gamma(y_t + 1) \Gamma(a_{t|t-1}^*)}.$$

A função de log-verossimilhança é dada por

$$\begin{aligned} \ln L(\varphi; \mathbf{Y}_n) = & \sum_{t=1}^n \ln \Gamma(a_{t|t-1}^* + y_t) - \ln y_t! - \ln \Gamma(a_{t|t-1}^*) + \\ & a_{t|t-1}^* \ln b_{t|t-1}^* - (a_{t|t-1}^* + y_t) \ln(1 + b_{t|t-1}^*), \end{aligned} \quad (6.17)$$

onde  $\varphi = (w, \beta)'$ .

Então, segue das propriedades da distribuição binomial negativa, que a média e variância da distribuição preditiva de  $y_{n+1} | \mathbf{Y}_n, \varphi$  são, respectivamente,

$$\tilde{y}_{n+1|n} = E(y_{n+1} | \mathbf{Y}_n, \varphi) = a_{n+1|n}^* / b_{n+1|n}^*$$

e

$$Var(y_{n+1} | \mathbf{Y}_n, \varphi) = a_{n+1|n}^* (1 + b_{n+1|n}^*) / (b_{n+1|n}^*)^2.$$

### 6.3.2 Modelo Gama

Suponha que a série temporal  $\{y_t\}$  é gerada de uma distribuição Gama com o parâmetro de forma  $\chi$  desconhecido e o parâmetro de escala  $\chi\mu_t$ . Logo,

$$p(y_t|\mu_t, \boldsymbol{\varphi}) = \frac{y_t^{\chi-1} \exp(-\mu_t \chi y_t)}{\Gamma(\chi)(\mu_t \chi)^{-\chi}}, \quad (6.18)$$

onde  $y_t > 0$ ,  $\mu_t = \lambda_t g(x_t, \boldsymbol{\beta})$  e  $y_t > 0; \forall t \leq n$ . O valor esperado de  $(y_t|\mu_t, \boldsymbol{\varphi})$  é  $1/\mu_t$ . Se  $\chi = 1$ ,  $(y_t|\mu_t, \boldsymbol{\varphi})$  tem distribuição exponencial com média  $1/\mu_t$ .

O modelo Gama pode ser escrito na forma da família Gama de modelos dinâmicos em que  $a(y_t, \boldsymbol{\varphi}) = y_t^{\chi-1} \chi^\chi / \Gamma(\chi)$ ,  $b(y_t, \boldsymbol{\varphi}) = \chi$  e  $c(y_t, \boldsymbol{\varphi}) = \chi y_t$ .

Pelo Teorema 1, dado a  $t$ -ésima observação e as funções  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$  devidamente discriminadas, a distribuição *a posteriori* de  $\mu_t|\mathbf{Y}_t$  é Gama com parâmetros

$$\begin{aligned} a_t^* &= a_{t|t-1}^* + \chi \\ b_t^* &= b_{t|t-1}^* + \chi y_t. \end{aligned}$$

Logo, tem-se que  $\lambda_t = \mu_t[g(x_t, \boldsymbol{\beta})]^{-1}|\mathbf{Y}_t \sim \text{Gama}(a_t, b_t)$  e as equações de atualização são dadas por:

$$\begin{aligned} a_t &= a_{t|t-1} + \chi \\ b_t &= b_{t|t-1} + \chi y_t g(x_t, \boldsymbol{\beta}). \end{aligned}$$

Substituindo as funções  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$  em (6.13), a distribuição preditiva  $y_t|\mathbf{Y}_{t-1}, \boldsymbol{\varphi}$  é chamada de Gama-Gama com parâmetros  $a_{t|t-1}^*, b_{t|t-1}^*/\chi$  e  $\chi$  e denotada  $G - G(a_{t|t-1}^*, b_{t|t-1}^*/\chi, \chi)$ . Sua densidade é dada por

$$p(y_t|Y_{t-1}, \boldsymbol{\varphi}) = \frac{\Gamma(\chi + a_{t|t-1}^*) y_t^{\chi-1}}{\Gamma(a_{t|t-1}^*) \Gamma(\chi) (b_{t|t-1}^*/\chi)^{-a_{t|t-1}^*} (y_t + b_{t|t-1}^*/\chi)^{\chi + a_{t|t-1}^*}},$$

se  $y_t > 0$  e 0, caso contrário.

A função de verossimilhança é o produto das funções de densidade preditivas dada por:

$$\begin{aligned} \ln L(\boldsymbol{\varphi}; \mathbf{Y}_n) &= \ln \prod_{t=1}^n p(y_t|\mathbf{Y}_{t-1}, \boldsymbol{\varphi}) = \sum_{t=1}^n \ln p(y_t|\mathbf{Y}_{t-1}, \boldsymbol{\varphi}) = \ln \Gamma(\chi + a_{t|t-1}^*) - \\ &\ln \Gamma(\chi) \Gamma(a_{t|t-1}^*) + a_{t|t-1}^* \ln(b_{t|t-1}^*/\chi) + \ln y_t^{(\chi-1)} - (\chi + a_{t|t-1}^*) \ln(y_t + b_{t|t-1}^*/\chi), \end{aligned}$$

onde  $\boldsymbol{\varphi} = (w, \beta, \chi)'$ .

Das propriedades da distribuição Gama-Gama, a média e a variância de  $y_{n+1}$  condicionada à informação até o tempo  $n$  são

$$\tilde{y}_{n+1|n} = E(y_{n+1}|\mathbf{Y}_n, \boldsymbol{\varphi}) = \frac{b_{n+1|n}^*}{a_{n+1|n}^* - 1}$$

e

$$Var(y_{n+1}|\mathbf{Y}_n, \boldsymbol{\varphi}) = \frac{(b_{n+1|n}^*)^2[\chi^2 + \chi(a_{n+1|n}^* - 1)]}{(a_{n+1|n}^* - 1)^2(a_{n+1|n}^* - 2)}.$$

### 6.3.3 Modelo Weibull

Se as observações no tempo  $t$  são geradas de uma distribuição Weibull tri-paramétrica (Ross, 2002) e os parâmetros  $\zeta_t = \zeta$  e  $\rho_t = \rho$  são invariantes no tempo e desconhecidos, então:

$$p(y_t|\mu_t, \boldsymbol{\varphi}) = \zeta\mu_t(y_t - \rho)^{\zeta-1} \exp[-\mu_t(y_t - \rho)^\zeta], \quad (6.19)$$

onde  $y_t > \rho$ ,  $\mu_t$ ,  $\zeta > 0$  e  $\mu_t = \zeta_t g(x_t, \boldsymbol{\beta})$ .

O modelo Weibull pode ser escrito na forma da família Gama de modelos dinâmicos onde  $a(y_t, \boldsymbol{\varphi}) = \zeta(y_t - \rho)^{\zeta-1}$ ,  $b(y_t, \boldsymbol{\varphi}) = 1$  e  $c(y_t, \boldsymbol{\varphi}) = (\mathbf{y}_t - \rho)^\zeta$ .

Pelo Teorema 1, a distribuição *a posteriori* de  $\mu_t|\mathbf{Y}_t$  é Gama com parâmetros

$$\begin{aligned} a_t^* &= a_{t|t-1}^* + 1, \\ b_t^* &= b_{t|t-1}^* + (y_t - \rho)^\zeta. \end{aligned}$$

Então, segundo o Teorema 1, tem-se que  $\zeta_t = \mu_t[g(x_t, \boldsymbol{\beta})]^{-1}|\mathbf{Y}_t \sim Gama(a_t, b_t)$  e as equações de atualização são dadas por:

$$a_t = a_{t|t-1} + 1, \quad (6.20)$$

$$b_t = b_{t|t-1} + (y_t - \rho)^\zeta g(x_t, \boldsymbol{\beta}). \quad (6.21)$$

Até o presente momento, não foi feita nenhuma consideração para dados que apresentam censura, os quais são muito comuns em análise de sobrevivência. Assim, se existe censura, pode-se assumir que  $y_t$  é observada se  $\delta_t = 1$  ou censurada à direita se  $\delta_t = 0$ . Desta forma, via teorema de Bayes, a equação de atualização  $a_t$  passa a ser  $a_t = a_{t|t-1} + \delta_t$ , onde  $\delta_t$  é um indicador de censura à direita.

Conhecendo as funções  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$  e utilizando o Teorema 1, a função de densidade preditiva de  $y_t | \mathbf{Y}_{t-1}, \varphi$  é dada por:

$$\frac{\Gamma(1 + a_{t|t-1}^*) \zeta(y_t - \rho)^{\zeta-1}}{\Gamma(a_{t|t-1}^*) (b_{t|t-1}^*)^{-a_{t|t-1}^*} [(y_t - \rho)^\zeta + b_{t|t-1}^*]^{1+a_{t|t-1}^*}},$$

onde  $y_t > \rho$  e  $a_{t|t-1}^*$  e  $b_{t|t-1}^*$  são dadas pelo Teorema 1.

A função de verossimilhança, que é o produto das funções de densidade preditivas, é dada por:

$$\ln L(\varphi; \mathbf{Y}_n) = \ln \prod_{t=1}^n p(y_t | \mathbf{Y}_{t-1}, \varphi) = \sum_{t=1}^n \ln \Gamma(1 + a_{t|t-1}^*) + \ln \zeta(y_t + \kappa)^{\zeta-1} - \ln \Gamma(a_{t|t-1}^*) + a_{t|t-1}^* \ln b_{t|t-1}^* - (1 + a_{t|t-1}^*) \ln[(y_t - \kappa)^\zeta + b_{t|t-1}^*],$$

em que  $\varphi = (w, \beta, \zeta, \rho)'$ .

### 6.3.4 Modelo Pareto

A distribuição de Pareto é uma distribuição que possui muitas aplicações em problemas econômicos, sociais e geofísicos. Se as observações no tempo  $t$  são geradas de uma distribuição Pareto com parâmetros  $\rho > 0$ , desconhecido e invariante no tempo e  $\mu_t > 0$ , então sua densidade é

$$p(y_t | \mu_t, \varphi) = \mu_t \rho^{\mu_t} y_t^{-\mu_t-1}, \quad (6.22)$$

onde  $y_t > \rho$  e  $\mu_t = \lambda_t g(x_t, \beta)$ .

O modelo Pareto pode ser escrito na forma da família Gama de modelos dinâmicos em que  $a(y_t, \varphi) = y_t^{-1}$ ,  $b(y_t, \varphi) = 1$  e  $c(y_t, \varphi) = \ln y_t - \ln \rho$ .

As distribuições *a priori* são as mesmas da família Gama de modelos dinâmicos. Quando a  $t$ -ésima observação é obtida e com as funções  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$  a distribuição *a posteriori* de  $\mu_t | Y_t$ , pelo Teorema 1, é Gama com parâmetros

$$\begin{aligned} a_t^* &= a_{t|t-1}^* + 1, \\ b_t^* &= b_{t|t-1}^* - \ln \rho + \ln y_t. \end{aligned}$$

Logo, fazendo a transformação inversa, tem-se que  $\lambda_t = \mu_t [g(x_t, \beta)]^{-1} | \mathbf{Y}_{t-1} \sim \text{Gama}(a_t, b_t)$  onde as equações de atualização são dadas por:

$$\begin{aligned} a_t &= a_{t|t-1} + 1, \\ b_t &= b_{t|t-1} + (\ln y_t - \ln \rho) g(x_t, \beta). \end{aligned}$$

Pelo Teorema 1, de posse das funções  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$  e  $c(\cdot, \cdot)$ , a função de densidade preditiva de  $y_t|Y_{t-1}$ ,  $\varphi$  é dada por

$$\frac{a_{t|t-1}^* y_t^{-1} [b_{t|t-1}^*]^{a_{t|t-1}^*}}{[b_{t|t-1}^* - \ln \rho + \ln y_t]^{a_{t|t-1}^* + 1}},$$

se  $y_t > \rho$  e 0, caso contrário.

Com as funções de densidade preditivas, é possível determinar a função de verossimilhança, que é dada por:

$$\begin{aligned} \ln L(\varphi; \mathbf{Y}_n) &= \ln \prod_{t=1}^n p(y_t|Y_{t-1}, \varphi) = \sum_{t=1}^n \ln p(y_t|Y_{t-1}, \varphi) = \\ &= \sum_{t=1}^n \ln(a_{t|t-1}^* y_t^{-1} (b_{t|t-1}^*)^{a_{t|t-1}^*}) - (a_{t|t-1}^* + 1) \ln[b_{t|t-1}^* - \ln \rho + \ln y_t], \end{aligned}$$

onde  $\varphi = (w, \beta, \rho)'$ .

### 6.3.5 Modelo Beta

Quando um parâmetro da distribuição Beta é igual a 1, pode-se escrever esse modelo na forma de FGMD. Esse modelo pode ser útil para modelar séries temporais que são proporções e probabilidades concentradas em torno de 1. Suponha que a série temporal  $\{y_t\}$  é gerada de uma distribuição Beta com parâmetros  $\mu_t$  e 1 cuja função de densidade é dada por

$$p(y_t|\mu_t, \varphi) = \mu_t y_t^{\mu_t-1}, \quad (6.23)$$

onde  $0 < y_t < 1$ .

Esse modelo também pertence à FGMD no qual  $a(y_t, \varphi) = y_t^{-1}$ ,  $b(y_t, \varphi) = 1$  e  $c(y_t, \varphi) = -\ln(y_t)$ . As equações de atualização e a função de densidade preditiva podem ser encontradas similarmente aos modelos anteriores, usando o Teorema 1. Neste caso,  $\varphi = (w, \beta)'$ .

### 6.3.6 Modelo Normal com média conhecida

A metodologia de FGMD pode ser aplicada também em modelos Gaussianos. Se as observações no tempo  $t$  são geradas de uma distribuição Normal com média  $d_t$  e parâmetro de precisão  $\mu_t > 0$ , então:

$$p(y_t|\mu_t, \varphi) = \frac{\mu_t^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{\mu_t(y_t - d_t)^2}{2}\right), \quad (6.24)$$



onde  $-\infty < y_t < \infty$  e  $\mu_t = \lambda_t g(x_t, \beta)$ .

O modelo Normal pode ser escrito na forma da família Gama de modelos dinâmicos em que  $a(y_t, \varphi) = (2\pi)^{-1/2}$ ,  $b(y_t, \varphi) = 1/2$  e  $c(y_t, \varphi) = (y_t - d_t)^2/2$ .

A distribuição *a priori* de  $\lambda_t | \mathbf{Y}_{t-1}$  é dada pelo item 2 do Teorema 1. Segundo o Teorema 1, a distribuição *a posteriori* de  $\mu_t | \mathbf{Y}_t$  também pode ser obtida, sendo Gama com parâmetros

$$\begin{aligned} a_t^* &= a_{t|t-1}^* + 1/2, \\ b_t^* &= b_{t|t-1}^* + (y_t - d_t)^2/2. \end{aligned}$$

Logo, usando novamente a propriedade de escala da distribuição Gama, tem-se que  $\lambda_t = \mu_t [g(x_t, \beta)]^{-1} | \mathbf{Y}_t \sim \text{Gama}(a_t, b_t)$  onde as equações de atualização são dadas por:

$$\begin{aligned} a_t &= a_{t|t-1} + 1/2, \\ b_t &= b_{t|t-1} + [(y_t - d_t)^2/2]g(x_t, \beta). \end{aligned}$$

A função de densidade preditiva  $y_t | \mathbf{Y}_{t-1}, \varphi$ , pelo Teorema 1, tem a seguinte forma:

$$\frac{\Gamma(a_{t|t-1}^* + 1/2)(2\pi)^{-1/2}(b_{t|t-1}^*)^{a_{t|t-1}^*}}{\Gamma(a_{t|t-1}^*)[(y_t - d_t)^2/2 + b_{t|t-1}^*]^{a_{t|t-1}^* + 1/2}},$$

onde  $-\infty < y_t < \infty$ .

Com as funções de densidade preditivas, é possível determinar a função de verossimilhança, que é dada por:

$$\begin{aligned} \ln L(\varphi; \mathbf{Y}_n) &= \ln \prod_{t=1}^n p(y_t | Y_{t-1}, \varphi) = \sum_{t=1}^n \ln p(y_t | Y_{t-1}, \varphi) = \\ &= \sum_{t=1}^n \ln \Gamma(a_{t|t-1}^* + 1/2)(2\pi)^{-1/2}(b_{t|t-1}^*)^{a_{t|t-1}^*} \\ &\quad - \ln \Gamma(a_{t|t-1}^*)[(y_t - d_t)^2/2 + b_{t|t-1}^*]^{a_{t|t-1}^* + 1/2}, \end{aligned}$$

em que  $\varphi = (w, \beta)'$ .

### 6.3.7 Transformações de modelos

Da mesma forma que o modelo Normal com média conhecida e evolução na variância foi construído, os modelos Log-normal e Normal Inversa podem ser desenvolvidos. Eles nada mais são que transformações apropriadas nas séries. O modelo baseado na distribuição do valor extremo também pode ser construído, uma vez que essa distribuição pode ser obtida a partir de uma transformação um-a-um de uma variável com distribuição exponencial (Smith & Miller, 1986).

Tabela 6.1: EMV e EB para o modelo Poisson.

	EMV	EB-Mediana	EB-Media	EB-Moda
$w = 0,90$	0,912 (0,003)	0,889 (0,003)	0,883 (0,003)	0,910 (0,004)
$\beta = 1,00$	0,990 (0,068)	1,038 (0,148)	1,039 (0,150)	1,005 (0,017)

## 6.4 Estudos de simulação

Nesta seção, simulações Monte Carlo são realizadas para o modelo de Poisson. O EMV e os estimadores Bayesianos são comparados quanto ao vício e ao EQM, assim como os intervalos de credibilidade e de confiança são comparados com respeito à amplitude e taxa de cobertura.

Os desempenhos do estimador de máxima verossimilhança (EMV) e dos estimadores Bayesianos - EB-média, EB-mediana e EB-moda - foram investigados para séries temporais de tamanho  $n = 100$ , geradas sob o modelo Poisson com uma covariável  $x_t = \cos(2\pi t/12)$ , para  $t = 1, \dots, n$  e parâmetros  $w = 0,90$  e  $\beta = 1$ . Duas cadeias de 2000 amostras foram geradas das quais as 1000 primeiras foram excluídas. O número de replicações Monte Carlo foi fixado em 500. O nível de confiança e a probabilidade de credibilidade dos intervalos de confiança e de credibilidade, respectivamente, foram fixados em 0,95.

O vetor  $\varphi$  de parâmetros do modelo Poisson é estimado maximizando a função de log-verossimilhança em (6.17), utilizando o método de otimização BFGS. Uma estimativa para a matriz de covariâncias de  $\hat{\varphi}$  é obtida através do método BFGS (ver Seção 3.1). Já os intervalos de credibilidade para os parâmetros são calculados como na Seção 4.1.3.

A Figura 6.1 apresenta uma série simulada sob o modelo Poisson com os mesmos valores dos parâmetros descritos anteriormente. Observa-se que a série oscila em torno de um nível médio igual a 3, aproximadamente. Nota-se também que a linha tracejada referente à média da distribuição preditiva do modelo Poisson acompanha bem o comportamento da série.

Na Tabela 6.1, o EMV e os estimadores Bayesianos são comparados quanto ao vício e ao EQM. Para  $w$ , todos os estimadores possuem valores de vício e EQM muito próximos. Já para  $\beta$ , o EMV e o EB-moda apresentam valores do EQM menores que os outros métodos.

Na Tabela 6.2, os intervalos de confiança e de credibilidade são compara-

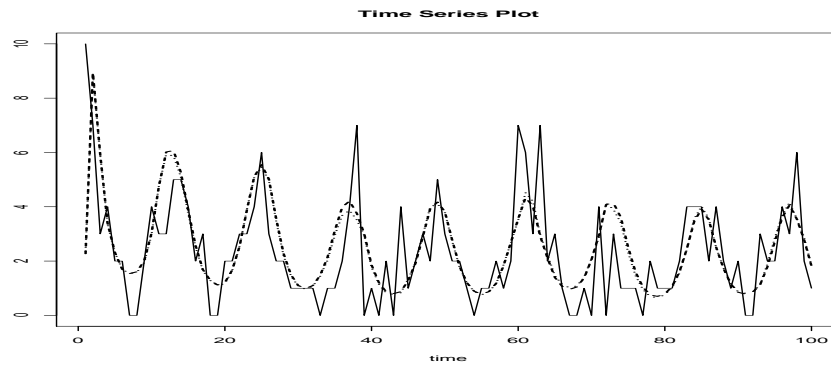


Figura 6.1: Série temporal simulada sob o modelo Poisson. A linha contínua representa a série temporal e as linhas tracejada e pontilhada as médias da distribuição preditiva dos ajustes clássico e Bayesiano, respectivamente.

Tabela 6.2: Intervalos de confiança e de credibilidade para o modelo Poisson.

		<b>Int. Cred.</b>	<b>Int. Assint.</b>
$w = 0,90$	limites	[0,771; 0,962]	[0,816; 1,013]
	amplitude	0,191	0,197
	cobertura	(0,962)	(0,922)
$\beta = 1,00$	limites	[0,787; 1,308]	[0,812; 1,207]
	amplitude	0,521	0,395
	cobertura	(0,956)	(0,944)

dos através da taxa de cobertura e amplitude. Para  $w$ , a taxa de cobertura do intervalo de credibilidade está mais próxima do nível nominal 0,95 assumido que o intervalo assintótico, e sua amplitude é ligeiramente menor. Com relação ao parâmetro  $\beta$ , ambos os intervalos têm taxas de cobertura próxima de 0,95, mas a amplitude do intervalo de confiança é bem menor. Nota-se que o limite superior do intervalo de confiança do parâmetro  $w$  ultrapassou o limite do espaço paramétrico.

## 6.5 Aplicação a séries reais

Nesta seção, o modelo Poisson é ajustado a dados mensais do número médio de pacientes internados com afecção das vias aéreas superiores e inferiores - AVASI - na cidade de São Paulo, no período de janeiro de 1997 a dezembro 2000. As covariáveis consideradas são a série contendo níveis de dióxido de enxofre,  $x_{1t} = SO_2$ , e uma série que introduz sazonalidade determinística no modelo,  $x_{2t} = \text{sen}(2\pi t/12)$ , onde  $t = 1, \dots, 48$ .

A Figura 6.2 mostra os gráficos dessas séries, onde pode-se perceber a presença de sazonalidade. A série AVASI apresenta valores mais altos nos meses de março, abril e maio de 2000, enquanto os níveis de concentração do poluente  $SO_2$  tem valor médio mais alto em 1997. Pelo gráfico de dispersão (Figura 6.3), parece existir uma relação entre essas séries, com indícios de aumento dos valores da variável AVASI, à medida que níveis de  $SO_2$  aumentam.

O modelo ajustado à série foi  $y_t \sim \text{Poisson}(\mu_t)$ , onde  $\mu_t = \lambda_t \exp(\beta_1 x_{1t} + \beta_2 x_{2t})$  e  $\lambda_t$  é o parâmetro responsável pela descrição do nível dinâmico.

A Figura 6.4 mostra o gráfico da série AVASI e os ajustes do modelo Poisson à mesma. Observa-se que as linhas pontilhadas e tracejadas têm um comportamento parecido e acompanham o comportamento da série.

Na Tabela 6.3 estão o EMV e as estimativas Bayesianas (EB-Mediana, EB-Média e EB-Moda) dos parâmetros  $w$ ,  $\beta_1$  e  $\beta_2$  do modelo Poisson. Os valores referentes ao critério de convergência de Gelman e Rubin (Gelman, 1996) para as 2 cadeias geradas de cada parâmetro estão próximos da unidade, indicando a convergência das mesmas. Pode-se verificar que as estimativas são bem similares em todos os métodos. Nota-se que, como esperado, o valor de  $\beta_1$  é positivo, evidenciando que o aumento em  $SO_2$  leva a um aumento em AVASI.

A Tabela 6.4 apresenta os intervalos de confiança assintótico e de cre-

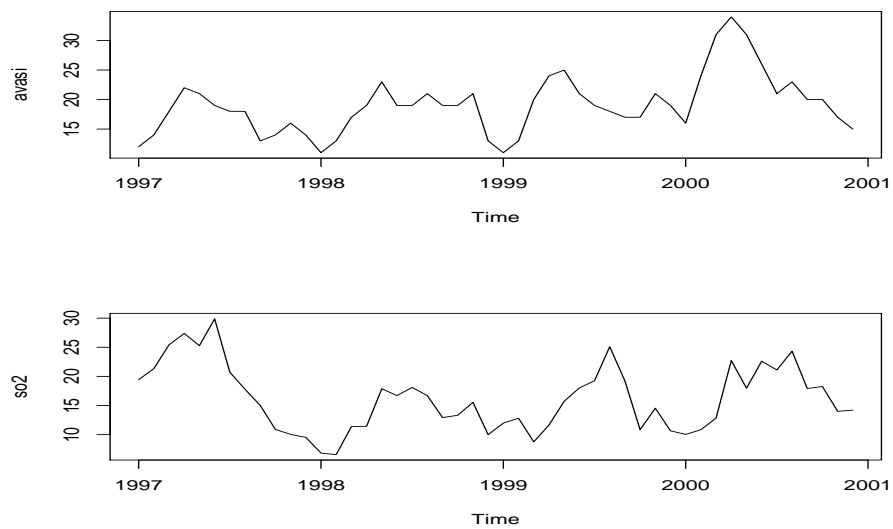


Figura 6.2: Gráficos das séries AVASI e SO2 nos anos de 1997 a 2000.

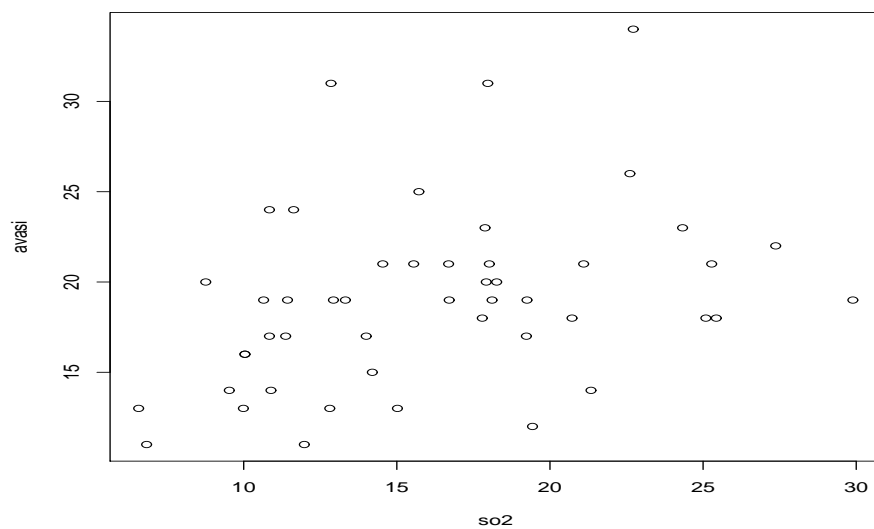


Figura 6.3: Gráfico de dispersão das séries AVASI e SO2.

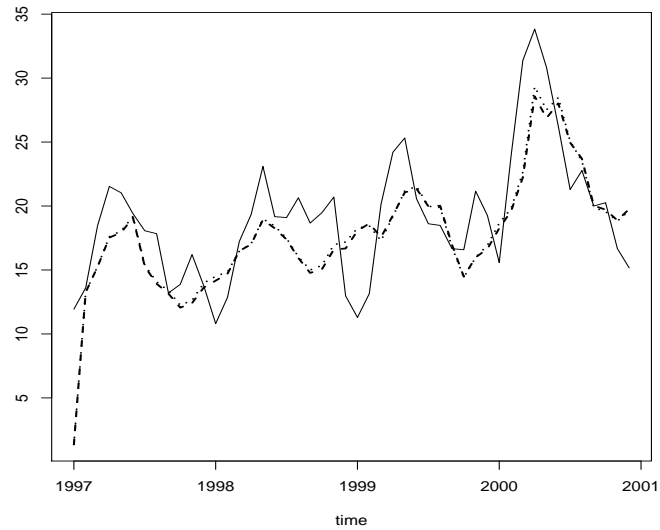


Figura 6.4: A linha contínua indica a série AVASI,  $y_t$ , e as linhas tracejada e pontilhada indicam as médias da distribuição preditiva do ajuste do modelo Poisson sob os enfoques clássico e Bayesiano, respectivamente.

Tabela 6.3: EMV e EB para os parâmetros do modelo Poisson ajustado à série AVASI.

	<b>EMV</b>	<b>EB-Mediana</b>	<b>EB-Media</b>	<b>EB-Moda</b>
$w$	0,842	0,806	0,796	0,836
$\beta_1$	0,016	0,015	0,016	0,017
$\beta_2$	0,161	0,160	0,162	0,160

Tabela 6.4: Intervalos de confiança e de credibilidade de 0,90 para os parâmetros do modelo Poisson ajustado à série AVASI.

	<b>Int. Assint.</b>	<b>Int. Cred.</b>
$w$	[0,820; 1,179]	[0,647; 0,905]
$\beta_1$	[-0,026; 0,059]	[0,003; 0,028]
$\beta_2$	[-0,295; 0,616]	[0,075; 0,258]

dibilidade para os parâmetros. Conclui-se que os parâmetros  $\beta_1$  e  $\beta_2$  são significativos ao nível de credibilidade de 0,90. A mesma conclusão não pode ser tomada, entretanto, analisando-se os intervalos assintóticos, já que o valor zero está contido em ambos os intervalos para  $\beta_1$  e  $\beta_2$ . O limite superior do intervalo assintótico para  $w$  ultrapassou o valor 1 (limite do espaço paramétrico). Vale ressaltar que se trata de um intervalo aproximado e que no seu cálculo utiliza-se uma aproximação para a matriz informação de Fisher, obtida através do método de otimização numérica BFGS.

Os resíduos do ajuste do modelo não são autocorrelacionados, evidenciando que o modelo conseguiu captar bem a estrutura de autocorrelação entre as observações da série.

# Referências Bibliográficas

- Alves, M., Gamerman, D. & Ferreira, M. A. R. (2009). Transfer functions in dynamic generalized linear models. A ser publicado em *Statistical Modeling*.
- Avriel, M. (2003). *Nonlinear Programming: Analysis and Methods*. Dover Publishing.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3a. ed. New Jersey: Prentice-Hall.
- Box, G.E.P. & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, **70**, 70-79.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden Day.
- Brockwell, P. J. & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*, Springer text in Statistics. New York: Springer.
- Broyden, C. G. (1970). The Convergence of a class of double-rank minimization algorithms, *Journal of the Institute of Mathematics and Its Applications*, **6**, 76-90.
- Campbell, D. T. & Stanley, J. C. (1968). The Connecticut crackdown on speeding: Time series data in quasi-experimental analysis. *Law and Society Review*, **3**, 33-53.
- Carvalho, C., Johannes, M., Lopes, H.F. & Polson, N. (2009). Particle learning and smoothing. *Technical Report*. University of Chicago Booth School of Business.



- Casella, G. & Berger, R. L. (2002). *Statistical Inference*, Duxbury: Pacific Grove.
- Cavaliere, G. & Taylor, R. A. M. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, **24**, 1, 43-71.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, **49**, 327-335.
- Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, **25**, 573-578.
- Chiogna, M. & Gaetan, C. (2002). Dynamic generalized linear models with application to environmental epidemiology. *Applied Statistics*, **51**, 453-468.
- Cox, D. R. (1981). Statistical analysis of time-series: some recent developments. *Scandinavian Journal of Statistics*, **8**, 93-115.
- Cramer, J.S. (1986). *Econometric Applications of Maximum Likelihood Methods*. Cambridge: Cambridge University Press.
- Davidson, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Applications*. Cambridge: Cambridge University Press.
- Davidson, A. C. & MacKinnon, R. (1997). Bootstrap testing in nonlinear models. *GREQAM Document de Travail* No. 97A05.
- de Jong, P. & Penzer, J. (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association*, **93**, 796-806.
- Dettea, H. & Weißbach, R. (2009). A bootstrap test for the comparison of nonlinear time series. *Computational Statistics & Data Analysis*, **53**, 4, 1339-1349.
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models.*, 2a. ed., New York: Chapman and Hall.
- Doornik, J. A. (1999). *Ox: An Object-Oriented Matrix Language*, 3a. ed. London: Timberlake Consultants Press.

- Durbin, J. & Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (com discussão). *Journal of the Royal Statistical Society, Series B*, **62**, 3-56.
- Durbin, J. & Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**, 1-26.
- Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.
- Fahrmeir, L. (1987). Regression models for nonstationary categorical time series. *Journal of Time Series Analysis*, **8**, 147-160.
- Fahrmeir, L. & Wagenpfeil, S. (1997). Penalized likelihood estimation and iterative Kalman smoothing for non-gaussian dynamic regression models. *Computational Statistics and Data Analysis*, **24**, 295-320.
- Feng, Z. D. & McCulloch, C. E. (1996). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society, Series B*, **3**, 609-617.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *Computer Journal*, **13**, 317-322.
- Fletcher, R. (1987). *Practical Methods of Optimization*, 2nd ed. New York: John Wiley & Sons.
- Franco, G. C. (1998). *Bootstrap em Modelos Estruturais: Construção de Intervalos de Confiança e Testes de Hipóteses*. Tese de doutorado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica, Rio de Janeiro.
- Franco, G. C., Koopman, S. J. & Souza, R. C. (1999). Bootstrap tests when parameters of nonstationary time series lie on the boundary of the parameter space. *Brazilian Journal of Probability and Statistics*, **13**, 41-54.
- Franco, G. C. & Souza, R. C. (2002). A comparison of methods for bootstrapping in the local level model. *Journal of Forecasting*, **21**, 27-38.

- Franco, G. C., Santos, T. R., Ribeiro, J. A. & Cruz, F. R. B. (2008). Confidence intervals for hyperparameters in structural models. *Communications in Statistics: Simulation and Computation*, **37**(3), 486-497.
- Fruhwirth-Schnatter, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Statistical Computation*, **4**, 259-269.
- Fuchun, L. & Tkacz, G. (2006). A consistent bootstrap test for conditional density functions with time-series data. *Journal of Econometrics*, **133**, 863-886.
- Gamerman, D. & West, M. (1987). An application of dynamic survival models in unemployment studies. *Statistician*, **36**, 269-274.
- Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics*, **40**, 63-79.
- Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, **85**, 215-227.
- Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2a. ed. London: Chapman and Hall.
- Glass, G.V. (1972). Estimating the effects of intervention into a nonstationary time series. *American Educational Research Journal*, **9**(3), 463-477.
- Gelman, A. (1996). *Inference and monitoring convergence*, in *Markov Chain Monte Carlo in Practice*, W.R. Gilks *et al.*, eds., Chapman and Hall, London, pp. 131-140.
- Gilks, W.R., Richardson, S. & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.
- Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimization*, New York: Academic Press.
- Godolphin, E. J. & Triantafyllopoulos, K. (2006). Decomposition of time series models in state-space form. *Computational Statistics and Data Analysis*, **50**, 2232-2246.

- Goldfarb, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computation*, **24**, 23-26.
- Grunwald, G. K., Raftery, A. E. & Guttorp, P. (1993). Time series of continuous proportions. *Journal of the Royal Statistical Society, Series B*, **55**, 103-116.
- Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, **14**, 125-129.
- Harrison, P. J. & Stevens, C. F. (1971). A Bayesian approach to short-term forecasting. *Operational Research Quarterly*, **22**, 341-362.
- Harrison, P. J. & Stevens, C. F. (1976). Bayesian forecasting (com discussão). *Journal of the Royal Statistical Society, Series B*, **38**, 205-247.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: University Press.
- Harvey, A. C. & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, **7**, 407-417.
- Harvey, A. C. & Koopman, S. J. (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics*, **10**, 377-389.
- Harvey, A. C. & Streibel, M. (1998). Tests for deterministic versus indeterminate cycles. *Journal of Time Series Analysis*, **19**, 823-826.
- Harvey, A. C. (1993). *Time Series Models*. Maylands: Prentice Hall.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, **57**, pp. 97-109.
- Hemming, K. & Shaw, J. E. H. (2002). A parametric dynamic survival model applied to breast cancer survival times. *Applied Statistics*, **51**, 421-435.
- Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *ONR Research Memorandum 52*, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon Press.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.*, **82**, 35-45.
- Kaufmann, H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. *Annals of Statistics*, **15**, 79-98.
- Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series. *Journal of the American Statistical Association*, **82**, 1032-1063.
- Knorr-Held, L. (1999). Conditional prior proposal in dynamic models. *Scandinavian Journal of Statistics*, **26**, 129-144.
- Koopman, S. J. (1992). *Diagnostic Checking and Intra-Daily Effects in Time Series Models*. Amsterdam: Thesis Publishers.
- Koopman, S. J. & Shephard, N. (1992). Exact score for time series models in state space form. *Biometrika*, **79**, 823-826.
- Koopman, S. J., Shephard, N. & Doornik, J. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, **2**, 113-166.
- LaMotte, L. R. & McWhorter, A. Jr. (1978). An exact test for the presence of random walk coefficients in a linear regression model. *Journal of the American Statistical Association*, **73**, 816-820.
- Lindsey, J. K. & Lambert, P. (1995). Dynamic generalized linear models and repeated measurements. *Journal of Statistical Planning and Inference*, **47**, 129-139.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, **21**, 1087-1091.
- Migon, H. S. & Gamerman, D. (1999). *Statistical Inference: an Integrated Approach*. London: Arnold.
- Morettin, P. A. & Toloi, C. M. (2004). *Análise de Séries Temporais*. São Paulo: Editora E. Blucher-Associação Brasileira de Estatística.

- Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, **55**, 299-305.
- Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135, 370-384.
- Nervole, M. & Wage, S. (1964). On the optimality of adaptive forecasting. *Management Science*, **10**, 207-229.
- Nyblom J. & Makelainen, T. (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, **78**, 384, 856-864.
- Palm, F. C., Smeekes, S. & Urbain, J. (2008). Bootstrap unit-root tests: C comparison and extensions. *Journal of Time Series Analysis*, **29**, 2, 371-401.
- Paparoditis, D. & Politis, D. N. (2005). Tests for Autoregressive Time Series. *Journal of the American Statistical Association*, **100**, 470, 545-553.
- Penzer, J. (2007). State space models for time series with patches of unusual observations. *Journal of Time Series Analysis*, **28**, 629-645.
- Pereira, P. L. V. & Migon, H. S. (1986). *Modelagem Estrutural - Abordagens Bayesianas e Clássicas*. Santa Maria: UFSM.
- Pereira, P. L. V. & Migon, H. S. (1985). *Procedimentos Bayesianos e Clássicos em Modelos Estruturais*. Rio de Janeiro: ENCE.
- Pérez-Alonso, A. (2007). A bootstrap approach to test the conditional symmetry in time series models. *Computational Statistics & Data Analysis*, **51**, 7, 3484-3504.
- Petris, G. (2008). dlm: MLE and Bayesian analysis of Dynamic Linear Models. Aplicativo disponível em <http://cran.r-project.org/web/packages/dlm/index.html>
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2005). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7-11. Aplicativo disponível em <http://www-fis.iarc.fr/coda/>.

- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1988). *Numerical Recipes in C*. New York: Cambridge University Press.
- Ravines, R. E. R., Schmidt, A. M. & Migon, H. S. (2008). Revisiting distributed lag models through a Bayesian perspective. *Applied Stochastic Models in Business Industry*, **22**, 193-210.
- R Development Core Team. (2005). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.
- Reis, E. A., Salazar, E. & Gamerman, D. (2006). Comparison of sampling schemes for dynamic linear models. *International Statistical Review*, **74**, 203-214.
- Rodriguez, A. & Ruiz, E. (2009). Bootstrap prediction intervals in state space models. *Journal of Time Series Analysis*, **30**, 167-178.
- Ross, S. (2002). *A First Course in Probability*, 6th ed., Prentice Hall.
- Salvador, M. & Gargallo, P. (2004). Automatic monitoring and intervention in multivariate dynamic linear models. *Computational Statistics and Data Analysis*, **47**, 401-431.
- Santos, T. R. & Franco, G. C. (2009). Inference for hyperparameters in structural models under Bayesian and classical perspectives. Preprint.
- Santos, T.R., Franco, G.C. & Gamerman, D. (2009). Comparison of classical and Bayesian approaches for intervention analysis in structural models. *Relatório Técnico do LES - UFRJ*.
- Santos, T.R. (2009). *Inferência sobre os Hiperparâmetros em Modelos Estruturais sob as Perspectivas Clássica e Bayesiana*. Dissertação de mestrado, EST-UFMG, Belo Horizonte.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, **24**, 647-656.
- Shao, J. & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer-Verlag.
- Shephard, N. & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653-667.

- Singh, A. C. and Roberts, G. R. (1992). State space modelling of cross-classified time series of counts. *International Statistical Review*, **60**, 321-36.
- Smith, R. L., & Miller, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society, Series B*, **48**, 79-88.
- Souza, R. C. S. (1989). *Modelos Estruturais para Previsão de Séries Temporais. Abordagens Clássica e Bayesiana*. 17o Colóquio Brasileiro de Matemática. Rio de Janeiro: IMPA.
- Spall, J. C. (1988). *Bayesian Analysis of Time Series and Dynamic Models*. New York: Marcel Dekker.
- Stoffer, D. S. & Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, **86**, 1024-1033.
- Theil, H. & Wage, S. (1964). Some observations on adaptative forecasting. *Management Science*, **10**, 198-206.
- Thombs, L. A. & Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, **85**, 486-492.
- Wall, K. D. & Stoffer, D. S. (2002). A state space approach to bootstrapping conditional forecasts in ARMA models. *Journal of Time Series Analysis*, **23**, 733-751.
- Wei, W. W. S. (1990). *Time Series Analysis - Univariate and Multivariate Methods*. New York: Addison-Wesley.
- West, M., Harrison, P. J. & Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (com discussão). *Journal of the American Statistical Association*, **80**, 387, 73-97.
- West, M. & Harrison, P. J. (1986). Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, **81**, 395, 741-750.



- West, M., Harrison, P. J. & Pole, A. S. (1988). *BATS: A user guide Bayesian Analysis of Time Series. Program manual*. Statistics Department, University of Warwick, Reino Unido.
- West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324-342.