

Universidade Federal do Rio de Janeiro

Guido Alberti Moreira

Modelos Dinâmicos para Estimação
de Ciclos: Um estudo sobre geração e ajuste de dados

Rio de Janeiro

2009

Livros Grátis

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

Milhares de livros grátis para download.

UFRJ

Modelos Dinâmicos para Estimação
de Ciclos: Um estudo sobre geração e ajuste de dados

Guido Alberti Moreira

Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Orientadora:

Marina Silva Paez

Rio de Janeiro

2009

Modelos Dinâmicos para Estimação
de Ciclos: Um estudo sobre geração e ajuste de dados

Guido Alberti Moreira

Orientadora:

Marina Silva Paez

Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Aprovada por:

Profa. Marina S. Paez - Orientadora

Prof. Hélio dos Santos Migon

Prof. Cristiano Augusto Coelho Fernandes

Rio de Janeiro

2009

FICHA CATALOGRÁFICA

Moreira, Guido Alberti

Modelos Dinâmicos para Estimação de Ciclos: Um estudo sobre geração e ajuste de dados / Guido Alberti Moreira - Rio de Janeiro: UFRJ/IM, 2008.

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

1. Modelos Dinâmicos 2. Inferência Bayesiana 3. Estimação de Ciclos Econômicos 4. Análise Espectral I. Paez, Marina Silva II. Universidade Federal do Rio de Janeiro, Instituto de Matemática III. Título

Resumo

Modelos Dinâmicos para Estimação de Ciclos: Um estudo sobre geração e ajuste de dados

Guido Alberti Moreira

Orientadora: Profa. Marina Silva Paez

Resumo da Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística, Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

É interessante fazer estudo de previsões a longo prazo via modelagem de ciclos econômicos. Através de modelos dinâmicos, pode-se definir claramente uma componente de ciclo independente da média. Ao ser capaz de estimar bem o ciclo, podemos fazer previsões sob o ponto de vista da inferência Bayesiana, em que calculamos distribuições preditivas para os dados a se prever. Além disso, para ser realista com dados econômicos, usamos uma componente de tendência que indica uma média crescente e assintótica. Combinadas, essas duas componentes podem iniciar um estudo sobre dados econômicos. Para fazer o estudo, geramos dados simulados artificialmente e tentamos ajustá-los. Este trabalho inclui um estudo sobre cuidados ao gerar a amostra artificial, além de ilustrar as tentativas de ajuste com diferentes configurações de informação *a priori*. Acrescentamos na conclusão uma proposta futura para inclusão de informação trazida por análise espectral da amostragem.

Palavras-chave: Inferência Bayesiana, Modelos Dinâmicos, Estimação de Ciclos Econômicos, Análise Espectral.

Abstract

Dynamic Models for Cycle Estimation:

A study on sampling and data fitting

Guido Alberti Moreira

Orientadora: Profa. Marina Silva Paez

Abstract da Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística, Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

It is an interesting matter to do long term forecast through the modeling of economic cycles. Through the use of dynamic models, defining a cycle component, independently from the mean, is actually easy. By successfully estimating the cycle, we can do Bayesian forecasts, where we evaluate forecasting distributions for the data set we want to predict. Furthermore, a realistic study on economic data sets includes a trend resulting of a growing, yet asymptotic mean, which we also estimate. When combined, these two components can initiate a study on economic data. For that purpose, we obtain a sample from the model and try to fit artificial data set. This dissertation includes a study on how to sample the artificial data, as well as illustrates attempts to fit the data with various configurations of prior information. We include in the conclusion a proposition to add information through spectral analysis of the sampling method.

Key-words: Bayesian Inference, Dynamic Models, Estimation of Economic Cycles, Spectral Analysis.

Agradecimentos

Aos meus professores da UFRJ, que, em sua eterna paciência, permitiram me encontrar no meu (des)caminho. Desses, alguns se destacam, pelo apoio, orientação e cobrança, como Alexandra, H. Migon, Dani e Marina.

Aos meus pais quem amo incondicionalmente, pelos seus ensinamentos, direcionamento, paciência, impaciência, preocupação e compreensão.

Aos meus amigos, que, mesmo me conhecendo, ficaram do meu lado. Pelo apoio que me deram em momentos de dificuldade, mesmo sem saber, em alguns casos, o que se passava. Enumerá-los seria injusto, mas eles sabem quem são. Serão sempre meus companheiros do que gosto de fazer, desde as noitadas de que sempre nos lembraremos, até as profundas e eternas conversas nórdicas sobre se um mago ganharia de um guerreiro num duelo, ou se, de fato, Darth Vader tem alguma relação com Jesus Cristo (ambos foram “concebidos” por uma Força maior). Mantenho a esperança de que algum dia poderei fazer para todos aquilo que fizeram para mim.

Outros focos também merecem minha gratidão: A portaria do clube Germania que não fiscaliza a entrada dos peladeiros de basquete todo domingo; A Blizzard Entertainment®, que me ensinou o quão divertido é matar um dragão gigante que cospe fogo; O Felipe, que consertou minha bicicleta tantas vezes quantas ela quebrou; O Dudá, o DINI e o Hajime que, comigo, tocamos o melhor do Rock sem preocupações com volume; Os meus vizinhos

que agüentaram dito Rock; Olívia, Fábio, Mau e Luiz que sediaram incontáveis seções de RPG regadas de abobrinha (conversas) e *junk food*; O ar-condicionado que faz do Rio de Janeiro um planeta habitável; Beethoven por entender do que são feitas as nossas víceras; O Rock, Jazz, Blues, pop, Bossa-Nova e todos os sons que compõem o nosso dia-a-dia.

Sumário

1	Introdução	1
1.1	Sumário da dissertação	3
2	Revisão Metodológica	4
2.1	Aspectos básicos da inferência Bayesiana	4
2.1.1	Teorema de Bayes	5
2.1.2	Estimação de Parâmetros	6
2.1.3	Distribuição Preditiva	8
2.2	Séries Temporais e Modelos Dinâmicos	9
2.2.1	Modelos Lineares Dinâmicos	11
2.2.2	Inferência	14
2.2.3	Previsão	17
2.3	Modelo Linear Dinâmico com tendência e ciclo	18
2.4	Aspectos Computacionais: Método MCMC	20
2.4.1	Amostrador de Gibbs	21
2.4.2	Método Metropolis-Hastings	23
2.4.3	Amostrando da Distribuição Preditiva	25
2.4.4	Cuidados gerais sobre MCMC	26

3	Geração e ajuste dos dados	27
3.1	Geração dos dados	27
3.2	Cálculo das condicionais completas e probabilidades de aceitação	32
3.3	Primeiro conjunto de dados	37
3.3.1	Geração dos dados	37
3.3.2	Ajuste	38
3.3.3	Previsão	48
3.4	Segundo conjunto de dados simulados	51
3.4.1	Geração dos dados	51
3.4.2	Ajuste	52
3.4.3	Previsão	59
3.5	Estudo da amplitude inicial do ciclo	61
4	Conclusão e continuidade	66
4.1	O problema	66
4.1.1	Geração dos dados	67
4.1.2	Ajuste	68
4.2	Periodograma: um Estudo pela Transformada de Fourier	69
4.2.1	O periodograma	69
4.2.2	Efeito do intervalo de tempo	71
4.2.3	Efeito do intervalo de amostragem	72
4.2.4	Estudo <i>a priori</i>	73
4.3	Projetos Futuros	74
	Referências Bibliográficas	76

Apêndices	77
Apêndice A - Código em R que gerou os dados simulados para o capítulo 4 . . .	78
Apêndice B - Introdução histórica	80
Apêndice C - Programa em Ox TM que gerou as amostras das <i>posterioris</i> e da preditiva	85

Capítulo 1

Introdução

Um problema interessante em Estatística é o estudo da evolução de certos dados no tempo. Esses dados são comumente chamados de séries temporais e existe uma grande quantidade de abordagens para o problema na literatura. Em muitos casos, a série temporal é obtida através de uma amostragem discreta num intervalo de tempo em que se assume que os dados sejam correlacionados entre si. No caso deste trabalho, estamos interessados no estudo de dados econômicos e sua previsão a longo prazo.

Este trabalho é inspirado no problema de previsão de séries com ciclo a longo prazo. A modelagem de ciclos econômicos visa prever desvios em uma atividade econômica de longo prazo, que costumam durar mais tempo que uma eventual sazonalidade, normalmente mais do que um ano. A série do Produto Interno Bruto de um país costuma apresentar esse efeito, como também as flutuações do valor de ações na bolsa de valores. Azevedo *et al.* (2006) abordam o problema usando um filtro multivariado para fazer previsões do ciclo financeiro da região do Euro. Em seu trabalho, alguns parâmetros, como o período do ciclo, são fixadas num valor arbitrário. O problema de previsão a longo prazo é bem interessante, e propomos aqui uma abordagem Bayesiana para o problema. O modelo que

usaremos será um modelo simples para previsão de longo prazo de dados com tendência estocástica que apresenta comportamento assintótico e um ciclo cujo efeito diminui com o tempo. Sob o ponto de vista Bayesiano, todos os parâmetros do modelo podem ser estimados através de suas *posteriors*, logo, a nossa proposta, embora simples, oferece alguma vantagem sobre a proposta de Azevedo *et al.* Por ser um modelo dinâmico, a nossa proposta também tem a vantagem de permitir que mais componentes sejam adicionados ao modelo, como por exemplo efeitos de covariáveis incluídos em Azevedo *et al.* Dessa forma, poderemos em trabalhos futuros incrementar o modelo e torná-lo viável para reais previsões de ciclos econômicos de longo prazo.

Além disso, a estatística Bayesiana tem como princípio a atualização de informação através da coleta dos dados, o que implica em que muitas vezes há informação que não está somente contida nos dados. Por essa via, há facilidades em algumas estimações e no estudo, além de possibilitar a estimação do modelo que propomos.

É muito comum, para o estudo de diferentes formas de modelagem, a geração de um conjunto de dados artificiais sob a hipótese de que estes seguem o modelo estudado, com o objetivo de estimar os parâmetros do modelos, que são supostamente desconhecidos. Como há muitos parâmetros que se relacionam inter-mutuamente, a escolha deles para a geração dos dados causa mudanças, embora sensíveis, na forma que os dados tomam. Isso se torna um problema, quando se quer estudar os dados numa certa forma, ou, mais genericamente, com certas características.

O objetivo deste trabalho é analisar a modelagem proposta na estimação de dados gerados artificialmente. Para a inferência, usaremos métodos de MCMC (Markov Chain Monte Carlo) (Gamerman e Lopes, 2006) para extrair amostras das *posteriors* dos parâmetros.

1.1 Sumário da dissertação

No capítulo 2 será feita uma revisão metodológica em que descrevemos os aspectos básicos da inferência Bayesiana, apresentamos os modelos lineares e os modelos dinâmicos lineares para modelar séries temporais. Também serão apresentados o modelo que usaremos e os métodos computacionais. Apresentaremos os métodos do amostrador de Gibbs e Metropolis-Hastings para se fazer inferência com MCMC. Além disso, mostraremos como será feita a aplicação dos métodos de MCMC em modelos dinâmicos.

No capítulo 3 serão apresentados os cálculos necessários para os métodos apresentados no capítulo 2. Também apresentaremos dois conjuntos de dados gerados pelo modelo proposto. Além disso, faremos alguns estudos sobre o ajuste e sobre a previsão dos dados e uma comparação entre os dois conjuntos de dados gerados e seus ajustes e previsões.

No capítulo 4 discutiremos os resultados obtidos no capítulo 3, além de propor algumas soluções para problemas apresentados. Serão levantadas algumas discussões sobre essas soluções. Também serão apresentadas algumas propostas para trabalhos futuros envolvendo modelos mais complexos, que tomam o modelo apresentado como base, para modelar ciclos econômicos, como uma alternativa para o modelo freqüentista proposto em Azevedo *et al.* (2006). Também é proposta uma forma de aumentar a informação *a priori* sob o parâmetro que determina o período do ciclo baseada em análise espectral.

Em seguida, as referências bibliográficas usadas neste trabalho são apresentadas.

No apêndice A, apresentamos o programa em R[®] que usamos para gerar os dados. Fizemos no apêndice B uma introdução histórica para estatística e inferência Bayesiana. Por fim, incluímos no apêndice C o programa em Ox[™] para gerar um conjunto de amostras das *posteriors*.

Capítulo 2

Revisão Metodológica

2.1 Aspectos básicos da inferência Bayesiana

A inferência estatística vem sendo utilizada há muito tempo pela humanidade. Uma introdução histórica sobre inferência é apresentada no apêndice B. O ponto de vista da inferência Bayesiana se caracteriza por tomar em consideração informações que não se limitam aos dados coletados, ou pelo menos, aos dados coletados mais recentemente. Mais do que isso, por tratar os parâmetros do modelo como variáveis aleatórias, a inferência Bayesiana permite que estruturas hierárquicas sejam criadas. Dessa maneira, pode-se criar estruturas diversas dando ao modelo muito mais versatilidade do que seria possível caso contrário. Um exemplo básico de modelo hierárquico é assumir que os dados tenham distribuição normal com variância conhecida, e a média seja aleatória, também seguindo uma distribuição normal. A inferência Bayesiana não é muito diferente disso, em que, para cada valor desconhecido num modelo, estende-se o modelo para colocar uma hierarquia estocástica sobre eles.

Através da inferência Bayesiana, a estimação e previsão não se limitam a um único

valor, mas são representadas por uma distribuição (*a posteriori*, no caso de estimação, e preditiva, no caso de previsão). Através dessas distribuições, pode-se estudar não só média e variância, mas também quantis, medidas de simetria e probabilidades, entre outros, o que torna a inferência muito mais completa do que se isso não fosse possível. Além disso, se obtivermos mais dados, pode-se fazer outra atualização. Isso faz do ponto de vista Bayesiano uma poderosa ferramenta para análise de dados.

O problema da inferência estatística é o de entender o comportamento de uma população, ou de algum conjunto de características de uma população. Para tal, associamos parâmetros matemáticos a diferentes aspectos da população e tentamos estudá-los. Seja θ o conjunto de parâmetros de interesse no modelo sob estudo. Na inferência frequentista, usa-se a informação da amostra. Na inferência Bayesiana, atribuímos especificações probabilísticas aos parâmetros. Essas especificações são sempre olhadas após, isto é, *a posteriori* à coleta de dados de uma amostra. Por isso, normalmente há duas distribuições probabilísticas que descrevem as quantidades de interesse. A distribuição *a priori* dos dados serem coletados, e a distribuição *a posteriori*, que é a distribuição atualizada.

2.1.1 Teorema de Bayes

A inferência Bayesiana se baseia na atualização da informação através do teorema de Bayes. A seguinte descrição define o teorema de Bayes no contexto da inferência Bayesiana segundo Migon e Gamerman (1999):

Seja \mathbf{H} a informação disponível inicialmente para alguma grandeza de interesse. Seja θ o vetor de parâmetros desconhecidos que assume valores em Θ . Supondo que a informação inicial pode ser expressa em termos probabilísticos $P(\theta|\mathbf{H})$, temos uma descrição completa da incerteza a respeito de θ .

Para atualizar a informação a respeito de $\boldsymbol{\theta}$, podemos observar uma amostra de um vetor aleatório \mathbf{X} relacionado com $\boldsymbol{\theta}$. Assim, a informação disponível para a inferência passará a ser $H^* = H \cup \{\mathbf{X} = \mathbf{x}\}$.

Precisamos ainda conhecer a distribuição amostral de \mathbf{x} para realizar a atualização. Essa distribuição, conhecida como *função de verossimilhança*, associa para cada valor de $\boldsymbol{\theta}$ a plausibilidade de \mathbf{x} ser observado. A partir da especificação dos elementos da inferência, o teorema de Bayes é especificado da forma:

$$p(\boldsymbol{\theta}|H^*) = p(\boldsymbol{\theta}|\mathbf{x}, H) = \frac{p(\boldsymbol{\theta}, \mathbf{x}|H)}{p(\mathbf{x}|H)} = \frac{p(\mathbf{x}|\boldsymbol{\theta}, H)p(\boldsymbol{\theta}|H)}{p(\mathbf{x}|H)}, \quad (2.1)$$

em que

$$p(\mathbf{x}|H) = \int_{\Theta} p(\boldsymbol{\theta}, \mathbf{x}|H)d\boldsymbol{\theta}.$$

Como $p(\mathbf{x}|H)$ não depende de $\boldsymbol{\theta}$ e como \mathbf{H} é comum a todos os termos, podemos reescrever o teorema (2.1) da seguinte forma:

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.2)$$

O resultado apresentado em (2.2) é conhecido como teorema de Bayes e se constitui como a base de todos os procedimentos da inferência Bayesiana.

2.1.2 Estimação de Parâmetros

A estimação de parâmetros pode ser feita tanto pontualmente como por intervalo.

Estimação pontual

Migon e Gamerman (1999) descrevem a estimação sob a perspectiva da inferência Bayesiana como um problema de decisão que se divide nos seguintes elementos: Espaço

de parâmetros Θ ; espaço de resultados do experimento Ω ; espaço de ações possíveis A . Essa estimação constitui-se basicamente por associar perdas a possíveis decisões errôneas e “decidir” por estimar o parâmetro com aquele valor que minimiza a perda atribuída, dadas as probabilidades dos erros.

Criamos uma função definida em Ω que assume valores em A . Chamamos essa função de $\delta : \Omega \rightarrow A$. δ é chamado de *regra de decisão*. Criamos, então, uma função L que mede a perda da decisão $\delta(x)$, $\forall x \in \Omega$ quando o valor verdadeiro do parâmetro é θ , $\forall \theta \in \Theta$. O risco de tomar uma decisão $\delta(x)$ é a esperança da perda $L(\delta, \theta)$ com relação à distribuição *a posteriori* de θ :

$$R(\delta) = E_{\theta|x}(L(\delta, \theta)).$$

O estimador pontual de θ é a decisão que tem risco esperado mínimo para aquela função perda. Dentre as funções perda mais usadas estão:

- Perda quadrática: $L(\delta, \theta) = (\delta - \theta)^2$
- Perda absoluta: $L(\delta, \theta) = |\delta - \theta|$
- Perda zero-um: $L(\delta, \theta) = \begin{cases} 1, & \text{se } \delta = \theta \\ 0, & \text{se } \delta \neq \theta. \end{cases}$

Vale notar que os estimadores pontuais resultantes das perdas quadrática, absoluta e zero-um são, respectivamente, a média, mediana e moda da distribuição *a posteriori* de θ .

Estimação por intervalo

Segundo Migon e Gamerman (1999), a forma mais adequada de avaliar a informação disponível a respeito de uma quantidade desconhecida θ é através da distribuição *a posteriori*. Muitas vezes, a estimação pontual não é suficiente para o pesquisador, pois não

contém informação sobre sua precisão. Uma alternativa é encontrar intervalos de credibilidade para θ . A definição do intervalo de credibilidade Bayesiano é:

Seja θ uma quantidade desconhecida em Θ . Uma região $C \subset \Theta$ é um intervalo de credibilidade Bayesiano $100(1 - \alpha)\%$ para θ se $p(\theta \in C|x) \geq 1 - \alpha$.

O intervalo de credibilidade Bayesiano $100(1 - \alpha)\%$ é denotado por $IC(1 - \alpha)\%$. O que a estimação por intervalo de credibilidade faz é dar um intervalo no qual θ se encontra, com probabilidade $1 - \alpha$. Essa interpretação só é possível devido à maneira que o estatístico Bayesiano vê o problema, ou seja, tratando cada parâmetro desconhecido como uma variável aleatória.

2.1.3 Distribuição Preditiva

Para fazer previsões sob o ponto de vista Bayesiano, a distribuição preditiva é a maneira natural de se abordar o problema.

Seja \mathbf{y} o conjunto de dados amostrados, e seja y_p o valor que se quer prever. A distribuição preditiva de y_p dado \mathbf{y} é:

$$\begin{aligned} p(y_p|\mathbf{y}) &= \int_{\Theta} p(y_p, \boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} = \int_{\Theta} p(y_p|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} = \\ &= \int_{\Theta} p(y_p|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \end{aligned} \tag{2.3}$$

na qual a última igualdade se dá por causa da independência entre y_p e \mathbf{y} dado $\boldsymbol{\theta}$.

De uma certa forma, pode-se ver a última integral como a esperança da distribuição paramétrica de y_p relativa à *posteriori* de $\boldsymbol{\theta}$, ou seja,

$$p(y_p|\mathbf{y}) = E_{\boldsymbol{\theta}|\mathbf{y}}[p(y_p|\boldsymbol{\theta})].$$

Uma vez que temos a distribuição preditiva, podemos fazer estimativas pontuais ou intervalos de credibilidade para as observações futuras, que podem servir como previsões.

2.2 Séries Temporais e Modelos Dinâmicos

Um dos problemas da estatística é o de se estudar séries de dados cujas distribuições não são constantes ao longo do tempo. Dados desse tipo formam uma classe de conjunto de dados chamada séries temporais. Há diversos métodos para se tratar com esses tipos de dados. Harrison e Stevens (1976) propõem o uso de uma ampla classe de modelos para tratar séries temporais, conhecidos como modelos dinâmicos. Alguns aspectos dos modelos dinâmicos que os caracterizam são a versatilidade e o fato de que eles possuem parâmetros que variam no tempo.

Os parâmetros dos modelos dinâmicos podem variar no tempo, ou seja, para cada nova observação, há um valor diferente para cada parâmetro. Isso dá possibilidade para modelar o progresso dos parâmetros, além dos dados, ou seja, pode-se ditar a estrutura de progressão do modelo para cada observação. Além disso, podemos montar uma estrutura de auto-dependência temporal para as componentes de forma variada e que facilite a estimação.

Há muitos tipos de dados cuja evolução no tempo é desejável acompanhar. Um dos motivos para isso é o desejo de se prever o comportamento desses dados estatisticamente, para se poder tomar decisões bem informadas. Alguns casos em que temos essa situação são, por exemplo, dados econômicos, dados de dispersão de fluido, dados de quantidade de poluição numa localidade, entre outros. Para modelar dados dessa natureza, torna-se crucial um método que seja versátil e eficaz.

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas, usualmente baseados em observações dependentes, que ocorrem em períodos de tempo igualmente espaçados. Esses modelos geralmente consideram que dados coletados com espaçamento menor são mais correla-

cionados do que dados coletados em instantes mais afastados no tempo. Essa é uma hipótese razoável, levando em conta efeitos cumulativos de variáveis omitidas do modelo, mas que podem estar afetando o processo.

Modelos dinâmicos são usados principalmente por sua versatilidade, pois há muitos outros modelos que podem ser vistos como um dos seus casos particulares. Em geral, deseja-se que um modelo de séries temporais seja capaz de incorporar variáveis explicativas, como se estivéssemos fazendo um modelo de regressão cujas variáveis não são independentes, mas evolutivas no tempo. Além disso, deseja-se poder acompanhar tendências e efeitos cíclicos de sazonalidade que alguma dessas variáveis explicativas não sejam capazes de explicar. Além disso, muitas vezes existe o interesse em modelar a variância do processo, ou compará-lo a um processo auto-regressivo. Todas essas questões podem ser resolvidas através de modelos dinâmicos.

O método computacional conhecido como MCMC (Markov Chain Monte Carlo) pode ser usado para estimar os parâmetros que compõem (F_t, V_t, G_t, W_t) . Esse método, entretanto, depende das componentes dos modelos dinâmicos, o que cria uma dependência mútua. Como, sob o ponto de vista Bayesiano, não se usa estimativas como substitutos de seus valores verdadeiros, não podemos usar a estimativa de um método para substituir no outro. Porém, nada impede que os façamos simultaneamente, pois o método MCMC é um método iterativo. Usando os dois métodos de inferência (os dois algoritmos relativos aos modelos dinâmicos e o MCMC), construímos um ciclo iterativo em que cada um dos métodos complementa o outro, permitindo-nos fazer inferência, não só sobre os parâmetros, mas também sobre valores a prever.

2.2.1 Modelos Lineares Dinâmicos

Essa classe de modelos para modelar séries temporais é apresentada em West e Harrison (1997). O modelo é definido pelas equações:

$$y_t = F_t' \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \quad (2.4)$$

$$\theta_t = G_t \theta_{t-1} + \nu_t, \quad \nu_t \sim N(0, W_t), \quad (2.5)$$

em que y_t é a observação no tempo t , F_t é um vetor de variáveis explicativas, θ_t é um vetor de coeficientes de regressão e parâmetros de estado, G_t é a matrix de evolução do modelo e $N(\mu, \sigma^2)$ denota a distribuição normal com média μ e variância σ^2 . Os ruídos brancos ϵ_t e ν_t são mutualmente independentes e V_t e W_t são suas respectivas variâncias. Na formulação geral dos modelos dinâmicos, ϵ_t e ν_t não precisam necessariamente ter distribuição normal. Neste trabalho assumimos normalidade para essas quantidades, pois essa hipótese oferece diversas facilidades, como a possibilidade de fazer inferência via Filtro de Kalman e FFBS (Forward Filtering Backwards Sampling, ver mais a frente). Um modelo dinâmico linear está completamente definido por $(F, G, V, W)_t$. O modelo dinâmico linear mais simples que pode se encontrar em séries temporais é o modelo de primeira ordem que é dado pelas equações

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t).$$

Um modelo dinâmico é inteiramente definido por quatro matrizes, que podem variar no tempo, (F_t, V_t, G_t, W_t) . Há uma grande variedade de escolhas para essas matrizes, cada uma representando um diferente modelo dinâmico. Determinadas escolhas dessas matrizes podem levar a outras classes de modelos conhecidas, como por exemplo os modelos auto-regressivos e o de médias móveis, o que as tornam, de certa forma, subclasses dos

modelos dinâmicos. Além disso, a classe permite diversos tipos de componentes, como uma estrutura de média, sazonalidade e a inclusão de covariáveis. Essas componentes podem ser especificadas de forma a serem independentes no modelo.

Esse modelo é conhecido como modelo de tendência linear constante. Para comparar com as equações (2.4)-(2.5), faça $F_t = G_t = 1, \forall t$. Esse modelo pode ser usado em problemas para séries que não apresentam sazonalidade e cujo nível muda suavemente no tempo.

Podemos facilmente especificar um modelo de tendência estocástica de crescimento linear. Para isso, devemos adicionar um novo parâmetro, β_t . Note que neste caso podemos fazer $F_t = (1, 0)'$ e $G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, levando ao modelo:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{1,t}, \quad \omega_{1,t} \sim N(0, W_{1,t})$$

$$\beta_t = \beta_{t-1} + \omega_{2,t}, \quad \omega_{2,t} \sim N(0, W_{2,t}).$$

O modelo de crescimento linear é o segundo modelo dinâmico não estacionário mais simples. Dessa forma, pode-se estudar a série sem necessariamente ser preciso fazer transformações nos dados para dar estacionaridade a ela. Através de superposição de modelos, pode-se colocar componentes de mais interesse, sem que sejam afetados pela média, se seu crescimento (ou decréscimo) for linear.

Outro modelo bastante utilizado é o que inclui uma componente sazonal s_t que tem um comportamento que se repete de p em p períodos. Se fizermos $F_t = (1, 0)'$ e $G_t =$

$\begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix}$, ficamos com o modelo:

$$y_t = s_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

$$s_t = s_{t-1} \cos \omega + s_{t-1}^+ \sin \omega + \omega_{1,t}, \quad \omega_{1,t} \sim N(0, W_{1,t})$$

$$s_t^+ = -s_{t-1} \sin \omega + s_{t-1}^+ \cos \omega + \omega_{2,t}, \quad \omega_{2,t} \sim N(0, W_{2,t}),$$

sendo ω um escalar entre 0 e 2π , e $T = \frac{2\pi}{\omega}$.

Modelos com componentes sazonais são muito úteis em diversos casos. Há situações que têm comportamento visível- e intuitivamente cíclico como eventos que se repetem em certas épocas do ano, como produção agrícola, quantidade de vendas em lojas, ou venda de passagens aéreas em temporadas de viagens. Há outros eventos que se repetem em outros períodos, como trimestrais, semanais, etc... Todos eles podem ser modelados por componente sazonais. Morettin e Tolo (2004) usam uma série do Índice de Produto Industrial do Brasil de janeiro de 1973 a dezembro de 1976 para exemplificar seu método de modelar sazonalidade.

Como foi visto nos casos acima, a escolha de F_t , G_t , V_t e W_t permite a definição de múltiplos modelos. Pode-se inclusive adicionar componentes independentes no modelo, basta definir F_t adequadamente e G_t como bloco diagonal. Por exemplo, podemos juntar o modelo de primeira ordem e o de sazonalidade no seguinte modelo:

$$y_t = \mu_t + s_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

$$\mu_t = \mu_{t-1} + \omega_{1,t}, \quad \omega_{1,t} \sim N(0, W_{1,t})$$

$$s_t = s_{t-1} \cos \omega + s_{t-1}^+ \sin \omega + \omega_{2,t}, \quad \omega_{2,t} \sim N(0, W_{2,t})$$

$$s_t^+ = -s_{t-1} \sin \omega + s_{t-1}^+ \cos \omega + \omega_{3,t}, \quad \omega_{3,t} \sim N(0, W_{3,t}).$$

O modelo acima é um modelo dinâmico definido em (2.4) e (2.5) com $F_t = (1, 1, 0)'$

e $G_t = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{array} \right)$. O modelo citado pode ser usado em situações em que deseja-se acompanhar o ciclo de um processo cuja média tem uma tendência estocástica crescente linear. Note que nos exemplos citados, F_t é um vetor de uns e zeros, pois todas as componentes do modelo são parâmetros de estado.

2.2.2 Inferência

Filtro de Kalman

Um dos principais aspectos dos modelos dinâmicos é o fato de que se pode fazer inferência sobre a distribuição atualizada de $\theta_t|y^t$, em que y^t é a informação até o tempo t . Para um modelo fechado, isto é, que só recebe informação dos dados, temos que $y^t = (y_1, \dots, y_t)'$. Essa inferência sequencial é baseada em três passos. Evolução, previsão e atualização.

Considerando o modelo linear dinâmico dado em (2.4) e (2.5), suponha que a distribuição de θ_{t-1} dado y^{t-1} seja Normal com média m_{t-1} e variância C_{t-1} . A equação (2.5) pode ser vista como $\theta_t|\theta_{t-1} \sim N(G_t\theta_{t-1}, W_t)$. Temos a evolução em:

$$\theta_t|y^{t-1} \sim N(a_t, R_t), \quad (2.6)$$

em que $a_t = G_t m_{t-1}$ e $R_t = G_t C_{t-1} G_t' + W_t$.

Podemos fazer previsão um passo a frente com a distribuição:

$$y_t|y^{t-1} \sim N(f_t, Q_t), \quad (2.7)$$

em que $f_t = F_t' a_t$ e $Q_t = F_t' R_t F_t + V_t$.

Por fim, a informação pode ser atualizada com:

$$\theta_t|y^t \sim N(m_t, C_t), \quad (2.8)$$

com $m_t = a_t + A_t e_t$ e $C_t = R_t - A_t A_t' Q_t$, em que $A_t = R_t F_t / Q_t$ e $e_t = y_t - f_t$.

Demonstração: Para provar isso, usamos de uma propriedade da distribuição Normal multivariada comumente chamada de reconstrução da conjunta:

Sejam dois vetores aleatórios x_1 de dimensão d_1 e x_2 de dimensão d_2 . Se $x_1|x_2 \sim N_{d_1}(\mu_1 + B_1(x_2 - \mu_2), B_2)$ para matrizes de escalares B_1 e B_2 com dimensões, respectivamente, $d_1 \times d_2$ e $d_1 \times d_1$, e $x_2 \sim N(\mu_2, \Sigma_{22})$, sendo que $N_d(\boldsymbol{\mu}, \Sigma)$ denota a distribuição Normal multivariada de dimensão d com média $\boldsymbol{\mu}$ e variância Σ , então

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N_d \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right],$$

em que $\Sigma_{11} = B_2 + B_1 \Sigma_{22} B_1'$ e $\Sigma_{21}' = \Sigma_{12} = B_1 \Sigma_{22}$.

Logo, para provar (2.6), faça $x_1 = \theta_t|y^{t-1}$ e $x_2 = \theta_{t-1}|y^{t-1}$. Note que $\theta_t|\theta_{t-1} \equiv \theta_t|\theta_{t-1}, y^{t-1}$, pois θ_t é independente de y^{t-1} dado θ_{t-1} , como pode ser visto em (2.5).

Temos então que $x_1|x_2 \sim N(G_t \theta_{t-1}, W_t)$ e $x_2 \sim N(m_{t-1}, C_{t-1})$. Como estamos interessados na distribuição de $\theta_t|y^{t-1}$, então precisamos encontrar a distribuição marginal de x_1 . Fazendo a analogia à reconstrução da conjunta, obtemos que essa distribuição é $N(\mu_1, \Sigma_{11})$, logo basta encontrar μ_1 e Σ_{11} . Pela analogia, $G_t \theta_{t-1} = \mu_1 + B_1(x_2 - m_{t-1})$ e $\Sigma_{11} = W_t + B_1 C_{t-1} B_1'$. Como desejamos encontrar a distribuição marginal de x_1 , devemos estipular B_1 de forma que μ_1 , a média marginal de x_1 , seja independente de x_2 . Fazemos, então, $B_1 = G_t$. Note que, como estamos tratando da distribuição condicional a x_2 , este é considerado conhecido, o que resulta em $\theta_{t-1}|y^{t-1} \equiv \theta_{t-1}$, portanto $G_t \theta_{t-1}$, que está nos dois lados da equação, são simplificados, levando a $\mu_1 = G_t m_{t-1} = a_t$. Como Σ_{11} só dependia de B_1 para ser inteiramente definido, e $B_1 = G_t$, então temos que

$\Sigma_{11} = W_t + G_t C_{t-1} G_t' = R_t$, como queríamos demonstrar.

A demonstração de (2.7) é análoga à anterior. Basta fazer $x_1 = y_t | y^{t-1}$ e $x_2 = \theta_t | y^{t-1}$. Novamente, dado θ_t , y_t é independente de y^{t-1} , como pode ser visto em (2.4). Temos, partindo de (2.4), que $x_1 | x_2 \sim N(F_t \theta_t, V_t)$ e $x_2 \sim N(a_t, R_t)$, como visto anteriormente.

Fazendo a analogia à reconstrução da conjunta, temos um procedimento análogo ao anterior. $F_t' \theta_t = \mu_1 + B_1(x_2 - a_t)$. Fazemos $B_1 = F_t'$ e chegamos a $\mu_1 = F_t' a_t = f_t$ e $\Sigma_{11} = V_t + F_t' R_t F_t = Q_t$, c.q.d.

Tanto a inferência Bayesiana quanto a freqüentista funcionam de maneira condicional a um elemento teórico crescer infinitamente, porém no caso da estimação Bayesiana, esse elemento é um aspecto computacional, em detrimento do tamanho da amostra, o que é uma grande vantagem, visto que todos os aspectos computacionais podem ser controlados. A inferência Bayesiana é definida por uma atualização da informação *a priori* dos dados a serem coletados: a informação *a posteriori*. Essa atualização é possível através do teorema de Bayes, apresentado em 2.1.1. Assumindo normalidade dos erros, a inferência sobre os modelos dinâmicos é constituída de dois algoritmos: o primeiro calcula as distribuições *a priori* e *a posteriori* dos parâmetros seqüencialmente para cada instante de tempo t , como se dados só tivessem sido observados até aquele ponto; o segundo algoritmo faz uma suavização da estimativa dos dados calculando a distribuição *a posteriori* das componentes usando o conjunto completo dos dados. Esses dois algoritmos, entretanto, são obtidos condicionalmente a (F_t, V_t, G_t, W_t) . Assim, no caso em que se deseja fazer inferência sobre essas matrizes também, esses dois algoritmos não são suficientes.

FFBS

Para se fazer inferência baseada no teorema de Bayes usando MCMC, precisa-se das distribuições de $\theta_t | y^n$, ou seja, dos parâmetros condicionais a todos os dados, mas isso

não é alcançável com o filtro de Kalman, pois sua análise é sequencial, isto é, só fornece as distribuições de $\theta_t|y^t$. Para resolver esse problema, existe uma técnica de suavização conhecida como *FFBS* que é sigla para *Forward Filtering Backwards Sampling* que nos permite encontrar essas distribuições. Temos

$$\theta_t|y^n \sim N(m_t^n, C_t^n), \quad (2.9)$$

em que

$$\begin{aligned} m_t^n &= m_t + C_t G'_{t+1} R_{t+1}^{-1} (m_{t+1}^n - a_{t+1}), \\ C_t^n &= C_t - C_t G'_{t+1} R_{t+1}^{-1} (R_{t+1} - C_{t+1}^n) R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

Com essas equações, podemos facilmente fazer inferência condicional a $(F, G, V, W)_t$ sobre um modelo dinâmico através do seguinte algoritmo:

- Amostre da distribuição de $\theta_n|y^n$ adquirida pelo filtro de Kalman e faça $t = n - 1$;
- Amostre θ_t de (2.9);
- Faça $t = t - 1$ e volte ao passo anterior a não ser que $t = 1$. Neste caso, pare.

Através desse algoritmo, obtemos uma amostra da posteriori de θ .

2.2.3 Previsão

Dado que se tem a distribuição de $\theta_t|y^t$, podemos fazer previsões k passos à frente. A distribuição preditiva $y_{t+k}|y^t$ é dada por:

$$y_{t+k}|y^t \sim N(f_t(k), Q_t(k)), \quad (2.10)$$

em que $f_t(k) = F'_t a_t(k)$, $Q_t(k) = F'_{t+k} R_t(k) F_{t+k} + V_{t+k}$, $a_t(k) = G_{t+k} a_t(k-1)$, $R_t(k) = G_{t+k} R_t(k-1) G'_{t+k} + W_{t+k}$ e $a_t(0) = m_t$ e $R_t(0) = C_t$.

Para o caso especial em que $F_t = F, \forall t$ e $G_t = G, \forall t$, temos que $f_t(k) = F'G^k m_t$ e $Q_t(k) = V_{t+k} + F'(G^k C_t G'^k + \sum_{i=0}^{k-1} G^i W_{t+k} G'^i)F$

Intuitivamente, a distribuição em (2.3) pode ser vista da seguinte maneira: A média $f_t(k)$ é, na verdade, a média de y_{t+k} dada toda a informação que temos no tempo t , pois se desconsiderarmos as parcelas estocásticas da equação do sistema para os tempos $t+1, \dots, t+k$ (cujos valores não sabemos), basta multiplicar as matrizes G apropriadas e aplicar na média de θ que conhecemos no tempo t para encontrar a média de θ no tempo $t+k$. Assim, basta aplicar, como na equação das observações, a matriz F' do tempo $t+k$ para encontrarmos a média da previsão.

O efeito das componentes aleatórias será passado adiante através da variância. Da mesma forma que a média, a variância que será passada para a equação das observações é modificada por F_{t+k} e, assim, somada à própria variância das observações. Note que a variância cresce quando k aumenta pois mais parcelas positivas vão sendo adicionadas.

Para uma visão mais aprofundada sobre modelos dinâmicos, ver Migon et. al. (2005) ou West e Harrison (1997).

2.3 Modelo Linear Dinâmico com tendência e ciclo

Nesta seção descrevemos um modelo dinâmico com uma componente de tendência de crescimento amortecido e uma componente de ciclo. A componente de ciclo funciona como uma componente sazonal cuja influência decai com o tempo. Faremos a superposição dessas componentes de maneira independente, ou seja, a matriz G , constante, será bloco-diagonal, em que cada bloco representa a submatriz G de sua respectiva componente. Os erros serão normalmente distribuídos e independentes em todos os casos, ou seja, ruídos brancos normais. Uma extensão interessante seria considerar as variâncias dos

erros variando no tempo, pois é comum que, à medida que a média de uma série aumenta, a sua variância aumente também. Mais especificamente, a variância é vista como uma porcentagem da média, de forma que a razão entre a média e variância (ou desvio padrão) fique constante. Neste trabalho supomos as variâncias constantes no tempo. Usando a notação dos modelos dinâmicos, o modelo proposto é especificado por:

$$\begin{aligned}
 F'_t = F' &= (1, 0, 1, 0), \\
 G_t = G &= \left(\begin{array}{cc|cc}
 1 & 1 & 0 & 0 \\
 0 & \phi & 0 & 0 \\
 \hline
 0 & 0 & \phi_c \cos \omega & \phi_c \sin \omega \\
 0 & 0 & -\phi_c \sin \omega & \phi_c \cos \omega
 \end{array} \right), \\
 V_t &= V, \\
 W_t = W &= \left(\begin{array}{cccc}
 W_1 & 0 & 0 & 0 \\
 0 & W_2 & 0 & 0 \\
 0 & 0 & W_3 & 0 \\
 0 & 0 & 0 & W_4
 \end{array} \right), \forall t,
 \end{aligned}$$

em que $\phi_c \in (0, 1)$ é um parâmetro que faz o efeito da matriz ciclo diminuir com o tempo e W_1, W_2, W_3 e W_4 são escalares positivos. O parâmetro ϕ_c controla o decaimento do efeito do ciclo. Em termos da modelagem, esse efeito é o que mais diferencia um efeito de ciclo de um sazonal, pois no caso de sazonalidade, não se observa esse decaimento. Outras diferenças que podemos destacar é o fato de que para se modelar sazonalidade, costuma-se usar diversos harmônicos, além de que seu período é fácil de estimar, pois costuma ser fácil de intuir. Colocando o modelo na forma de equações de observações e

do sistema, temos que, para $t = 1, \dots, n$:

$$y_t = \mu_t + c_t + \epsilon_t, \epsilon_t \sim N(0, V), \quad (2.11)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{1,t}, \omega_{1,t} \sim N(0, W_1),$$

$$\beta_t = \phi\beta_{t-1} + \omega_{2,t}, \omega_{2,t} \sim N(0, W_2),$$

$$c_t = \phi_c \cos\omega c_{t-1} + \phi_c \sin\omega c_{t-1}^+ + \omega_{3,t}, \omega_{3,t} \sim N(0, W_3),$$

$$c_t^+ = -\phi_c \sin\omega c_{t-1} + \phi_c \cos\omega c_{t-1}^+ + \omega_{4,t}, \omega_{4,t} \sim N(0, W_4),$$

em que c_t é a componente de ciclo do modelo e c_t^+ é a variável auxiliar do ciclo. O parâmetro ω tem uma relação inversamente proporcional ao período de repetição do ciclo, mais especificamente, se p é o período, então $\omega = \frac{2\pi}{p}$.

A estimação dos parâmetros ϕ , ϕ_c e ω é de grande interesse para o pesquisador, pois são eles que dão forma à série gerada e ao modelo. Se conhecessemos os verdadeiros valores desses parâmetros, teríamos todo o conhecimento necessário sobre o modelo, pois são esses aspectos dos dados que estamos interessados em descobrir. Sendo assim, a estimação adequada desses parâmetros é muito importante. Os parâmetros das variâncias afetam diretamente a estimação de ϕ , ϕ_c e ω , além de serem importantes para a precisão das previsões, por isso devem ser bem estimados também.

2.4 Aspectos Computacionais: Método MCMC

A inferência Bayesiana traz muitas facilidades quando se consegue determinar a distribuição *a posteriori* dos parâmetros, pois nesse caso não é necessário depender de estimativas ou de tamanhos de amostras grandes para se fazer inferência. Entretanto, nem sempre essa distribuição pode ser encontrada. Isso acontece quando a parte do denominador de (2.1) não tem solução analítica.

Apesar de, às vezes, não ser possível encontrar a função de distribuição *a posteriori*, é possível encontrar integrais da mesma usando um método de aproximação de integrais muito prático e conveniente chamado integração por Monte Carlo, dado pela simples equação:

$$\int_{\omega} f(x)p(x)dx \approx \frac{\sum_{i=1}^n f(x_i)}{n}, \quad (2.12)$$

em que $f(x)$ é uma função qualquer e x_i são amostras da distribuição $p(x)$ no espaço ω^1 . A facilidade desse método de integração é o fato de que existem diversos algoritmos para se extrair amostras de distribuições das quais não se tem a forma fechada. Além disso, encontrar integrais da *posteriori* é suficiente em muitos casos, pois momentos e probabilidades são obtidas através de integrais. Sendo assim, encontrar uma boa amostra da *posteriori* se tornou o objetivo maior da inferência Bayesiana.

Nasce assim o método que ficou conhecido como MCMC (Markov Chain Monte Carlo), em que se faz aproximações de integrais da *posteriori* usando a integral de Monte Carlo, e se extrai amostras da distribuição de interesse usando uma cadeia de Markov. Este trabalho usa duas maneiras de conseguir essa cadeia de Markov. O método Metropolis-Hastings e um caso particular desse, conhecido por amostrador de Gibbs. Apesar do amostrador de Gibbs ter aparecido cronologicamente posteriormente ao Metropolis-Hastings, apresentá-lo-ei primeiro.

2.4.1 Amostrador de Gibbs

O amostrador de Gibbs foi proposto por Geman e Geman (1984) num problema de Mecânica Estatística em que a distribuição de Gibbs² era a de interesse. Eles propuseram

¹Mais detalhes em Gamerman e Lopes (2006).

² $f(x_1, \dots, x_n) \propto \exp[-\frac{1}{kT}E(x_1, \dots, x_n)]$, em que k é uma constante positiva, T é a temperatura e E é a energia do sistema, função positiva.

uma amostragem explorando as distribuições condicionais completas através de um algoritmo iterativo que define uma cadeia de Markov. Esse trabalho não foi associado inicialmente a uma solução para problemas da estatística Bayesiana, mas sim a uma finalidade de processamento de imagens. No entanto, Gelfand e Smith (1990) compararam-no com outros métodos de simulação estocástica, o que mostrou a potencialidade do amostrador para problemas da inferência Bayesiana.

O amostrador de Gibbs é centrado na amostragem iterativa das distribuições condicionais completas das variáveis. Suponha um conjunto de variáveis aleatórias $\theta = (\theta_1, \dots, \theta_n)$. A distribuição condicional completa de $\theta_i, i = 1, \dots, n$ é $F(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$, ou, em outra notação, $F(\theta_i | \theta_{-i})$, ou seja, a distribuição condicional de θ_i dado o resto do espaço paramétrico. O amostrador de Gibbs é essencialmente uma cadeia de Markov cujo núcleo de transição é a composição das distribuições condicionais completas. Suponha que a distribuição de interesse seja a *posteriori* de θ dada por $\pi(\theta|x)$. Suponha também que as distribuições condicionais completas $\pi(\theta_i | \theta_{-i}, x), i = 1, \dots, n$ estejam disponíveis e que se saiba gerar amostras delas. O algoritmo de amostragem da cadeia de Markov é:

1. Inicie o contador de iteração $j = 1$ e escolha um valor arbitrário para iniciar a cadeia

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)}).$$

2. Com base no valor de θ no passo $(j - 1)$, obtenha $\theta^{(j)}$ da seguinte maneira: Para

$i = 1, \dots, n$, gere um valor para $\theta_i^{(j)}$ de

$$\theta_i^{(j)} \sim \theta_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_n^{(j-1)}$$

3. Faça $j = j + 1$ e volte para o passo 2 até conseguir convergência.

Pode ser mostrado que a distribuição de equilíbrio dessa cadeia de Markov é a distribuição *a posteriori* de θ , e sendo assim, teoricamente, para um número de iterações suficientemente grande, os pontos gerados pela cadeia podem ser vistos como uma amostra da distribuição *a posteriori* de θ . É claro que não se pode esperar um número infinito de iterações para conseguir convergência, mas há métodos de se analisar se as amostras que estão sendo extraídas a partir de certo ponto seguem a distribuição de interesse. Para ser mais específico, esses métodos verificam se a cadeia já convergiu a partir de um determinado ponto arbitrário.

2.4.2 Método Metropolis-Hastings

O método Metropolis-Hastings, como é usado hoje, provém de um algoritmo proposto em Metropolis *et al.* (1953) e generalizado em Hastings (1970). O método é usado atualmente para complementar o amostrador de Gibbs quando uma condicional completa *a posteriori* se mostra difícil de amostrar. Para resolver o problema, gera-se uma distribuição arbitrária, ou distribuição proposta, e se aceita ou rejeita o valor gerado segundo uma probabilidade de aceitação que depende da distribuição de interesse e da distribuição proposta. O amostrador de Gibbs pode ser visto, de uma certa forma, como um caso particular do método Metropolis-Hastings em que a probabilidade de aceitação é sempre 1.

Seja a distribuição *a posteriori* de θ , $\pi(\theta|x)$, a distribuição de interesse. Suponha que as condicionais completas *a posteriori* $\pi(\theta_i|\theta_{-i}, x)$, $i = 1, \dots, n$ estejam disponíveis, mas que não sabemos extrair amostras diretamente delas. Usa-se uma distribuição arbitrária, chamada distribuição proposta, $q(\phi|\theta_i^{(j-1)})$ para gerar um valor proposto para $\theta_i^{(j)}$, para $i = 1, \dots, n$. Aceitamos ϕ com uma determinada probabilidade α . Caso aceito, $\theta_i^{(j)}$ assume

o valor do valor proposto, caso contrário, fazemos $\theta_i^{(j)} = \theta_i^{(j-1)}$. Assim, o algoritmo é:

1. Inicie o contador de iteração $j = 1$ e escolha um valor arbitrário para iniciar a cadeia $\theta^{(0)}$
2. Faça $i = 1$
3. Gere um novo valor ϕ da densidade proposta $q(\phi|\theta_i^{(j-1)})$
4. Calcule a probabilidade de aceitação $\alpha(\phi, \theta_i^{(j-1)}) = \min \left\{ 1, \frac{\pi(\phi|\theta_{-i}^{(j-1)}, x)q(\theta_i^{(j-1)}|\phi)}{\pi(\theta_i^{(j-1)}|\theta_{-i}^{(j-1)}, x)q(\phi|\theta_i^{(j-1)})} \right\}$.
Se o passo é aceito, então $\theta_i^{(j)} = \phi$, senão, $\theta_i^{(j)} = \theta_i^{(j-1)}$ e a cadeia não se move
5. Faça $i = i + 1$ e volte para o passo 3, a não ser que $i = n$
6. Atualize o contador j para $j + 1$ e volte para o passo 2 até conseguir convergência.

O algoritmo Metropolis-Hastings pode ser usado em diversas situações em que se deseja gerar amostras de distribuições das quais normalmente não se sabe amostrar diretamente. Ele tem a vantagem de que não se precisa conhecer as constantes das distribuições, que são simplificadas no cálculo da probabilidade α . Entretanto, o método pede uma distribuição arbitrária, que pode ajudar ou atrapalhar a sua eficiência. Quanto mais próxima a distribuição proposta for da distribuição de interesse, maior será a probabilidade de aceitação. Além disso, há alguns casos em que a escolha da distribuição proposta simplifica bastante os cálculos da probabilidade de aceitação. A seguir apresentamos duas propostas comumente utilizadas.

Proposta Simétrica em torno do passo anterior

Quando a distribuição proposta é simétrica, e o ponto de simetria é o passo anterior $\theta_i^{(j-1)}$, a probabilidade de aceitação pode ser simplificada para:

$$\alpha = \min \left\{ 1, \frac{\pi(\phi|\theta_{-i}^{(j-1)}, x)}{\pi(\theta_i^{(j-1)}|\theta_{-i}^{(j-1)}, x)} \right\}, \quad (2.13)$$

pois $q(\theta_i^{(j-1)}|\phi) = q(\phi|\theta_i^{(j-1)})$.

Priori como proposta

A distribuição *a priori* é usada como proposta. Nesse caso, a proposta não depende do passo anterior, portanto, $q(\phi|\theta_i^{(j-1)}) = q(\phi)$. Além disso, no cálculo da probabilidade de aceitação, as *prioris* se cancelam (a proposta e o fator da *priori* na condicional completa), e α será a razão das funções de verossimilhança:

$$\alpha = \min \left\{ 1, \frac{f(x|\phi, \theta_{-i}^{(j-1)})}{f(x|\theta_i^{(j-1)}, \theta_{-i}^{(j-1)})} \right\} \quad (2.14)$$

2.4.3 Amostrando da Distribuição Preditiva

Se não temos a distribuição *a posteriori* na forma fechada, então não há como calcular a distribuição preditiva analiticamente. Em casos em que precisamos usar métodos de MCMC para fazer inferência sobre a distribuição *a posteriori* de $\boldsymbol{\theta}$, é possível também se extrair amostras da distribuição preditiva.

Para amostrar da distribuição preditiva, tratamos os valores a se prever y_p como parâmetros do modelo, sobre os quais também queremos fazer inferência. As condicionais completas são fáceis de obter, pois $p(y_p|\boldsymbol{\theta}, \mathbf{y}) = p(y_p|\boldsymbol{\theta})$ é, ao mesmo tempo, a condicional completa de y_p e o próprio modelo. Portanto, se a cada iteração j extraírmos uma amostra de $p(y_p|\boldsymbol{\theta}^{(j)})$, teremos uma amostra dos “parâmetros” $y_p^{(j)}$. Dessa forma, temos também

uma cadeia de Markov para a distribuição preditiva e podemos fazer integrais de Monte Carlo sobre ela.

2.4.4 Cuidados gerais sobre MCMC

É importante chamar atenção para alguns cuidados que se deve tomar ao utilizar o método MCMC.

A distribuição na qual se está interessado quando se faz MCMC é a distribuição *a posteriori* dos parâmetros, ou seja, a distribuição limite da cadeia de Markov. Como não podemos extrair infinitas amostras da cadeia, deve haver um ponto a partir do qual se aceita que as amostras extraídas são, de fato, da *posteriori*. As iterações antes desse ponto são chamadas de *burn-in*, e são comumente descartadas.

Para aproximação de integrais por Monte Carlo, a amostra deve ser aleatória, e portanto, não deve haver auto-correlação. Porém, por ser uma cadeia de Markov, amostras extraídas por MCMC podem ter uma auto-correlação forte. Para contornar isso, uma solução bem comum é se usar *thinning*, ou seja, pegar uma amostra de n em n iterações (com n escolhido de forma que os pontos da amostra sejam aproximadamente não correlacionados) e descartar o resto. Vale lembrar que, depois de descartar amostras de *Burn-in* e *thinning*, deve haver dados suficientes para fazer uma integral de Monte Carlo aceitável, pois quanto maior a amostra, melhor é a aproximação.

Para mais sobre MCMC, ver Gamerman e Lopes (2006).

Capítulo 3

Geração e ajuste dos dados

No capítulo 2, revimos a metodologia que usaremos. Neste capítulo, aplicaremos o que foi discutido em dados simulados artificialmente do modelo (2.11). Faremos primeiro um estudo sobre a geração dos dados, em seguida apresentaremos os cálculos para encontrar as distribuições condicionais completas dos parâmetros que estimamos. Geraremos, então, dois conjuntos de dados e aplicaremos os métodos de ajuste e previsão por MCMC para verificar se conseguimos estimar os parâmetros do modelo e prever seus valores futuros.

3.1 Geração dos dados

Para estudar o comportamento da estimação do modelo, geramos dois conjuntos de dados de tamanho n dados do próprio modelo e tentamos ajustá-los. Assumindo que esses dados gerados são dados reais sobre os quais não temos nenhuma informação, temos que supor não conhecer os parâmetros com os quais os dados foram gerados e tentar estimá-los, com a expectativa de que essa estimação fique próxima dos valores com os quais geramos os dados.

Trabalhando com séries temporais, nosso principal objetivo é fazer previsões. Para

isso, uma prática comum para o caso de dados simulados artificialmente é usar um subconjunto reduzido dos dados de tamanho k , com $k < n$. Assim, ajustamos o modelo para $t = 1, \dots, n - k$, assumindo não conhecer os dados em $t = n - k + 1, \dots, n$. Fazemos então a previsão deste último conjunto de pontos com os parâmetros estimados pelo primeiro conjunto. Dessa forma, podemos comparar as previsões com os dados que, de fato, foram gerados, o que nos dá um indicativo da capacidade preditiva do modelo.

A geração dos dados assume que os valores dos parâmetros são arbitrários, talvez com algumas restrições advindas de alguns conhecimentos *a priori* sobre os parâmetros, como, por exemplo, sabe-se que os valores de W_1, W_2, W_3 e W_4 são normalmente menores que V . Também sabe-se que os valores de ϕ e ϕ_c costumam ser próximos de 1, ou pelo menos, maiores que 0.8.

Porém, ao gerar os dados, notamos que há diversos cuidados a mais a se tomar. Um deles, é a relação entre n e ϕ_c , pois a informação do ciclo que decai segundo dita ϕ_c pode chegar a ser menor que a informação do erro, ditado por W_3 e W_4 , em um período de tempo t pequeno. De uma forma genérica, a quantidade de informação do ciclo que se mantém num certo instante t_1 , a partir de um instante de tempo inicial t_0 , é $\phi_c^{t_1 - t_0}$, ou seja, se $\phi_c = 0.8$, então em apenas 20 passos, a quantidade de informação no modelo do ciclo será somente $\approx 1\%$ da quantidade de informação inicial. Em outras palavras, se o ciclo causa uma variação de 100 unidades nos dados em $t_0 = 1$, então em $t_1 = 21$, essa variação será de pouco mais de 1 unidade. Isso significa que, em qualquer conjunto de dados, qualquer efeito de ciclo não será mais notado a partir de certo t_1 , em que t_1 depende de ϕ_c, W_3 e W_4 . Essa situação não só deixa a previsão mais difícil, como tira toda a utilidade da componente de ciclo para a previsão, pois para qualquer $t \geq t_1$, o ciclo não afetará mais os dados. Dessa forma, ao se escolher o tamanho de amostra n , é

importante ter o valor de ϕ_c em mente. Para ilustrar a quantidade de informação mantida para diferentes valores de ϕ_c ao longo do tempo, podemos ver a figura 3.1.

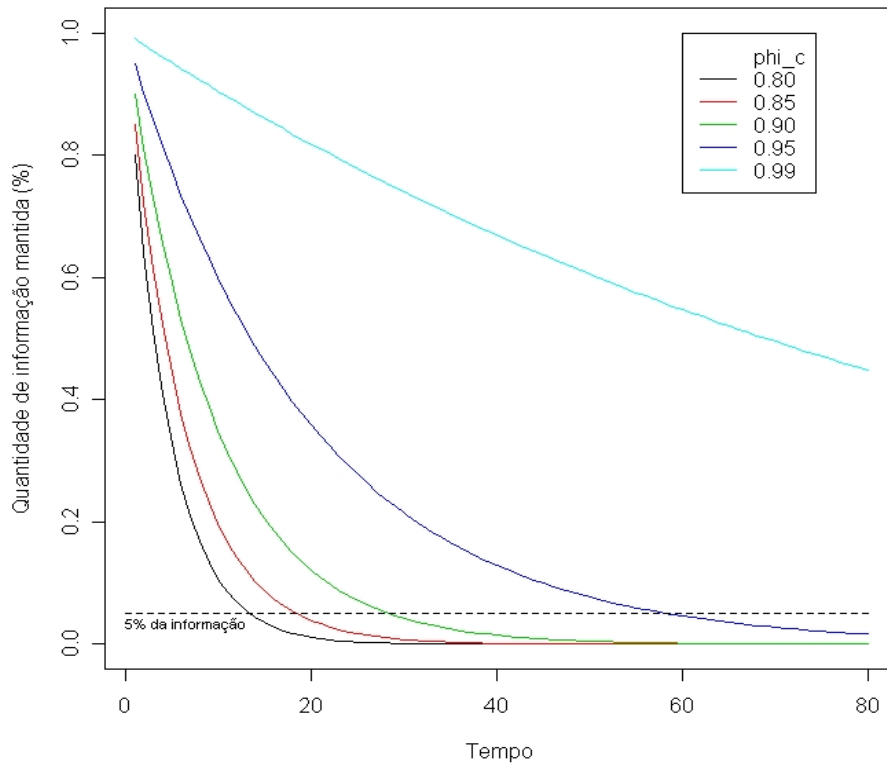


Figura 3.1: Informação mantida no tempo para diversos valores de ϕ_c .

Um outro aspecto importante ao gerar os dados é a especificações dos valores iniciais de β_0 e c_0 .¹ Esses valores são usados na equação do sistema para gerar μ_1 , β_1 , c_1 e c_1^+ e, portanto, definem diretamente os valores iniciais da série. Esses valores devem ser escolhidos de acordo com os valores de V , W_1 , W_2 , W_3 e W_4 , de forma que a série tenha crescimento amortecido e componente cíclica aparente, o que chamaremos de forma desejada. Teoricamente, esses valores deveriam ser arbitrários, mas vimos ao longo deste trabalho que a escolha desses valores pode facilmente gerar séries cuja forma não é a desejada.

¹A rigor, o valor de c_0^+ também é igualmente importante, mas esse valor pode ser incorporado a c_0 .

Para entender isso, imagine uma situação extrema em que o amplitude inicial do ciclo é menor que o intervalo de 90% de densidade de seu ruído branco. É fácil intuir que, nesse caso, qualquer movimento cíclico seria pequeno demais para ser percebido, e seria confundido com o ruído branco, inflando sua variância. O mesmo aconteceria se a amplitude do ciclo fosse muito pequena relativamente à variação dos dados, representada por V , pois o movimento cíclico seria, novamente, confundido com o ruído branco, desta vez, dos dados. Da mesma forma, se o valor de β_0 for excessivamente pequeno, os valores de β_t que se seguem seriam um valor deflacionado por ϕ desse β original, fazendo com que seu ruído branco passasse a ter mais influência sobre seu comportamento do que o próprio β_t , dando assim estacionaridade à série, ou, pelo menos, tirando sua propriedade assintótica.

Uma vez que β_0 é escolhido apropriadamente, o valor de ϕ deve ser escolhido. Ele se torna importante, pois a forma da média da série é muito bem definida por ϕ . Esse valor também depende de n . Um valor de ϕ próximo de 1 aproxima a tendência para um crescimento linear estocástico, enquanto um valor menor aumenta o comportamento assintótico da série, de forma que um valor de ϕ pequeno faz com que a média fique praticamente constante já para valores pequenos de tempo. Para ilustrar isso, apresentamos duas figuras (3.2 e 3.3). A primeira usa $n = 200$, e a segunda, $n = 50$.

É importante ressaltar que, por conta dos diferentes valores de ϕ , a taxa de crescimento das curvas é diferente e singular para cada uma delas. Por isso, para poder comparar as diferentes curvas, colocamos todas na mesma escala. Para fazer isso, dividimos todos os valores das séries pelos seus respectivos máximos e multiplicamo-los pelo valor máximo de todas as séries. Note que o objetivo dos gráficos é mostrar o efeito de ϕ para a forma global da série, o que não muda com o re-escalonamento.

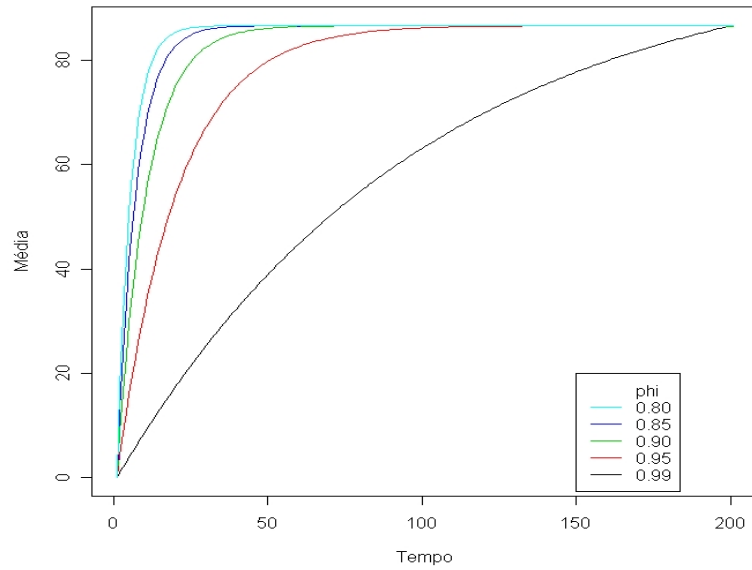


Figura 3.2: Forma da média para diferentes valores de ϕ .

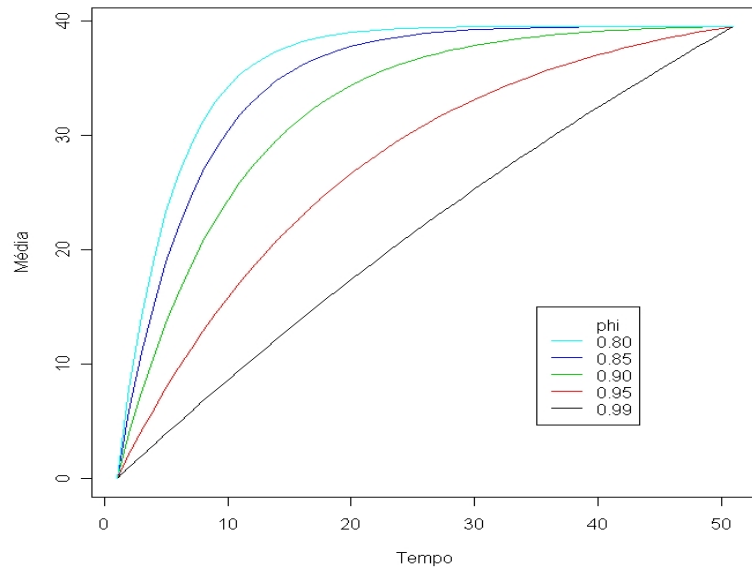


Figura 3.3: Forma da média para diferentes valores de ϕ .

Uma vez que todos esses parâmetros estão bem definidos, resta-nos especificar as variâncias. Como vimos, há alguns valores iniciais que dependem das variâncias. É importante, como já dito, que a variância das observações seja maior que a variância do sistema, pois, caso contrário, a maior parte da variação viria das componentes do sistema.

Um dos maiores problemas disso é a pouca confiabilidade das previsões, pois a cada passo a mais que se faz previsão, deve-se incluir uma parcela da variância da previsão que depende de W . Além disso, grandes variações da média alteram a própria tendência, em alguns casos alterando a forma da série.

Em nossos estudos simulados, verificamos que os valores de W_1 e W_2 devem ser muito menores que os de V . Especificamente, esses valores devem ser próximos a $\frac{V}{1000}$ para podermos obter uma forma desejada para a série. Isso não significa que não conseguimos obter a forma desejada com valores maiores, mas sim que, para obtê-la, precisamos de algumas triagens do programa, ou seja, a probabilidade da série ter a forma desejada não é alta com valores de W_1 e W_2 maiores que $\frac{V}{1000}$.

Nas seções 3.3 e 3.4, apresentamos resultados obtidos com a simulação de dois conjuntos de dados distintos. No Apêndice A, encontra-se o programa em R[®] usado para gerar os dados para o primeiro conjunto de dados.

3.2 Cálculo das condicionais completas e probabilidades de aceitação

Para calcular as condicionais completas, precisamos da função de verossimilhança, dada na fórmula (3.1), e das distribuições *a priori* dos parâmetros. A função de verossimilhança é encontrada através do produto, para todo t , da função originada pela equação

das observações (2.11) se virmos y_t como uma normal com média $\mu_t + c_t$ e variância V :

$$p(y_t|\mu_t, c_t, V) = \frac{1}{\sqrt{2\pi V}} \exp \left[-\frac{1}{2V} (y_t - \mu_t - c_t)^2 \right]$$

↓

$$f_n(\mathbf{y}|\boldsymbol{\mu}, \mathbf{c}, V) = (2\pi V)^{-\frac{n}{2}} \exp \left[-\frac{1}{2V} \sum_{t=1}^n (y_t - \mu_t - c_t)^2 \right]. \quad (3.1)$$

A equação do sistema do modelo (2.5) não está incluída na informação contida na função de verossimilhança, logo, sua forma deve ser introduzida na especificação das *prioris*, mais especificamente, sua estrutura como vista em (2.11) será incluída nas *prioris*.

A distribuição *a priori* conjunta dos parâmetros é necessária para o cálculo, no entanto usamos o fato de que, se X e Y são variáveis aleatórias, então $p(X, Y) = p(X|Y)p(Y)$. Usando essa propriedade, podemos fazer o produto das distribuições condicionais pelas marginais *a priori* e, assim, obter a *priori* conjunta.

As *prioris* de μ_t , β_t , c_t e c_t^+ são necessárias para os cálculos das condicionais completas de outros parâmetros, mesmo que suas condicionais completas sejam encontradas diretamente através do algoritmo FFBS. Os parâmetros das variâncias sofrerão uma reparametrização, pois é mais fácil modelá-los dessa maneira. Para cada variância, definimos um parâmetro τ , como por exemplo, $\tau_V = \frac{1}{V}$. O valor absoluto do Jacobiano dessas transformações é τ^{-2} . Como as *prioris* serão as mesmas (exceto pelos hiperparâmetros)

em todos os casos, só precisamos fazer estes cálculos uma vez. As *prioris* são:

$$\begin{aligned}
\pi(\boldsymbol{\mu}|\boldsymbol{\beta}, W_1) &= (2\pi W_1)^{-\frac{n}{2}} \exp \left[-\frac{1}{2W_1} \sum_{t=1}^n (\mu_t - \mu_{t-1} - \beta_{t-1})^2 \right], \\
\pi(\boldsymbol{\beta}|\phi, W_2) &= (2\pi W_2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t - \phi\beta_{t-1})^2 \right], \\
\pi(\mathbf{c}, \mathbf{c}^+|\phi_c, \omega, W_3, W_4) &= (2\pi W_3)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2W_3} \sum_{t=1}^n [c_t - \phi_c(\cos\omega c_{t-1} + \sin\omega c_{t-1}^+)]^2 \right\} \times \\
&\quad (2\pi W_4)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2W_4} \sum_{t=1}^n [c_t^+ - \phi_c(-\sin\omega c_{t-1} + \cos\omega c_{t-1}^+)]^2 \right\}.
\end{aligned} \tag{3.2}$$

Para os parâmetros τ_V , τ_{W_1} , τ_{W_2} , τ_{W_3} e τ_{W_4} , usamos *prioris* Gama Invertida, para ϕ e ϕ_c , usamos Beta e para ω usamos Uniforme nos seu intervalo de variação, ou seja, $(0, 2\pi)$. Uma das facilidades trazidas pelo amostrador de Gibbs e pelo método Metropolis-Hastings é o fato de que não é necessário ter as condicionais completas de forma fechada, sendo suficiente conhecer o núcleo das distribuições. Se reconhecemos o núcleo de uma distribuição conhecida, na literatura, podemos utilizar o amostrador de Gibbs com essa distribuição, pois o conhecimento da distribuição implica no conhecimento da constante, enquanto no método Metropolis-Hastings as constantes das distribuições são anuladas na fração da probabilidade de aceitação, tornando seu conhecimento igualmente dispensável no cálculo das condicionais completas. Por isso, ao calculá-las, basta encontrar uma proporcionalidade, em detrimento de uma igualdade. Para os cálculos a seguir, denotaremos R (resto) como o vetor paramétrico retirado o parâmetro do qual estamos calculando a condicional completa:

A distribuição condicional completa para $\tau_V = \frac{1}{V}$ é dada por:

$$\begin{aligned}
\pi(\tau_V|R) &\propto \tau_V^{\frac{n}{2}} \tau_V^{\alpha_V-1} \tau_V^{-2} \exp \left\{ \tau_V \left[-\frac{1}{2} \sum_{t=1}^n (y_t - \mu_t - c_t)^2 \right] \right\} \exp(-\tau_V \beta_V) = \\
&= \tau_V^{\alpha_V + \frac{n}{2} - 2 - 1} \exp \left\{ -\tau_V \left[\frac{1}{2} \sum_{t=1}^n (y_t - \mu_t - c_t)^2 + \beta_V \right] \right\}.
\end{aligned}$$

Podemos reconhecer esse núcleo como uma distribuição Gama, logo V segue distribuição Gama Invertida, que denotamos por $V \sim GI(\alpha_V + \frac{n}{2} - 2, \beta_V + \frac{1}{2} \sum_{t=1}^n (y_t - \mu_t - c_t)^2)$.

O cálculo das distribuições condicionais completas para $\tau_{W_1}, \dots, \tau_{W_4}$ é análogo e essas distribuições são dadas por:

$$\begin{aligned}
\pi(\tau_{W_1}|R) &\propto \tau_{W_1}^{\frac{n}{2}} \tau_{W_1}^{\alpha_{W_1}-1} \tau_{W_1}^{-2} \exp \left\{ \tau_{W_1} \left[-\frac{1}{2} \sum_{t=1}^n (\mu_t - \mu_{t-1} - \beta_{t-1})^2 \right] \right\} e^{-\tau_{W_1} \beta_{W_1}} = \\
&= \tau_{W_1}^{\alpha_{W_1} + \frac{n}{2} - 2 - 1} \exp \left\{ -\tau_{W_1} \left[\frac{1}{2} \sum_{t=1}^n (\mu_t - \mu_{t-1} - \beta_{t-1})^2 + \beta_{W_1} \right] \right\}, \\
\pi(\tau_{W_2}|R) &\propto \tau_{W_2}^{\frac{n}{2}} \tau_{W_2}^{\alpha_{W_2}-1} \tau_{W_2}^{-2} \exp \left\{ \tau_{W_2} \left[-\frac{1}{2} \sum_{t=1}^n (\beta_t - \phi \beta_{t-1})^2 \right] \right\} e^{-\tau_{W_2} \beta_{W_2}} = \\
&= \tau_{W_2}^{\alpha_{W_2} + \frac{n}{2} - 2 - 1} \exp \left\{ -\tau_{W_2} \left[\frac{1}{2} \sum_{t=1}^n (\beta_t - \phi \beta_{t-1})^2 + \beta_{W_2} \right] \right\}, \\
\pi(\tau_{W_3}|R) &\propto \tau_{W_3}^{\frac{n}{2}} \tau_{W_3}^{\alpha_{W_3}-1} \tau_{W_3}^{-2} \exp \left\{ \tau_{W_3} \left[-\frac{1}{2} \sum_{t=1}^n (c_t - \phi_c (\cos \omega c_{t-1} + \sin \omega c_{t-1}^+))^2 \right] \right\} e^{-\tau_{W_3} \beta_{W_3}} = \\
&= \tau_{W_3}^{\alpha_{W_3} + \frac{n}{2} - 2 - 1} \exp \left\{ -\tau_{W_3} \left[\frac{1}{2} \sum_{t=1}^n (c_t - \phi_c (\cos \omega c_{t-1} + \sin \omega c_{t-1}^+))^2 + \beta_{W_3} \right] \right\}, \\
\pi(\tau_{W_4}|R) &\propto \tau_{W_4}^{\frac{n}{2}} \tau_{W_4}^{\alpha_{W_4}-1} \tau_{W_4}^{-2} \exp \left\{ \tau_{W_4} \left[-\frac{1}{2} \sum_{t=1}^n (c_t^+ - \phi_c (-\sin \omega c_{t-1} + \cos \omega c_{t-1}^+))^2 \right] \right\} e^{-\tau_{W_4} \beta_{W_4}} = \\
&= \tau_{W_4}^{\alpha_{W_4} + \frac{n}{2} - 2 - 1} \exp \left\{ -\tau_{W_4} \left[\frac{1}{2} \sum_{t=1}^n (c_t - \phi_c (-\sin \omega c_{t-1} + \cos \omega c_{t-1}^+))^2 + \beta_{W_4} \right] \right\},
\end{aligned}$$

ou seja, as distribuições condicionais completas de todos esses parâmetros são Gammas Invertidas com parâmetros dados respectivamente por $(\alpha_{W_1} + \frac{n}{2} - 2, \beta_{W_1} + \sum_{t=1}^n (\mu_t - \mu_{t-1} - \beta_{t-1})^2)$, $(\alpha_{W_2} + \frac{n}{2} - 2, \beta_{W_2} + \sum_{t=1}^n (\beta_t - \phi \beta_{t-1})^2)$, $(\alpha_{W_3} + \frac{n}{2} - 2, \beta_{W_3} + \sum_{t=1}^n (c_t - \phi_c (\cos \omega c_{t-1} + \sin \omega c_{t-1}^+))^2)$ e $(\alpha_{W_4} + \frac{n}{2} - 2, \beta_{W_4} + \sum_{t=1}^n (c_t^+ - \phi_c (-\sin \omega c_{t-1} + \cos \omega c_{t-1}^+))^2)$.

Para o cálculo da distribuição condicional completa de ϕ , devemos notar que os fatores

dos quais ϕ faz parte são a *priori* de β e a *priori* de ϕ , ou seja,

$$\begin{aligned}\pi(\beta) &= \left(\exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t - \phi\beta_{t-1})^2 \right] \right) e \\ \pi(\phi) &= (\phi^{\alpha_\phi - 1} (1 - \phi)^{\beta_\phi - 1}), \text{ e portanto} \\ p(\phi|R) &\propto \exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t - \phi\beta_{t-1})^2 \right] \phi^{\alpha_\phi - 1} (1 - \phi)^{\beta_\phi - 1}.\end{aligned}$$

Para a amostragem desse parâmetro, utilizamos passos de Metropolis-Hasting, usando a *priori* como proposta. A probabilidade de aceitação α é o mínimo entre 1 e²:

$$\begin{aligned}& \frac{\exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t - \phi_p \beta_{t-1})^2 \right]}{\exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t - \phi_j \beta_{t-1})^2 \right]} = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [(\beta_t - \phi_p \beta_{t-1})^2 - (\beta_t - \phi_j \beta_{t-1})^2] \right\} = \\ &= \exp \left[-\frac{1}{2W_2} \sum_{t=1}^n (\beta_t^2 - 2\phi_p \beta_t \beta_{t-1} + \phi_p^2 \beta_{t-1}^2 - \beta_t^2 + 2\phi_j \beta_t \beta_{t-1} - \phi_j^2 \beta_{t-1}^2) \right] = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [2\beta_t \beta_{t-1} (\phi_j - \phi_p) + \beta_{t-1}^2 (\phi_p^2 - \phi_j^2)] \right\} = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [2\beta_t \beta_{t-1} (\phi_j - \phi_p) + \beta_{t-1}^2 (\phi_p + \phi_j)(\phi_p - \phi_j)] \right\} = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [(\phi_j - \phi_p)(2\beta_t \beta_{t-1} - \beta_{t-1}^2 (\phi_j + \phi_p))] \right\} = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [\beta_{t-1} (\phi_j - \phi_p)(2\beta_t - \beta_{t-1} (\phi_j + \phi_p))] \right\} = \\ &= \exp \left\{ -\frac{1}{2W_2} \sum_{t=1}^n [\beta_{t-1} (\phi_p - \phi_j)(\beta_{t-1} (\phi_p + \phi_j) - 2\beta_t)] \right\}\end{aligned} \tag{3.3}$$

Para o cálculo da distribuição condicional completa de ϕ_c , devemos notar que ϕ_c aparece na expressão das *prioris* de c, c_+ e na *priori* de ϕ_c , de forma que a condicional completa de ϕ_c é dada por:

$$p(\phi_c|R) \propto \exp \left\{ \begin{array}{l} -\frac{1}{2W_3} \sum_{t=1}^n [c_t - \phi_c (\cos \omega c_{t-1} - \text{sen} \omega c_{t-1}^+)]^2 - \\ -\frac{1}{2W_4} \sum_{t=1}^n [c_t^+ + \phi_c (\text{sen} \omega c_{t-1} - \cos \omega c_{t-1}^+)]^2 \end{array} \right\} \phi_c^{\alpha_{\phi_c} - 1} (1 - \phi_c)^{\beta_{\phi_c} - 1}.$$

²Considere ϕ_p como o passo proposto e ϕ_j como o passo anterior da cadeia

A probabilidade de aceitação do Metropolis é o mínimo entre 1 e:

$$\exp \left\{ \begin{array}{l} -\frac{1}{2W_3} \sum_{t=1}^n [(c_t - \phi_p(\cos\omega c_{t-1} - \text{sen}\omega c_{t-1}^+))^2 - \\ \quad - (c_t - \phi_j(\cos\omega c_{t-1} - \text{sen}\omega c_{t-1}^+))^2] - \\ -\frac{1}{2W_4} \sum_{t=1}^n [(c_t^+ + \phi_p(\text{sen}\omega c_{t-1} - \cos\omega c_{t-1}^+))^2 - \\ \quad - (c_t^+ + \phi_j(\text{sen}\omega c_{t-1} - \cos\omega c_{t-1}^+))^2] \end{array} \right\}.$$

O cálculo da distribuição condicional completa de ω é bem parecido com o de ϕ_c , pois os fatores de que fazem parte só diferem na *priori*. Como a *priori* de ω é uniforme, então,

$$p(\omega|\boldsymbol{\theta}_{-\omega}) \propto \exp \left\{ \begin{array}{l} -\frac{1}{2W_3} \sum_{t=1}^n [c_t - \phi(\cos\omega c_{t-1} - \text{sen}\omega c_{t-1}^+)]^2 - \\ -\frac{1}{2W_4} \sum_{t=1}^n [c_t^+ + \phi(\text{sen}\omega c_{t-1} - \cos\omega c_{t-1}^+)]^2 \end{array} \right\}.$$

Isso nos leva à probabilidade de aceitação para ω dada pelo mínimo entre 1 e:

$$\exp \left\{ \begin{array}{l} -\frac{1}{2W_3} \sum_{t=1}^n [(c_t - \phi_c(\cos\omega_p c_{t-1} - \text{sen}\omega_p c_{t-1}^+))^2 - \\ \quad - (c_t - \phi_c(\cos\omega_j c_{t-1} - \text{sen}\omega_j c_{t-1}^+))^2] - \\ -\frac{1}{2W_4} \sum_{t=1}^n [(c_t^+ + \phi_c(\text{sen}\omega_p c_{t-1} - \cos\omega_p c_{t-1}^+))^2 - \\ \quad - (c_t^+ + \phi_c(\text{sen}\omega_j c_{t-1} - \cos\omega_j c_{t-1}^+))^2] \end{array} \right\}$$

3.3 Primeiro conjunto de dados

3.3.1 Geração dos dados

Usando o *software* R[®], geramos uma amostra de tamanho $n = 100$ do modelo (2.11). Os dados foram gerados levando em consideração as observações feitas na seção 3.1. A seguir, vemos os valores dos parâmetros que usamos para gerar os dados. Note que, como já foi dito, foi necessário especificar valores bem pequenos para W_1 , W_2 , W_3 e W_4 .

Os parâmetros com que geramos os dados são:

$$n = 100,$$

$$\phi = 0.95,$$

$$\phi_c = 0.95,$$

$$p = 12 \Rightarrow \omega \approx 0.5236,$$

$$V = 1.5,$$

$$W_1 = W_2 = \frac{V}{1000},$$

$$W_3 = W_4 = \frac{V}{100},$$

$$\beta_0 = 1,$$

$$c_0 = 5.$$

Com esses parâmetros, as formas dos dados e do ciclo são razoavelmente próximas das desejadas, ou seja, a tendência de crescimento estocástico amortecido com componente de ciclo aparente. A tendência é claramente crescente no começo (até $t \approx 40$) e segue aparentemente constante para o restante da série, mostrando seu comportamento assintótico. O ciclo decai até $t \approx 50$ e sua amplitude não muda a partir desse ponto. As figuras 3.4 e 3.5 mostram esse comportamento. A figura 3.4 mostra os dados gerados. A linha preta representa a série enquanto a linha vermelha representa a componente de média dos dados. A figura 3.5 mostra a componente de ciclo separadamente.

3.3.2 Ajuste

Para o ajuste do modelo, vamos primeiramente assumir que alguns parâmetros são conhecidos. Dessa forma, podemos analisar quais parâmetros interferem na estimação de outros. Como o filtro de Kalman, e conseqüentemente o algoritmo FFBS, depende desses

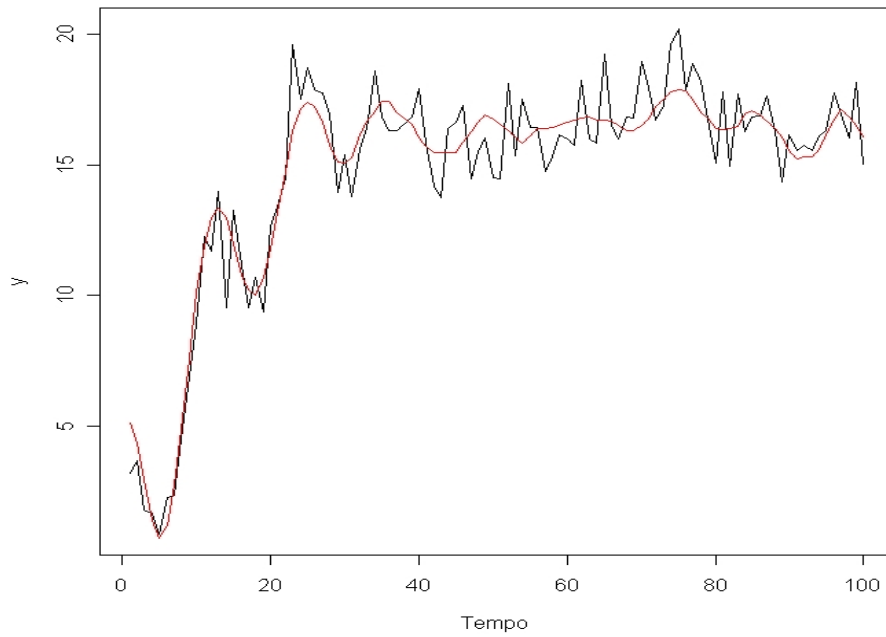


Figura 3.4: Dados gerados artificialmente.

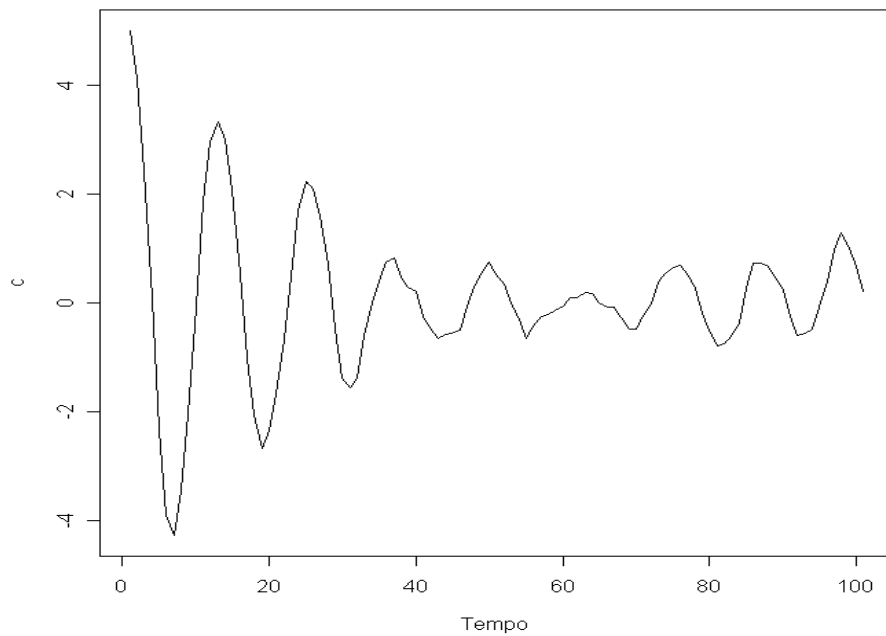


Figura 3.5: Ciclo gerado.

valores, então estudaremos tanto as componentes do modelo quanto seus parâmetros através de MCMC, ou seja, através de suas *posteriors*.

Ajuste com variâncias e ω conhecidos

Vamos assumir inicialmente que conhecemos todas as variâncias e ω , ou seja, só precisamos estimar ϕ , ϕ_c e as componentes μ_t , β_t , c_t e c_t^+ para todo t . Em outras palavras, vamos assumir conhecer as variâncias intrínsecas do sistema e o período do ciclo, faltando estimar o amortecimento do crescimento da média e o decaimento do ciclo. Uma vez encontrada a função de verossimilhança, devemos atribuir informação *a priori* para os parâmetros. Algumas *prioris* podem ser especificadas de forma que as contas sejam facilitadas, mas no caso de ϕ e ϕ_c , isso não é verdade. A distribuição Beta, que varia no intervalo $(0, 1)$ é ideal para modelar esses parâmetros, mas suas condicionais completas não têm forma conhecida, portanto, foi necessário usar passos do método Metropolis-Hastings. As distribuições propostas para o algoritmo foram as próprias *prioris*.

De início, tentamos usar o mínimo de informação possível *a priori*. Para ϕ e ϕ_c , isso significa usar uniformes no intervalo $(0, 1)$, que podem ser encontradas fazendo as *prioris* como $\text{Beta}(1, 1)$. Após as gerações das cadeias de Markov, obtivemos como resultado *a posteriori*, valores para esses parâmetros muito abaixo dos seus valores verdadeiros, em diversos casos (ou configurações), com estimativas abaixo de 0.5. Como os valores desses parâmetros em geral são próximos de 1, poderíamos supor que essas estimativas não são adequadas, mesmo se não tivéssemos seus verdadeiros valores. Decidimos, então, incluir essa informação *a priori*. Para tal, usamos parâmetros para as distribuições Beta tais que a média e desvio padrão *a priori* de ϕ e ϕ_c fossem, respectivamente, 0.9 e 0.1. Todos os resultados que se seguem usam essa *priori*, tanto para ϕ , quanto para ϕ_c .

Uma vez que as amostras das condicionais completas estão sendo usadas no algoritmo FFBS, que por sua vez está gerando amostras da posteriori de $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, \mathbf{c} e \mathbf{c}^+ , temos especificado o algoritmo MCMC. Sendo mais específico, se notarmos $(\mu_t, \beta_t, c_t, c_t^+)$ por

θ_t , uma forma geral do algoritmo é:

1. Iniciar a cadeia com valores arbitrários para os parâmetros e fazer $j = 1$;
2. Passar o filtro de Kalman com os parâmetros da última iteração e amostrar da distribuição de $\theta_n^{(j)}|y^n$;
3. Usar o algoritmo FFBS e amostrar da distribuição de $\theta_t^{(j)}|y^n, \forall t$;
4. Amostrar das condicionais completas dos parâmetros para obter $\phi^{(j)}$ e $\phi_c^{(j)}$;
5. Atualizar $j = j + 1$ e voltar para o passo 2 até conseguir convergência.

Inicialmente, fizemos 25000 iterações, com *burn-in* de 5000 e *thining* de 20, ou seja, adicionamos um ponto à amostra a cada 20 iterações. Assim, tivemos amostras de tamanho 1000 para usar em integrais de Monte Carlo. Todos os cálculos de MCMC foram feitos com a linguagem de programação OxTM.³ Apesar do *thining*, observamos que as cadeias de ϕ e ϕ_c tinham forte auto-correlação. Por isso, aumentamos o número de iterações para 505000 com mesmo *burn-in* e *thining* de 500. Dessa vez, obtivemos uma auto-correlação quase nula para os dois parâmetros.

Uma vez que as amostras *a posteriori* são obtidas, podemos estimar a média dos dados, que chamaremos de ajuste. Para estimar essa média pontualmente, usaremos as médias *a posteriori* de cada ponto t , ou seja, a média de y_t é estimada por $\bar{\mu}_t + \bar{c}_t, \forall t$, em que $\bar{\mu}_t$ e \bar{c}_t são as médias *a posteriori* de μ_t e c_t , respectivamente. Na figura 3.6, apresentamos os dados (linha preta) e o ajuste (linha vermelha).

Como podemos ver por essa figura, o ajuste parece seguir o movimento dos dados adequadamente. Também é fácil ver um movimento cíclico no ajuste. Para melhor visualizar

³O programa que gerou as cadeias se encontram no Apêndice C.

o ajuste do ciclo, podemos ver o ajuste deste como uma componente separada na figura 3.7.

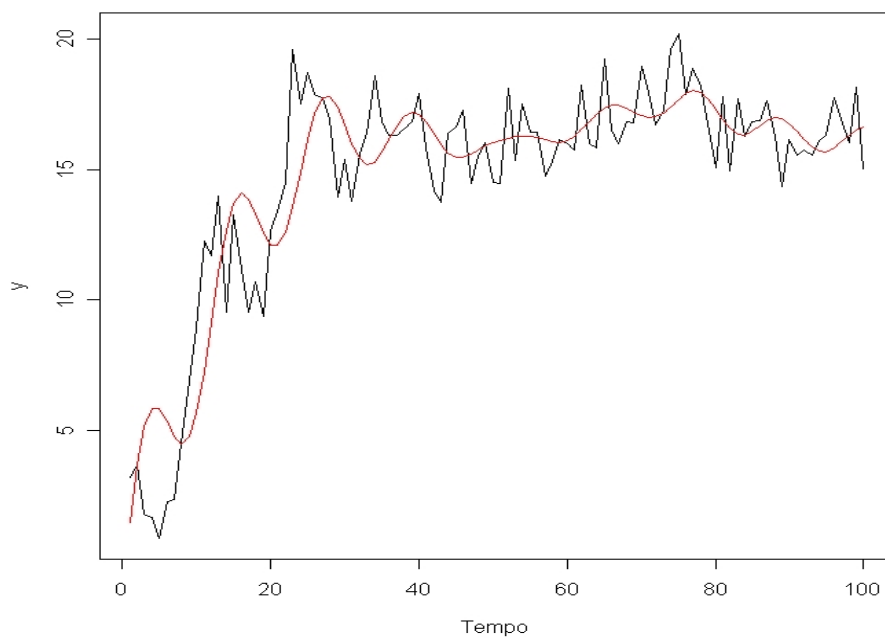


Figura 3.6: Ajuste da série.

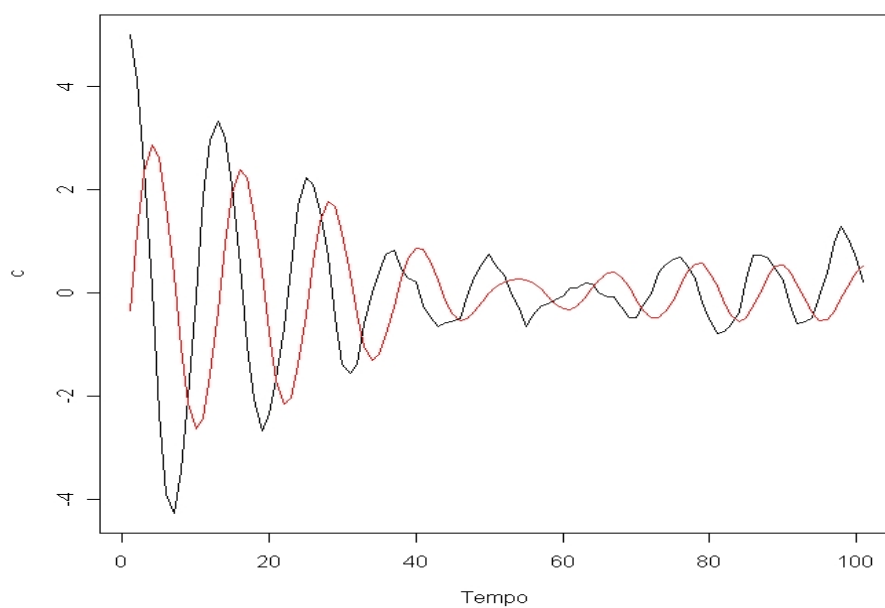


Figura 3.7: Estimação do ciclo.

O ciclo estimado claramente segue os movimentos do ciclo gerado, porém com alguma defasagem. Esse comportamento é esperado, dado que diversos parâmetros estão fixos nos seus valores verdadeiros.

Ajuste com ω conhecido

Numa situação real, o valor das variâncias não é conhecido. Logo para continuar esta análise, vamos assumir que não conhecemos os valores verdadeiros das variâncias, e apenas fixaremos o parâmetro ω no seu valor verdadeiro. É também verdade que não se sabe o valor verdadeiro de ω numa situação real, mas para que possamos ver o comportamento do modelo no ajuste, assumimos conhecê-lo por agora.

No caso das variâncias, podemos usar *prioris* que facilitem os cálculos das condicionais completas por causa da suposição de normalidade, tanto na equação das observações, quanto do sistema. Se usarmos distribuições *prioris* Gama Invertida para as variâncias, então as condicionais completas também terão distribuição Gama Invertida, que são fáceis de amostrar. Sendo assim, é possível utilizar o amostrador de Gibbs para amostrar desses parâmetros.

Como estamos assumindo que não sabemos os valores verdadeiros das variâncias, vamos assumir pouca informação *a priori*, especificando ambos os parâmetros das distribuições Gama Invertida como iguais 1, o que implica em média e variância infinitas. As figuras 3.8 e 3.9 ilustram o ajuste dos dados e a estimação do ciclo, respectivamente.

O que vemos nessas figuras ilustra um fato recorrente nas tentativas de diferentes quantidades de informação *a priori* observado no primeiro e no segundo conjunto de dados, como veremos, ou seja, o conhecimento ou não das variâncias afeta pouco o ajuste dos dados. Podemos ver na tabela a seguir a comparação dos valores verdadeiros das variâncias com seus intervalos de credibilidade 95% *a posteriori*:

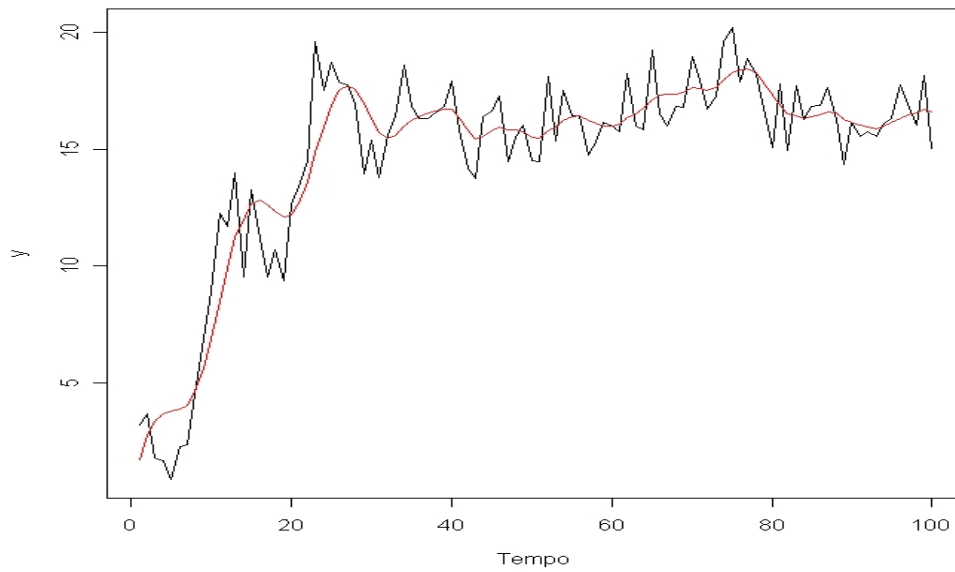


Figura 3.8: Ajuste da série.

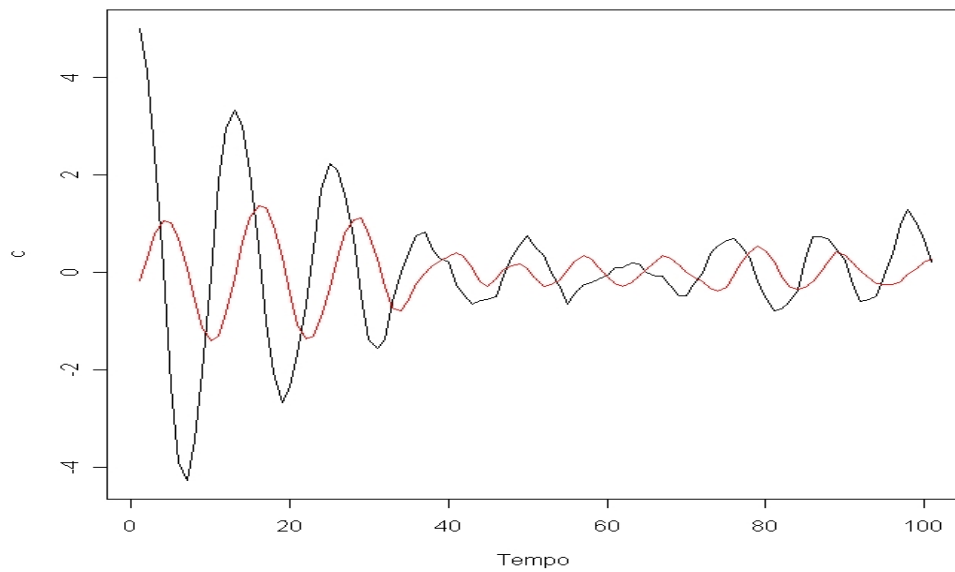


Figura 3.9: Estimação do ciclo.

A tabela 3.1 mostra que todas as variâncias foram superestimadas em todos os casos. No entanto, o ajuste foi quase idêntico ao anterior. A seguir, supomos ω também desconhecido.

Parâmetro	Valor verdadeiro	Limite Inferior	Limite Superior
V	1.5	2.2469	8.6027
W_1	0.0015	0.1347	0.8195
W_2	0.0015	0.0936	0.3829
W_3	0.0015	0.1433	1.3402
W_4	0.0015	0.1198	0.6782

Tabela 3.1: Intervalos de credibilidade 95% *a posteriori* e valores verdadeiros.

Ajuste com variâncias conhecidas

Primeiramente, fixamos as variâncias nos seus valores verdadeiros, e tentamos estimar ω . Como o resultado foi muito parecido com o caso em que tentamos estimar todos os parâmetros, mostraremos somente este.

Como o parâmetro ω varia no intervalo $(0, 2\pi)$, a distribuição *a priori* mais simples que podemos usar para modelá-lo é uma Uniforme que varia nesse intervalo, o que significa que estamos assumindo o mínimo de informação *a priori*. A condicional completa de ω não tem forma conhecida, então usamos o método Metropolis-Hastings com proposta igual à *priori* para extrair amostras de sua *posteriori*. Assim, chegamos ao ajuste visto na figura 3.10

À primeira vista, o ajuste parece captar bem o movimento da série. A figura 3.11 mostra que a componente de ciclo não foi captada pelo modelo. O que observamos é que a estimação do parâmetro ω define fortemente o comportamento do ciclo estimado. Isso é intuitivo, pois ω está diretamente relacionado com o período do ciclo, e se ele é mal estimado, então o modelo pode confundir a tendência com o ciclo.

Tentamos diversas *prioris* para o modelo para estimar ω , mas mesmo quando usamos a

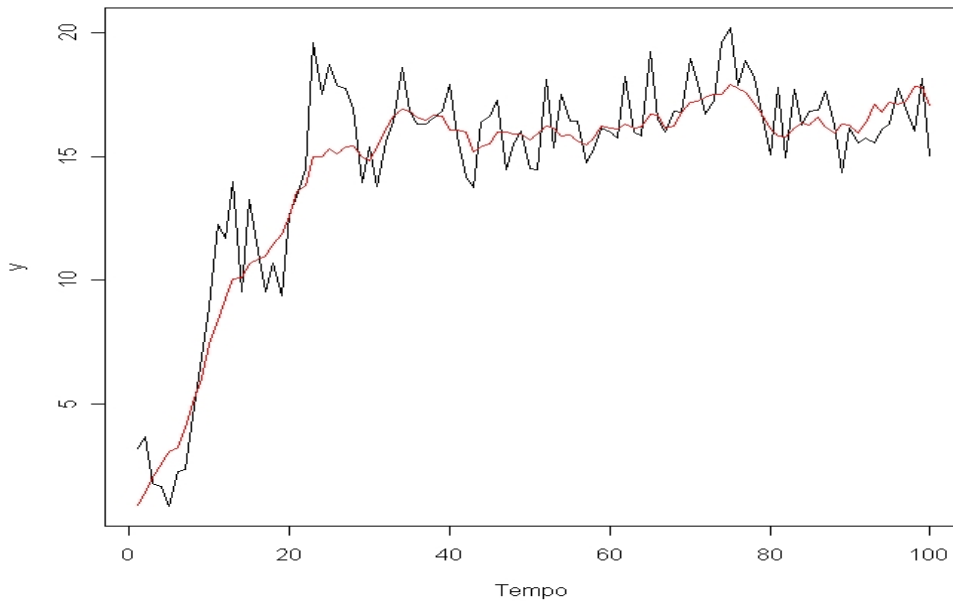


Figura 3.10: Ajuste da série.

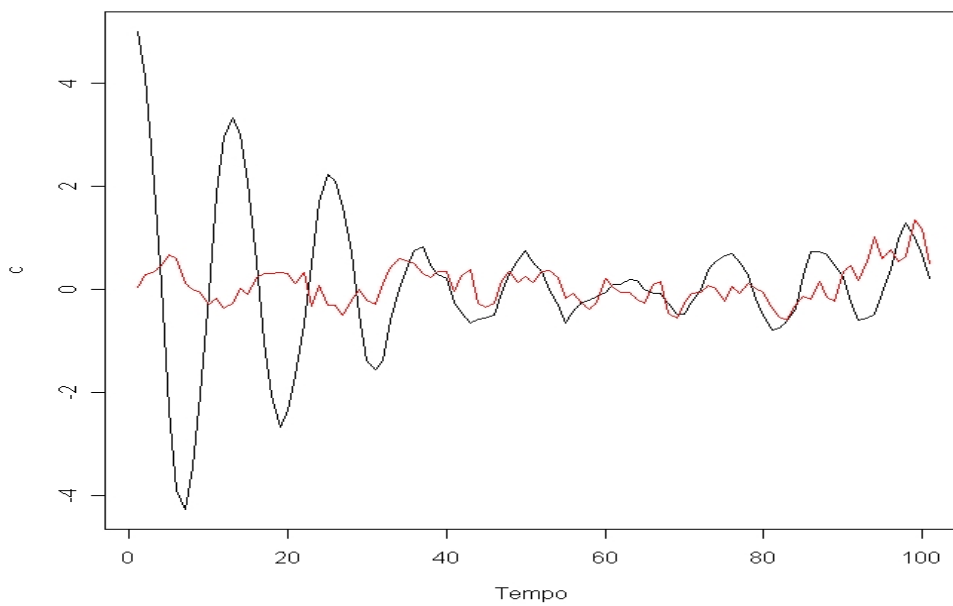


Figura 3.11: Estimação do ciclo.

informação máxima, ou seja, todos os outros parâmetros fixos nos seus valores verdadeiros, o período não foi estimado de maneira conclusiva. Em todos os casos, a distribuição de ω é bimodal, sendo que a maioria dos pontos amostrados da *posteriori* são maiores que 5.8

ou menores que 0.2. Esses valores podem ser traduzidos para períodos do ciclo de ≈ 1.08 e ≈ 31 , respectivamente.

Como o período usado para gerar os dados é 12, sabemos que a distribuição *a posteriori* de ω não está variando próximo de seu valor verdadeiro, o que implica diretamente em que o parâmetro não está sendo bem estimado. Para melhorar isso, podemos fazer modificações na *priori* do próprio ω para tentar aumentar a densidade *a posteriori* em torno do seu valor verdadeiro. Como estamos usando *priori* Uniforme para ω , a única coisa que podemos mudar é o seu intervalo de variação, limitando *a priori* os possíveis valores que o parâmetro pode assumir. O que se espera disso é que, como os valores que a cadeia de Markov assumiam anteriormente não são mais possíveis, deveria-se observar outra forma para a cadeia. É preciso cautela ao se limitar *a priori* dessa maneira, pois estamos potencialmente eliminando o verdadeiro valor do parâmetro da estimação. Por isso, a *priori* mais restrita que usamos foi uma Uniforme entre 0.26 e 3.11. Esses valores equivalem a períodos de ≈ 2 e ≈ 24 . Essa escolha é intuitiva, pois um período menor do que duas unidades de medida é difícil de conceber e um período de até 24 implica em pelo menos 4 repetições da parte cíclica.

Os resultados de todas as *prioris* Uniformes restritas tiveram resultado análogo ao anterior, com distribuição bimodal com focos perto dos limites superior e inferior. Entre as cadeias de Markov de ω resultantes das diferentes *prioris*, a cadeia que melhor ilustra o comportamento de todas elas, inclusive do caso irrestrito é o caso da $U(0.26, 3.11)$, que pode ser vista na figura 3.12.

Mesmo numa situação em que fixamos todos os outros parâmetros nos seus valores verdadeiros (ou seja, $\theta_{-\omega}^{(j)} = \theta_{-\omega}, \forall j$), não pudemos estimar ω eficientemente. Como este trabalho se motiva em problemas de previsão, não nos delongaremos demais na parte de

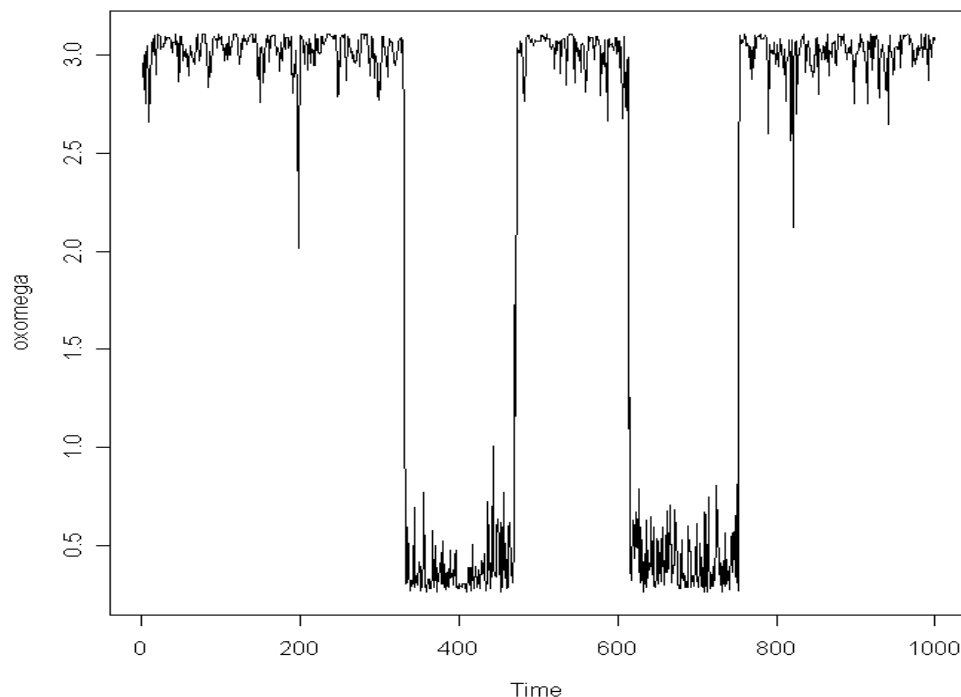


Figura 3.12: Cadeia de Markov de ω .

ajuste.

3.3.3 Previsão

O maior interesse no caso de séries temporais é fazer previsão. Quando estamos tratando de dados simulados, retiramos algumas observações do conjunto de dados \mathbf{y} . Fazemos isso para poder comparar as previsões com os valores verdadeiros e, assim, estudar a capacidade preditiva do modelo. Para este exercício, retiramos as 40 observações finais da série gerada. Amostrando da distribuição preditiva (2.3), podemos fazer inferência sobre os valores a prever. Podemos fazer, por exemplo, para cada ponto a ser previsto, um intervalo de credibilidade de 95%, ao invés de analisar apenas a média. Mais uma vez, assumiremos, a princípio, o conhecimento das variâncias e de ω para facilitar as previsões. Na figura 3.13, a linha preta representa os dados, a linha azul representa o

ajuste dos dados até o tempo $t = 60$, ou seja, até o ponto em que assumimos conhecê-los, e a linha vermelha representa as previsões (linha cheia para a média e linha tracejada para o intervalo de credibilidade de 95%).

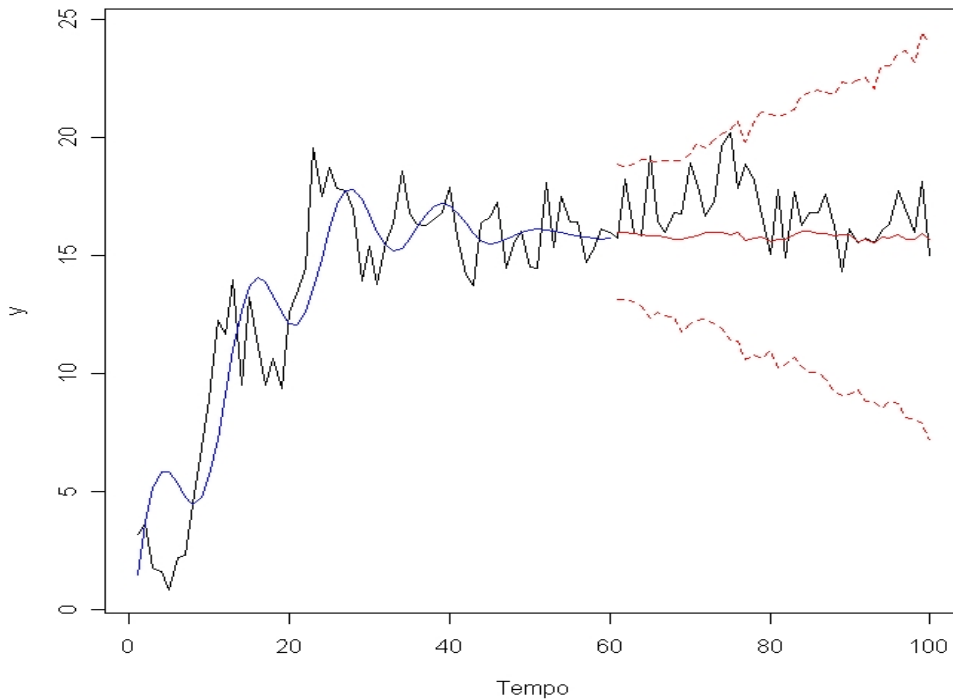


Figura 3.13: Previsão para os últimos 40 dados.

Com a excessão de um ponto, os dados obtidos a partir de $t = 60$ se encontraram dentro do intervalo de credibilidade da distribuição preditiva, o que é esperado para um intervalo 95%. Pode-se dizer que o intervalo é muito largo e que as previsões não assumem um comportamento cíclico, mas vale lembrar, como foi visto na figura 3.1, em $t = 60$, a amplitude do ciclo é menor que 5% da amplitude original. Dessa forma a componente do ciclo já se comporta pouco regida pelo comportamento cíclico, e mais pela componente aleatória $\omega_{3,t}$. Não obstante, note que à medida que se tenta prever mais à frente, ou seja, quanto maior for k ,⁴ maior o intervalo de credibilidade. Esse é um comportamento

⁴Como em 2.10

intuitivo, pois, geralmente, quanto mais à frente se tenta prever, mais imprecisa é a previsão. Em seguida, passamos para o caso em que estimamos as variâncias e só ω é fixo no seu valor verdadeiro. A figura 3.14 ilustra o ajuste e a previsão desse caso.

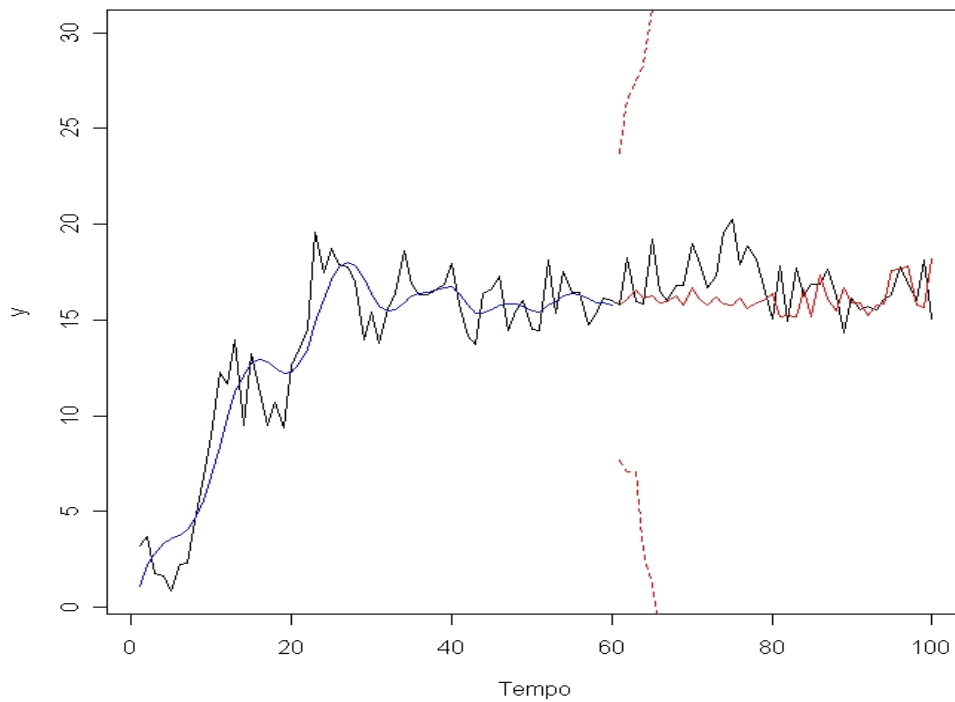


Figura 3.14: Previsão para os últimos 40 dados.

Apesar de, na figura 3.14, a média da distribuição preditiva estar bem próxima dos dados, podemos ver que o intervalo de credibilidade está bastante amplo, tanto que em $t \approx 65$, os limites do intervalo saem da janela de gráfico (decidimos não ampliá-la pois isso dificultaria a visualização dos dados e das previsões pontuais). O significado disso é que o modelo prevê que aquelas observações têm 95% de probabilidade de estar nesse intervalo, ou seja, mesmo que as médias das distribuições preditivas tenham ficado perto dos valores que elas deveriam prever, as incertezas (ou variâncias) associada a essas previsões são grandes.

Esse comportamento das previsões é resultado direto da superestimação das variâncias.

Mesmo com diversas configurações de distribuições *a priori*, esse cenário não melhorou muito, piorando ainda mais quando tentamos estimar ω . Deve-se levar em conta que quando $t > 60$, tanto o efeito do ciclo quanto de β_t estão muito pequenos, restando poucas componentes para influenciar a previsão. As parcelas de variâncias são, portanto, bastante influentes, o que pode causar dificuldades na previsão.

A seguir, geramos outro conjunto de dados para fazer uma análise análoga e complementativa. Assim, encerramos a análise sobre o primeiro conjunto de dados. Tentaremos gerar o segundo conjunto de dados de tal forma que as variâncias não sejam tão influentes quanto no caso dos dados usados nesta seção.

3.4 Segundo conjunto de dados simulados

3.4.1 Geração dos dados

Para continuar o estudo sobre o modelo, geramos outro conjunto de dados com os parâmetros redefinidos em outros valores para poder ter outra base de estudo para o modelo. Para este exemplo, usamos $n = 200$, por isso, como discutido no capítulo 3, na seção sobre a geração dos dados, os valores de ϕ e ϕ_c não podem ser pequenos. Para os valores de W_3 e W_4 , usamos os maiores valores que possível desde que a série mantivesse a sua forma. Os valores escolhidos foram:

$$\begin{aligned}
n &= 200 \\
\phi &= 0.99 \\
\phi_c &= 0.99 \\
p = 20 &\Rightarrow \omega \approx 0.3142 \\
V &= 20 \\
W_1 = W_2 &= \frac{V}{1000} \\
W_3 = W_4 &= \frac{V}{20} \\
\beta_0 &= 5 \\
c_0 &= 30
\end{aligned}$$

Para $\phi_c = 0.99$, a quantidade de informação mantida em $t = n = 200$ é 13%, ou seja, no tempo máximo, ainda há informação significativa da amplitude original. A série e a componente de ciclo vista independentemente resultantes são vistos nas figuras 3.15 e 3.16, respectivamente:

3.4.2 Ajuste

No ajuste deste capítulo, procedemos de forma análoga à do capítulo anterior, e usamos as mesmas *prioris* para a análise.

Ajuste com variâncias e ω conhecidas

Fizemos uma progressão de informação análoga à do capítulo anterior, ou seja, começamos com as variâncias e o período do ciclo conhecidos e fixos nos seus valores verdadeiros. O ajuste dos dados e estimação do ciclo podem ser vistos nas figuras 3.17 e 3.18 a seguir. Na

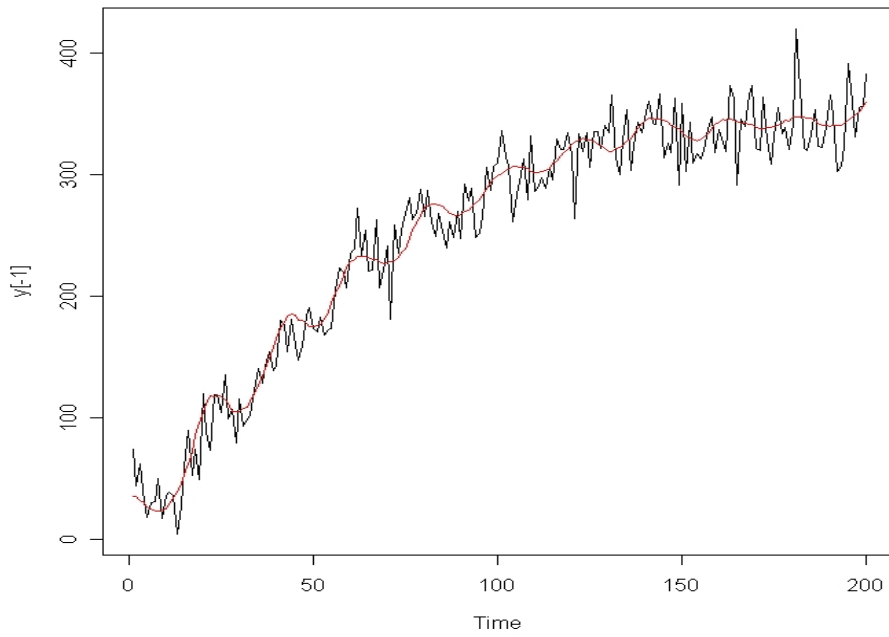


Figura 3.15: Dados gerados artificialmente.

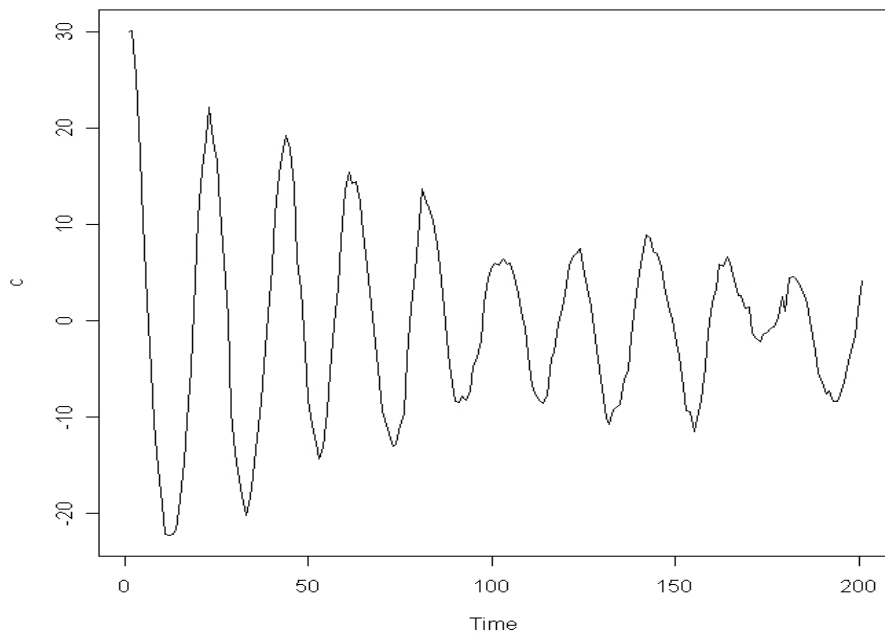


Figura 3.16: Ciclo gerado.

figura 3.17, a linha vermelha representa o ajuste e as linhas tracejadas azuis representam o intervalo de credibilidade 95% da média estimada.

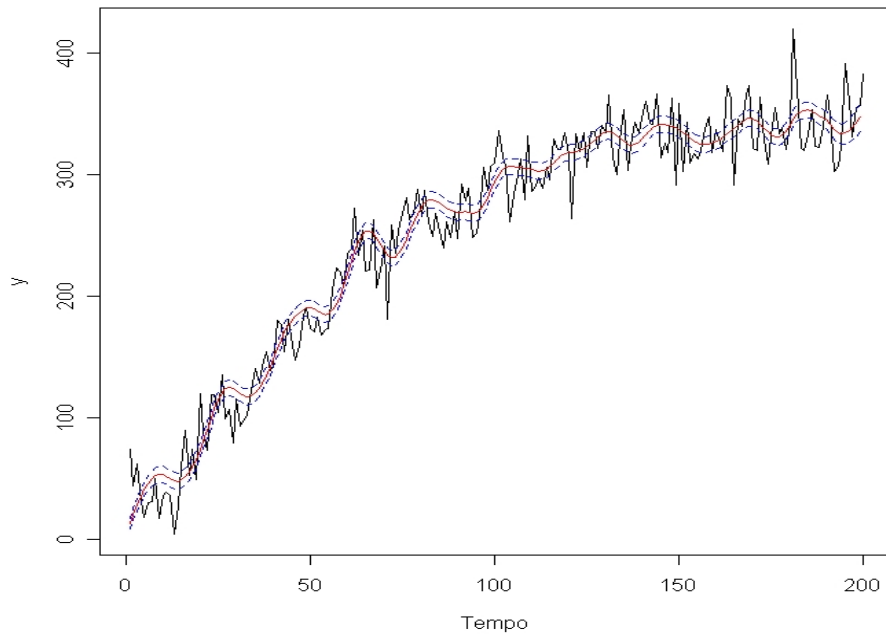


Figura 3.17: Ajuste da série.

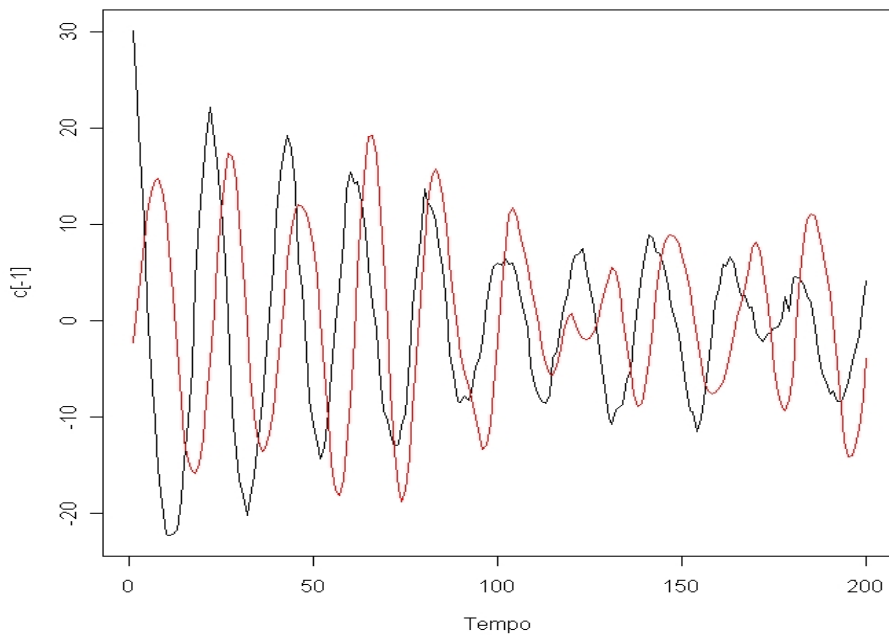


Figura 3.18: Estimativa do ciclo.

Como esperado, o modelo ajusta os dados bem quando os parâmetros são conhecidos.

Vale notar que o intervalo de credibilidade da média é pequeno e os dados em geral não

se encontram dentro dele, mas isso se deve ao fato de que a variância V não está sendo contada nesse intervalo, pois é um intervalo de estimação da média somente.

Ajuste com variâncias conhecidas

Em seguida, fixamos as variâncias nos seus valores verdadeiros, e tentamos estimar ω . Como já foi dito, fizemos o mesmo com o primeiro conjunto de dados, mas os resultados não foram muito diferentes de outra situação. Já neste caso, o ajuste tem um comportamento interessante. Para analisar os resultados, podemos observar a figura 3.19.

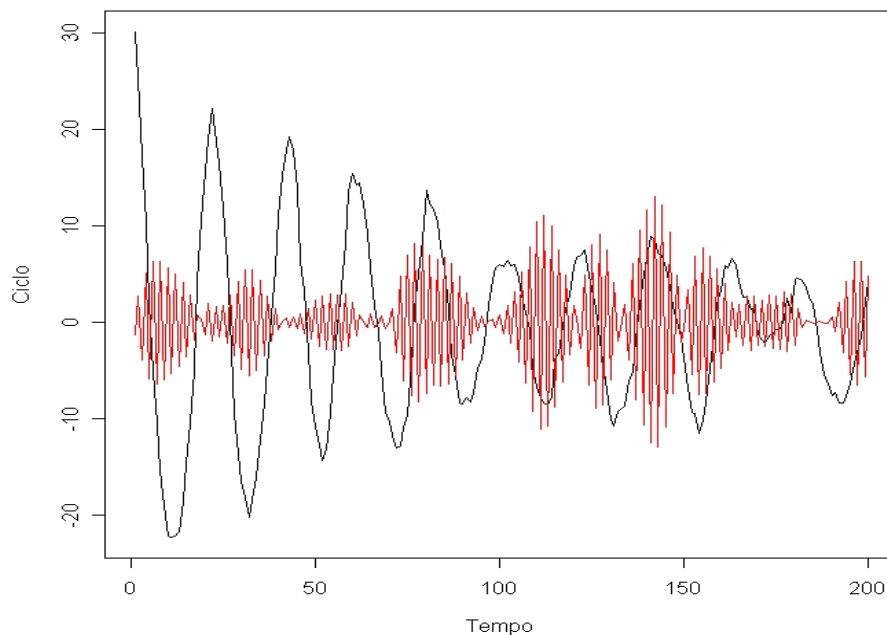


Figura 3.19: Estimativa do ciclo.

À primeira vista, o gráfico da figura 3.19 é de difícil compreensão. A linha vermelha ilustra a tentativa de ajuste da componente de ciclo, porém essa estimativa não teve bons resultados. O período do ciclo é muito sub-estimado e a tentativa do modelo de alcançar a amplitude do ciclo resultou no efeito que é visto na figura, ou seja, sem sucesso nem capacidade de ser interpretado como ciclo. Devido à forma da estimação do ciclo,

a visualização do ajuste é difícil, pois há muitos trechos em que o movimento intenso da estimação do ciclo faz que o gráfico fique confuso. Dessa forma, decidimos mostrar somente a componente μ estimada com os dados na figura 3.20. A figura mostra que a componente de tendência μ ajusta bem a tendência dos dados, porém vimos na figura 3.19 que o ciclo está sendo mal estimado, especialmente o período, que é um parâmetro que também não foi bem ajustado para o primeiro conjunto de dados.

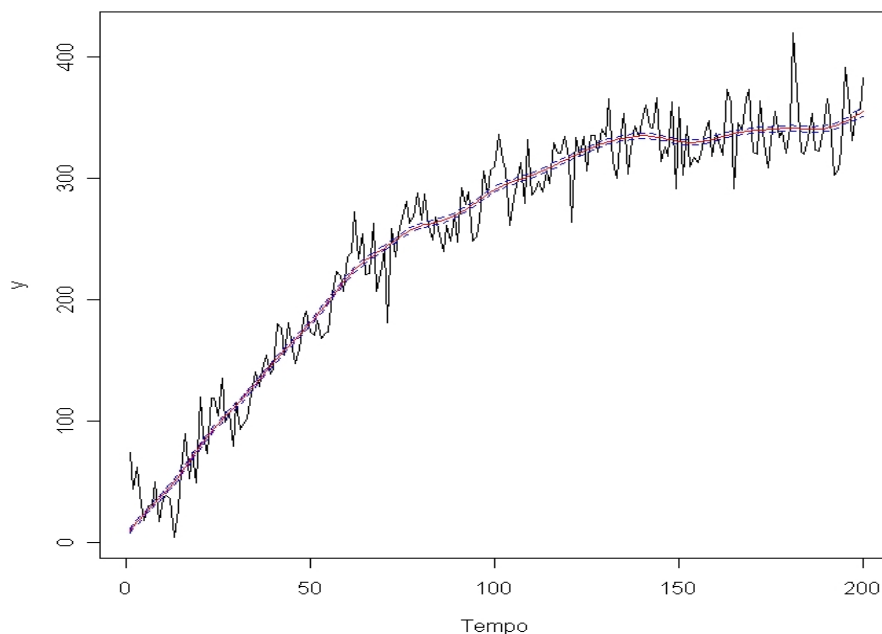


Figura 3.20: Componente de tendência estimada.

Ajuste com todos os parâmetros desconhecidos

Fizemos também o ajuste no caso em que conhecemos somente ω , mas como no primeiro conjunto de dados, o resultado foi bem próximo ao caso em que conhecemos tanto ω quanto as variâncias. Por fim, fizemos o ajuste no caso em que assumimos não conhecer nenhum parâmetro e tentamos estimar todos. Tentamos usar *prioris* com mais informações para as variâncias, com média próxima de seus valores verdadeiros e variância

não muito grande para facilitar a estimação. O resultado desse ajuste pode ser visto na figura 3.21

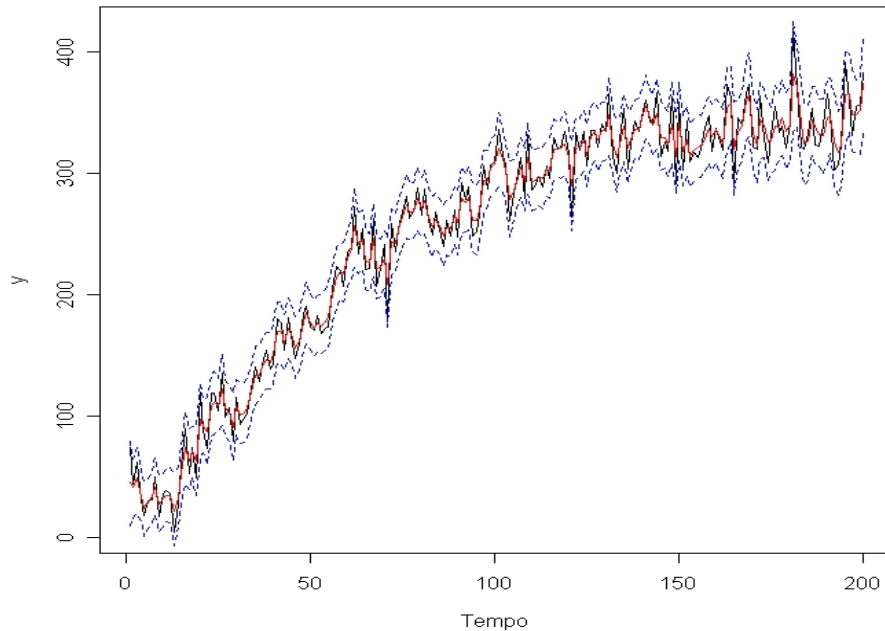


Figura 3.21: Ajuste da série.

Apesar do ajuste não ter sido muito suave, aparentemente os dados foram bem estimados. No entanto, devemos estudar a estimativa do ciclo para melhor julgar o ajuste.

A figura 3.22 mostra que o ciclo não foi bem estimado. Mais uma vez, parece que a má estimação do período do ciclo, ou seja, do parâmetro ω , influi fortemente na estimação da componente de ciclo. Isso foi um impedimento para os ajustes de ambos os conjuntos de dados. Diferentemente do primeiro conjunto de dados, as cadeias de Markov para ω não tiveram comportamento bimodal. Em vez disso, não obtivemos convergência da cadeia em nenhuma das tentativas, mesmo quando fizemos 505000 iterações como no primeiro conjunto de dados. Portanto, sua estimação *a posteriori* foi infrutífera. A cadeia de Markov de ω pode ser vista na figura 3.23 no caso em que fizemos 505000 iterações. Vemos que a cadeia não convergiu. Isso mostra a grande dificuldade em estimar o período

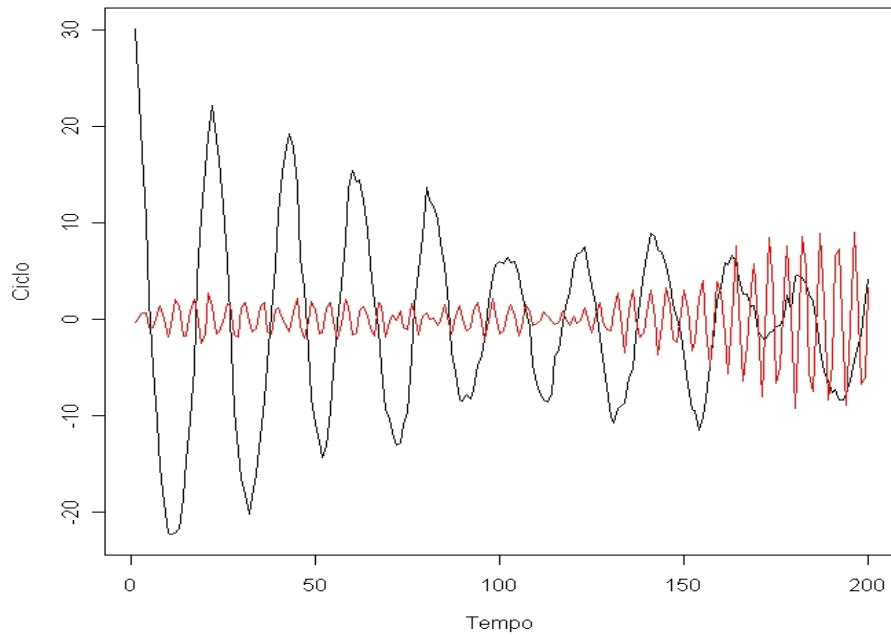


Figura 3.22: Estimativa do ciclo.

de um ciclo em uma série temporal. Falaremos mais sobre os possíveis motivos da não convergência de ω na última seção do capítulo 4.

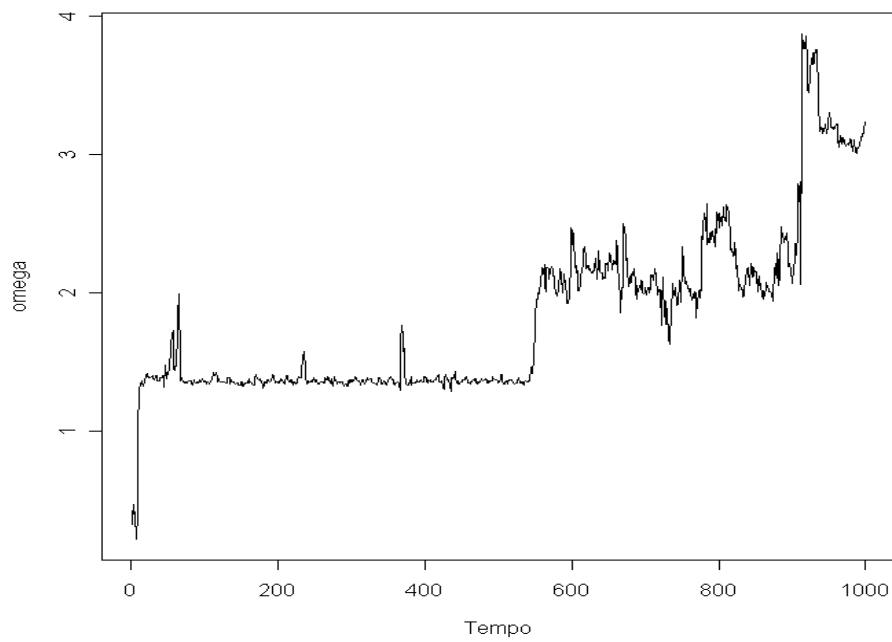


Figura 3.23: Cadeia de Markov de ω .

Para prosseguir com o estudo do segundo conjunto de dados, passamos para as previsões.

3.4.3 Previsão

Retiramos os últimos 50 dados da série para tentar prevê-los. Também começamos com todos os parâmetros V , W_1 , W_2 , W_3 , W_4 e ω fixos nos seus valores verdadeiros, pois, como vimos no capítulo anterior, essa configuração traz o melhor resultado em termos de intervalo de credibilidade. Na figura 3.24, vemos o resultado dessa previsão.

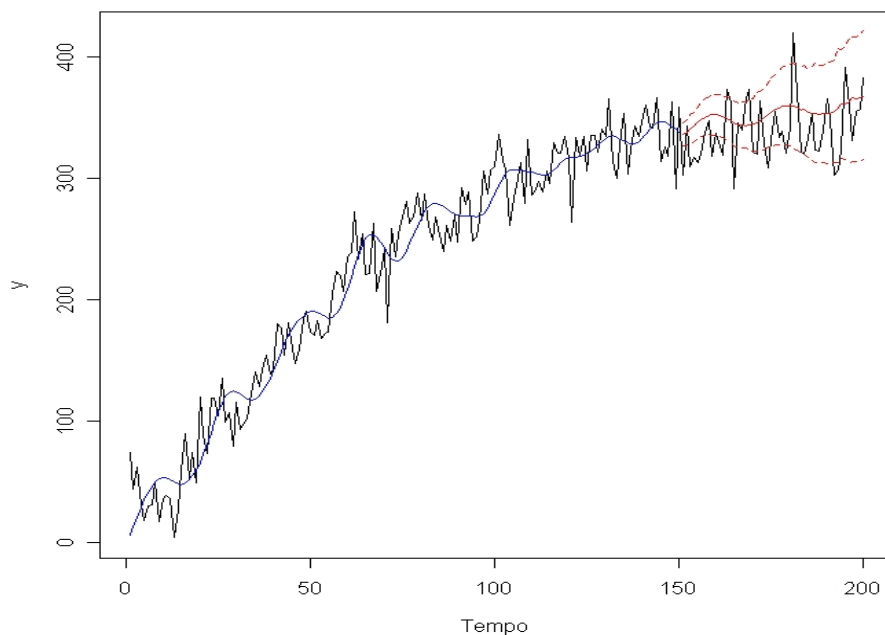


Figura 3.24: Previsão para os últimos 50 dados.

Apesar de os intervalos de credibilidade estarem pequenos, ou seja, há pouca margem de erro para os valores previstos, podemos ver que diversos valores a prever se encontraram fora dos intervalos. Isso significa que, para cada um desses pontos, o modelo previu que os valores tinham menos de 5% de probabilidade de acontecer, logo, se tantos pontos se encontraram fora dos intervalos, as previsões não foram muito confiáveis. No entanto,

antes dos valores previstos, pode-se ver que o ajuste está bastante suave e a componente de ciclo se mostra bem destacada.

É intuitivo pensar que estimar as variâncias pode aumentar os intervalos de credibilidade e, assim, englobar os valores que ficaram de fora. Fazemos então uma tentativa em que só ω está fixo no seu valor verdadeiro. O resultado da previsão está apresentado na figura 3.25

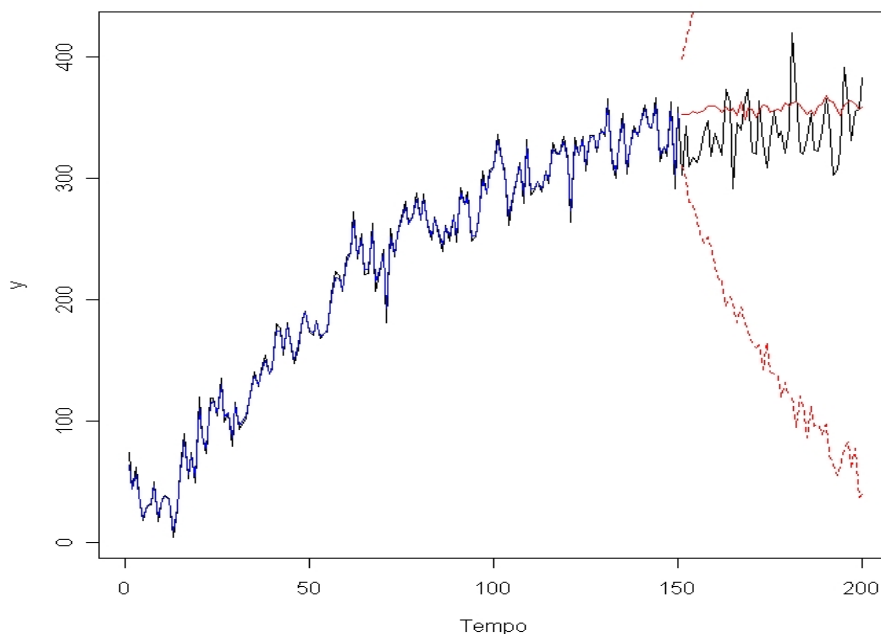


Figura 3.25: Previsão para os últimos 50 dados.

Como vemos na figura 3.25, os intervalos, mais uma vez, ficaram demasiadamente abertos, apesar do ajuste seguir os dados rigorosamente. Apesar de ω estar fixo no seu valor verdadeiro, as previsões não mostraram qualquer comportamento cíclico. Isso levanta algumas questões sobre o estudo do ciclo. Vejamos o ciclo estimado em $t = 1, \dots, 150$ para a configuração em que temos V , W_1 , W_2 , W_3 , W_4 e ω fixos nos seus valores verdadeiros:

Vemos pela figura 3.26 que a amplitude do ciclo está mais próxima da verdadeira,

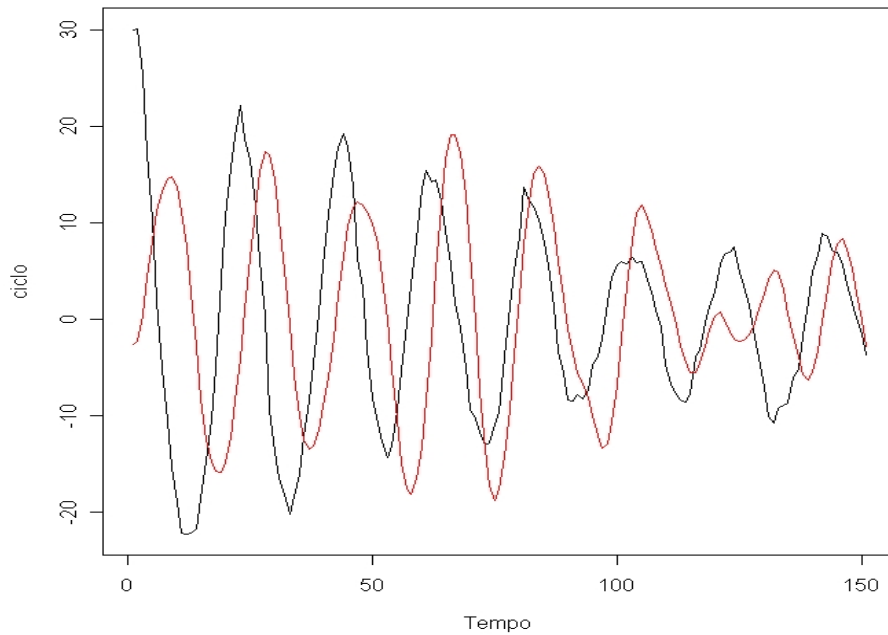


Figura 3.26: Estimativa do ciclo em $t = 1, \dots, 150$.

apesar de estar um pouco defasada. Entretanto, mesmo com as condições tão favoráveis quanto esta última, o modelo não conseguiu estimar bem o início da componente de ciclo. Decidimos, portanto, investigar o motivo disso ter ocorrido.

3.5 Estudo da amplitude inicial do ciclo

Com base no segundo conjunto de dados, estudamos o motivo para a má estimação da amplitude inicial do ciclo. O que define a estimação do início da série são as *prioris* de μ_1 , β_1 , c_1 e c_1^+ , ou seja, a informação que temos em $t = 0$. Portanto, a maneira da qual definimos m_0 e C_0 (usados em (2.6) no filtro de Kalman) dá ao modelo maneiras de prever as componentes de maneira mais eficiente. Como definimos $m'_0 = c(0, 0, 0, 0)$ e

$$C_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
 então era muito pouco provável que o modelo conseguisse estimar

os valores iniciais β_0 e c_0 corretamente, pois nos dois conjuntos de dados gerados, os valores verdadeiros desses parâmetros foram grandes. Por exemplo, no caso do segundo conjunto, tivemos $\beta_0 = 5$ e $c_0 = 30$, logo tentamos definir C_0 de forma que, *a priori*, esses valores

tivessem densidade mais alta. Mais especificamente, definimos $C_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 300 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

Dessa maneira, decidimos refazer a previsão com nenhum parâmetro fixo no seu valor verdadeiro, ou seja, estimamos todos eles e verificar o quanto a mudança de C_0 influi nas estimações e na previsão. Podemos ver a estimativa do ciclo com essa configuração na figura 3.27

Podemos ver que, mesmo com uma configuração de *prioris* que inicialmente resultou em estimativas ruins, o aumento da variância *a priori* dos valores iniciais da série melhora consideravelmente o ajuste do ciclo. Note que, nesse caso, a cadeia de Markov de ω também não convergiu, e, apesar disso, o modelo ajustou o ciclo de forma comparável ao caso em que ω é conhecido e fixo no seu valor verdadeiro. As previsões dos últimos 50 dados, por sua vez, mostram grande melhora também, como é visto na figura 3.28.

Mesmo que as previsões, neste último caso, não tenham um comportamento cíclico, podemos ver que os intervalos de credibilidade das previsões estão nitidamente menores do que no caso de sua contra-parte no modelo em que C_0 era uma matriz identidade. No caso, uma boa estimativa de ω poderia dar um ciclo estimado mais próximo de seu valor

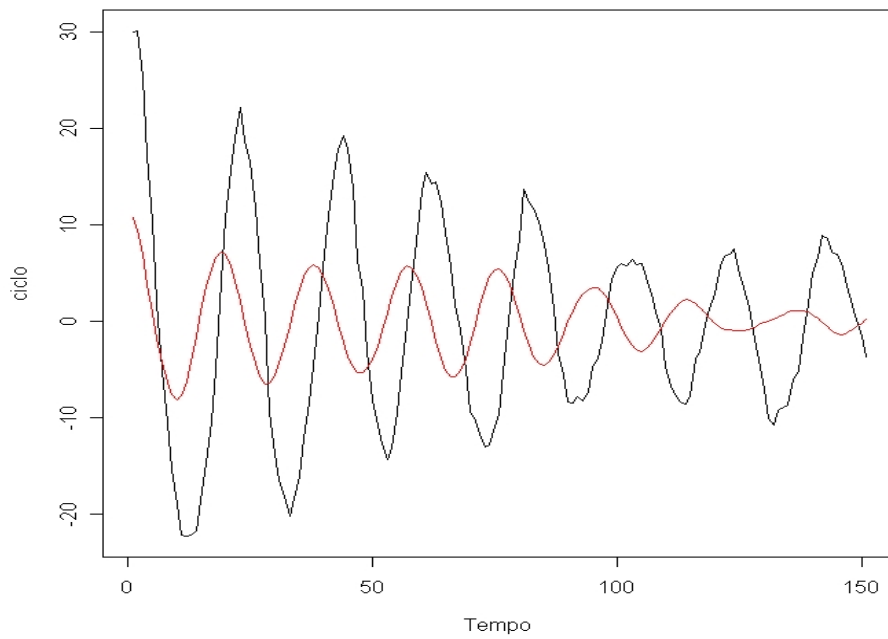


Figura 3.27: Estimativa do ciclo em $t = 1, \dots, 150$.

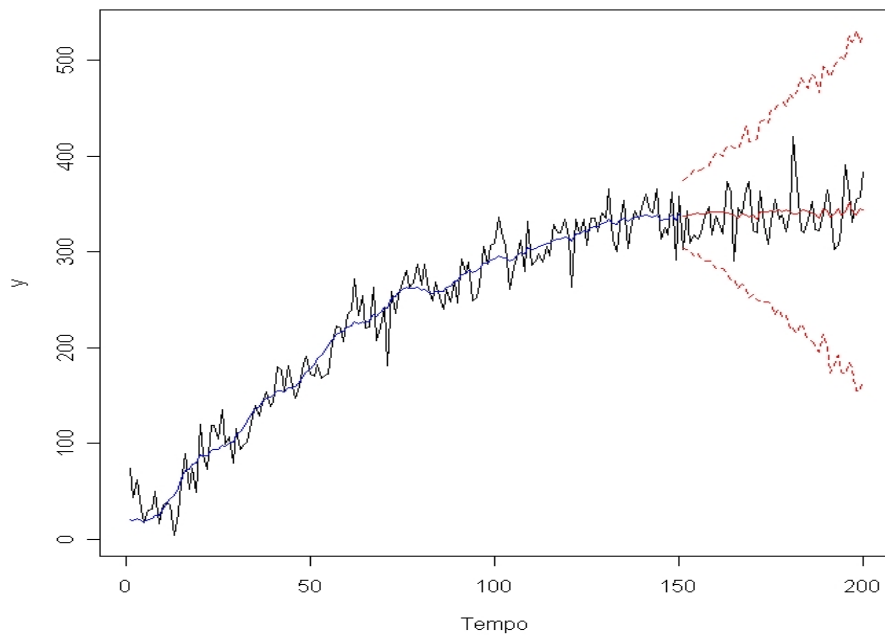


Figura 3.28: Previsão para os últimos 50 dados.

real e, talvez, melhorar as previsões. Certamente, isso abre possibilidades para discussões sobre essas estimações.

Por enquanto, podemos saber o resultado com um período bem estimado, pois conhecemos o valor verdadeiro do ω . Logo podemos fixá-lo em seu valor verdadeiro e refazer a previsão. Neste caso, todos os parâmetros foram estimados como no caso anterior, exceto por ω , que foi fixo no seu valor verdadeiro. Podemos ver as previsões na figura 3.29 e a estimação do ciclo na figura 3.30.

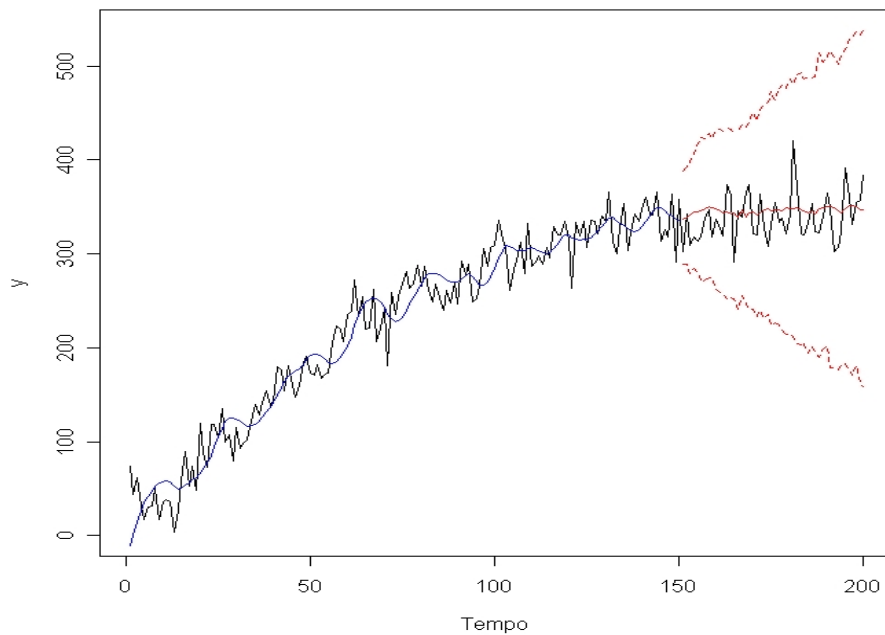


Figura 3.29: Previsão para os últimos 50 dados.

Podemos ver que com ω conhecido, as previsões não mudam muito comparadas ao caso em que estimamos o parâmetro. Entretanto, a estimação do ciclo mostra que a amplitude inicial do ciclo foi bem estimada, o que, somado com o conhecimento prévio do período, nos dá um ciclo bem estimado. Porém, podemos notar que a estimação do ciclo está um pouco defasada com relação ao período verdadeiro. Por fim, é fácil perceber que o conhecimento de ω foi de grande valia para a estimação do ciclo, portanto sua estimação é importante nos casos em que não o conhecemos previamente. Como já foi dito, podem-se seguir discussões sobre essa estimação, tanto com relação à restrição de seu intervalo

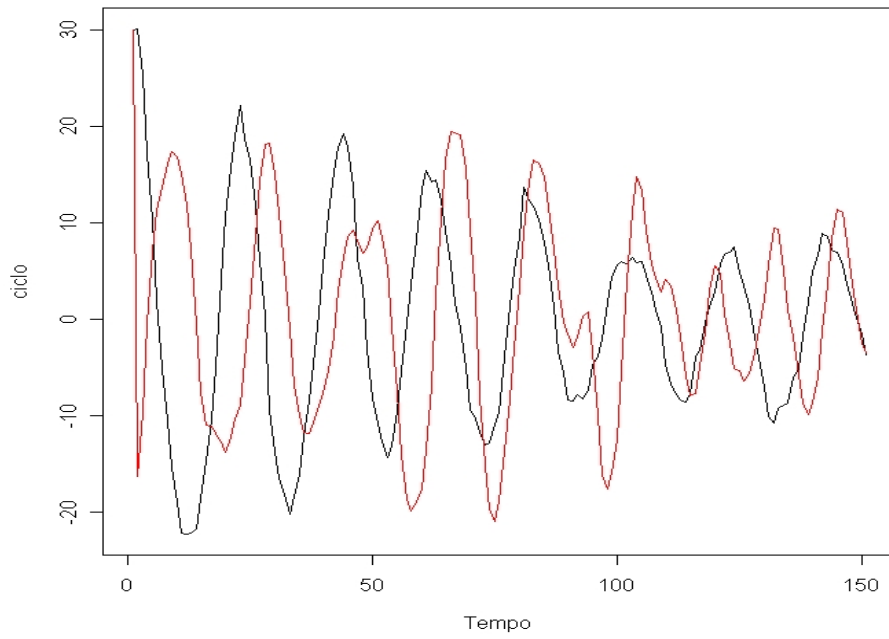


Figura 3.30: Previsão para os últimos 50 dados.

de variação, quanto sobre outras *prioris* que não a Uniforme.

Capítulo 4

Conclusão e continuidade

4.1 O problema

Os modelos dinâmicos são uma classe de modelos que dá a quem os usa uma ferramenta poderosa de modelagem. Entre suas vantagens estão a sua ampla versatilidade e a possibilidade de modelar séries não estacionárias sem a necessidade de transformação de dados, pois pode-se adicionar componentes independentes, sendo elas, por exemplo, uma componente de média, e outra, uma componente estacionária. O estudo dessa classe de modelos sob o ponto de vista Bayesiano adiciona ainda outras vantagens, das quais podemos destacar o uso de informação cuja fonte não se limita aos dados.

Neste trabalho, tentamos estimar um modelo com uma componente de tendência de crescimento amortecido e de ciclo, esta com a característica de se comportar como uma componente sazonal, com a diferença de que o efeito do ciclo decai com o passar do tempo. Para este trabalho, nos inspiramos no estudo que Azevedo *et al.* (2006) fizeram com séries financeiras da região do Euro para previsões a longo prazo. Nesse estudo, os autores usam valores arbitrários para alguns parâmetros, por exemplo, a periodicidade do

ciclo. O valor escolhido por eles, de fato, basea-se em algumas suposições fundadas em seu trabalho, entre outros. Em contrapartida, acreditamos que se esse período for introduzido como um parâmetro desconhecido, ele pode ser estimado. Entretanto alguma informação *a priori* deve ser incluída para que essa estimação seja feita de forma satisfatória.

Com o uso de modelos dinâmicos, e principalmente, análise Bayesiana, essa abordagem pode ser simplificada. Através do conceito de atualização da informação da *priori* para a *posteriori* via teorema de Bayes, podemos fazer inferência de forma consistente sobre o parâmetro. Como é visto neste estudo, entretanto, essa estimação não é trivial.

4.1.1 Geração dos dados

O primeiro impedimento que alcançamos foi, também, o mais inesperado. O nosso estudo foi todo feito com base em dados simulados artificialmente para estudar as capacidades de ajuste e predição do modelo. Em geral, a etapa de geração dos dados acontece rapidamente e sem maiores preocupações. No nosso caso, vimos que os parâmetros com os quais se gera a série podem, muito facilmente, distorcer a forma dos dados, frequentemente tornando-os impróprios para a análise desejada. Muito tempo e esforço foram gastos nessa etapa, até que pudemos encontrar todas as dependências que influem nessa geração, em que a última etapa foi a especificação dos valores iniciais das componentes.

Ao encontrar as relações a que os parâmetros precisam obedecer, dois conjuntos de dados foram simulados. As duas séries foram geradas com parâmetros diferentes, mas mantiveram as relações discutidas no final do terceiro capítulo deste trabalho.

4.1.2 Ajuste

Muitos problemas em inferência estatística têm como foco o ajuste e previsão, uma vez decidido o modelo. Para o estatístico Bayesiano, isso está intimamente relacionado com a escolha das distribuições *a priori* dos parâmetros. Neste trabalho, usamos diversas configurações de distribuições *a priori* para tentar fazer com que nosso modelo pudesse, de forma satisfatória, se ajustar aos dados gerados. Fizemos estimações que usaram inicialmente a suposição de riqueza de informação sobre o modelo, e progredimos para o caso em que a informação *a priori* era mínima. Como fizemos o mesmo estudo com dois conjuntos de dados, podemos comparar os resultados alcançados nos dois casos.

Vimos que no caso do ajuste, o conhecimento ou a boa estimação das variâncias foi de pouca relevância para a inferência, enquanto na previsão, conhecer as variâncias influenciou bastante o intervalo de credibilidade dos valores a prever. No segundo conjunto de dados, vimos que *prioris* mais informativas melhoravam essa estimação. Como sabemos que a variância das observações costuma ser maior que as das componentes, pudemos especificar *prioris* que levassem essa informação em consideração, assim como no caso de ϕ e ϕ_c , o conhecimento *a priori* de que eles assumem valores próximos de 1 foi muito importante para sua estimação. Em todos os casos, a distribuição *a priori* que suportava esse fato foi instrumental em sua estimação, que passou a ser sempre muito próxima dos valores verdadeiros.

A estimação do período, por sua vez, foi insatisfatória. No primeiro conjunto de dados, a cadeia de Markov cuja distribuição limite era a *posteriori* de ω , teve comportamento bimodal em todas as tentativas. No caso do segundo conjunto de dados, a cadeia de Markov não obteve convergência em nenhuma das configurações.

Como para os outros parâmetros, uma solução simples para isso seria usar uma *pri-*

ori que não fosse uma Uniforme em $(0, 2\pi)$, ou seja, uma *priori* mais informativa. No entanto, há uma falta de distribuições que tenham o intervalo $(0, 2\pi)$ como domínio, e o desenvolvimento de uma está fora do escopo desta dissertação. No entanto, há algumas considerações a se fazer com relação à interpretação de dados periódicos sob o ponto de vista físico.

4.2 Periodograma: um Estudo pela Transformada de Fourier

4.2.1 O periodograma

O desenvolvimento de modelos matemáticos para estudar fenômenos da natureza e da humanidade é já intrínseco nas ciências da matemática e da natureza devido à já tradicional maneira de concepção de problemas dessas origens. Essa modelagem é, por princípio, uma idealização do que seria a representação fiel do que é observado no universo em aspectos matemáticos, ou, até um certo ponto, filosóficos. No entanto, por mais que se tente, a matemática não é capaz de incorporar todos os aspectos e formas externas por causa da sua própria natureza, ou seja, a de uma linguagem fechada. Mais do que isso, a estimação de modelos matemáticos incorpora incerteza a eles que é naturalmente propagada para a sua idealização. Porém, como pode ser visto em Priestley (1989), o estudo através de análise espectral pode trazer algumas melhorias nesse aspecto.

O nosso modelo assume, além dos aspectos estatísticos, como normalidade, que os dados são extraídos em intervalos constantes Δt iguais a 1. Além disso, há um intervalo de tempo ΔT do qual os dados são extraídos, que, se o tempo inicial for 0, então temos que $\Delta T = n\Delta t$. Esses dois fatos, mesmo que sejam corretamente assumidos no modelo, geram

alguns efeitos. Esses efeitos, não só podem, como efetivamente atrapalham o estudo sobre período ou frequência nos dados. Esses efeitos podem ser estudados através da função conhecida como periodograma. Os dados gerados pelo nosso modelo contêm um efeito periódico associado a uma frequência, a qual chamaremos de frequência natural (dos dados). Ao estudar o periodograma, veremos que, na própria forma com que os dados são amostrados (ou gerados, no caso), surgem algumas frequências diferentes da natural. Essas falsas frequências não só dificultam a estimação da frequência natural, como podem fazer com que a estimação seja focada sobre elas, em detrimento da frequência natural. Como consequência, podemos ter uma estimação que não se foca num valor só, como foi visto no capítulo 4.

Para entender isso, vejamos primeiro a série de Fourier. Uma função $f(x)$ pode ser escrita da forma:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \operatorname{sen}(nx), \quad (4.1)$$

em que a_0 , a_n e b_n são constantes. É fácil ver que funções harmônicas estão intimamente ligadas à série de Fourier. Por exemplo, digamos que $f(x) = a \cos 2x + b \operatorname{sen} 2x + c$, com a , b e c reais. É direto que, em (4.1), $a_0 = c$, $a_n = a$ e $b_n = b$, para $n = 2$, e $a_n = b_n = 0$, para $n \neq 2$. Se b_n puder assumir valores complexos e se renomearmos n como x e x como k em que k e x são contínuos, vemos que pode haver relação com a transformada de Fourier:

$$F(k) = \int_{-\infty}^{+\infty} f(x) e^{ikx} dx, \quad (4.2)$$

em que i é tal que $i^2 = -1$ ¹. Diz-se que a transformada de Fourier representa um mapeamento da função $f(x)$ no espaço de frequências. O módulo de $F(k)$, ou seu quadrado, é

¹Mesmo que não seja importante para o argumento, é interessante notar que, se $f(x)$ é uma função de densidade de probabilidade, então sua transformada de Fourier é a função característica da distribuição de x

também chamado de Espectro de Potências, ou Periodograma.

É de interesse para esta discussão estudarmos a transformada de uma classe específica de função, conhecida como “trem de pulso”. Seja a função:

$$f(x) = \begin{cases} 1, & \text{se } x \in (x_0 - e, x_0 + e) \\ 0, & \text{caso contrário} \end{cases}$$

em que x_0 é um valor qualquer no domínio de f e $e < \infty$. A transformada de Fourier dessa função, usando a definição (4.2) é:

$$F(k) = \int_{x_0-e}^{x_0+e} e^{ikx} dx = \frac{e^{ik(x_0+e)} - e^{ik(x_0-e)}}{ik}.$$

Desenvolvendo essa expressão, temos

$$F(k) = -\frac{2\text{sen}ke}{k}(\text{cos}kx_0 - i\text{sen}kx_0),$$

que nos leva ao periodograma:

$$P(k) = \left| \frac{2\text{sen}ke}{k} \right|. \quad (4.3)$$

Note que o periodograma não depende de x_0 . A figura 4.1 mostra o comportamento do periodograma de uma função “trem de pulso” com $e = 1$ em escala logarítmica.

Seja c uma constante escalar tal que $\int_{-\infty}^{+\infty} cP(k)dk = 1$, logo podemos dizer que $cP(k)$ é a função de densidade de probabilidade das frequências geradas pela função “trem de pulso”.

4.2.2 Efeito do intervalo de tempo

Como já foi discutido, a própria forma como os dados são obtidos causa um efeito que dificulta a estimação da frequência natural da série. Nesta seção e na próxima, estudaremos os efeitos causados pelo fato de que os dados são obtidos num intervalo limitado de tempo e de forma discretizada.

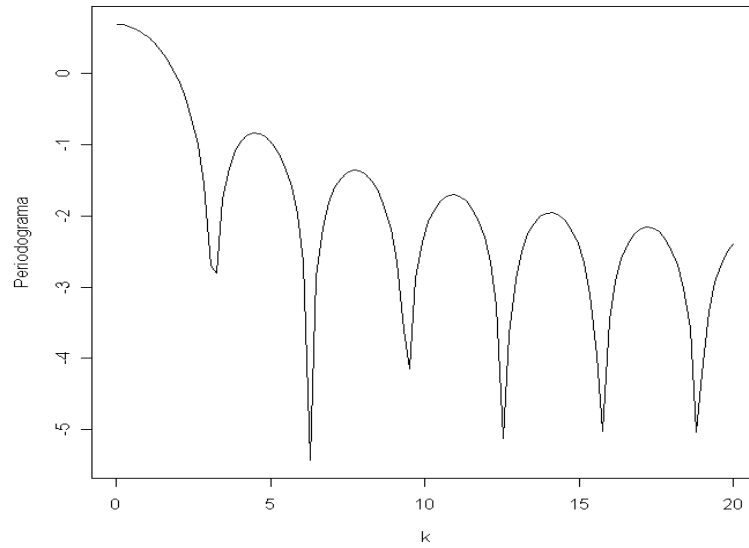


Figura 4.1: Periodograma de uma função “trem de pulso” com $e = 1$ em escala logarítmica.

Um conjunto de dados obtidos no tempo, qualquer que seja ele, será obtido num período limitado de tempo. Digamos que, a título de exemplo, dados de um certo fenômeno são obtidos de forma contínua e com variância nula durante um espaço de tempo ΔT . Além disso, digamos que os valores obtidos para todas as medidas seja unitário. Em outras palavras, temos uma função em que seu valor é unitário para todo o domínio da medida. Lembremos, mais uma vez, que os valores obtidos têm erro nulo. Essa função pode ser entendida como uma função “trem de pulso” vista acima, e seu periodograma pode ser colocado em (4.3) fazendo $e = \frac{\Delta T}{2}$:

$$P_W(k) = \left| \frac{2 \operatorname{sen} k \frac{\Delta T}{2}}{k} \right|. \quad (4.4)$$

Essa função também é chamada “Função Janela” ou “Efeito Janela” (*windowing*).

4.2.3 Efeito do intervalo de amostragem

Cada dado obtido numa série temporal é amostrado, isto é, a função experimental é discretizada nos pontos $j\Delta t$, com j inteiro e Δt o intervalo da amostragem (que no caso

dos nossos dados é 1). Isso é equivalente a associar uma função pulso a cada valor medido. Digamos que uma medida tenha valor y_j , em que j é o índice de um ponto qualquer no domínio de amostragem. O intervalo na abcissa em que a função pulso associada a y_j não é 0 será $(t_j - \frac{\Delta t}{2}, t_j + \frac{\Delta t}{2})$, sendo t_j o tempo em que y_j foi amostrado. Consideramos, para efeito de exemplo, que a série $y_j, j = 1, \dots, n$ não contenha qualquer período. O periodograma decorrente das medidas será uma combinação linear de funções do tipo (4.3):

$$P_y(k) = \left| \frac{2 \operatorname{sen} k \frac{\Delta t}{2}}{k} \right| \sum_j y_j.$$

4.2.4 Estudo *a priori*

Vimos que, apenas proveniente das condições de obtenção dos dados, encontramos um periodograma que é a decorrente dos efeitos provenientes do intervalo de tempo em que os dados foram obtidos, assim como sua discretização amostral. Esses efeitos criam freqüências falsas que dificultam, ou em alguns casos impossibilitam, a estimação da freqüência natural do sistema.

Através dos periodogramas resultantes, podemos combiná-los e, assim, ter uma idéia de quais freqüências podem ser falsas e que atrapalham a estimação. Podemos, então, incluir esse conhecimento como uma informação *a priori*. Mais especificamente, se, de alguma forma, invertermos o periodograma para que, *a priori*, as probabilidades das falsas freqüências sejam diminuídas, poderíamos encontrar mais facilidade ao estimar a freqüência natural.

Para entender isso melhor, se visualizarmos o gráfico do periodograma como sendo uma *priori* “invertida”, ou seja, com as regiões de maiores ordenadas como sendo de menor densidade, então teríamos uma *priori* que ajudaria a, digamos, filtrar as freqüências falsas

de forma a facilitar a estimação. Com a álgebra adequada, poderíamos encontrar uma distribuição *a priori* que incluisse essa informação.

Formalidade teórica

O estudo do periodograma assume uma série estacionária. Apesar de que isso claramente não é verdade para o nosso caso, acreditamos que esse tipo de análise ainda é válida para o caso de séries não estacionárias.

4.3 Projetos Futuros

Conseguimos através deste trabalho propor um modelo que pode ser aprimorado para poder prever ciclos econômicos a longo prazo. Há algumas coisas que podemos fazer para dar continuidade a esse processo.

Uma possibilidade para um projeto futuro seria incluir uma componente de sazonalidade separada do ciclo, como é observado em dados econômicos, ou incluir covariáveis e uma resposta multivariada como é feito em Azevedo *et al.* (2006). É claramente uma maneira mais realista de modelar os dados e pode incrementar as interpretações que se dá aos parâmetros, além de se poder fazer previsões com os dados reais, ao invés de dados simulados.

Por outro lado, podemos levar adiante o estudo de periodogramas e incrementar a estimação de ω via distribuição *a priori*. Isso, não só melhoraria bastante o estudo sobre o modelo que propomos, mas poderia também possibilitar uma maneira de estimar frequências, abrindo assim, uma nova porta para o progresso da estimação Bayesiana. Muitos cálculos e verificações hão de ser feitas para esse estudo. O maior desafio será, certamente, o caso de mostrar que a *a priori* por periodograma é válida mesmo quando a

série não é estacionária, sem necessidade de transformar os dados.

Supondo que estejamos corretos nessa suposição, é fácil entender que a estimação da frequência da série será melhorada. Não só isso, mas poderemos abrir mais uma possibilidade de vertentes de trabalho baseados na inferência Bayesiana, o que é, no mínimo, inspirador.

Referências Bibliográficas

- [1] Azevedo, J. V., Koopman S. J. e Rua A. (2006). Tracking the business cycle of the Euro area: a multivariate model-based band-pass filter. *Journal of Business and Economic Statistics* Vol. 24, 278-290.
- [2] Gamerman, D., Lopes H. F. (2006) *Markov Chain Monte Carlo*, 2nd ed. Boca Raton: Chapman & Hall/CRC
- [3] Gelfand, A. e Smith, A. M. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- [4] Geman, S. e Geman, D. (1984). Stochastic relation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, 721-741.
- [5] Harrison, P. J. e Stevens, C. F. (1976). Bayesian Forecasting. *Journal of Royal Statistical Society, Series B*, 38, 205-247.
- [6] Hastings, W. K. (1970). Monte Carlo samplings using Markov chains and their applications. *Biometrika*, Vol. 57, 97-109.
- [7] Jaynes, E.T. (1986), *Maximum Entropy and Bayesian Methods in Applied Statistics*, p. 1, James H. Justice.

- [8] Metropolis, N. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953). Equation of station calculations by fast computing machine. *Journal of Chemical Physics*. Vol. 21, 1087-1091.
- [9] Migon, H. S., Gamerman, D., Lopes, H. F. e Ferreira, M. A. R. (2005) Bayesian Dynamic Models. *In Dey, D. and Rao, C. R. (Eds.) Handbook of Statistics*, Vol. 25, Ch. 19, 557-592, ISBN 0169-7161
- [10] Migon, H. S. e Gamerman, D. (1999). *Statistical Inference an Integrated Approach*, London: Arnold.
- [11] Morettin, P. A. e Tolo, C. M. C. (2004) *Análise de Séries Temporais*, São Paulo: Edgard Blücher LTDA.
- [12] Priestley, M. B. (1989). *Spectral Analysis and Time Series*. London: Academic Press Limited.
- [13] West, M. e Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. London: Springer Verlag
- [14] http://en.wikipedia.org/wiki/Thomas_Bayes
- [15] http://en.wikipedia.org/wiki/Jeffreys_prior
- [16] http://en.wikipedia.org/wiki/Markov_chain#History_2
- [17] R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing
- [18] Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com

Apêndices

Apêndice A - Código em R que gerou os dados simulados para o capítulo 4

```
n=100  
t=n + 1  
y=numeric(t)  
mu=numeric(t)  
beta=numeric(t)  
c=numeric(t)  
c1=numeric(t)  
phi=0.95  
p=12  
omega=2 * pi / p  
phic=0.95  
V=sqrt(1.5)  
W1=V/sqrt(1000)  
W2=V/sqrt(1000)  
W3=V/sqrt(100)
```

```

mu[1]=0
beta[1]=1
c[1]=5
y[1]=mu[1] + c[1] + rnorm(1,0,V)
for (i in 2:t)
{
  c[i]=phic * (cos(omega) * c[i-1] + sin(omega) * c1[i-1]) + rnorm(1,0,W3)
  c1[i]=phic * (-sin(omega) * c[i-1] + cos(omega) * c1[i-1]) + rnorm(1,0,W3)
  beta[i]=phi * beta[i-1] + rnorm(1,0,W2)
  mu[i]=mu[i-1] + beta[i-1] + rnorm(1,0,W1)
  y[i]=mu[i] + c[i] + rnorm(1,0,V)
}

```

Apêndice B - Introdução histórica

As decisões que tomamos todos os dias, em todas as instâncias, estão intimamente ligadas às inferências que fazemos diante das diferentes situações. A decisão de levar o guarda-chuva, por exemplo, ao sairmos de casa, é uma em função do que esperamos que o tempo vai se comportar. Comemos, ou não, um prato de comida acreditando, pelo aspecto, odor e procedência, que não vai nos fazer mal. Compramos uma determinada mercadoria pela expectativa que temos do bem que ela nos fará.

O maior problema da inferência é que ela decorre do fato que nossa informação a respeito do objeto de interesse é incompleta. Aliás, a inferência só existe porque a informação é incompleta. A chance de sucesso na inferência está diretamente ligada à completeza da informação sobre o objeto de interesse. Se tivermos toda a informação sobre um determinado tema, nossa decisão será baseada na certeza do que se passará a seguir, e a inferência será feita com 100% de chance de acerto, isto é, não será necessário inferir, pois a decisão é consequência lógica.

É sabido que Heródoto em 500 a.C. já discutia a política das decisões dos reis persas. Ele ressaltou que a decisão era “sábia” se as evidências disponíveis indicavam que a melhor decisão era aquela, mesmo que isso levasse a consequências desastrosas. Analogamente, uma decisão era “equivocada” se as evidências assim indicavam, mesmo que tal decisão levasse a bons resultados (Jaynes, 1986).

A operação de inferência com base em dados disponíveis obtidos do passado nos parece tão natural que sabemos inconscientemente que essas observações de Heródoto podem ser reproduzidas em bases matemáticas e, como consequência óbvia, aplicadas a todas as atividades intelectuais humanas, desde engenharia, medicina, política e até o estudo do cosmos.

James Bernoulli publicou em 1713 a primeira representação matemática da inferência em sua obra “Ars Conjectandi”, em que a “arte da conjectura” é apresentada em bases científicas. Ele enumerou um conjunto básico de casos “igualmente possíveis” (x_1, x_2, \dots, x_n) , que seriam chamados eventos ou proposições. É definido, assim, um “campo de discurso” ou “espaço da hipótese” H_0 . Introdz-se, então, uma proposição A , verdadeira em um certo subconjunto $H(A)$ com M pontos de H_0 e falsa para o resto. O número M , ou seja, o “número de vezes” em que A pode ser verdadeira, chama-se **multiplicidade** de A , e a **probabilidade** de A é definida como $p(A) = \frac{M}{N}$.

O próprio Bernoulli intuiu os pontos fortes e fracos desse método, uma vez que não é em qualquer caso que essa argumentação é eficaz. Enquanto que, para uma situação de jogo de dados ou a bola de uma roleta, o cálculo da probabilidade é perfeitamente exequível, há casos em que o universo estudado é parcial. Como ele próprio considerou “Qual mortal poderá determinar o número de doenças?”, para denotar a dificuldade em se estimar a probabilidade de uma doença. A conclusão que paralisou Bernoulli carecia da base do que chamamos, hoje, inferência: a incompletude da informação. A postura moderna é uma adaptação à realidade: complementamos H_0 da melhor forma possível com aquilo que conhecemos, e com base nisso são feitos cálculos posteriores. Se não temos como obter *todas as doenças que existem*, consideramos apenas aquelas que conhecemos. Sabemos que não é correto, mas supomos que esse é o *mais perto* que podemos obter, logo, a inferência que daí decorre é a mais próxima da verdade.

Na primeira definição de H_0 , de fato, não temos informação suficiente para avaliar que a probabilidade obtida representa a verdadeira probabilidade, no mundo real. Nem mesmo sabemos se a suposição do espaço “igualmente possível” é correta. Pode haver algum tipo de “simetria” desconhecida no espaço de eventos que torne algumas configurações

mais prováveis do que outras. Mas, ao obtermos as conclusões retiradas da hipótese H_0 poderemos constatar o quão errôneas elas são, comparando com o que podemos observar, e assim obter pistas de onde estão os erros infringidos na hipótese inicial. Refazemos as considerações, partindo, então, para a hipótese atualizada H_1 e nos aproximamos mais do mundo observado, e assim, sucessivamente.

Diante da situação de incompletude, uma vez que não tinha informação sobre todas as doenças, Bernoulli tratou de melhorar suas H_n (n atualizações depois da suposição inicial) em observando a proporção de doentes que melhoravam com a prescrição de um remédio (Jaynes, 1989). Bernoulli estabeleceu a conexão entre probabilidade e frequência, hoje conhecida como a lei fraca dos grandes números.

Thomas Bayes (1702-1761) teve publicada *post-mortem* em 1764, sua obra “*Essay Towards Solving a Problem in Doctrine of Chances*” na *Philosophical Transactions of the Royal Society of London*. Era um clérigo presbiteriano e matemático amador. Nessa obra encontra-se o primeiro enunciado do princípio, na época chamado probabilidade inversa, do que é hoje conhecido por Teorema de Bayes, muito embora sua formulação tenha pouco a ver com o que modernamente é apresentado. Além disso, sua obra tem pouco ou quase nada do que se chama hoje “estatística Bayesiana”, sendo ela assim cunhada como expressão de homenagem ao pioneiro, não ao(s) construtor(es) posterior(es).

Marquês de Laplace (1749-1827) publicou em 1774 seu primeiro trabalho retomando o princípio de Bayes, embora aparentemente que ele não tinha tido contato com o trabalho de Bayes. Após esse trabalho, Laplace aplicou o princípio para resolver problemas em diferentes áreas, desde astronomia até jurisprudência. Chegou mesmo a estimar estatisticamente os campos na astronomia que mais necessitavam de análise para neles trabalhar. Matemático de extrema inteligência, Laplace inaugurou os procedimentos de inferência

a priori e *a posteriori*, hoje largamente utilizados na astronomia. Seu livro de dois volumes em 1812 sobre a teoria da probabilidade é considerado uma das mais completas e atualizadas obras a respeito.

Em 1939, Sir Harold Jeffreys (1891-1989) retoma a estatística Bayesiana e a descreve de maneira mais clara que Laplace. Ele estabelece critérios para a hipótese H_0 hoje conhecida como *Jeffreys prior*, ou *a priori* de Jeffreys.

Richard T. Cox (1898-1991) publicou, em 1946, no *American Journal of Physics*, o artigo “Probability, Frequency and Reasonable Expectation”, em que, através da suposição que a plausibilidade de uma afirmação pode ser expressa em números reais, mostrou que o princípio de Bayes é logicamente consistente.

Na obra seminal de Claude Shannon (1916-2001) em 1948, “The Mathematical Theory of Communication”, o conceito de entropia, originada na termodinâmica e física estatística, foi estendida à Teoria das Comunicações, criada pelo próprio. A chamada “Entropia de Gibbs” foi deduzida por Shannon como a expressão do número de bits mínimo para a informação em uma mensagem ser completa. Da mesma forma, a relação mede a estimativa da informação perdida dada uma variável aleatória desconhecida.

Da extensão do conceito de entropia para a Teoria da Informação, pode-se, finalmente, dar base racional ao princípio da hipótese H_0 . O conceito de máxima entropia entra na estatística Bayesiana a ponto dos dois conceitos se confundirem. Toda inferência, na estatística Bayesiana, parte da análise da máxima entropia.

Andrey Markov (1856-1922) em conferências na Universidade de São Petersburgo (1906) desenvolveu os conceitos fundamentais do que ficaram conhecidos como Cadeias de Markov, e mais tarde generalizado por Kolmogorov (1936). Em 1984, Geman, S. e Geman, D. formularam um método que ficou conhecido como amostrador de Gibbs (gerador

de uma cadeia de Markov) para estudar a distribuição de Gibbs para uso em problemas de processamento de imagens. Alguns anos depois, o amostrador de Gibbs foi reconhecido como um poderoso método para análises Bayesianas. Fica fechado, assim, o ciclo que consolidou a estatística Bayesiana nos moldes em que é praticada hoje.

Encontramos a estatística Bayesiana aplicada a virtualmente todas as áreas de análise experimental e fenomenológica modernas. É utilizada na economia, finanças, criptografia, restauração de dados, modelos em ciências dos materiais, meteorologia, medicina sanitária, simulações físicas, aviação, cristalografia e mecânica quântica, entre muitos outros. É mais que um mero método de análise de dados, é também um instrumento poderoso em instâncias de apoio à decisão. Invade praticamente todos os campos de análise do saber humano.

Apêndice C - Programa em OxTM que gerou as amostras das *posterioris* e da *preditiva*

```
#include <oxstd.h>

#include <oxprob.h>

#include <oxdraw.h>

#include <oxfloat.h> // constantes pre-definidas

// programa para amostrar da posteriori dos parâmetros de um
// modelo dinâmico de crescimento amortecido

samplewmedia(alpha,betaw,mu,beta,phi,i){

// i == 1 para w1, i == 2 para w2

    decl j;

    decl w,a,b,T,ss=0;

    T=sizeof(mu)-1;

    if (i == 1)

    for (j=0;j<T;j++)

        ss+=(mu[j+1]-mu[j]-beta[j])^2;

    else if (i == 2)

    for (j=0;j<T;j++)

        ss+=(beta[j+1]-beta[j]*phi)^2;

    a=alpha+T/2;
```

```

b=betaw+ss/2;

w=1/rangamma(1,1,a,b);

return(w);

} // fim samplewmedia

samplewiclo(alpha,beta,c,cmais,omega,phic,co,s,i){
// i == 3 para w3, i == 4 para w4

decl j;

decl w,a,b,T,ss=0;

T=sizeof(c)-1;

if (i == 3)
    for (j=0;j<T;j++)
        ss+=(c[j+1]-phic*(c[j]*co+cmais[j]*s))^2;

else if (i == 4)
    for (j=0;j<T;j++)
        ss+=(cmais[j+1]-phic*(-c[j]*s+cmais[j]*co))^2;

a=alpha+T/2;

b=beta+ss/2;

w=1/rangamma(1,1,a,b);

return(w);

```

```

} // fim samplewcciclo

samplev(alpha,beta,y,mu,c) {

    decl j;

    decl v,a,b,T,ss=0;

    T=sizeof(y);

    for (j=0;j<T;j++)

        ss+=(y[j]-mu[j+1]-c[j+1])^2;

    a=alpha+T/2;

    b=beta+ss/2;

    v=1/rangamma(1,1,a,b);

    return(v);

} // fim samplev

samplephi(alphaphi,betaphi,w2,beta,phi) {

// Distribuição proposta é a priori (Beta)

    decl T,v,U,phiprop,phipasso,alpha,ss=0,j,deno;

    T=sizeof(beta)-1;

    phiprop=ranbeta(1,1,alphaphi,betaphi);

    // Gerando valor proposto

    deno=-1/(2*w2);

```

```

for (j=0;j<T;j++)
    ss+=beta[j] * (phiprop - phi) * (beta[j] * (phiprop + phi) - 2 * beta[j+1]);

alpha=min(1,exp(deno*ss));

// Probabilidade de aceitação do Metropolis-Hastings
U=ranbinomial(1,1,1,alpha);
phipasso=U*phiprop + (1-U)*phi;

return(phipasso);

} // fim samplephi

samplephic(alphaphi,beta,w3,w4,c,cmais,co,s,phi) {
// Distribuição proposta é a priori (Beta)

decl T,U,phiprop,phipasso,alpha,ss3=0,ss4=0,j,deno3,deno4;

T=sizeof(c)-1;

phiprop=ranbeta(1,1,alphaphi,beta);

// Gerando valor proposto
deno3=-1/(2*w3);
deno4=-1/(2*w4);

for (j=0;j<T;j++){
    ss3+=(c[j+1]-phiprop*(c[j]*co+cmais[j]*s))^2-(c[j+1]-phi*(c[j]*co+cmais[j]*s))^2;
    ss4+=(cmais[j+1]-phiprop*(-c[j]*s+cmais[j]*co))^2-(cmais[j+1]
-phi*(-c[j]*s+cmais[j]*co))^2;
}

```

```

alpha=min(1,exp(deno3*ss3 + deno4*ss4));

// Probabilidade de aceitação do Metropolis-Hastings

U=ranbinomial(1,1,1,alpha);

phipasso=U*phiprop + (1-U)*phi;

return(hipasso);

} // fim samplephic

sampleomega(w3,w4,c,cmais,co,s,phic,omega){
// Distribuição proposta é a priori U[0,2 * Pi]

decl T,U,omegapasso,alpha,ss3=0,ss4=0,j,copr,spr,omegaprop,deno3,deno4;

T=sizeof(c)-1;

omegaprop = ranu(1,1) * M_2PI;

copr=cos(omegaprop);

spr=sin(omegaprop);

deno3=-1/(2*w3);

deno4=-1/(2*w4);

for (j=0;j<T;j++){

    ss3+=(c[j+1]-phic*(c[j]*copr+cmajis[j]*spr))^2-(c[j+1]-phic*(c[j]*co+cmajis[j]*s))^2;

    ss4+=(cmajis[j+1]-phic*(-c[j]*spr+cmajis[j]*copr))^2-(cmajis[j+1]

- phic*(-c[j]*s+cmajis[j]*co))^2;

}

alpha=min(1,exp(deno3*ss3 + deno4*ss4));

```

```

// Probabilidade de aceitação do Metropolis-Hastings

U=ranbinomial(1,1,1,alpha);

omegapasso=U*omegaprop + (1-U)*omega;

return(omegapasso);

} // fim sampleomega

ffbs_theta(m,C,G,W,n,a,R){

    decl t;

    decl saida=zeros(n+1,4);

    decl med=zeros(1,4), vari=unit(4), winv=unit(4);

    med = m[ ][n];

    vari = C[n][ ][ ];

    winv=invert(W);

    saida[n][ ] = (choleski(vari) * rann(4,1) + med)';

    for (t=(n-1);t>-1;t-)

    {

        med=invert(G' * winv * G + invert(C[t][ ][ ])) * (G' * winv * saida[t+1][ ]' +

invert(C[t][ ][ ])) * m[ ][t]);

        vari = invert(G' * winv * G + invert(C[t][ ][ ]));

        saida[t][ ] = (choleski(vari) * rann(4,1) + med)';

        // Para efeito de MCMC, amostra-se da distribuição calculada com // FFBS

    }

```

```

return(saida);

} // fim ffbs_theta

samplepred(k1,k2,F,G,m,v,C,W) {
// Amostra da distribuição preditiva de  $y_{-(t+k1)}$  até  $y_{-(t+k2)}$ 

decl k;

decl Gk = unit(4), Gk1 = zeros(4), fq = zeros((k2-k1+1),2), pw=zeros(4), saida=zeros(1,(k2-
k1+1));

for (k=0;k<(k1-1);k++)
{
Gk = G^k;
pw += Gk * W * Gk';
}

if (k1 > 1)
Gk *= G;

for (k=k1;k<=k2;k++){
Gk1 = Gk;
Gk *= G;
pw += Gk1 * W * Gk1';
fq[(k-k1)][0] = F' * Gk * m;
fq[(k-k1)][1] = sqrt(F' * (Gk * C * Gk' + pw) * F + v);
// Raiz quadrada para a decomposicao de choleski

```

```

}

saida = (diag(fq[ ][1]) * rann((k2-k1+1),1) + fq[ ][0]);

return(saida);

} // fim samplepred

main(){
    // Declaracao de variaveis

    ranseed(34748576);

    decl time=timer();

    // le dados

    decl y = loadmat("ycurto.dat");

    decl n=sizeof(y); // numero de tempos observados

    // define numero de iteracoes, burnin e thinning

    decl niter=505001;

    decl thin=50;

    decl burnin=5000;

    // tamanho do vetor de dados a ser previsto

    decl p = 40;

    decl i,ll,k;

    // parametros do modelo

```



```

decl mu=zeros(n+1,1), beta=zeros(n+1,1), c=zeros(n+1,1), cmais=zeros(n+1,1);

decl theta=zeros(n+1,4); // parametro auxiliar para usar o ffbs

decl v,w1,w2,w3,w4;

decl phi,phic,omega;

decl co,s; // cosseno e seno de omega

decl ypred=zeros(p,1);

// parametros do filtro de Kalman

decl F=zeros(4,1);

F[0]=1; F[3]=1;

decl G=unit(4);

G[0][1]=1;

decl W=unit(4);

decl Q=ones(n,1);

decl a=zeros(4,n),m=zeros(4,n+1),A=zeros(4,n);

decl R=new array[n],C=new array[n+1];

for (i=0;i<=n;i++)

C[i]=unit(4);

decl e=ones(n,1),f=ones(n,1);

// decl delta=0.9;

// valores iniciais da cadeia do MCMC

v=1;

w1=1;

w2=1;

```

```

w3=1;

w4=1;

phi=0.5;

phic=0.5;

omega=1;

co=cos(omega);

s=sin(omega);

// parametros das prioris

decl alphav = 1,betav = 1;

decl alphaw1 = 1,betaw1 = 1;

decl alphaw2 = 1,betaw2 = 1;

decl alphaw3 = 1,betaw3 = 1;

decl alphaw4 = 1,betaw4 = 1;

decl alphaphi = 17.1, betaphi = 0.9;

decl alphaphic = 7.2, betaphic = 0.8;

// criando arquivos com amostras

decl fpmu=fopen("smu.txt","w");

decl fpbeta=fopen("sbeta.txt","w");

decl fpc=fopen("sc.txt","w");

decl fpcmais=fopen("scmais.txt","w");

decl fpphi=fopen("sphi.txt","w");

decl fpphic=fopen("sphic.txt","w");

decl fpomega=fopen("somega.txt","w");

```

```

decl fpw1=fopen("sw1.txt","w");

decl fpw2=fopen("sw2.txt","w");

decl fpw3=fopen("sw3.txt","w");

decl fpw4=fopen("sw4.txt","w");

decl fpv=fopen("sv.txt","w");

decl fppreditiva=fopen("spreditiva.txt","w");

// loop MCMC
for (ll=0;ll<niter;ll++){

    v=samplev(alphav,betav,y,mu,c);

    w1=samplewmedia(alphaw1,betaw1,mu,beta,phi,1);

    w2=samplewmedia(alphaw2,betaw2,mu,beta,phi,2);

    w3=samplewciclo(alphaw3,betaw3,c,cmais,omega,phic,co,s,3);

    w4=samplewciclo(alphaw4,betaw4,c,cmais,omega,phic,co,s,4);

    //phi=0.95;

    //phic=0.95;

    phi=samplephi(alphaphi,betaphi,w2,beta,phi);

    phic=samplephic(alphaphic,betaphic,w3,w4,c,cmais,co,s,phic);

    //omega=sampleomega(w3,w4,c,cmais,co,s,phic,omega);

    omega=M_PI/6;

    co=cos(omega); s=sin(omega);

    // filtro de Kalman

    G[1][1]=phi;

    G[2][2]=phic*co;

```

```

G[2][3]=phic*s;
G[3][2]=-phic*s;
G[3][3]=phic*co;
W[0][0]=w1;
W[1][1]=w2;
W[2][2]=w3;
W[3][3]=w4;
for (i=0;i<n;i++){
    // priori
    a[ ][i]=G * m[ ][i];
    R[i]=W + G * C[i] * G';
    // previsao
    f[i]=F'*a[ ][i];
    Q[i]=v + F' * R[i] * F;
    // posteriori
    e[i]=y[i]-f[i];
    A[ ][i]=(R[i] * F) / Q[i];
    m[ ][i+1]=a[ ][i] + e[i] * A[ ][i];
    C[i+1]=R[i] - Q[i] * ( A[ ][i] * (A[ ][i])' );
} // fim loop Kalman

theta=ffbs_theta(m,C,G,W,n,a,R);

mu[ ]=theta[ ][0];
beta[ ]=theta[ ][1];
c[ ]=theta[ ][2];

```

```

cmais[ ]=theta[ ][3];

ypred=samplepred(1,p,F,G,m[ ][n],v,C[n],W);

println("iteracao: ",ll);

println(" v: ",v);

println(" w1: ",w1);

println(" w2: ",w2);

println(" w3: ",w3);

println(" w4: ",w4);

println(" phi: ",phi);

println(" phic: ",phic);

println(" omega: ",omega);

println("\n\n\n\n");

k=imod((ll+1),thin);

if ((k<=0) && (ll>burnin)) {

    fprintf(fpmu,mu[]);

    fprintf(fpbeta,beta[]);

    fprintf(fpc,c[]);

    fprintf(fpcmais,cmais[]);

    fprintf(fpphi,phi);

    fprintf(fpphic,phic);

    fprintf(fpomega,omega);

    fprintf(fpw1,w1);

    fprintf(fpw2,w2);

```

```

        fprintf(fpw3,w3);

        fprintf(fpw4,w4);

        fprintf(fpv,v);

        fprintf(fppreditiva,ypred);

    } // Fim if

} // fim loop MCMC

println("time lapsed: ", timespan(time), "\n");

fclose(fpmu);

fclose(fpbeta);

fclose(fpc);

fclose(fpcmais);

fclose(fpphi);

fclose(fpphic);

fclose(fpomega);

fclose(fpw1);

fclose(fpw2);

fclose(fpw3);

fclose(fpw4);

fclose(fpv);

fclose(fppreditiva);

} // fim programa

```

Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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