

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 18.04.2000

Assinatura: Leonilce Mena

## Processos com Parâmetros Aleatórios para Modelos de Séries Temporais

*Leonilce Mena*

**Orientador: *Prof. Dr. Marinho Gomes de Andrade Filho***

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Área: Ciências de Computação e Matemática Computacional.

**USP – São Carlos  
Abril de 2000**

“Tudo posso nAquele que me fortalece”  
(FL 4:13)

Aos meus pais,  
João e Angelina,  
Aos meus irmãos

- A Deus, simplesmente, por tudo.
- Aos meus pais, João e Angelina, que, com amor e carinho, tornaram possível esta minha caminhada.
- Aos meus irmãos, em especial à minha irmã Dalva, e à minha sobrinha Érica que, com paciência e confiança, tornaram mais branda minha ausência no convívio familiar.
- Ao professor Dr. Marinho G. de Andrade Filho, pela orientação, pelo estímulo, e pelo apoio no decorrer deste trabalho.
- À professora Dr<sup>a</sup> Maria Creusa Breta Salles, pelas valiosas sugestões e comentários dados no início do meu curso de Pós-Graduação e no meu exame de qualificação.
- Ao professor Dr. Jorge Alberto Achcar, pelos valiosos comentários e sugestões no exame de qualificação.
- Aos amigos, colegas e funcionários do ICMC-USP.
- À CAPES pelo apoio financeiro.
- À UEM, especialmente aos meus amigos e colegas do DES, pelo apoio e incentivo, em particular, à Terezinha pela contribuição no início deste trabalho.
- Em especial, ao meu amigo Ulisses, pelos preciosos ensinamentos e apoio em programação computacional.
- Aos amigos e colegas que, ao folhearem este exemplar, sintam o prazer de terem apoiado e contribuído de alguma forma na realização deste trabalho.

Meus Agradecimentos.

# Sumário

<b>1</b>	<b>Processo Auto Regressivo com Parâmetros Aleatórios</b>	<b>1</b>
1.1	Introdução . . . . .	1
1.2	Descrição do Modelo . . . . .	5
1.2.1	Modelo para $\theta_t$ Aleatório . . . . .	5
<b>2</b>	<b>Modelo Hierárquico para um Processo Auto-Regressivo de Primeira Ordem com Variância Conhecida</b>	<b>8</b>
2.1	Inferência para o Processo de Primeira Ordem com Variância Conhecida	9
2.1.1	A Distribuição <i>a Posteriori</i> de $\lambda$ e $\theta_t$ . . . . .	9
2.1.2	Distribuição Condicional para $\theta_t$ . . . . .	13
2.2	Distribuição Preditiva de $\theta_{t+1}$ e de $y_{t+1}$ . . . . .	15
2.2.1	Densidade Preditiva de $\theta_{t+1}$ . . . . .	15
2.2.2	Densidade Preditiva de $y_{t+1}$ . . . . .	16
2.2.3	Intervalo de Credibilidade de $y_{t+1}$ . . . . .	17
2.2.4	Densidade Preditiva de $\theta_{t+k}$ e $y_{t+k}$ . . . . .	17
2.3	Resultados . . . . .	20
<b>3</b>	<b>Modelo Hierárquico para Processo de Primeira Ordem com Variância Desconhecida</b>	<b>22</b>
3.1	Inferência para os Parâmetros do Modelo . . . . .	22
3.1.1	A Distribuição Conjunta <i>a Posteriori</i> de $\lambda$ e $\tau$ . . . . .	22
3.1.2	Distribuição Condicional para $\lambda$ e $\tau$ . . . . .	27
3.1.3	Distribuição Condicional <i>a Posteriori</i> de $\theta_t$ . . . . .	27
3.2	Densidade Preditiva para $\theta_{t+1}$ e $y_{t+1}$ . . . . .	31
3.2.1	Densidade Preditiva para $\theta_{t+1}$ . . . . .	31

3.2.2	Distribuição Preditiva para $y_{t+1}$ . . . . .	32
3.2.3	Densidade Preditiva para $\theta_{t+k}$ e $y_{t+k}$ . . . . .	33
3.3	Resultados . . . . .	34
<b>4</b>	<b>Modelo Dinâmico para um Processo-Auto Regressivo de Primeira Ordem</b>	<b>37</b>
4.1	O Modelo Filtro de Kalman e Suas Soluções . . . . .	38
4.2	Adaptação do Filtro de Kalman . . . . .	43
4.2.1	Inferência para um Modelo Filtro de Kalman Adaptado . . . . .	44
4.2.2	Cálculo da Função de Verossimilhança . . . . .	45
4.2.3	Cálculo de <i>a Posteriori</i> de $\theta_t$ dado $y^{(t)}$ e $\lambda$ . . . . .	47
4.2.4	O Comportamento de $\theta_t$ . . . . .	48
4.3	Densidade Preditiva para $\theta_{t+1}$ e $y_{t+1}$ . . . . .	51
4.3.1	Densidade Preditiva para $\theta_{t+1}$ . . . . .	51
4.3.2	Densidade Preditiva para $y_{t+1}$ . . . . .	52
4.4	Resultados . . . . .	53
<b>5</b>	<b>Aplicação - Estudo de Casos</b>	<b>55</b>
5.1	Modelo Hierárquico com $\tau^2$ Conhecida . . . . .	56
5.2	Modelo Hierárquico com $\tau^2$ Desconhecida . . . . .	62
5.3	Modelo Dinâmico com Variância Conhecida . . . . .	70
<b>6</b>	<b>Considerações Finais</b>	<b>77</b>
6.1	Conclusão . . . . .	77
6.2	Proposta Futura . . . . .	77
	<b>Anexos</b>	<b>79</b>
	<b>Referências Bibliográficas</b>	<b>93</b>

## Lista de Abreviaturas

ARG = Auto-Regressivo Generalizado

MCMC = Monte Carlo em Cadeia de Markov

BOVESPA = Bolsa de Valores de São Paulo

Z.= Zootécnica

Dist.= Distribuição

## Resumo

Este trabalho apresenta uma abordagem bayesiana para fazer inferência sobre os parâmetros de modelos auto-regressivos. Neste contexto, quando os parâmetros variam de forma aleatória e independente adotamos um modelo hierárquico para descrever a densidade *a posteriori* dos parâmetros. Uma segunda abordagem supõe que os parâmetros variam de acordo com um modelo auto-regressivo de primeira ordem, nesse caso a abordagem proposta é vista como uma extensão do filtro de Kalman onde as variâncias dos ruídos são conhecidas. Os modelos foram analisados usando-se técnicas de simulação de Monte Carlo e a geração de amostras das densidades *a posteriori* permitiram fazer previsões de séries através das densidades preditivas. Ilustrações de séries financeiras com dados reais são apresentadas e avaliadas pela qualidade da previsão obtida, salientando-se o modelo que melhor representa os dados.

## Abstract

This work deals with the Bayesian method to make inferences on the parameters of autoregressive models. When in this context the parameters of these models vary randomly and independently, a hierarchical was adopted to obtain a posteriori density of parameters. Another approach of the same method presupposes that the parameters of model vary according to a first-order autoregressive model and is regarded as an extension of Kalman's filter in which the variances of noises are known. Both models were analysed through Monte Carlo's simulation techniques and the resulting samples of a posteriori densities allow to calculate a data series through predictable densities. Examples of a finance series with actual data are provided and the two models are evaluated through their predicting qualities thus revealing the most appropriate.



# Capítulo 1

## Processo Auto-Regressivo com Parâmetros Aleatórios

### 1.1 Introdução

O propósito deste trabalho é formular um procedimento bayesiano, utilizando simulação de Monte Carlo em Cadeia de Markov (MCMC), para análise e previsão de séries temporais modeladas por processos auto-regressivos com coeficientes aleatórios denominados processos auto-regressivos generalizados (ARG).

A análise e previsão de séries temporais é aplicada em diversas áreas, tais como: na Engenharia, na Economia, na Teoria de Controle e na Sociologia, onde o conjunto de observações é ordenado no tempo.

Box e Jenkins (1994) aplicaram modelos auto-regressivos e auto-regressivo-média móvel em diversas áreas para modelar alguns fenômenos por meio de modelos de coeficientes constantes.

Muitos métodos podem ser utilizados para estimar os parâmetros de um processo auto-regressivo proposto para uma série temporal, tais como: método dos mínimos quadrados e método de máxima verossimilhança, já amplamente estudados (Box e Jenkins, 1994). Um método alternativo de análise e inferência para modelos de séries temporais são os métodos Bayesianos aplicados para séries com coeficientes constantes (Broemeling e Cook, 1993); neste trabalho, vamos aplicar o método bayesiano para séries com coeficientes aleatórios.

Para qualquer investigação, os modelos com coeficientes constantes não captam toda a

variabilidade provocada pelos fenômenos aleatórios inerentes; por essa razão, os modelos com coeficientes aleatórios são mais recomendados. Um modelo dessa classe é o modelo auto-regressivo generalizado (ARG).

Recentemente, alguns autores têm estudado modelos para análise e previsão de séries temporais com coeficientes aleatórios no contexto bayesiano, onde podemos destacar, entre outros, os seguintes trabalhos sobre o assunto encontrados na literatura: Liu(1995), Singpurwalla e Soyer (1985), Nicholls e Quinn (1980), Guyton (1986), Diaz(1990).

Nicholls e Quinn (1980) introduzem o estudo de modelo auto-regressivo generalizado de primeira ordem com parâmetros variáveis e independentes para modelos e previsão de séries temporais.

Singpurwalla e Soyer (1985) determinam o coeficiente aleatório de processo auto regressivo de primeira ordem para descrever o crescimento ou decrescimento de confiabilidade de software, introduzindo um modelo e algumas ramificações.

Guyton (1986) desenvolve a metodologia para a classe de processo auto-regressivo de coeficientes aleatórios, visando a previsão de séries temporais. Descreve o procedimento para o cálculo dos estimadores de máxima verossimilhança para os parâmetros e desenvolve as condições necessárias e suficientes para existência e estacionariedade desses modelos.

Diaz (1990) apresenta uma solução bayesiana para o problema de previsão de séries temporais, utilizando *priori* não-informativa e *priori* informativa, normal-gama, cujo procedimento é baseado na densidade preditiva da observação futura.

Liu (1995) faz uma comparação entre o procedimento de aproximação bayesiana e o procedimento de aproximação de estimadores de máxima verossimilhança para coeficientes aleatórios dos modelos auto-regressivo-integrado- média-móvel (ARIMA).

Segundo Meinhold e Singpurwalla (1983), o modelo auto-regressivo de coeficientes aleatórios pode ser desenvolvido aplicando uma adaptação do modelo filtro de Kalman, sob o enfoque bayesiano, empregando alguns resultados da estatística multivariada muito conhecidos, devido sua similaridade com modelos de regressão linear , para fazer inferência sobre os parâmetros aleatórios e previsão de valores futuros um passo à frente.

As duas teorias, clássica e bayesiana, têm princípios básicos diferentes; porém existe alguma similaridade entre elas no desempenho de previsões futuras. Isso foi demonstrado por Liu (1995), que explorou analiticamente esta similaridade para o modelo auto-

regressivo-média-móvel com coeficientes aleatórios.

A teoria bayesiana considera os coeficientes como variáveis aleatórias segundo alguma distribuição *a priori* própria. Logo, o procedimento bayesiano tem a vantagem de que qualquer informação disponível, antes da coleta dos dados, pode ser incorporada aos dados e a distribuição *a posteriori* pode ser obtida, em alguns casos, eliminando-se o uso de resultados assintóticos.

Soyer (1985) introduziu procedimentos bayesianos para modelos auto-regressivos com coeficientes aleatórios, utilizando aproximação numérica para obter os resultados computacionais das densidades *a posteriori*, mas o uso dessa aproximação é limitada pelo número de parâmetros do modelo.

Com a finalidade de apresentar um procedimento alternativo, aplicamos simulação de Monte Carlo em Cadeia de Markov, utilizando algoritmos amostrador de Gibbs e Metropolis-Hasting para fazer inferência dos parâmetros do processo auto-regressivo com coeficientes aleatórios proposto por Soyer (1985).

Neste trabalho, descrevemos ramificações de um processo auto-regressivo de primeira ordem de coeficientes aleatórios e variantes no tempo, assumindo uma estrutura de dependência dos componentes não observáveis do modelo ( $\theta_t$ 's), fazendo a descrição do modelo para uma série temporal em função desses componentes quando  $\theta_t$  é expresso em função dos valores dos mesmos no instante anterior.

Este trabalho está centrado nos procedimentos pelos quais se obtêm estimadores atualizados das componentes não observáveis ( $\theta_t$ ), a todo instante de tempo, a partir da informação dada pelo componente observável do sistema ( $y_t$ ), apresentando duas alternativas de procedimento.

Descrevemos a utilização e a aplicação de um Modelo Hierárquico para modelar uma série temporal onde as componentes não observáveis têm uma densidade de probabilidade normal cuja média também tem uma distribuição normal, assegurando-se, assim, que os coeficientes sejam variáveis mas não independentes. Descrevemos também a aplicação de um Modelo Dinâmico em problemas de previsões, onde a série é modelada por uma média que varia no tempo, superposta a um ruído aditivo. Essa média é por hipótese, uma combinação linear de funções conhecidas cujos coeficientes são desconhecidos.

No Capítulo 2, apresentamos uma abordagem bayesiana através de simulação de Monte Carlo em Cadeia de Markov, aplicando o algoritmo Amostrador de Gibbs para

fazer inferência dos parâmetros do modelo Hierárquico, onde a variância do ruído é conhecida, porém considerando a estrutura de dependência dos coeficientes do processo em questão.

No Capítulo 3, estendemos o desenvolvimento do modelo Hierárquico mantendo a estrutura de dependência dos coeficientes aleatórios e assumindo a variância do ruído ( $\tau^2$ ) desconhecida. Consideramos nossa incerteza sobre essa variância por uma densidade *a priori*, o que nos conduz a distribuições *a posteriori* que não podem ser identificadas como uma distribuição conhecida; utilizamos, então, o algoritmo Metropolis-Hasting para simular amostras *a posteriori* e assim fazemos a inferência dos parâmetros do processo e as previsões de observações futuras.

No Capítulo 4, consideramos uma ramificação do processo auto-regressivo de primeira ordem, assumindo uma dependência dos coeficientes aleatórios que nos leva a um modelo de filtro de Kalman adaptado, (Soyer, 1985), termo usado na literatura por engenheiros, onde alguns ou todos os parâmetros da equação de observação ou da equação de estados são estimados dos dados.

Embora o modelo filtro de Kalman adaptado considerado por nós seja uma generalização do filtro de Kalman Ordinário, sua análise produz dificuldades técnicas no sentido de que não é possível encontrar uma forma fechada para o filtro; assim, aplicamos simulação de Monte Carlo em Cadeia de Markov, utilizando os algoritmos amostrador de Gibbs com Metropolis-Hasting para fazer inferência dos parâmetros.

A técnica de análise apresentada nesta dissertação representa a contribuição desse trabalho para o importante tópico que trata dos Modelos Auto-Regressivos para análise e previsão de Séries Temporais.

No Capítulo 5, ilustramos a utilidade de nossa técnica, aplicando o processo auto-regressivo de primeira ordem com coeficiente aleatório para modelagem de séries de dados reais, na área financeira. Em nossa análise, mostramos o quanto esses modelos produzem proveitosas informações na previsão de observações futuras.

Finalmente, fazemos a comparação desses modelos em termos de resultados obtidos através da distribuição preditiva, identificando o modelo que melhor descreve o conjunto de dados financeiros. Apresentamos as nossas conclusões e sugestões para trabalhos futuros.

## 1.2 Descrição do Modelo

Os modelos que estamos trabalhando consideram um processo estocástico  $X_t; t = 1, 2, \dots$  com distribuição log-normal denotado por :

$$X_t = X_{t-1}^{\theta_t} ; \quad t = 1, 2, \dots \quad (1.1)$$

onde  $\theta_t$  é um coeficiente cujos valores descrevem um crescimento ou decrescimento no processo.

Para produzir uma generalização no modelo, introduzimos um erro multiplicativo  $\delta_t$ , assim temos:

$$X_t = X_{t-1}^{\theta_t} \delta_t ; \quad t = 1, 2, \dots \quad (1.2)$$

Convenientemente, assumimos que  $\delta_t$  também tem distribuição log-normal com média zero e variância  $\tau^2$  e, em alguns problemas, é conveniente supor que o expoente  $\theta_t$  tem característica aleatória. (Diaz,1990; Liu,1995).

Se aplicarmos o logaritmo natural, em ambos os lados, em (1.2) e seja  $\epsilon_t = \log(\delta_t)$  então nosso modelo torna-se :

$$y_t = \theta_t y_{t-1} + \epsilon_t ; \quad t = 1, 2, \dots \quad (1.3)$$

onde,  $y_t \stackrel{\text{def}}{=} \log(X_t)$  com  $y_t$  e  $\epsilon_t$  normalmente distribuídos,  $\epsilon_t$  tem média zero e variância  $\tau^2$ .

A partir desse ponto, enfocamos as variáveis  $y_t$ 's ao invés das variáveis  $X_t$ 's, com o objetivo de estimar  $\theta_t$  e de fazer previsão para observações futuras  $y_{t+1}, y_{t+2}, \dots$ . A previsão de  $\theta_t$  permite analisar o comportamento do processo que depende dos valores dos  $\theta_t$ 's , ou seja, o crescimento do processo ocorre quando  $\theta_t \geq 1$  e o decrescimento ocorre quando  $\theta_t < 1$ . Um caso particular ocorre quando  $\theta_t = 0$ , pois teremos da eq 1.3 que  $y_t = \epsilon_t$ , o que implica  $y_t$  não depender de  $y_{t-1}$ , ou seja,  $y_t$  é um ruído branco.

### 1.2.1 Modelo para $\theta_t$ Aleatório

Soyer (1985) considera o processo de crescimento ou decrescimento de confiabilidade como um problema de séries temporais, apresentando assim um processo auto-regressivo

de parâmetros aleatórios para modelar crescimento de confiabilidade de software dado por:

$$\begin{aligned} y_t &= \theta_t y_{t-1} + v_t \\ v_t &\sim N(0, \sigma_1^2) \\ \theta_t &\sim N(\lambda, \sigma_2^2) \end{aligned} \quad (1.4)$$

onde,  $v_t$  e  $\omega_t$  são independentes e  $\sigma_1^2$ ,  $\sigma_2^2$ , e  $\lambda$  são conhecidos.

Desenvolve uma extensão desse modelo introduzindo uma incerteza em  $\lambda$  e aplica o procedimento bayesiano empírico (Morris, 1983) para fazer inferência sobre os parâmetros do processo. Descreve a incerteza sobre  $\lambda$  considerando uma distribuição *a priori*  $\lambda \sim N(m_0, S_0)$  para  $\lambda$  assim o modelo torna-se:

$$\begin{aligned} y_t &= \theta_t y_{t-1} + v_t & v_t &\sim N(0, \sigma_1^2) \\ \theta_t &= \lambda + \omega_t & \omega_t &\sim N(0, \sigma_2^2) \\ & & \lambda &\sim N(m_0, S_0) \end{aligned} \quad (1.5)$$

com  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $m_0$  e  $S_0$  conhecidos

Dado os dados  $y^{(t)} = (y_1, y_2, \dots, y_t)$ , obtém-se a distribuição *a posteriori* para  $\theta_t$  dado  $y^{(t)}$  como:

$$(\theta_t | y^{(t)}) \sim N(\hat{\theta}_t, \hat{\Sigma}_t) \quad (1.6)$$

onde

$$\begin{aligned} \hat{\theta}_t &= \frac{\sigma_1^2 m_{t-1} + \sigma_2^2 y_t y_{t-1}}{\sigma_2^2 y_{t-1}^2 + \sigma_1^2} \\ \hat{\Sigma}_t &= \frac{\sigma_1^4 S_t}{(\sigma_2^2 y_{t-1}^2 + \sigma_1^2)^2} + \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 y_{t-1}^2 + \sigma_1^2} \\ S_t &= \left( \frac{1}{S_0} + \sum_{i=1}^t \frac{y_{i-1}^2}{r_i} \right)^{-1} \quad ; \quad r_i = \sigma_2^2 y_{i-1}^2 + \sigma_1^2 \end{aligned}$$

Compara esses resultados com os resultados de mínimos quadrados para o modelo determinístico obtido por Anderson(1978), discutindo a relação entre os dois processos.

Neste trabalho, assumindo os  $\theta_t$ 's com densidade de probabilidade normal de média

$\lambda$  e variância  $\gamma^2$ , a estratégia para assegurar que os  $\theta_t$ 's sejam dependentes é assumirmos também uma distribuição de probabilidade para  $\lambda$ , isto é, que  $\lambda$  tem distribuição normal com média  $m$  e variância  $S^2$ .

Podemos, assim, resumir nosso modelo da seguinte maneira :

$$\begin{aligned} y_t &= \theta_t y_{t-1} + \epsilon_t \\ \theta_t &= \lambda + \omega_t \\ \lambda &\sim N(m, S^2) ; m \text{ e } S^2 \text{ conhecidos} \end{aligned} \tag{1.7}$$

Na equação (1.3), assumimos  $\theta_t$  independente de  $\epsilon_t$ , então determinaremos  $y^{(t)} = (y_1, \dots, y_t)$ . Dado  $y^{(t)}$ , nosso objetivo é fazer inferências sobre  $\lambda$ ,  $\theta_t$  e fazer previsão de valores futuros de  $y_{t+k}$ ,  $k \geq 1$ .

Nos capítulos subseqüentes, notaremos que  $E(\theta_t | y^{(t)})$  fornece informações sobre o comportamento do processo na transição do estágio  $(t - 1)$  para o estágio  $(t)$ , enquanto que  $E(\lambda | y^{(t)})$  fornece informações sobre o comportamento global do processo. Assim, por exemplo, se  $(E(\theta_t | y^{(t)}) \geq 1)$ , podemos avaliar que há um crescimento para vários valores; caso contrário, se  $(E(\theta_t | y^{(t)}) < 1)$ , dizemos que há um decrescimento do processo.

O propósito aqui é desenvolver este modelo e calcular as estimativas dos coeficientes do modelo para fazer previsões das observações futuras.





## Capítulo 2

# Modelo Hierárquico para um Processo Auto-Regressivo de Primeira Ordem com Variância Conhecida

Os Modelos Hierárquicos Bayesianos têm sido utilizados em aplicações nas mais diversas áreas do conhecimento.

A análise desses modelos torna-se, geralmente, intratável quando não existe conjugação entre distribuição *a priori* e a função de verossimilhança. Nesses casos, métodos de aproximação numérica (Soyer, 1985) ou de simulação devem ser utilizados para a obtenção das distribuições *a posteriori* de interesse.

Neste capítulo, através de simulação de Monte Carlo (MCMC), utilizando o algoritmo amostrador de Gibbs (Casella e George, 1992), <sup>podemos obter as</sup> obtemos as distribuições *a posteriori* das variáveis de interesse e <sup>fazemos</sup> inferência sobre os parâmetros e a previsão de observações futuras para um processo auto-regressivo de primeira ordem, com parâmetros aleatórios e variantes no tempo ajustado por um Modelo Hierárquico, considerando a variância do ruído conhecida.

## 2.1 Inferência para o Processo de Primeira Ordem com Variância Conhecida

Inicialmente, desenvolvemos a extensão do modelo considerado por:

$$y_t = \theta_t y_{t-1} + \epsilon_t ; \quad (2.1)$$

$$\theta_t = \lambda + \omega_t$$

onde,  $\epsilon_t$  e  $\omega_t$  são independentes e normalmente distribuídos  $\epsilon_t \sim N(0, \tau^2)$  e  $\omega_t \sim N(0, \gamma^2)$ . Esse modelo implica que  $\theta_t \sim N(\lambda, \gamma^2)$ . Assumimos a incerteza sobre  $\lambda$  considerando  $\lambda \sim N(m, S^2)$  com  $\lambda$  independente de  $\omega_t$ .

Seja a série  $y^{(t)} = (y_1, y_2, \dots, y_t)$  e dado  $y^{(t)}$ , o objetivo é fazer inferência sobre  $\theta_t$ ,  $\lambda$  e previsão de  $\theta_{t+1}$  e da observação futura  $y_{t+1}$ .

Considerando que  $E(\theta_t | y^{(t)})$  nos dá informação sobre o comportamento do processo no tempo  $t$ , enquanto que  $E(\lambda | y^{(t)})$  nos dá informação sobre o comportamento global do processo. Assumindo que  $\lambda$  tem distribuição normal  $N(m, S^2)$  com  $m$  e  $S^2$  conhecidos, o modelo considerado é estendido para:

$$\begin{aligned} y_t &= \theta_t y_{t-1} + \epsilon_t ; \quad \epsilon_t \sim N(0, \tau^2) \quad \tau^2 \text{ conhecido} \\ \theta_t &= \lambda + \omega_t ; \quad \omega_t \sim N(0, \gamma^2) \quad \gamma^2 \text{ conhecido} \\ \lambda &\sim N(m, S^2) \quad m \text{ e } S^2 \text{ conhecidos} \end{aligned} \quad (2.2)$$

### 2.1.1 A Distribuição *a Posteriori* de $\lambda$ e $\theta_t$

Denotando a função densidade de  $\theta_t$  dado  $y^{(t)}$ , por  $p(\theta_t | y^{(t)})$ , vamos calcular essa densidade considerando a função densidade de probabilidade:

$$f(y_t | y_{t-1}, \theta_t, \tau) \propto \tau^{-1} \exp \left\{ -\frac{1}{2\tau^2} \left( y_t - \theta_t y_{t-1} \right)^2 \right\} \quad (2.3)$$

E, considerando a densidade *a priori* para  $\lambda$ :

$$\Pi(\lambda) \propto S^{-1} \exp \left\{ -\frac{1}{2S^2} (\lambda - m)^2 \right\} \quad (2.4)$$

Aplicando o teorema de Bayes, temos que densidade de probabilidade para  $\lambda$ :

$$p(\lambda | y^{(t)}) = \frac{p(y^{(t)} | \lambda) \Pi(\lambda)}{\int p(y^{(t)} | \lambda) \Pi(\lambda) d\lambda} \quad (2.5)$$

onde  $p(y^{(t)}|\lambda)$  é a função de verossimilhança de  $\lambda$ .

Assim, temos a densidade *a posteriori* para  $\theta_t$  dada por:

$$p(\theta_t|y^{(t)}) = \int p(\theta_t|y^{(t)}, \lambda)p(\lambda|y^{(t)})d\lambda \quad (2.6)$$

Podemos determinar  $p(\lambda|y^{(t)})$  por:

$$p(y^{(t)}|\lambda) = \int \dots \int p(y_1, \dots, y_t|\theta_1, \dots, \theta_t, \lambda) p(\theta_1, \dots, \theta_t|\lambda) d\theta_1, \dots, d\theta_t$$

Contudo, dado  $\lambda$  e considerando os  $\theta_t$ 's independentes, assim :

$$p(\theta_1, \dots, \theta_t|\lambda) = \prod_{i=2}^t p(\theta_i|\lambda)$$

onde  $p(\theta_t|\lambda) \sim N(\lambda, \gamma^2)$  é a função densidade de  $\theta_t$  dado  $\lambda$ , portanto:

$$p(y^{(t)}|\lambda) = \int \dots \int p(y^{(t)}|\theta^{(t)}) \left[ \prod_{i=2}^t p(\theta_i|\lambda) \right] d\theta_1, \dots, d\theta_t \quad (2.7)$$

Por outro lado, usando a regra da multiplicação de probabilidades e a propriedade de Markov para o processo auto-regressivo de primeira ordem, podemos escrever:

$$p(y^{(t)}|\theta^{(t)}, \lambda) = p(y_1|\theta_1, \lambda)p(y_2|y_1, \theta_2, \lambda)\dots p(y_t|y_{t-1}, \theta_t, \lambda)$$

Na seqüência, para simplificar essa expressão e também obter outros valores de interesse, temos que:

$$p(y^{(t)}|\theta^{(t)}, \lambda) = \prod_{i=2}^t p(y_i|y_{i-1}, \theta_i, \lambda)p(y_1|\theta_1, \lambda) \quad (2.8)$$

Dada a seqüência  $\theta^{(t)}$ ,  $y_t$  é independente de  $\lambda$  e dependente de  $y_{t-1}$  e  $\theta_t$ , onde  $\theta^{(t)} = (\theta_1, \theta_2, \dots, \theta_t)$ . Desse modo,

$$p(y^{(t)}|\theta^{(t)}, \lambda) = \prod_{i=2}^t p(y_i|y_{i-1}, \theta_i) \quad (2.9)$$

Então, a função de verossimilhança de  $\lambda$  é:

$$p(y^{(t)}|\lambda) = \prod_{i=2}^t \int p(y_i|y_{i-1}, \theta_i)p(\theta_i|\lambda)d\theta_i \quad (2.10)$$

Sendo:

$$p(y_t|y_{t-1}, \lambda) = \int p(y_t|y_{t-1}, \theta_t)p(\theta_t|\lambda)d\theta_t \quad (2.11)$$

Podemos, então, escrever a eq(2.10) como:

$$p(y^{(t)}|\lambda) = \prod_{i=2}^t p(y_i|y_{i-1}, \lambda) \quad (2.12)$$

Da eq(2.3) implica também que:

$$(y_t|y_{t-1}, \theta_t, \tau^2) \sim N(\theta_t y_{t-1}, \tau^2)$$

$$(\theta_t|\lambda, \gamma^2) \sim N(\lambda, \gamma^2) \quad \text{com } \gamma^2 \text{ conhecido}$$

Podemos, agora, obter  $p(y_t|y_{t-1}, \lambda)$ , usando a estrutura linear gaussiana. Note que, desde que  $(\theta_t|\lambda, \gamma^2) \sim N(\lambda, \gamma^2)$ , podemos escrever:

$$\theta_t = \lambda + \omega_t \quad \omega_t \sim N(0, \gamma^2) \quad (2.13)$$

$$y_t = \theta_t y_{t-1} + \epsilon_t \quad \epsilon_t \sim N(0, \tau^2) \quad (2.14)$$

Pela substituição da expressão (2.13) em (2.14), podemos escrever:

$$y_t = (\lambda + \omega_t)y_{t-1} + \epsilon_t$$

onde  $\lambda$  é independente de  $\omega_t$  e de  $\epsilon_t$ , e  $E(\epsilon_t \omega_t) = 0$

$$y_t - \lambda y_{t-1} = \omega_t y_{t-1} + \epsilon_t \implies (\omega_t y_{t-1} + \epsilon_t) \sim N(0, (y_{t-1}^2 \gamma^2 + \tau^2))$$

$$E(y_t|y_{t-1}, \lambda) = \lambda y_{t-1}$$

$$Var(y_t|y_{t-1}, \lambda) = y_{t-1}^2 \gamma^2 + \tau^2$$

Logo, dado  $\lambda$  e  $y_{t-1}$ , temos:

$$(y_t|y_{t-1}, \lambda) \sim N[\lambda y_{t-1}, (\gamma^2 y_{t-1}^2 + \tau^2)] \quad (2.15)$$

Então, a função de verossimilhança de  $\lambda$  (2.12) é da forma :

$$p(y^{(t)}|\lambda) = \prod_{i=2}^t \exp \left\{ -\frac{1}{2} \frac{(y_i - \lambda y_{i-1})^2}{r_i} \right\} \quad (2.16)$$

onde

$$r_i = \gamma^2 y_{i-1}^2 + \tau^2$$

A distribuição *a posteriori* de  $\lambda$  agora é obtida inserindo a eq(2.16) na eq(2.5), isto é:

$$p(\lambda|y^{(t)}) \propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=2}^t \frac{(y_i - \lambda y_{i-1})^2}{r_i} + \frac{(\lambda - m)^2}{S^2} \right] \right\} \quad (2.17)$$

ou seja,

$$p(\lambda|y^{(t)}) \propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=2}^t \frac{y_{i-1}^2}{r_i} \left( \lambda - \frac{y_i}{y_{i-1}} \right)^2 + \frac{(\lambda - m)^2}{S^2} \right] \right\}$$

Usando a identidade (Box e Tiao, 1973):

$$A(z - a)^2 + B(z - b)^2 = (A + B)(z - c)^2 + \frac{AB}{A + B}(a - b)^2 \quad (2.18)$$

onde

$$c = \frac{Aa + Bb}{A + B}$$

fazendo:

$$A = \sum_{i=2}^t \frac{y_{i-1}^2}{r_i} \quad ; \quad B = \frac{1}{S^2} \quad ; \quad a = \frac{y_i}{y_{i-1}} \quad ; \quad b = m$$

$$A + B = \sum_{i=2}^t \frac{y_{i-1}^2}{r_i} + \frac{1}{S^2} \quad ; \quad AB = \sum_{i=2}^t \frac{y_{i-1}^2}{S^2 r_i}$$

$$Aa + Bb = \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i} + \frac{m}{S^2}$$

Denominando:

$$m_t = \left( \frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i} \right) S_t \quad e \quad S_t = \left( \frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i} \right)^{(-1)}$$

Verificamos, então, que a distribuição *a posteriori* de  $\lambda$  dado  $y^{(t)}$  é da forma:

$$p(\lambda|y^{(t)}) \propto \exp \left\{ -\frac{1}{2S_t} (\lambda - m_t)^2 \right\} \quad (2.19)$$

logo,

$$(\lambda|y^{(t)}) \sim N(m_t, S_t)$$

A média e a variância *a posteriori* de  $\lambda$  também podem ser obtidas recursivamente.

Para isso, podemos escrever  $m_t$  e  $S_t$  como:

$$m_t = \frac{S_{t-1}y_t y_{t-1} + m_{t-1}r_t}{S_{t-1}y_{t-1}^2 + r_t} \quad (2.20)$$

$$S_t = \frac{S_{t-1}r_t}{S_{t-1}y_{t-1}^2 + r_t} \quad (2.21)$$

### 2.1.2 Distribuição Condicional para $\theta_t$

Em seguida, calcularemos  $p(\theta_t|y^{(t)}, \lambda)$ . Desde que, dado  $y_t, y_{t-1}$  e  $\lambda$ ,  $\theta_t$  é independente de  $y^{(t-2)}$ , isto é:

$$p(\theta_t|y^{(t)}, \lambda) = p(\theta_t|y_t, y_{t-1}, \lambda)$$

Podemos ver isso através do teorema de Bayes, assim temos:

$$p(\theta_t|y_t, y^{(t-1)}, \lambda) = \frac{p(y_t|\theta_t, y^{(t-1)}, \lambda)p(\theta_t|y^{(t-1)}, \lambda)}{p(y_t|y^{(t-1)}, \lambda)}$$

Mas, dado  $\theta_t$  e  $y_{t-1}$ ,  $y_t$  é independente de  $y^{(t-2)}$  e  $\lambda$ . Além disso, condicionado a  $\lambda$   $\theta_t$  é independente de  $y^{(t-1)}$  e  $y_t$  é independente de  $y^{(t-2)}$ . Então:

$$p(\theta_t|y_t, y^{(t-1)}, \lambda) = \frac{p(y_t|\theta_t, y_{t-1})p(\theta_t|\lambda)}{p(y_t|y_{t-1}, \lambda)} \quad (2.22)$$

Similarmente,

$$p(\theta_t|y_t, y_{t-1}, \lambda) = \frac{p(y_t|\theta_t, y_{t-1}, \lambda)p(\theta_t|y_{t-1}, \lambda)}{p(y_t|y_{t-1}, \lambda)}$$

Portanto,

$$p(\theta_t|y_t, y_{t-1}, \lambda) = \frac{p(y_t|\theta_t, y_{t-1})p(\theta_t|\lambda)}{p(y_t|y_{t-1}, \lambda)}$$

Então,

$$p(\theta_t|y^{(t)}, \lambda) = p(\theta_t|y_t, y_{t-1}, \lambda)$$

Se ignorarmos a constante de proporcionalidade em (2.22) e se substituirmos a expressão apropriada no numerador, onde  $(y_t|y_{t-1}, \lambda, \theta_t, \tau^2) \sim N(\theta_t y_{t-1}, \tau^2)$ , temos:

$$p(\theta_t|y_t, y_{t-1}, \lambda) \propto \exp\left\{-\frac{1}{2}\left[\frac{(y_t - \theta_t y_{t-1})^2}{\tau^2} + \frac{(\theta_t - \lambda)^2}{\gamma^2}\right]\right\} \quad (2.23)$$

Aplicando a identidade (2.18), temos:

$$p(\theta_t|y_t, y_{t-1}, \lambda) \propto \exp\left\{-\frac{1}{2\hat{\Sigma}_t}(\theta_t - \hat{\theta}_t)^2\right\} \quad (2.24)$$

onde

$$\hat{\theta}_t = \frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{r_t}$$

$$\hat{\Sigma}_t = \frac{\tau^2 \gamma^2}{r_t} \quad ; \quad r_t = \gamma^2 y_{t-1}^2 + \tau^2$$

Logo,

$$(\theta_t|y^{(t)}, \lambda) = (\theta_t|y_t, y_{t-1}, \lambda) \sim N(\hat{\theta}_t, \hat{\Sigma}_t) \quad (2.25)$$

A distribuição *a posteriori*  $(\theta_t|y^{(t)})$  é calculada de (2.19) e (2.24), onde temos:

$$\theta_t = \hat{\theta}_t + \mu_t \quad ; \quad \mu_t \sim N(0, \hat{\Sigma}_t)$$

$\mu_t$  e  $\lambda$  são independentes.

Então:

$$E(\theta_t|y^{(t)}) = E[E(\theta_t|y^{(t)}, \lambda)]$$

$$E(\theta_t|y^{(t)}) = \frac{\tau^2 E(\lambda|y^{(t)}) + \gamma^2 y_t y_{t-1}}{r_t}$$

Logo,

$$E(\theta_t|y^{(t)}) = \frac{\tau^2 m_t + \gamma^2 y_t y_{t-1}}{r_t} \quad (2.26)$$

Usando que :

$$V(\theta_t|y^{(t)}) = E[V(\theta_t|y^{(t)}, \lambda)] + V[E(\theta_t|y^{(t)}, \lambda)]$$

Temos que:

$$V(\theta_t|y^{(t)}) = E\left[\frac{\tau^2\gamma^2}{r_t}\right] + V\left[\frac{\tau^2\lambda + \gamma^2 y_t y_{t-1}}{r_t}\right]$$

$$V(\theta_t|y^{(t)}) = \frac{\tau^2\gamma^2}{r_t} + \frac{\tau^4}{r_t^2} V(\lambda|y^{(t)})$$

$$V(\theta_t|y^{(t)}) = \frac{\tau^2\gamma^2}{r_t} + \frac{\tau^4}{r_t^2} S_t \quad (2.27)$$

Logo, a densidade *a posteriori* para  $(\theta_t|y^{(t)})$  é:

$$(\theta_t|y^{(t)}) \sim N(\theta_t^* , \Sigma_t^*) \quad (2.28)$$

Onde  $\theta_t^*$  e  $\Sigma_t^*$  são dados por (2.26)e (2.27), respectivamente.

## 2.2 Distribuição Preditiva de $\theta_{t+1}$ e de $y_{t+1}$

Dado os dados  $y^{(t)}$ , queremos fazer, no tempo  $t$ , previsões para  $\theta_{t+1}$  e  $y_{t+1}$

### 2.2.1 Densidade Preditiva de $\theta_{t+1}$

A distribuição preditiva de  $\theta_{t+1}$  é dada como segue:

$$p(\theta_{t+1}|y^{(t)}) = \int p(\theta_{t+1}|y_t, \lambda)p(\lambda|y^{(t)})d\lambda$$

Considerando que, dado  $\lambda$ ,  $\theta_{t+1}$  é independente de  $y_t$ , assim

$$p(\theta_{t+1}|y^{(t)}) = \int p(\theta_{t+1}|\lambda)p(\lambda|y^{(t)})d\lambda \quad (2.29)$$

Análogo a eq(2.13), podemos escrever:

$$\theta_{t+1} = \lambda + \omega_{t+1} , \text{ onde } \omega_{t+1} \sim N(0, \gamma^2)$$

e  $\lambda$  é independente de  $\omega_{t+1}$  e

$$E(\theta_{t+1}|y^{(t)}) = E(\lambda + \omega_{t+1}|y^{(t)})$$

da (2.19) temos que:

$$E(\theta_{t+1}|y^{(t)}) = E(\lambda|y^{(t)}) = m_t$$



e 
$$V(\theta_{t+1}|y^{(t)}) = V(\lambda|y^{(t)}) + V(\omega_{t+1}|y^{(t)})$$

logo, 
$$(\theta_{t+1}|y^{(t)}) = S_t + \gamma^2$$

onde

$$m_t = \frac{S_{t-1}y_t y_{t-1} + m_{t-1}r_t}{S_{t-1}y_{t-1}^2 + r_t}$$

$$S_t = \frac{S_{t-1}r_t}{S_{t-1}y_{t-1}^2 + r_t}$$

$$r_t = \gamma^2 y_{t-1}^2 + \tau^2$$

Logo,

$$(\theta_{t+1}|y^{(t)}) \sim N(m_t, S_t + \gamma^2) \quad (2.30)$$

Portanto;  $E(\theta_{t+1}|y^{(t)}) = m_t$  é a média *a posteriori* de  $(\lambda|y^{(t)})$  ver(2.20)

## 2.2.2 Densidade Preditiva de $y_{t+1}$

Da mesma forma, a densidade preditiva de  $y_{t+1}$  dado  $y^{(t)}$  é:

$$p(y_{t+1}|y^{(t)}) = \int p(y_{t+1}|y^{(t)}, \lambda) p(\lambda|y^{(t)}) d\lambda$$

porém, dado  $\lambda$ ,  $y_{t+1}$  é independente de todos  $y$ 's anteriores a  $y_t$  e

$$p(y_{t+1}|y^{(t)}) = \int p(y_{t+1}|y_t, \lambda) p(\lambda|y^{(t)}) d\lambda \quad (2.31)$$

Usando eq(2.15), podemos escrever:

$$y_{t+1} = \lambda y_t + \mu_{t+1}, \text{ onde } \mu_{t+1} \sim N(0, r_{t+1})$$

$\lambda$  é independente de  $\mu_{t+1}$  e

$$r_{t+1} = \gamma^2 y_t^2 + \tau^2$$

Então,

$$E(y_{t+1}|y^{(t)}) = E(y_t \lambda | y^{(t)}) + E(\mu_{t+1} | y^{(t)}) \implies E(y_{t+1}|y^{(t)}) = m_t y_t$$

$$V(y_{t+1}|y^{(t)}) = V(y_t \lambda | y^{(t)}) + V(\mu_{t+1} | y^{(t)})$$

$$V(y_{t+1}|y^{(t)}) = y_t^2 V(\lambda|y^{(t)}) + V(\mu_{t+1}|y^{(t)})$$

$$V(y_{t+1}|y^{(t)}) = y_t^2 S_t + r_{t+1}$$

logo ,

$$(y_{t+1}|y^{(t)}) \sim N(m_t y_t, y_t^2 S_t + r_{t+1}) \quad (2.32)$$

### 2.2.3 Intervalo de Credibilidade de $y_{t+1}$

O intervalo com  $(1 - \alpha)100\%$  de credibilidade para  $y_{t+1}$  é dado por :

$$m_t y_t \pm z_{\frac{\alpha}{2}} \sqrt{y_t^2 S_t + r_{t+1}} \quad (2.33)$$

onde  $z_{\frac{\alpha}{2}}$  é o percentil  $100(1 - \frac{\alpha}{2})\%$  da distribuição normal padronizada.

### 2.2.4 Densidade Preditiva de $\theta_{t+k}$ e $y_{t+k}$

Quando fazemos predição para  $k$ -passos , para  $k \geq 2$  , a estrutura linear descrita anteriormente desaparece e temos que considerar a distribuição preditiva da média de  $y_{t+k}$ . Para observarmos isso, consideramos:

$$y_{t+2} = \theta_{t+2} y_{t+1} + \epsilon_{t+2} \quad (2.34)$$

para o tempo  $t$ , depois da observação  $y_t$ , a predição de  $\theta_{t+2}$  é descrita por eq(2.30), ou seja :

$$(\theta_{t+k}|y^{(t)}) \sim N(m_t, (S_t + \gamma^2)) \quad \text{para todo } k \geq 1 \quad (2.35)$$

isto porque  $(\theta_{t+k})$  é uma seqüência permutável.

Por eq(2.34) temos que a densidade preditiva  $p(y_{t+2}|y^{(t)})$  não pode ser uma normal; no entanto, podemos obter a média preditiva de  $y_{t+k}$  dado  $y^{(t)}$ , denotado por  $E(y_{t+k}|y^{(t)})$ , para  $k \geq 2$  . Se na condição (2.34) considerarmos dado  $y^{(t)}$  e  $\lambda$ , então:

$$E(y_{t+2}|y^{(t)}, \lambda) = E(\theta_{t+2} y_{t+1}|y^{(t)}, \lambda)$$

Além disso, dado  $\lambda$  e  $y^{(t)}$ ,  $y_{t+1}$  é independente de  $\theta_{t+2}$ , logo

$$E(\theta_{t+2} y_{t+1}|y^{(t)}, \lambda) = E(\theta_{t+2}|y^{(t)}, \lambda) E(y_{t+1}|y^{(t)}, \lambda)$$

sendo:

$$E(\theta_{t+2}|y^{(t)}, \lambda) = \lambda$$

$$E(y_{t+1}|y^{(t)}, \lambda) = E(\theta_t y_t | y^{(t)}, \lambda)$$

$$E(\theta_{t+2} y_{t+1} | y^{(t)}, \lambda) = \lambda E(\theta_t y_t | y^{(t)}, \lambda)$$

$$E(\theta_{t+2} y_{t+1} | y^{(t)}, \lambda) = \lambda y_t \lambda$$

portanto,

$$E(y_{t+2}|y^{(t)}, \lambda) = \lambda^2 y_t \tag{2.36}$$

Sendo,

$$E(y_{t+2}|y^{(t)}) = E[E(y+2)|y^{(t)}, \lambda] = E(\lambda^2|y^{(t)}) y_t$$

Como  $\lambda|y^{(t)} \sim N(m_t, S_t)$ , a função geradora de momentos de  $\lambda$  é dada por:

$$M_\lambda(x) = e^{(m_t x + \frac{S_t x^2}{2})}$$

e

$$E(\lambda^2|y^{(t)}) = \left. \frac{d^2 M_\lambda(x)}{dx^2} \right|_{x=0}$$

$$E(\lambda^2|y^{(t)}) = \left[ S_t e^{(m_t x + \frac{S_t x^2}{2})} + (m_t + S_t x)(m_t + S_t x) e^{(m_t x + \frac{S_t x + S_t x^2}{2})} \right]$$

para  $x = 0$

$$E(\lambda^2|y^{(t)}) = S_t + m_t^2$$

logo,

$$E(y_{t+2}|y^{(t)}) = (S_t + m_t^2) y_t \tag{2.37}$$

Usando (2.31), podemos escrever:

$$E(y_{t+2}|y^{(t)}) = m_t E(y_{t+1}|y^{(t)}) + S_t y_t$$

Denotando

$$\hat{y}_{t+1} = E(y_{t+1}|y^{(t)})$$

temos:

$$\hat{y}_{t+2} = m_t \hat{y}_{t+1} + S_t y_t \quad (2.38)$$

Similarmente, temos que a média preditiva de  $y_{t+3}$  dado  $y^{(t)}$  é :

$$E(y_{t+3}|y^{(t)}, \lambda) = E(\theta_{t+3} y_{t+2} | y^{(t)}, \lambda)$$

$$E(\theta_{t+3} y_{t+2} | y^{(t)}, \lambda) = E(\theta_{t+3} | y^{(t)}, \lambda) E(y_{t+2} | y^{(t)}, \lambda)$$

$$E(\theta_{t+3} | y^{(t)}, \lambda) E(y_{t+2} | y^{(t)}, \lambda) = \lambda \lambda^2 y_t$$

logo,

$$\hat{y}_{t+3} = E(\lambda^3 | y^{(t)}) y_t$$

onde

$$E(\lambda^3 | y^{(t)}) = \left. \frac{d^3 M_\lambda(x)}{dx^3} \right|_{x=0}$$

$$E(\lambda^3 | y^{(t)}) = S_t (m_t + S_t x) e^{(m_t x + \frac{S_t x^2}{2})} + (2m_t S_t + 2S_t x) e^{(m_t x + \frac{S_t x^2}{2})} + \\ (m_t^2 + 2m_t S_t x + S_t x^2) (m_t + S_t x) e^{(m_t x + \frac{S_t x^2}{2})}$$

para  $x = 0$

$$E(\lambda^3 | y^{(t)}) = S_t m_t + 2m_t S_t + m_t^2 (m_t) = 3m_t S_t + m_t^3$$

Assim,

$$\hat{y}_{t+3} = E(\lambda^3 | y^{(t)}) y_t = (m_t^3 + 3m_t S_t) y_t ,$$

pode ser escrito como:

$$\hat{y}_{t+3} = m_t (m_t^2) y_t + 2S_t m_t y_t$$

Portanto,

$$\hat{y}_{t+3} = m_t \hat{y}_{t+2} + 2S_t \hat{y}_{t+1} \quad (2.39)$$

Generalizando, temos que a média preditiva de  $y_{t+k}$  pode ser escrita como:

$$\begin{aligned} \hat{y}_{t+k} &= E(\lambda^k | y^{(t)}) y_t \\ \hat{y}_{t+k} &= m_t \hat{y}_{t+k-1} + (k-1) S_t \hat{y}_{t+k-2} \end{aligned} \quad (2.40)$$

## 2.3 Resultados

Resumindo os resultados desse modelo, temos:

Modelo:

$$y_t = \theta_t y_{t-1} + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \tau^2)$$

$$\theta_t = \lambda + \omega_t \quad ; \quad \omega_t \sim N(0, \gamma^2)$$

$$\lambda \sim N(m, S^2)$$

com ,  $\epsilon_t$  e  $\omega_t$  independentes e  $E(\omega_t \omega_{t+k}) = 0$  ,  $k \neq 0$  ;  $E(\epsilon_t \epsilon_{t+k}) = 0$  ,  $k \neq 0$

As densidades *a posteriori* são:

$$\lambda | y^{(t)} \sim N(m_t, S_t)$$

com,

$$m_t = \frac{S_{t-1} y_t y_{t-1} + m_{t-1} r_t}{S_{t-1} y_{t-1}^2 + r_t} \quad ; \quad S_t = \frac{S_{t-1} r_t}{S_{t-1} y_{t-1}^2 + r_t}$$

e

$$\theta_t | y^{(t)} \sim N(\theta_t^* , \Sigma_t^*)$$

$$\theta_t^* = \frac{\tau^2 m_t + \gamma^2 y_t y_{t-1}}{r_t} \quad ; \quad \Sigma_t^* = \frac{\tau^4}{r_t^2} S_t + \frac{\tau^2 \gamma^2}{r_t}$$

As densidades preditivas para  $\theta_{t+k}$ ,  $k \geq 1$  são dadas por:

$$\theta_{t+k} \sim N(m_t, S_t + \gamma^2)$$

As previsões para  $y_{t+k}$  dado  $y^{(t)}$  podem ser feitas por:

$$\hat{y}_{t+k} = m_t \hat{y}_{t+k-1} + (k-1)S_t \hat{y}_{t+k-2} \quad ; \quad k \geq 1$$

Esses resultados serão utilizados no capítulo 5 para fazer inferência e previsões de dados de uma séries temporais de índice de preços de ações e do preço da arroba do boi gordo.

No próximo capítulo, as técnicas utilizadas nesse modelo serão estendidas para o processo de primeira ordem, quando a variância  $\tau^2$  do ruído  $\epsilon_t$ , é considerada desconhecida.

# Capítulo 3

## Modelo Hierárquico para Processo de Primeira Ordem com Variância Desconhecida

### 3.1 Inferência para os Parâmetros do Modelo

Neste capítulo, desenvolvemos um modelo Hierárquico para um processo auto-regressivo de primeira ordem, com a variância do ruído ( $\tau^2$ ) desconhecida. Descrevemos nossa incerteza sobre essa variância considerando uma distribuição *a priori*  $IG(\alpha, \beta)$  para  $\tau$ .

Esse processo é uma generalização do modelo descrito no capítulo anterior, ou seja:

$$\begin{aligned}y_t &= \theta_t y_{t-1} + \epsilon_t ; \quad \epsilon_t \sim N(0, \tau^2) \quad \tau^2 \text{ desconhecido} \\ \theta_t &= \lambda + \omega_t ; \quad \omega_t \sim N(\lambda, \gamma^2) \quad \gamma^2 \text{ conhecido} \\ \lambda &\sim N(m, S^2) \quad m \text{ e } S^2 \text{ conhecidos}\end{aligned} \tag{3.1}$$

Seja a série  $y^{(t)} = (y_1, y_2, \dots, y_t)$  e a seqüência  $\theta^{(t)} = (\theta_1, \dots, \theta_t)$  e dado  $y^{(t)}$ , nosso objetivo é fazer inferência sobre  $\theta_t$ ,  $\tau$ ,  $\lambda$  e previsão para  $\theta_{t+1}$  e  $y_{t+1}$ .

#### 3.1.1 A Distribuição Conjunta *a Posteriori* de $\lambda$ e $\tau$

Se considerarmos a distribuição de  $(\theta_t, \tau)$  dado  $y^{(t)}$  e  $\lambda$  por  $p(\theta_t, \tau | y^{(t)}, \lambda)$ , e aplicando o mesmo raciocínio utilizado no caso da variância conhecida, podemos calcular essa

densidade considerando a função densidade de probabilidade:

$$f(y_t|\theta_t, \tau) \propto \tau^{-1} \exp\left\{-\frac{1}{2\tau^2} \left(y_t - \theta_t y_{t-1}\right)^2\right\} \quad (3.2)$$

E, considerando as densidades *a priori* para  $\lambda$  como uma normal  $N(m, S^2)$  e para  $\tau$  como uma gama inversa  $IG(\alpha, \beta)$ :

$$\Pi_1(\tau) \propto \tau^{-(\alpha+1)} \exp\left\{-\frac{\beta}{\tau}\right\} \quad (3.3)$$

$$\Pi_2(\lambda) \propto S^{-1} \exp\left\{-\frac{1}{2S^2}(\lambda - m)^2\right\}$$

Aplicando o teorema de Bayes, temos que densidade de probabilidade para  $\lambda$  e  $\tau$ :

$$p(\lambda, \tau|y^{(t)}) = \frac{p(y^{(t)}|\lambda, \tau)\Pi(\lambda)\Pi(\tau)}{\int p(y^{(t)}|\lambda, \tau)\Pi(\lambda)\Pi(\tau)d\lambda d\tau} \quad (3.4)$$

onde  $p(y^{(t)}|\lambda, \tau)$  é a função de verossimilhança de  $(\lambda, \tau)$  :

$$p(y^{(t)}|\lambda, \tau) = \int \dots \int p(y^{(t)}|\theta^{(t)}, \lambda, \tau)p(\theta^{(t)}|\lambda, \tau)d\theta_1, \dots, d\theta_t \quad (3.5)$$

Considerando os  $\theta_i$ 's independentes, dados  $\lambda$  e  $\tau$  temos que:

$$p(\theta^{(t)}|\lambda, \tau) = \prod_{i=2}^t p(\theta_i|\lambda, \tau) \quad (3.6)$$

onde  $p(\theta^{(t)}|\lambda, \tau)$  é a função densidade de  $\theta^{(t)}$  dados  $\lambda$  e  $\tau$ .

Podemos determinar  $p(y^{(t)}|\lambda, \tau)$  por:

$$p(y^{(t)}|\lambda, \tau) = \int \dots \int p(y^{(t)}|\theta^{(t)}, \lambda, \tau) \prod_{i=2}^t p(\theta_i|\lambda, \tau)d\theta_i \quad (3.7)$$

Por outro lado, usando a regra da multiplicação de probabilidades e a propriedade de Markov para o processo auto-regressivo de primeira ordem, podemos escrever:

$$p(y^{(t)}|\theta^{(t)}, \lambda, \tau) = p(y_1|\theta_1, \lambda, \tau)p(y_2|y_1, \theta_2, \lambda, \tau)\dots p(y_t|y_{t-1}, \theta_t, \lambda, \tau) \quad (3.8)$$

Na seqüência, para simplificar esta expressão, podemos escrever:

$$p(y^{(t)}|\theta^{(t)}, \lambda, \tau) = \prod_{i=2}^t p(y_i|y_{i-1}, \theta_i, \lambda, \tau) \quad (3.9)$$



Dada a seqüência  $\theta^{(t)}$ ,  $y_t$  é independente de  $\lambda$  e dependente de  $y_{t-1}$ ,  $\theta_t$  e  $\tau$ . Desse modo, temos que a função de verossimilhança é:

$$p(y^{(t)}|\lambda, \tau) = \prod_{i=2}^t \int p(y_i|y_{i-1}, \theta_i, \tau) p(\theta_i|\lambda, \tau) d\theta_i \quad (3.10)$$

Sendo:

$$p(y_t|y_{t-1}, \lambda, \tau) = \int p(y_t|y_{t-1}, \theta_t, \tau) p(\theta_t|\lambda, \tau) d\theta_t \quad (3.11)$$

Podemos, então, escrever a eq(3.10) como:

$$p(y^{(t)}|\lambda, \tau) = \prod_{i=2}^t p(y_i|y_{i-1}, \lambda, \tau) \quad (3.12)$$

Da eq(3.2) implica também que:

$$(y_t|y_{t-1}, \theta_t, \tau) \sim N(\theta_t y_{t-1}, \tau^2)$$

e

$$(\theta_t|\lambda, \gamma^2) \sim N(\lambda, \gamma^2) \quad \text{com } \gamma^2 \text{ conhecido}$$

Podemos, agora, obter  $p(y_t|y_{t-1}, \lambda, \tau)$ , usando a estrutura linear gaussiana. Note que, desde que  $(\theta_t|\lambda, \gamma^2) \sim N(\lambda, \gamma^2)$ , podemos escrever:

$$\theta_t = \lambda + \omega_t \quad , \quad \text{onde } \omega_t \sim N(0, \gamma^2) \quad (3.13)$$

$$y_t = \theta_t y_{t-1} + \epsilon_t \quad , \quad \text{onde } \epsilon_t \sim N(0, \tau^2) \quad (3.14)$$

Pela substituição da expressão (3.13) em (3.14), podemos escrever:

$$y_t = (\lambda + \omega_t)y_{t-1} + \epsilon_t \quad , \quad \text{onde } \lambda \text{ independente de } \omega_t \text{ e de } \epsilon_t \text{ e } E(\epsilon_t \omega_t) = 0$$

$$y_t - \lambda y_{t-1} = \omega_t y_{t-1} + \epsilon_t \implies (\omega_t y_{t-1} + \epsilon_t) \sim N(0, (y_{t-1}^2 \gamma^2 + \tau^2))$$

$$E(y_t|y_{t-1}, \lambda, \tau) = \lambda y_{t-1} \quad \text{e} \quad V(y_t|y_{t-1}, \lambda, \tau) = y_{t-1}^2 \gamma^2 + \tau^2$$

Logo, dado  $\lambda$ ,  $\tau$  e  $y_{t-1}$ , temos:

$$(y_t|y_{t-1}, \lambda, \tau) \sim N(\lambda y_{t-1}, (\gamma^2 y_{t-1}^2 + \tau^2)) \quad (3.15)$$

Então, a função de verossimilhança de (3.12) é da forma :

$$p(y^{(t)}|\lambda, \tau) = \prod_{i=2}^t \frac{1}{(r_i(\tau))^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \lambda y_{i-1})^2}{r_i(\tau)} \right\} \quad \text{onde} \quad r_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2 \quad (3.16)$$

A distribuição conjunta *a posteriori* de  $\lambda$  e  $\tau$  agora é obtida inserindo a eq(3.16) em eq(3.4), isto é:

$$p(\lambda, \tau|y^{(t)}) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=2}^t \frac{(y_i - \lambda y_{i-1})^2}{r_i(\tau)} \right\} \exp \left\{ -\frac{(\lambda - m)^2}{S^2} \right\} \exp \left\{ -\frac{\beta}{\tau} \right\} \quad (3.17)$$

ou seja,

$$p(\lambda, \tau|y^{(t)}) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left( \lambda - \frac{y_i}{y_{i-1}} \right)^2 + \frac{(\lambda - m)^2}{S^2} \right] \right\} \exp \left\{ -\frac{\beta}{\tau} \right\}$$

Usando a identidade (2.18), fazendo:

$$A = \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \quad ; \quad B = \frac{1}{S^2} \quad ; \quad z = \lambda \quad ; \quad a = \frac{y_t}{y_{t-1}} \quad ; \quad b = m$$

$$AB = \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau) S^2} \quad ; \quad A + B = \left( \frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \right)$$

$$Aa + Bb = \left( \frac{\lambda}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i(\tau)} \right)$$

Denominando:

$$S_t(\tau) = \left( \frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \right)^{-1} \quad ; \quad m_t(\tau) = \left( \frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i(\tau)} \right) S_t(\tau)$$

Podemos reescrever a *posteriori* conjunta como:

$$p(\lambda, \tau | y^{(t)}) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2S_t(\tau)} \left(\lambda - m_t(\tau)\right)^2\right\} \times \quad (3.18)$$

$$\exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Aqui, observamos que a densidade conjunta *a posteriori* não tem uma desidade padrão; no entanto, podemos determinar a distribuição marginal para  $\tau$  por integração na variável  $\lambda$

A distribuição marginal para  $\tau$  é dada por:

$$p(\tau | y^{(t)}) = \int_{-\infty}^{+\infty} p(\lambda, \tau | y^{(t)}) d\lambda$$

isto é

$$p(\tau | y^{(t)}) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\} \times$$

$$\exp\left\{-\frac{\beta}{\tau}\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2S_t(\tau)} \left(\lambda - m_t(\tau)\right)^2\right\} d\lambda$$

ou seja,

$$p(\tau | y^{(t)}) \propto \frac{S_t(\tau) \tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Assim, a distribuição marginal de  $\tau$  pode ser escrita como:

$$p(\tau | y^{(t)}) \propto \frac{\left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)}\right)^{-1} \tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\} \quad (3.19)$$

onde

$$r_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

Notamos da (3.19) que uma análise dos momentos *a posteriori* para  $\tau$  não é possível sem utilização de técnicas de aproximação numérica ou simulação de Monte Carlo (Gamerman, 1996). Podemos avaliar as densidades *a posteriori* marginais usando simulação de Monte Carlo em cadeia de Markov a partir das condicionais .

### 3.1.2 Distribuição Condicional para $\lambda$ e $\tau$

A distribuição condicional *a posteriori* para  $\lambda$  dado  $y^{(t)}$  e  $\tau$  é da forma:

$$p(\lambda|y^{(t)}, \tau) \propto \exp \left\{ -\frac{1}{2S_t(\tau)} (\lambda - m_t(\tau))^2 \right\} \quad (3.20)$$

onde

$$S_t(\tau) = \left( \frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \right)^{-1} \quad ; \quad m_t(\tau) = \left( \frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i(\tau)} \right) S_t(\tau)$$

$$r_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

ou seja ,

$$(\lambda|y^{(t)}, \tau) \sim N(m_t(\tau) , S_t(\tau))$$

A distribuição condicional *a posteriori* para  $\tau$  dado  $y^{(t)}$   $\lambda$  é por:

$$p(\tau|y^{(t)}, \lambda) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp \left\{ -\frac{\beta}{\tau} \right\} \exp \left\{ -\frac{1}{2S_t(\tau)} (\lambda - m_t(\tau))^2 \right\} \times \exp \left\{ -\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left( \frac{y_i}{y_{i-1}} - m \right)^2 \right\} \quad (3.21)$$

### 3.1.3 Distribuição Condicional *a Posteriori* de $\theta_t$

Calcularemos  $p(\theta_t|y^{(t)}, \lambda, \tau)$ . Desde que, dado  $y_t, y_{t-1}, \lambda$  e  $\tau$ ,  $\theta_t$  é independente de  $y^{(t-2)}$ , ou seja:

$$p(\theta_t|y^{(t)}, \lambda, \tau) = p(\theta_t|y_t, y_{t-1}, \lambda, \tau)$$

Através do teorema de Bayes, temos:

$$p(\theta_t|y_t, y^{(t-1)}, \lambda, \tau) = \frac{p(y_t|\theta_t, y^{(t-1)}, \lambda, \tau)p(\theta_t|y^{(t-1)}, \lambda, \tau)}{p(y_t|y^{(t-1)}, \lambda, \tau)}$$

Mas, dado  $\theta_t$  e  $y_{t-1}$ ,  $y_t$  é independente de  $y^{(t-2)}$ ,  $\lambda$ . Além disso, condicionado a  $\lambda$  e  $\tau$ ,  $\theta_t$  é independente de  $y^{(t-1)}$ ,  $y_t$  é independente de  $y^{(t-2)}$ . Então:

$$p(\theta_t|y_t, y^{(t-1)}, \lambda, \tau) = \frac{p(y_t|\theta_t, y_{t-1}, \lambda, \tau)p(\theta_t|\lambda, \tau)}{p(y_t|y_{t-1}, \lambda, \tau)} \quad (3.22)$$

Similarmente,

$$p(\theta_t|y_t, y_{t-1}, \lambda, \tau) = \frac{p(y_t|\theta_t, y_{t-1}, \lambda, \tau)p(\theta_t|y_{t-1}, \lambda, \tau)}{p(y_t|y_{t-1}, \lambda, \tau)}$$

Então,

$$p(\theta_t|y^{(t)}, \lambda, \tau) = p(\theta_t|y_t, y_{t-1}, \lambda, \tau)$$

Se ignorarmos a constante de proporcionalidade em (3.22) e se substituirmos a expressão apropriada no numerador, onde  $(y_t|y_{t-1}, \lambda, \theta_t, \tau) \sim N(\theta_t y_{t-1}, \tau^2)$ , temos:

$$p(\theta_t|y_t, y_{t-1}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2}\left[\frac{(y_t - \theta_t y_{t-1})^2}{\tau^2} + \frac{(\theta_t - \lambda)^2}{\gamma^2}\right]\right\} \quad (3.23)$$

ou seja,

$$p(\theta_t|y_t, y_{t-1}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2}\left[\frac{y_{t-1}^2}{\tau^2}\left(\theta_t - \frac{y_t}{y_{t-1}}\right)^2 + \frac{(\theta_t - \lambda)^2}{\gamma^2}\right]\right\}$$

Aplicando a identidade (2.18), temos:

$$p(\theta_t|y_t, y_{t-1}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2\hat{\Sigma}_t(\tau)}\left(\theta_t - \hat{\theta}_t(\tau)\right)^2\right\} \quad (3.24)$$

onde

$$\hat{\theta}_t(\tau) = \frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{r_t(\tau)}$$

$$r_t(\tau) = \gamma^2 y_{t-1}^2 + \tau^2 \quad ; \quad \hat{\Sigma}_t(\tau) = \frac{\tau^2 \gamma^2}{r_t(\tau)}$$

Logo,

$$(\theta_t|y^{(t)}, \lambda, \tau) = (\theta_t|y_t, y_{t-1}, \lambda) \sim N(\hat{\theta}_t(\tau), \hat{\Sigma}_t(\tau)) \quad (3.25)$$

A distribuição *a posteriori*  $(\theta_t|y^{(t)})$  pode ser calculada de(3.19)e (3.25), onde temos:

$$\theta_t = \hat{\theta}_t + \mu_t ; \quad \mu_t \sim N(0, \hat{\Sigma}_t(\tau))$$

com  $\mu_t$  e  $\lambda$  independentes, temos:

$$\begin{aligned} E(\theta_t|y^{(t)}) &= E[E(\theta_t|y^{(t)}, \lambda, \tau)] \\ E(\theta_t|y^{(t)}) &= E\left[\frac{\tau^2\lambda + \gamma^2y_t y_{t-1}}{\gamma^2y_{t-1}^2 + \tau^2}\right] \end{aligned} \quad (3.26)$$

Usando que :

$$V(\theta_t|y^{(t)}) = E[V(\theta_t|y^{(t)}, \lambda, \tau)] + V[E(\theta_t|y^{(t)}, \lambda, \tau)]$$

Temos que:

$$V(\theta_t|y^{(t)}) = E\left[\frac{\tau^2\gamma^2}{\gamma^2y_{t-1}^2 + \tau^2}\right] + V\left[\frac{\tau^2\lambda + \gamma^2y_t y_{t-1}}{\gamma^2y_{t-1}^2 + \tau^2}\right] \quad (3.27)$$

onde

$$V\left[\frac{\tau^2\lambda + \gamma^2y_t y_{t-1}}{\gamma^2y_{t-1}^2 + \tau^2}\right] = E\left[\frac{\tau^2\lambda + \gamma^2y_t y_{t-1}}{\gamma^2y_{t-1}^2 + \tau^2}\right]^2 - \left[E(\theta_t|y^{(t)})\right]^2$$

O valor esperado e a variância nas expressões(3.26) e (3.27) são calculados em relação às variáveis aleatórias  $\tau$  e  $\lambda$ .

Nesse caso, podemos avaliar as expressões  $E(\theta_t|y^{(t)})$  e  $V(\theta_t|y^{(t)})$  usando os estimadores de Monte Carlo com as amostras geradas de a posteriori conjunta  $p(\lambda, \tau|y^{(t)})$  pelos algoritmos Amostrador de Gibbs [Casella e George, 1992) e Metropolis-Hansting [Chib e Greenberg, 1995), esquematizado da seguinte forma:

- i) inicialize o contador de iterações da cadeia  $j = 1$  e arbitre o valor inicial:

$$\lambda^{(0)}, \tau^{(0)}$$

- ii) obtenha um novo valor a partir da função de transição:

$$(\tau^*) \sim GI(\alpha, \beta)$$

- iii) calcule a probabilidade de aceitação do novo valor gerado:

$$\alpha(\tau^{(j)}, \tau^*) = \begin{cases} \min\left\{1, \frac{\Psi(\tau^*)}{\Psi(\tau^{(j)})}\right\} & \text{se } \Psi(\tau^{(j)}) > 0, \\ 1 & \text{caso contrário} \end{cases}$$

onde

$$\Psi(\tau) = \frac{\exp\left\{-\frac{1}{2S_t(\tau)}(\lambda - m_t(\tau))^2\right\}}{(\tau_t(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

- iv) gere  $u$  de Uniforme  $[0, 1]$  e faça:

$$\tau^{(j+1)} = \begin{cases} \tau^* & \text{se } u \leq \alpha(\tau, \tau^*), \\ \tau^{(j)} & \text{caso contrário.} \end{cases}$$

- v) gere

$$\lambda^{(j)} \sim N(m_t(\tau), S_t(\tau))$$

onde

$$S_t(\tau) = \left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)}\right)^{-1} \quad ; \quad m_t(\tau) = \left(\frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{\tau_i(\tau)}\right) S_t(\tau)$$

$$\tau_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

- vi) atualize o contador de iteração  $j$  para  $j+1$  e retorne a (ii) até que a convergência seja diagnosticada.

Determine:

$$\hat{\lambda} = \frac{1}{M} \sum_{j=1}^M \lambda^{(j)} \quad ; \quad \widehat{S}_\lambda^2 = \frac{1}{M-1} \sum_{j=1}^M \left[ (\lambda^{(j)})^2 - \hat{\lambda} \right]^2 \quad (3.28)$$

$$\hat{\tau} = \frac{1}{M} \sum_{j=1}^M \tau^{(j)} \quad ; \quad \widehat{S}_\tau^2 = \frac{1}{M-1} \sum_{j=1}^M \left[ (\tau^{(j)})^2 - \hat{\tau} \right]^2 \quad (3.29)$$

onde  $M$  é o tamanho da cadeia gerada para  $\lambda^{(j)}$  e  $\tau^{(j)}$ .

Assim, as estimativas de  $\theta_t$  e  $\Sigma_t$  são dadas por:

$$\hat{\theta}_t = E(\theta_t | y^{(t)}) = \frac{1}{M} \sum_{j=1}^M \left[ \frac{\lambda^{(j)} \tau^{(j)2} + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^{(j)2}} \right] \quad (3.30)$$

e

$$\hat{\Sigma}_t = V(\theta_t | y^{(t)}) = \frac{1}{M} \sum_{j=1}^M \left[ \frac{\tau^{(j)2} \gamma^2}{\gamma^2 y_{t-1}^2 + \tau^{(j)2}} \right] + \hat{V} \quad (3.31)$$

onde

$$\hat{V} = \frac{1}{M} \sum_{j=1}^M \left[ \frac{\lambda^{(j)} \tau^{(j)2} + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^{(j)2}} \right]^2 - (\hat{\theta}_t)^2$$

para a amostra gerada  $(\tau^{(j)}, \lambda^{(j)}, j = 1, 2, \dots, M)$ .

## 3.2 Densidade Preditiva para $\theta_{t+1}$ e $y_{t+1}$

Dado os dados  $y^{(t)}$ , queremos, no tempo  $t$ , fazer previsões para  $\theta_{t+1}$  e  $y_{t+1}$ .

### 3.2.1 Densidade Preditiva para $\theta_{t+1}$

A distribuição preditiva de  $\theta_{t+1}$  é encontrada de modo similar ao utilizado no caso de variância conhecida, ou seja:

$$p(\theta_{t+1} | y^{(t)}) = \int \int p(\theta_{t+1} | y_t, \lambda, \tau) p(\lambda, \tau | y^{(t)}) d\lambda d\tau \quad (3.32)$$

Dado  $\lambda$  e  $\tau$ ,  $\theta_{t+1}$  independe de  $y_t$ , então de (3.13), podemos escrever:

$$\theta_{t+1} = \lambda + \omega_{t+1} \quad \text{onde} \quad \omega_{t+1} \sim N(0, \gamma^2)$$

e  $\lambda$  é independente de  $\omega_{t+1}$  e

$$E(\theta_{t+1} | y^{(t)}) = E[(\lambda + \omega_{t+1}) | y^{(t)}]$$

isto é

$$E(\theta_{t+1} | y^{(t)}) = E(\lambda | y^{(t)}) \quad (3.33)$$



e

$$V(\theta_{t+1}|\mathbf{y}^{(t)}) = V[(\lambda + \omega_{t+1})|\mathbf{y}^{(t)}]$$

$$V(\theta_{t+1}|\mathbf{y}^{(t)}) = V(\lambda|\mathbf{y}^{(t)}) + V(\omega_{t+1}|\mathbf{y}^{(t)}) = V(\lambda|\mathbf{y}^{(t)}) + \gamma^2$$

logo

$$V(\theta_{t+1}|\mathbf{y}^{(t)}) = V(\lambda|\mathbf{y}^{(t)}) + \gamma^2 \quad (3.34)$$

Onde  $E(\lambda|\mathbf{y}^{(t)})$  e  $V(\lambda|\mathbf{y}^{(t)})$  são estimados por simulação de Monte Carlo, conforme (3.28).

### 3.2.2 Distribuição Preditiva para $y_{t+1}$

Da mesma forma, a densidade preditiva de  $y_{t+1}$  dado  $\mathbf{y}^{(t)}$  é dada por :

$$p(y_{t+1}|\mathbf{y}^{(t)}) = \int \int p(y_{t+1}|\mathbf{y}^{(t)}, \lambda, \tau) p(\lambda, \tau|\mathbf{y}^{(t)}) d\lambda d\tau$$

Além disso, dado  $\lambda$  e  $\tau$ ,  $y_{t+1}$  é independente de todos  $y$ 's anteriores a  $y_t$ , e

$$p(y_{t+1}|\mathbf{y}^{(t)}) = \int \int p(y_{t+1}|y_t, \lambda, \tau) p(\lambda, \tau|\mathbf{y}^{(t)}) d\lambda d\tau$$

então, usando a (3.15), podemos escrever:

$$y_{t+1} = \lambda y_t + \omega_{t+1} y_t + \epsilon_{t+1}$$

como  $\lambda$  é independente de  $\omega_{t+1}$  e de  $\epsilon_{t+1}$

Então,

$$E(y_{t+1}|\mathbf{y}^{(t)}) = E(\lambda y_t|\mathbf{y}^{(t)}) + E(\omega_{t+1}|\mathbf{y}^{(t)}) + E(\epsilon_{t+1}|\mathbf{y}^{(t)})$$

logo

$$E(y_{t+1}|\mathbf{y}^{(t)}) = y_t E(\lambda|\mathbf{y}^{(t)}) \quad (3.35)$$

e

$$V(y_{t+1}|\mathbf{y}^{(t)}) = V(\lambda y_t|\mathbf{y}^{(t)}) + V(\omega_{t+1}|\mathbf{y}^{(t)}) + V(\epsilon_{t+1}|\mathbf{y}^{(t)})$$

ou seja

$$V(y_{t+1}|y^{(t)}) = V(\lambda|y^{(t)})y_t^2 + \gamma^2 y_t^2 + \tau^2 \quad (3.36)$$

Onde  $E(\lambda|y^{(t)})$  e  $V(\lambda|y^{(t)})$  são estimados por simulação de Monte Carlo em Cadeia de Markov, pelas eq (3.28) e (3.29).

Usando esses estimadores, temos:

$$\begin{aligned} E(y_{t+1}|y^{(t)}) &= y_t \hat{\lambda} \\ V(y_{t+1}|y^{(t)}) &= S_{\lambda}^2 y_t^2 + \gamma^2 y_t^2 + \tau^2 \end{aligned} \quad (3.37)$$

### 3.2.3 Densidade Preditiva para $\theta_{t+k}$ e $y_{t+k}$

Quando fazemos predição para  $k$ -passos, para  $k \geq 2$ , a estrutura linear descrita anteriormente desaparece e temos que considerar a distribuição preditiva de  $y_{t+k}$  e, para isso, procedemos, de modo análogo, ao caso de variância conhecida, ou seja, considerando:

$$y_{t+2} = \theta_{t+2} y_{t+1} + \epsilon_{t+2} \quad (3.38)$$

Onde

$$(E\theta_{t+k}|y^{(t)}) = E(\lambda|y^{(t)}) \quad k \geq 1 \quad (3.39)$$

Pela (3.32) temos que a densidade preditiva de  $p(y_{t+2}|y^{(t)})$  não pode ser uma densidade normal e que podemos obter a média preditiva de  $y_{t+k}$  dado  $y^{(t)}$ , denotada por  $E(y_{t+k}|y^{(t)})$ , para  $k \geq 2$ . Então, na (3.33), considerando dado  $y^{(t)}$ ,  $\tau$  e  $\lambda$ , podemos escrever:

$$E(y_{t+2}|y^{(t)}, \theta_t, \lambda, \tau) = E(\theta_{t+2} y_{t+1} | y^{(t)}, \lambda, \tau),$$

Além disso, dado  $\lambda$ ,  $\tau$  e  $y_t$ ,  $y_{t+1}$  é independente de  $\theta_{t+2}$ ; logo,

$$E(\theta_{t+2} y_{t+1} | y^{(t)}, \lambda, \tau) = E(\theta_{t+2} | y^{(t)}, \lambda, \tau) E(y_{t+1} | y^{(t)}, \lambda, \tau)$$

então,

$$E(y_{t+2} | y^{(t)}, \lambda, \tau) = \lambda E(\theta_t y_t | y^{(t)}, \lambda, \tau)$$

$$E(y_{t+2} | y^{(t)}, \lambda, \tau) = \lambda y_t E(\theta_t | y^{(t)}, \lambda, \tau)$$

$$E(y_{t+2}|y^{(t)}, \lambda, \tau) = \lambda^2 y_t$$

$$E(y_{t+2}|y^{(t)}) = E[E(y_{t+2}|y^{(t)}, \lambda, \tau)]$$

$$E(y_{t+2}|y^{(t)}) = E(\lambda^2 y_t)$$

Portanto, podemos concluir que:

$$E(y_{t+2}|y^{(t)}) = E(\lambda^2) y_t \quad (3.40)$$

Considerando a amostra gerada por MCMC  $\lambda^{(j)}$ , temos:

$$E(\lambda^2) = \frac{1}{M} \sum_{j=1}^M \lambda_j^2$$

No caso geral, temos que a previsão para  $y_{t+k}$  pode ser calculada como:

$$E(y_{t+k}|y^{(t)}) = E(\lambda^k) y_t \quad (3.41)$$

onde

$$E(\lambda^k) = \frac{1}{M} \sum_{j=1}^M \lambda_j^k$$

### 3.3 Resultados

Resumindo os resultados desse modelo, temos:

Modelo:

$$y_t = \theta_t y_{t-1} + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \tau^2) \quad ; \quad \tau^2 \text{ desconhecido}$$

$$\theta_t = \lambda + \omega_t \quad ; \quad \omega_t \sim N(\lambda, \gamma^2) \quad ; \quad \gamma^2 \text{ conhecido}$$

$$\lambda \sim N(m, S^2) \quad ; \quad m \text{ e } S^2 \text{ conhecidos}$$

A distribuição conjunta *a posteriori* para  $\lambda$  e  $\tau$

$$p(\lambda, \tau | y^{(t)}) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2S_t(\tau)} \left(\lambda - m_t(\tau)\right)^2\right\} \times$$

$$\exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Distribuição marginal para  $\tau$

$$p(\lambda, \tau | \mathbf{y}^{(t)}) \propto \frac{\left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)}\right)^{-1} \tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Distribuição condicional para  $\lambda$

$$p(\lambda | \mathbf{y}^{(t)}, \tau) \propto \exp\left\{-\frac{1}{2S_t(\tau)} (\lambda - m_t(\tau))^2\right\}$$

onde

$$S_t(\tau) = \left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)}\right)^{-1} \quad ; \quad m_t(\tau) = \left(\frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{r_i(\tau)}\right) S_t(\tau)$$

$$r_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

Distribuição condicional para  $\tau$

$$p(\tau | \mathbf{y}^{(t)}, \lambda) \propto \frac{\tau^{-(\alpha+1)}}{(r_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2S_t(\tau)} (\lambda - m_t(\tau))^2\right\} \times$$

$$\exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{r_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Distribuição condicional para  $\theta_t$

$$p(\theta_t | y_t, y_{t-1}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2\hat{\Sigma}_t} (\theta_t - \hat{\theta}_t)^2\right\}$$

onde

$$E(\theta_t | \mathbf{y}^{(t)}) = E\left[\frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2}\right]$$

$$V(\theta_t | \mathbf{y}^{(t)}) = E\left[\frac{\tau^2 \gamma^2}{\gamma^2 y_{t-1}^2 + \tau^2}\right] + V\left[\frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2}\right]$$

e

$$V \left[ \frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2} \right] = E \left[ \frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2} \right]^2 - \left[ E(\theta_t | y^{(t)}) \right]^2$$

para a amostra gerada  $(\tau^{(j)}, \lambda^{(j)}, j = 1, 2, \dots, M)$ .

Os valores esperados da preditiva para  $\theta_{t+k}$  e  $y_{t+k}$ ,  $k \geq 1$  são dados por:

$$E(\theta_{t+k} | y^{(t)}) = E(\lambda | y^{(t)})$$

$$V(\theta_{t+k} | y^{(t)}) = V(\lambda | y^{(t)}) + \gamma^2$$

$$E(y_{t+k} | y^{(t)}) = E(\lambda^k) y_t$$

onde

$$\hat{\lambda} = \frac{1}{M} \sum_{j=1}^M \lambda^{(j)} \quad ; \quad \widehat{S}_{\lambda}^2 = \frac{1}{M-1} \sum_{j=1}^M \left[ (\lambda^{(j)})^2 - \hat{\lambda} \right]^2$$

$$\hat{\tau} = \frac{1}{M} \sum_{j=1}^M \tau^{(j)} \quad ; \quad \widehat{S}_{\tau}^2 = \frac{1}{M-1} \sum_{j=1}^M \left[ (\tau^{(j)})^2 - \hat{\tau} \right]^2$$

onde  $M$  é o tamanho da amostra gerada para  $\lambda^{(j)}$  e  $\tau^{(j)}$ .

Esses resultados serão aplicados no capítulo 5, no estudo de caso de séries financeiras, onde analisamos o índice de preços das Ações da Vale do Rio Doce, negociadas pela Bovespa em dólares, no período de 02 de janeiro de 1996 a 01 de fevereiro de 1999, e o preço da arroba do boi gordo, negociado pela Tortuga Cia Zootécnica Agrária, em dólares, de janeiro de 1998 a agosto de 1999.



## Capítulo 4

# Modelo Dinâmico para um Processo-Auto Regressivo de Primeira Ordem

Neste capítulo, apresentamos uma ramificação do processo auto-regressivo de primeira ordem com coeficiente aleatório. Essa ramificação aplica-se quando assumimos uma estrutura de dependência dos  $\theta_t$ 's tão forte quanto sua permutabilidade.

Descrevemos o modelo para uma série temporal em função de seus componentes não observáveis  $\theta_t$ 's quando  $\theta_t$  é expresso em função dos valores desses mesmos componentes no instante anterior.

Concentramo-nos no procedimento pelo qual obtemos estimadores atualizados dos componentes não observáveis a todo instante de tempo a partir da informação dada pelo componente observável do sistema, ou seja,  $y_t$ . Dada a natureza linear dinâmica do modelo, esta atualização seqüencial pode ser obtida representando o modelo sob a forma de espaço de estados e, então, utilizar o Filtro de Kalman, cujo procedimento é recursivo.

O filtro de Kalman proporciona a elaboração de um algoritmo de estimação recursivo, que tem sido utilizado com sucesso por engenheiros, na teoria de controle, e, mais recentemente por estatísticos e econometristas.

Descrevemos a utilização do filtro de Kalman em problemas de previsão, onde a série temporal é modelada por uma média, que varia no tempo, superposta a um ruído aditivo. Essa média é por hipótese, uma combinação linear de funções conhecidas cujos coeficientes são desconhecidos. Desse modo, a série pode ser representada por um sis-

tema linear cujo vetor de estados é formado pelos coeficientes desconhecidos e pelo valor da média do processo, no instante  $t$ . Nessas circunstâncias, o filtro de Kalman pode ser utilizado para obter estimativas ótimas do vetor de estados, com a vantagem de permitir a variação dos coeficientes através do tempo.

## 4.1 O Modelo Filtro de Kalman e Suas Soluções

Seja  $y^{(t)} = (y_t, y_{t-1}, \dots, y_1)$ , denotando os valores observados de uma variável de interesse para diferentes tempos, assumindo que  $y_t$  depende de uma quantidade não observável  $\theta_t$  ( $y_t$  e  $\theta_t$  podem ser escalares ou vetores). Aqui vamos considerar  $\theta_t$  como um vetor de parâmetros e nosso objetivo é fazer inferência sobre  $\theta_t$ .

A relação entre  $y_t$  e  $\theta_t$  é linear e especificada pela equação das observações e a característica dinâmica da variável não-observável  $\theta_t$  é dada pela equação de estado.

$$y_t = F_t \theta_t + v_t \quad (4.1)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (4.2)$$

onde

$F_t$ : uma quantidade conhecida, denominada na literatura como matriz de medida;

$v_t$ : o erro observado assumido, normalmente distribuído com média zero e variância  $V_t$  conhecida ;

$G_t$ : quantidade conhecida, denominada matriz de transição;

$w_t$ : erro do sistema com distribuição normal de média zero e variância  $W_t$  conhecida.

Por hipótese,  $v_t$  e  $w_t$  são considerados independentes entre si e entre estágios subsequentes.

O objetivo aqui é analisar o comportamento do sistema descrito pelas equações (4.1) e (4.2) para avaliar a mudança de comportamento do vetor de parâmetros  $\theta_t$  com o tempo.

Para atingir esse objetivo, vamos empregar uma técnica conhecida como “Filtro de Kalman” (Meinhold e Singpurwala, 1983) que permite atualizar passo-a-passo o valor das médias e a matriz de covariância de  $\theta_t$ , à medida que é feita uma nova observação  $y_t$ . Esse procedimento de inferência de  $\hat{\theta}_t$  (média de  $\theta_t$  condicionada às observações  $y^{(t)}$ ) e da matriz de covariância  $\Sigma_t$  nos permite fazer previsões para a observação futura  $y_{t+1}$ .



Como mostra Meinhold e Singpurwala(1983), essa inferência pode ser obtida aplicando diretamente o teorema de Bayes:

$$p(\theta_t|y^{(t)}) \propto p(y_t|\theta_t, y^{(t-1)})p(\theta_t|y^{(t-1)}) \quad (4.3)$$

onde  $p(y_t|\theta_t, y^{(t-1)})$  denota a função de verossimilhança da observação  $y_t$  e  $p(\theta_t|y^{(t-1)})$  representa a distribuição *a priori* para  $\theta_t$

O procedimento recursivo pode ser melhor explicado se considerarmos o tempo  $(t-1)$  e observarmos os dados até esse tempo:  $y^{(t-1)} = (y_{t-1}, \dots, y_1)$ .

No tempo  $(t-1)$ , nosso conhecimento sobre  $\theta_{t-1}$  é dado pela distribuição de probabilidade de  $\theta_{t-1}$ , ou seja :

$$(\theta_{t-1}|y^{(t-1)}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1}) \quad (4.4)$$

que representa a distribuição *a posteriori* para  $\theta_{t-1}$ .

Esse resultado é importante para observarmos que o procedimento recursivo é iniciado no tempo zero com  $\theta_0$  e  $\Sigma_0$ , nossa melhor suposição sobre a média e a variância de  $\theta_1$ , respectivamente.

Consideramos agora o desenvolvimento para o tempo  $t$ , porém em dois estágios:

Antes da observação  $y_t$  e após a observação  $y_t$ .

1- Antes da observação  $y_t$ , nossa melhor escolha para  $\theta_t$  é governada pela eq(4.2).

Logo,  $(\theta_t|y^{(t-1)})$  é uma combinação linear de duas normais independentes e, portanto, também tem distribuição normal com média e variância dadas por:

$$E(\theta_t|y^{(t-1)}) = E(G_t\theta_{t-1}|y^{(t-1)}) + E(w_t|y^{(t-1)})$$

sendo  $w_t$  e  $y^{(t-1)}$  independentes, então:

$$E(\theta_t|y^{(t-1)}) = G_t E(\theta_{t-1}|y^{(t-1)})$$

denotando por  $\hat{\theta}_t = E(\theta_t|y^{(t-1)})$ , temos:

$$\hat{\theta}_t = G_t \hat{\theta}_{t-1}$$

$$V(\theta_t|y^{(t-1)}) = V(G_t\theta_{t-1}|y^{(t-1)}) + V(w_t|y^{(t-1)})$$

lembrando que  $w_t \sim N(0, W_t)$

$$V(\theta_t|y^{(t-1)}) = G_t' V(\theta_{t-1}|y^{(t-1)}) G_t + W_t$$

chamando

$$V(\theta_t | y^{(t-1)}) = \hat{\Sigma}_t$$

$$\Sigma_t = G_t' \Sigma_{t-1} G_t + W_t$$

temos que:

$$(\theta_t | y^{(t-1)}) \sim N(G_t \hat{\theta}_{t-1}, \hat{\Sigma}_t) \quad (4.5)$$

2- Após a observação  $y_t$ , nosso objetivo é calcular a distribuição *a posteriori* de  $\theta_t$ , usando a relação (4.3). Para isso, precisamos conhecer a função de verossimilhança  $L(\theta_t | y^{(t)})$ , ou equivalentemente  $p(y_t | \theta_t, y^{(t-1)})$ , cuja determinação é feita pelo seguinte desenvolvimento:

Seja  $e_t$  o erro de predição de  $y_t$  dado  $y^{(t-1)}$ , onde:

$$\hat{y}_t = E(F_t G_t \theta_{t-1} | y^{(t-1)}) + E(v_t | y^{(t-1)})$$

sendo  $v_t$  e  $y^{(t-1)}$  independentes, então:

$$\hat{y}_t = F_t G_t E(\theta_{t-1} | y^{(t-1)}) = F_t G_t \hat{\theta}_{t-1}$$

podemos, então, escrever:

$$e_t = y_t - \hat{y}_t = y_t - F_t G_t \hat{\theta}_{t-1} \quad (4.6)$$

Desde que  $F_t$ ,  $G_t$  e  $\hat{\theta}_{t-1}$  sejam todos conhecidos, observar  $y_t$  é equivalente a observar  $e_t$ , assim a (4.3) pode ser reescrita como:

$$p(\theta_t | y_t, y^{(t-1)}) = p(\theta_t | e_t, y^{(t-1)}) \propto p(e_t | \theta_t, y^{(t-1)}) p(\theta_t | y^{(t-1)}) \quad (4.7)$$

onde  $p(e_t | \theta_t, y^{(t-1)})$  é a função de verossimilhança.

Usando o fato de que  $y_t = F_t \theta_t + v_t$ , a eq(4.6) torna-se:

$$e_t = F_t \theta_t + v_t - F_t G_t \hat{\theta}_{t-1}$$

$$e_t = F_t (\theta_t - G_t \hat{\theta}_{t-1}) + v_t$$

assim, temos a média de  $e_t$  dada por:

$$E(e_t | \theta_t, y^{(t-1)}) = E[F_t (\theta_t - G_t \hat{\theta}_{t-1}) + v_t | \theta_t, y^{(t-1)}]$$

$$E(e_t|\theta_t, y^{(t-1)}) = E[F_t(\theta_t - G_t\hat{\theta}_{t-1})|\theta_t, y^{(t-1)}] + E(v_t|\theta_t, y^{(t-1)})$$

$$E(e_t|\theta_t, y^{(t-1)}) = F_t(\theta_t - G_t\hat{\theta}_{t-1})$$

A variância é dada por:

$$V(e_t|\theta_t, y^{(t-1)}) = E[e_t - E(e_t|\theta_t, y^{(t-1)})]^2 = E(v_t^2|\theta_t, y^{(t-1)})$$

Lembrando da eq (4.1) que  $v_t \sim N(0, V_t)$  temos:

$$V(v_t|\theta_t, y^{(t-1)}) = V_t$$

Então, a função de verossimilhança é descrita por:

$$(e_t|\theta_t, y^{(t-1)}) \sim N[F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t] \quad (4.8)$$

Aplicando o teorema de Bayes, obtemos:

$$p(\theta_t|y_t, y^{(t-1)}) = \frac{p(e_t|\theta_t, y^{(t-1)})p(\theta_t|y^{(t-1)})}{\int \int p(e_t, \theta_t|y^{(t-1)})d\theta_t} \quad (4.9)$$

Essa equação descreve melhor nosso conhecimento sobre  $\theta_t$ , no tempo  $t$ . Sendo que após o cálculo de  $p(\theta_t|y_t, y^{(t-1)})$  podemos voltar em eq(4.6) para o próximo ciclo do procedimento recursivo.

O esforço requerido para obter  $p(\theta_t|y^{(t)})$  através da eq(4.9) pode ser evitado usando o resultado de estatística multivariada, muito conhecido, dado por Johnson e Wichern(1992, pp.133-139) e algumas propriedades padrão da distribuição normal.

Sejam  $X_1$  e  $X_2$  com distribuição normal bivariada de médias  $\mu_1$  e  $\mu_2$ , respectivamente, e matriz de covariância-variância  $\Sigma$  denotada por :

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right] \quad (4.10)$$

Aplicando a eq(4.10), a distribuição condicional de  $X_1$  dado  $X_2$  é descrita por:

$$\left( X_1 | X_2 = x_2 \right) \sim N \left( \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{11} \right) \quad (4.11)$$

onde a quantidade  $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$  é denominada função de regressão e  $\Sigma_{12}\Sigma_{22}^{-1}$  é o coeficiente de mínimos quadrados da regressão de  $X_1$  em  $x_2$ . Assim, temos como resultado:

$$X_2 \sim N(\mu_2, \Sigma_{22})$$

Fazendo a correspondência entre  $X_1$  e  $X_2$  e  $e_t$  e  $\theta_t$  e, respectivamente, denotando esta correspondência por:

$$X_1 = e_t \quad ; \quad X_2 = \theta_t$$

Desde que  $(\theta_t | y^{(t-1)}) \sim N(G_t \hat{\theta}_{t-1}, \Sigma_t)$ , vindo a eq(4.5), temos que:

$$\mu_2 = G_t \hat{\theta}_{t-1} \quad ; \quad \Sigma_{22} = \Sigma_t$$

Na eq(4.5), substituindo  $X_1$ ,  $X_2$ ,  $\mu_2$  e  $\Sigma_{22}$  por  $e_t$ ,  $\theta_t$ ,  $G_t \hat{\theta}_{t-1}$  e  $\Sigma_t$ , respectivamente, e lembrando o resultado da eq(4.8):

$$(e_t | \theta_t, y^{(t-1)}) \sim N[F_t(\theta_t - G_t \hat{\theta}_{t-1}), V_t]$$

temos :

$$\mu_1 + \Sigma_{12}S_t^{-1}(\theta_t - G_t \hat{\theta}_{t-1}) = F_t(\theta_t - G_t \hat{\theta}_{t-1})$$

desde que,

$$\mu_1 = 0 \quad e \quad \Sigma_{12} = F_t \Sigma_t$$

similarmente temos:

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - F_t \Sigma_t F_t' = V_t$$

Desde que  $\Sigma_{11} = V_t + F_t S_t F_t'$ . Podemos concluir que a distribuição conjunta de  $\theta_t$  e  $e_t$  dado  $y^{(t-1)}$  pode ser escrita como:

$$\left( \theta_t \quad e_t \mid y^{(t-1)} \right)' \sim N \left[ \left( G_t \hat{\theta}_{t-1} \quad 0 \right)' ; \begin{pmatrix} \Sigma_t & \Sigma_t F_t' \\ F_t \Sigma_t & V_t + F_t \Sigma_t F_t' \end{pmatrix} \right] \quad (4.12)$$

Fazendo  $e_t$  a variável condicional e identificando a eq(4.12) com eq(4.10), obtemos, através da eq(4.11), o resultado:

$$(\theta_t | e_t, y^{(t-1)}) \sim N \left( G_t \hat{\theta}_{t-1} + \Sigma_t F_t' (V_t + F_t \Sigma_t F_t')^{-1} e_t ; \Sigma_t - \Sigma_t F_t' (V_t + F_t \Sigma_t F_t')^{-1} F_t \Sigma_t \right) \quad (4.13)$$

Esta é a distribuição *a posteriori* desejada.

Agora, resumamos os resultados obtidos para salientar os elementos do procedimento recursivo:

Após o tempo  $t - 1$ , temos a distribuição *a posteriori* para  $\theta_{t-1}$ :

$$(\theta_{t-1} | y^{(t-1)}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1})$$

a previsão  $E(y_t | y^{(t-1)})$ ;

$$\hat{y}_t = F_t G_t \hat{\theta}_{t-1}$$

o erro de previsão;

$$e_t = y_t - F_t G_t \hat{\theta}_{t-1}$$

a média e a variância *a posteriori* de  $\theta_t$ ;

$$\hat{\theta}_t = G_t \hat{\theta}_{t-1} + \Sigma_t F_t' (V_t + F_t \Sigma_t F_t')^{-1} e_t \quad (4.14)$$

$$\hat{\Sigma}_t = \Sigma_t - \Sigma_t F_t' (V_t + F_t \Sigma_t F_t')^{-1} F_t \Sigma_t \quad (4.15)$$

distribuição *a posteriori* para  $\theta_t$ :

## 4.2 Adaptação do Filtro de Kalman

No modelo do filtro de Kalman ordinário, as quantidades ( $F_t$ ,  $G_t$ ,  $V_t$  e  $W_t$ ) são todas consideradas conhecidas. Elas podem variar no tempo, mas somente se a natureza da variação também for considerada conhecida. Em muitas aplicações é difícil especificar algum ou todos esses parâmetros, devido à informação inadequada sobre o processo ou porque os parâmetros são permutáveis, mas a natureza exata dessa permuta é desconhecida. Nesses casos, o filtro de Kalman é aplicado para estimar os parâmetros desconhecidos em cada estágio. Na literatura, esse procedimento é conhecido como Filtro de Kalman adaptado, cujos estimadores dos parâmetros podem ser obtidos na forma do estimador bayesiano.

### 4.2.1 Inferência para um Modelo Filtro de Kalman Adaptado

A estrutura de dependência imposta para a seqüência  $\theta_t$ , no modelo filtro de Kalman, é especificada via equação de estado. O modelo descreve agora os  $\theta_t$ 's permutáveis de um estágio para outro. Além disso, na prática, a natureza exata da permuta é desconhecida e, conseqüentemente,  $G_t$  é desconhecido. Desde que não é possível especificar um modelo definitivo para  $\theta_t$ , é razoável assumir  $G_t$  desconhecido, mas invariante no tempo. Se considerarmos todas as quantidades como escalares e fazendo  $G_t = \lambda$ ,  $F_t = y_{t-1}$ ,  $V_t = \epsilon_t$  e  $W_t = \omega_t$ , podemos escrever o modelo como:

$$\begin{aligned} y_t &= \theta_t y_{t-1} + \epsilon_t \\ \theta_t &= \lambda \theta_{t-1} + \omega_t \end{aligned} \quad (4.16)$$

onde  $\begin{pmatrix} \epsilon_t \\ \omega_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} \right]$ ,  $\sigma^2$  e  $\gamma^2$  conhecidos

Por hipótese,  $\epsilon_t$  e  $\omega_t$  são independentes entre si e entre estágios e  $\lambda$  desconhecido é independente de  $\omega_t$ .

Desde que  $\lambda$  é desconhecido, assumimos uma distribuição *a priori* para o mesmo. Ainda que (4.16) seja uma generalização do Filtro de Kalman, sua análise produz dificuldades técnicas, no sentido de que a solução recursiva natural do mesmo é perdida.

Dado os dados  $y^{(t)} = (y_t, y_{t-1}, \dots, y_1)$ , nosso objetivo é fazer inferência sobre  $\theta_t$  e a previsão futura de  $y_{t+1}$ . Então, precisamos obter a densidade *a posteriori*  $p(\theta_t | y^{(t)})$  e a densidade preditiva para  $p(y_{t+1} | y^{(t)})$ .

Podemos escrever a distribuição *a posteriori* de  $\theta_t$  como:

$$p(\theta_t | y^{(t)}) = \int p(\theta_t | y^{(t)}, \lambda) p(\lambda | y^{(t)}) d\lambda \quad (4.17)$$

onde a densidade  $p(\theta_t | y^{(t)}, \lambda)$  é dada pela solução do filtro de Kalman. E a distribuição *a posteriori* para  $\lambda$  dado  $y^{(t)}$  é:

$$p(\lambda | y^{(t)}) = \frac{p(\lambda) L(\lambda | y^{(t)})}{\int p(\lambda) L(\lambda | y^{(t)}) d\lambda} \quad (4.18)$$

onde  $p(\lambda)$  é a distribuição *a priori* para  $\lambda$  e  $L(\lambda | y^{(t)})$  é a função de verossimilhança de  $\lambda$  dado  $y^{(t)}$ .

## 4.2.2 Cálculo da Função de Verossimilhança

A função de verossimilhança  $L(\lambda|\mathbf{y}^{(i)})$ , em (4.18), é obtida pelo seguinte procedimento:

No modelo Filtro de Kalman, a densidade preditiva de  $p(\mathbf{y}_i|\mathbf{y}^{(i-1)})$  é dado por:

$$p(\mathbf{y}_i|\mathbf{y}^{(i-1)}, \lambda) = \int p(\mathbf{y}_i|\theta_i, \mathbf{y}^{(i-1)}, \lambda)p(\theta_i|\mathbf{y}^{(i-1)}, \lambda)d\theta_i \quad (4.19)$$

Da (4.16) temos que  $(\theta_i|\mathbf{y}^{(i-1)}, \lambda)$  é uma combinação linear de duas normais independentes e, portanto, também tem distribuição normal de média e variância dada por:

$$E(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = E(\lambda\theta_{i-1}|\mathbf{y}^{(i-1)}) + E(\omega_i\mathbf{y}_i^{(i-1)})$$

$$E(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = \lambda E(\theta_{i-1}|\mathbf{y}^{(i-1)})$$

Da eq(4.4) temos que  $E(\theta_{i-1}|\mathbf{y}^{(i-1)}) = \hat{\theta}_{i-1}$ . Assim:

$$E(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = \lambda\hat{\theta}_{i-1}$$

e

$$V(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = V(\lambda\theta_{i-1}|\mathbf{y}^{(i-1)}, \lambda) + V(\omega_i\mathbf{y}_i^{(i-1)}, \lambda)$$

$$V(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = \lambda^2 V(\theta_{i-1}|\mathbf{y}^{(i-1)}) + \gamma^2$$

também, da eq(4.4), temos  $V(\theta_{i-1}|\mathbf{y}^{(i-1)}) = \Sigma_{i-1}$ , logo,

$$V(\theta_i|\mathbf{y}^{(i-1)}, \lambda) = \lambda^2 \Sigma_{i-1} + \gamma^2$$

portanto,

$$(\theta_i|\mathbf{y}^{(i-1)}, \lambda) \sim N(\lambda\hat{\theta}_{i-1}, \lambda^2 \Sigma_{i-1} + \gamma^2) \quad (4.20)$$

denotando  $r_i = \lambda^2 \Sigma_{i-1} + \gamma^2$  temos que:

$$p(\theta_i|\mathbf{y}^{(i-1)}, \lambda) \propto r_i^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2r_i} \left( \theta_i - \lambda\hat{\theta}_{i-1} \right)^2 \right\}$$

A (4.20) é a distribuição *a priori* para  $\theta_i$ . E a distribuição *a posteriori* para  $\theta_{i-1}$ , no tempo  $(i-1)$ , após a observação  $\mathbf{y}_{i-1}$ , é dada pela eq(4.4), isto é:

$$(\theta_{i-1}|\mathbf{y}^{(i-1)}, \lambda) \sim N(\hat{\theta}_{i-1}, \Sigma_{i-1}) \quad (4.21)$$

Da eq(4.16) podemos tirar também que:

$$y_i = \theta_i y_{i-1} + \epsilon_i$$

$$E(y_i | \theta_i, y^{(i-1)}, \lambda) = E(\theta_i y_{i-1} + \epsilon_i | \theta_i, y^{(i-1)}, \lambda)$$

$$E(y_i | \theta_i, y^{(i-1)}, \lambda) = \theta_i y_{i-1}$$

e

$$V(y_i | \theta_i, y^{(i-1)}, \lambda) = V(\theta_i y_{i-1} + \epsilon_i | \theta_i, y^{(i-1)}, \lambda)$$

$$V(y_i | \theta_i, y^{(i-1)}, \lambda) = V(\theta_i y_{i-1} | \theta_i, y^{(i-1)}, \lambda) + V(\epsilon_i)$$

$$V(y_i | \theta_i, y^{(i-1)}, \lambda) = \sigma^2$$

Logo,

$$(y_i | \theta_i, y^{(i-1)}, \lambda) \sim N(\theta_i y_{i-1}, \sigma^2) \quad (4.22)$$

isto é,

$$p(y_i | \theta_i, y^{(i-1)}, \lambda) \propto \sigma^{-1} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - \theta_i y_{i-1})^2 \right\}$$

Temos também que  $p(y_i | y^{(i-1)}, \lambda)$  pode ser expressa calculando-se:

$$E(y_i | y^{(i-1)}, \lambda) = E(\lambda \theta_{i-1} y_{i-1} + \omega_i y_{i-1} + \epsilon_i | y^{(i-1)}, \lambda)$$

$$E(y_i | y^{(i-1)}, \lambda) = E(\lambda \theta_{i-1} y_{i-1} | y^{(i-1)}, \lambda) + E(\omega_i y_{i-1} + \epsilon_i | y^{(i-1)}, \lambda)$$

$$E(y_i | y^{(i-1)}, \lambda) = \lambda E(\theta_{i-1} | y^{(i-1)}) y_{i-1}$$

$$E(y_i | y^{(i-1)}, \lambda) = \lambda \hat{\theta}_{i-1} y_{i-1}$$

$$V(y_i | y^{(i-1)}, \lambda) = V(\lambda \theta_{i-1} y_{i-1} | y^{(i-1)}, \lambda) + V(\omega_i y_{i-1} + \epsilon_i | y^{(i-1)}, \lambda)$$



$$V(y_i|y^{(i-1)}, \lambda) = \lambda^2 V(\theta_{i-1}|y^{(i-1)})y_{i-1}^2 + V(\omega_i)y_{i-1}^2 + \sigma^2$$

$$V(y_i|y^{(i-1)}, \lambda) = \lambda^2 \Sigma_{i-1} y_{i-1}^2 + \gamma^2 y_{i-1}^2 + \sigma^2$$

$$V(y_i|y^{(i-1)}, \lambda) = y_{i-1}^2 (\lambda^2 \Sigma_{i-1} + \gamma^2) + \sigma^2$$

$$V(y_i|y^{(i-1)}, \lambda) = y_{i-1}^2 r_i + \sigma^2$$

Assim,

$$(y_i|y^{(i-1)}, \lambda) \sim N[(\lambda \hat{\theta}_{i-1} y_{i-1}), y_{i-1}^2 r_i + \sigma^2] \quad (4.23)$$

Ou seja,

$$p(y_i|y^{(i-1)}, \lambda) \propto (y_{i-1}^2 r_i + \sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(y_{i-1}^2 r_i + \sigma^2)} \left( y_i - \lambda \hat{\theta}_{i-1} y_{i-1} \right)^2 \right\} \quad (4.24)$$

Podemos, então, escrever a função de verossimilhança de  $\lambda$  como:

$$L(\lambda|y^{(i-1)}) = \prod_{i=1}^t p(y_i|y^{(i-1)}, \lambda)$$

### 4.2.3 Cálculo de a Posteriori de $\theta_t$ dado $y^{(t)}$ e $\lambda$

Vamos determinar agora a densidade  $p(\theta_t|y^{(t)}, \lambda)$  que é o segundo termo na eq(4.17).

Dado  $\lambda$ , a densidade  $p(\theta_t|y^{(t)}, \lambda)$  é obtida pela solução do filtro de Kalman, eq(4.13).

Relacionando os termos nas equações (4.12) e (4.13) com as variáveis do problema:

$$G_t \hat{\theta}_{t-1} = \lambda \hat{\theta}_{t-1} \quad F_t' \Sigma_t F_t = r_t y_{t-1}^2 \quad V_t + F_t \Sigma_t F_t' = \sigma^2 + r_t y_{t-1}^2$$

Podemos concluir que a distribuição conjunta de  $\theta_t$  e  $e_t$  pode ser escrita como:

$$(\theta_t|e_t, y^{(t-1)}, \lambda) \sim N \left( \lambda \hat{\theta}_{t-1} + r_t y_{t-1} (\sigma^2 + y_{t-1}^2 r_t)^{-1} e_t; r_t - y_{t-1}^2 r_t^2 (\sigma^2 + y_{t-1}^2 r_t)^{-1} \right) \quad (4.25)$$

Denominando:

$$\hat{\theta}_t = \lambda \hat{\theta}_{t-1} + \frac{r_t y_{t-1}}{\sigma^2 + y_{t-1}^2 r_t} \left( y_t - y_{t-1} \lambda \hat{\theta}_{t-1} \right)$$

ou ainda,

$$\hat{\theta}_t = \frac{\lambda \hat{\theta}_{t-1} \sigma^2 + r_t y_t y_{t-1}}{\sigma^2 + y_{t-1}^2 r_t} \quad (4.26)$$

onde

$$r_t = \lambda^2 \Sigma_{t-1} + \gamma^2$$

e

$$\hat{\Sigma}_t = r_t - r_t^2 y_{t-1}^2 (\sigma^2 + y_{t-1}^2 r_t)^{-1}$$

$$\hat{\Sigma}_t = \frac{r_t \sigma^2}{\sigma^2 + y_{t-1}^2 r_t} \quad (4.27)$$

Então, a distribuição *a posteriori* procurada é:

$$p(\theta_t | y^{(t)}, \lambda) \propto \Sigma_t^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2 \hat{\Sigma}_t} \left( \theta_t - \hat{\theta}_t \right)^2 \right\} \quad (4.28)$$

Onde  $\hat{\theta}_t$  e  $r_t$  são funções de  $\lambda$ .

Se analisarmos a eq(4.28), observamos que  $\hat{\theta}_t$  é o valor de peso na previsão do erro.

A quantidade  $\hat{\Sigma}_t$  é denominada “matriz de ganho do Filtro de Kalman” e é o peso que multiplica o erro de previsão. Observamos também que a variância *a posteriori* de  $\theta_t$ ,  $\hat{\Sigma}_t$  não é influenciada pelo dado observado  $y_t$ .

#### 4.2.4 O Comportamento de $\theta_t$

Agora queremos determinar o comportamento de  $\theta_t$ , para isso, temos da eq(4.17) que:

$$p(\theta_t, \lambda | y^{(t)}) = p(\theta_t | y^{(t)}, \lambda) p(\lambda | y^{(t)})$$

onde

$$p(\lambda | y^{(t)}) \propto p(\lambda) L(\lambda | y^{(t)})$$

e  $p(\lambda)$  é *a priori* para  $\lambda$ , com distribuição normal, com média  $m$  e variância  $S^2$ , ou seja:

$$p(\lambda) \sim N(m, S^2)$$

e

$$p(\lambda|y^{(t)}) \propto \prod_{i=1}^t (y_{i-1}^2 r_i + \sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(y_{i-1}^2 r_i + \sigma^2)} \left( y_i - \lambda \hat{\theta}_{i-1} y_{i-1} \right)^2 \right\} \times \quad (4.29)$$

$$S^{-1} \exp \left\{ -\frac{1}{2S^2} (\lambda - m)^2 \right\}$$

e  $p(\theta_i|y^{(t)}, \lambda)$  é dada por:

$$p(\theta_i|y^{(t)}, \lambda) \propto \hat{\Sigma}_i^{-1} \exp \left\{ -\frac{1}{2\hat{\Sigma}_i} (\theta_i - \hat{\theta}_i)^2 \right\}$$

sendo que da eq(4.26)

$$\hat{\theta}_i = \lambda \hat{\theta}_{i-1} + \frac{r_i y_{i-1}}{\sigma^2 + y_{i-1}^2 r_i} (y_i - y_{i-1} \lambda \hat{\theta}_{i-1})$$

e da eq(4.27)

$$\hat{\Sigma}_i = r_i - r_i^2 y_{i-1}^2 (\sigma^2 + y_{i-1}^2 r_i)^{-1}$$

Os valores de  $E(\theta_i|y^{(t)})$  e  $E(\lambda|y^{(t)})$  são calculados usando os estimadores MCMC com amostras geradas pelos algoritmos amostrador de Gibbs (Casella e George, 1992) e Metropolis-Hansting (Chib e Greenberg, 1995). Esse algoritmo é utilizado no contexto desse trabalho dentro do seguinte esquema:

- i- Inicia  $(y_0, \dots, y_t)$  ;  $\theta_0$  ;  $\lambda^{(0)}$  ;  $\hat{\theta}_0 = \theta_0$
- ii- Calcula;

$$\hat{\theta}_1 = \lambda^{(0)} \hat{\theta}_0 + \frac{r_1 y_0}{\sigma^2 + y_0^2 r_1} (y_1 - y_0 \lambda^{(0)} \hat{\theta}_0)$$

$$\hat{\theta}_2 = \lambda^{(0)} \hat{\theta}_1 + \frac{r_2 y_1}{\sigma^2 + y_1^2 r_2} (y_2 - y_1 \lambda^{(0)} \hat{\theta}_1)$$

⋮

$$\hat{\theta}_t = \lambda^{(0)} \hat{\theta}_{t-1} + \frac{r_t y_{t-1}}{\sigma^2 + y_{t-1}^2 r_t} (y_t - y_{t-1} \lambda^{(0)} \hat{\theta}_{t-1})$$

- iii- para calcular  $\hat{\theta}_t$  e  $\Sigma_t$ , gera-se:

$$\theta_t^{(0)} \sim N(\hat{\theta}_t, \hat{\Sigma}_t) ; \quad \theta_t^{(1)}$$

iv- com  $(y_0, \dots, y_t)$  ;  $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_{t-1}$  usando a eq(4.29), gera-se:

$$\lambda^{(1)} \sim p(\lambda)L(\lambda|y^{(t)})$$

com:

$$P_0(\lambda) \sim N(m_0, S_0^2)$$

$$\Psi(\lambda|y^{(t)}) = L(\lambda|y^{(t)})$$

$$\Psi(\lambda|y^{(t)}) = \prod_{i=1}^t (y_{i-1}\tau_i + \sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(y_{i-1}\tau_i + \sigma^2)} \left( y_i - \lambda \hat{\theta}_{i-1} y_{i-1} \right)^2 \right\}$$

v- gera o candidato:

$$\xi \sim N(m_0, S_0^2)$$

vi- aceita o valor calculado com probabilidade  $\alpha = \min\{1, p\}$

$$p = \left\{ \frac{\Psi(\xi|y^{(t)})}{\Psi(\lambda^{(j-1)}|y^{(t)})}, \Psi(\lambda^{(j-1)}|y^{(t)}) \neq 0 \right\}$$

gera-se  $u$  de Uniforme  $[0, 1]$  e faça:

$$\lambda^{(j+1)} = \begin{cases} \xi & \text{se } u \leq p, \\ \lambda^{(j)} & \text{caso contrário.} \end{cases}$$

vii- volta para (ii) e repita o procedimento até obter a convergência, onde  $j = 1, \dots, M$

$$\theta_t^{(j)}, j = 1, \dots, M$$

viii- calcula;

$$E(\lambda|y^{(t)}) = \frac{1}{M} \sum_{j=1}^M \lambda^{(j)}$$

$$E(\theta_t|y^{(t)}) = \frac{1}{M} \sum_{j=1}^M \theta_t^{(j)}$$

onde  $M$  é o tamanho da amostra gerada.

## 4.3 Densidade Preditiva para $\theta_{t+1}$ e $y_{t+1}$

### 4.3.1 Densidade Preditiva para $\theta_{t+1}$

A distribuição preditiva para  $\theta_{t+1}$  é dada por:

$$p(\theta_{t+1}|y^{(t)}) = \int p(\theta_{t+1}|y_t, \lambda)p(\lambda|y^{(t)})d\lambda$$

Considerando que dado  $\lambda$ ,  $\theta_{t+1}$  é independente de  $y_t$ , podemos escrever a equação anterior como:

$$p(\theta_{t+1}|y^{(t)}) = \int p(\theta_{t+1}|\lambda)p(\lambda|y^{(t)})d\lambda \quad (4.30)$$

Como o procedimento utilizado neste modelo para obter os estimadores é um procedimento recursivo da eq(4.16), temos que:

$$\theta_{t+1} = \lambda\theta_t + \omega_{t+1} \quad \text{onde} \quad \omega_{t+1} \sim N(0, \gamma^2)$$

mas, por hipótese, é independente de  $y^{(t)}$  e entre estágios, então:

$$(\theta_{t+1}|y^{(t)}, \lambda) = (\lambda\theta_t|y^{(t)}, \lambda) + (\omega_{t+1}|y^{(t)}, \lambda)$$

$$E(\theta_{t+1}|y^{(t)}, \lambda) = E(\lambda\theta_t|y^{(t)}, \lambda) + E(\omega_{t+1}|y^{(t)}, \lambda)$$

$$E(\theta_{t+1}|y^{(t)}, \lambda) = \lambda E(\theta_t|y^{(t)}, \lambda)$$

$$E(\theta_{t+1}|y^{(t)}, \lambda) = \lambda\hat{\theta}_t$$

denotando  $\hat{\theta}_{t+1} = \lambda\hat{\theta}_t$ , onde  $\hat{\theta}_t$  é dada pela eq(4.26), ou seja:

$$\hat{\theta}_t = \frac{\lambda\hat{\theta}_{t-1}\sigma^2 + r_t y_t y_{t-1}}{\sigma^2 + y_{t-1}^2 r_t} \quad \text{sendo} \quad r_t = \lambda^2 \Sigma_{t-1} + \gamma^2$$

A variância de  $\theta_{t+1}$  dado  $y^{(t)}$  e  $\lambda$  é descrita por:

$$V(\theta_{t+1}|y^{(t)}, \lambda) = V(\lambda\theta_t|y^{(t)}, \lambda) + V(\omega_{t+1}|y^{(t)}, \lambda)$$

$$V(\theta_{t+1}|y^{(t)}, \lambda) = \lambda^2 V(\theta_t|y^{(t)}, \lambda) + \gamma^2$$

$$V(\theta_{t+1}|\mathbf{y}^{(t)}, \lambda) = \lambda^2 \hat{\Sigma}_t + \gamma^2$$

fazendo,

$$V(\theta_{t+1}|\mathbf{y}^{(t)}, \lambda) = \Sigma_{t+1} = \lambda^2 \hat{\Sigma}_t + \gamma^2$$

onde da (4.27):

$$\hat{\Sigma}_t = \frac{r_t \sigma^2}{\sigma^2 + y_{t-1}^2 r_t}$$

### 4.3.2 Densidade Preditiva para $y_{t+1}$

Temos que a densidade preditiva de  $y_{t+1}$  dado  $\mathbf{y}^{(t)}$  é dada por:

$$p(y_{t+1}|\mathbf{y}^{(t)}) = \int p(y_{t+1}|\mathbf{y}^{(t)}, \lambda) p(\lambda|\mathbf{y}^{(t)}) d\lambda$$

Mas, dado  $\lambda$ ,  $y_{t+1}$  é independente de todos os  $y$ 's anteriores a  $y_t$ . Assim, temos que:

$$p(y_{t+1}|\mathbf{y}^{(t)}) = \int p(y_{t+1}|y_t, \lambda) p(\lambda|\mathbf{y}^{(t)}) d\lambda$$

Da eq(4.16), temos:

$$E(y_t|\mathbf{y}^{(t-1)}, \lambda) = E[(\lambda\theta_{t-1} + \omega_{t-1})y_{t-1} + \epsilon_t|\mathbf{y}^{(t-1)}, \lambda]$$

Podemos calcular o valor esperado e a variância de  $y_{t+1}$ , dado  $\mathbf{y}^{(t)}$ , usando o mesmo raciocínio, isto é:

$$E(y_{t+1}|\mathbf{y}^{(t)}, \lambda) = E[(\lambda\theta_t + \omega_t)y_t + \epsilon_{t+1}|\mathbf{y}^{(t)}, \lambda]$$

como, por hipótese,  $\omega_t$  e  $\epsilon_t$  são independentes entre si e entre estágios, então:

$$E(y_{t+1}|\mathbf{y}^{(t)}, \lambda) = E(\lambda\theta_t y_t|\mathbf{y}^{(t)}, \lambda) + E(\omega_t y_t|\mathbf{y}^{(t)}, \lambda) + E(\epsilon_{t+1}|\mathbf{y}^{(t)}, \lambda)$$

$$E(y_{t+1}|\mathbf{y}^{(t)}, \lambda) = \lambda y_t E(\theta_t|\mathbf{y}^{(t)}, \lambda)$$

$$E(y_{t+1}|\mathbf{y}^{(t)}, \lambda) = \lambda y_t \hat{\theta}_t$$

E a variância:

$$V(y_{t+1}|\mathbf{y}^{(t)}, \lambda) = V(\lambda\theta_t y_t|\mathbf{y}^{(t)}, \lambda) + V(\omega_t y_t|\mathbf{y}^{(t)}, \lambda) + V(\epsilon_t|\mathbf{y}^{(t)}, \lambda)$$

$$V(y_{t+1}|y^{(t)}, \lambda) = \lambda^2 y_t^2 V(\theta_t|y^{(t)}, \lambda) + y_t^2 V(\omega_t y^{(t)}, \lambda) + \sigma^2$$

$$V(y_{t+1}|y^{(t)}, \lambda) = \lambda^2 y_t^2 \hat{\Sigma}_t + y_t^2 \gamma^2 + \sigma^2$$

$$V(y_{t+1}|y^{(t)}, \lambda) = y_t^2 (\lambda^2 \hat{\Sigma}_t + \gamma^2) + \sigma^2$$

Assim,

$$(y_{t+1}|y^{(t)}, \lambda) \sim N[\lambda y_t \hat{\theta}_t; y_t^2 (\lambda^2 \hat{\Sigma}_t + \gamma^2) + \sigma^2] \quad (4.31)$$

## 4.4 Resultados

Resumindo os resultados desse modelo, temos:

Modelo:

$$y_t = \theta_t y_{t-1} + \epsilon_t$$

$$\theta_t = \lambda \theta_{t-1} + \omega_t$$

onde  $\begin{pmatrix} \epsilon_t \\ \omega_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma^2 & 0 \\ 0 & \gamma^2 \end{pmatrix}\right]$ ,  $\sigma^2$  e  $\gamma^2$  conhecidos

Função de Verossimilhança:

$$p(y_i|y^{(i-1)}, \lambda) \propto \prod_{i=1}^t (y_{i-1}^2 r_i + \sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2(y_{i-1}^2 r_i + \sigma^2)} (y_i - \lambda \hat{\theta}_{i-1} y_{i-1})^2\right\}$$

Densidade *a posteriori* para  $\theta_t$ :

$$p(\theta_t|y^{(t)}, \lambda) \propto \hat{\Sigma}_t^{-1} \exp\left\{-\frac{1}{2\hat{\Sigma}_t} (\theta_t - \hat{\theta}_t)^2\right\}$$

Densidade *a posteriori* para  $\lambda$ :

$$p(\lambda|y^{(t)}) \propto \prod_{i=1}^t (y_{i-1}^2 r_i + \sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2(y_{i-1}^2 r_i + \sigma^2)} (y_i - \lambda \hat{\theta}_{i-1} y_{i-1})^2\right\} \times$$

$$S^{-1} \exp\left\{-\frac{1}{2S^2} (\lambda - m)^2\right\}$$

Os valores esperados da preditiva de  $\theta_{t+1}$  são dados por:

$$E(\theta_{t+1}|y_t, \lambda) = \lambda \hat{\theta}_t$$

$$V(\theta_{t+1}|y_t, \lambda) = \lambda^2 \hat{\Sigma}_t + \gamma^2$$

A preditiva para  $y_{t+1}$ :

$$(y_{t+1}|y^{(t)}, \lambda) \sim N[\lambda y_t \hat{\theta}_t ; y_t^2 (\lambda^2 \hat{\Sigma}_t + \gamma^2) + \sigma^2]$$

Obs: Como o procedimento para calcular  $E(\theta_{t+k-1} y_{t+k-1} | y^{(t)}, \lambda)$  e  $V(\theta_{t+k-1} y_{t+k-1} | y^{(t)}, \lambda)$  envolve aproximação numérica, omitimos aqui o cálculo de preditiva de  $y_{t+k}$  para  $k > 1$ .



# Capítulo 5

## Aplicação - Estudo de Casos

Nos capítulos 2, 3 e 4, apresentamos os modelos hierárquico e dinâmico para modelagem do processo auto-regressivo com coeficiente aleatório.

Neste capítulo, ilustramos a aplicação do processo auto-regressivo de coeficientes aleatórios, sob enfoque bayesiano, apresentando os resultados obtidos pelos modelos propostos para estimação e previsão dos dados de duas série reais.

Primeiramente, consideramos a série correspondente ao preço de ações da Siderúrgica Vale do Rio Doce, em dólares, negociadas no período de 02 de janeiro de 1996 a 01 de fevereiro de 1999, pela BOVESPA (Bolsa de Valores de São Paulo).

Posteriormente, consideramos uma série que representa o preço da arroba do boi gordo, em dólares, negociadas pela Tortuga Cia Zootécnica Agrária, no período de janeiro de 1998 a agosto de 1999.

A partir das condicionais *a posteriori*, aplicando os algoritmos Gibbs e Metropolis-Hasting, geramos amostras das séries supracitadas, considerando 5 cadeias com 3000 iterações cada uma, e para garantir a independência entre os valores simulados de cada parâmetro, tomamos 50% das iterações, isto é, desprezamos 1500 valores e dessemos um salto a cada 5 valores, selecionando, assim, uma amostra com 300 valores em cada cadeia, totalizando uma amostra final de 1500 valores.

Utilizando essa amostra, aproximamos densidade preditiva dos parâmetros de interesse de cada modelo, através dos estimadores de Monte Carlo e verificamos graficamente qual o modelo que melhor representa os dados, em função dos valores previstos obtidos.

Para cada modelo proposto, os resultados obtidos são apresentados através de tabelas e gráficos, conforme veremos a seguir.

## 5.1 Modelo Hierárquico com $\tau^2$ Conhecida

Aqui, apresentamos o estudo de caso de dois conjuntos de dados aplicando o modelo proposto por (3.2), onde a variância do ruído ( $\tau^2$ ) é conhecida.

A partir das distribuições *a posteriori* condicionais, segundo o algoritmo Gibbs, geramos amostras para cada um dos parâmetros de interesse. Uma vez obtida essas amostras, fizemos a previsão dos valores observados para os últimos 30 dias do período considerado, cuja análise apresentamos graficamente.

Na Tabela 5.1, podemos observar os hiperparâmetros baseados em uma análise preliminar dos dados da série do preço das ações da Siderúrgica Vale do Rio Doce, negociadas em dólares, pela Bovespa, no período de 02 de janeiro de 1996 a 01 de fevereiro de 1999, considerando *a priori* informativa normal para  $\lambda$  e  $\theta_t$ .

**Tabela 5.1: Valores *a Priori* para ajustar a Série.**

$m$	$S$	$\gamma$	$\tau$
1.0	0.5	0.65	0.33

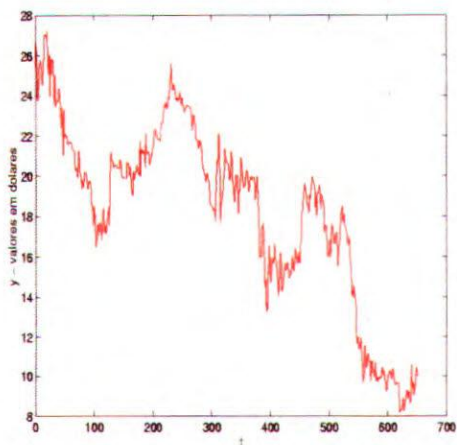
$m$  = média de  $\lambda$

$S$  = desvio padrão de  $\lambda$

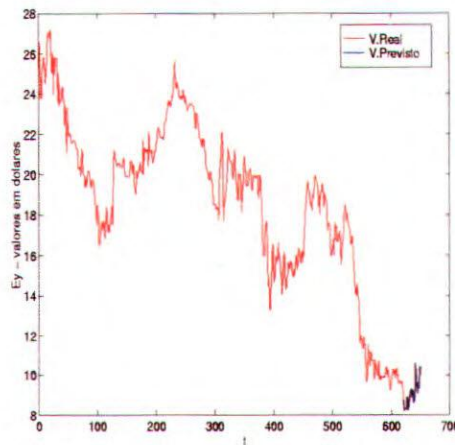
$\gamma$  = desvio padrão de  $\omega_t$

$\tau$  = desvio padrão de  $\epsilon_t$

A figura 5.1 ilustra o comportamento da série e a figura 5.2 apresenta a previsão para os últimos 30 dias.



**Fig. 5.1: Evolução de preço diário de ações da Vale do Rio Doce.**

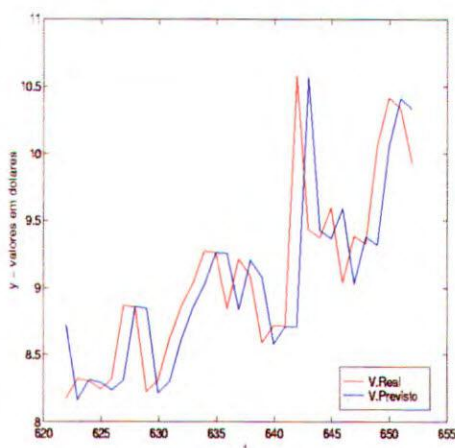


**Fig. 5.2: Preços reais e previstos de 30 dias, de ações da Vale do Rio Doce.**

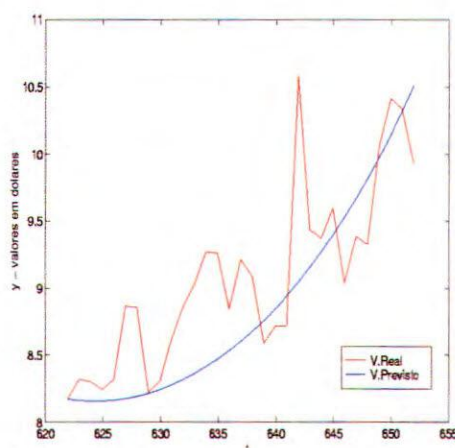
Pelas figuras 5.1 e 5.2, podemos observar que tanto os preços reais quanto os valores previstos não apresentam estacionariedade.

Com o objetivo de comparar a evolução de preços reais com os preços previstos, calculamos a previsão desses preços para 30 dias e apresentamos os resultados obtidos nas figuras 5.3 e 5.4.

Na figura 5.3, visualizamos o comportamento dos valores previstos considerando um passo (dia) à frente, ou seja,  $(y_{t+k}|y^{(t+k-1)})$  ;  $k = 1, \dots, 30$ , enquanto que na figura 5.4, visualizamos a tendência da previsão e dos valores reais para os 30 dias, considerando os dados observados até o instante  $t$ , isto é,  $(y_{t+k}|y^{(t)})$ , onde os 30 últimos dias foram deixados fora do conjunto de observações usadas no ajuste do modelo.



**Fig. 5.3 : Comparação diária de Preços de ações da Vale do Rio Doce.**



**Fig. 5.4: Tendência de preços de ações da Vale do Rio Doce.**

Conforme podemos observar nas figuras 5.3 e 5.4, os valores previstos, obtidos a partir da amostra gerada, possuem, basicamente, a mesma tendência que os valores reais; houve apenas uma diferença de locação nos valores previstos, que pode ser vista na figura 5.3.

Com o intuito de comparar os últimos 30 valores reais de  $y_t$  com os valores previstos, registramos na tabela 5.2, os valores de  $y_t$ ,  $\theta_{t+k}$ ,  $y_{t+k}$  e o erro percentual de previsão.

**Tabela 5.2: Valores Previstos, um passo à frente, para a série de preços de ações da Vale do Rio Doce, no período de 02/01/96 a 01/02/99.**

$k$	$y_t$	$E(\theta_{t+k} y^{(t)})$	$E(y_{t+k} y^{(t+k-1)})$	$e_t(\%)$
1	8.1668	0.9987	8.1668	0.00
2	8.3154	0.9987	8.1562	0.20
3	8.2968	0.9987	8.1512	0.12
4	8.2393	0.9987	8.1517	0.23
5	8.3123	0.9987	8.1578	0.96
6	8.8657	0.9988	8.1693	0.87
7	8.8541	0.9988	8.1865	0.04
8	8.2184	0.9987	8.2091	0.13
9	8.3013	0.9988	8.2374	0.53
10	8.6175	0.9988	8.2714	0.76
11	8.8554	0.9988	8.3110	0.89
12	9.0247	0.9989	8.3564	1.08
13	9.2677	0.9989	8.4077	0.99
14	9.2593	0.9989	8.4649	0.46
15	8.8430	0.9989	8.5281	0.83
16	9.2118	0.9989	8.5976	0.58
17	9.0842	0.9989	8.6733	0.11
18	8.5851	0.9988	8.7555	0.05
19	8.7167	0.9989	8.8443	0.16
20	8.7161	0.9989	8.9399	1.72
21	10.5737	0.9992	9.0424	0.39
22	9.4319	0.9990	9.1521	0.24
23	9.3710	0.9990	9.2692	0.36
24	9.5964	0.9991	9.3939	0.41
25	9.0351	0.9990	9.5266	0.16
26	9.3847	0.9990	9.6673	0.34
27	9.3234	0.9990	9.8166	0.25
28	10.0604	0.9991	9.9745	0.41
29	10.4134	0.9992	10.1416	0.18
30	10.3368	0.9992	10.3181	0.37

Fonte: Valores de  $y_t$  fornecidos pela BOVESPA.

Temos também que o erro médio de previsão é dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|y_{obs} - \hat{y}|}{y_{obs}} 100 = 0.461\%$$

Como podemos verificar, esse modelo apresenta um erro médio percentual de previsão inferior a 5%; assim, podemos concluir que o mesmo é adequado para fazer previsões de observações futuras.

Agora, fazemos um estudo de caso de uma série real que representa o preço da arroba de boi gordo, em dólares, no período de janeiro de 1998 a agosto de 1999, negociado pela Tortuga Cia Zootécnica Agrária.

Na Tabela 5.3, visualizamos os valores para os hiperparâmetros, baseados em uma análise preliminar dos dados, considerando priori informativa normal para  $\lambda$  e  $\theta_t$ .

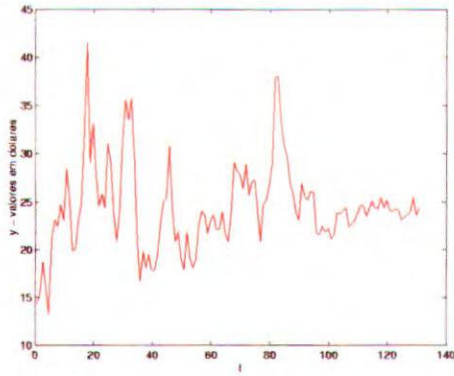
**Tabela 5.3: Valores das Densidades *a Priori* para ajustar a Série.**

$m$	$S$	$\gamma$	$\tau$
1.0	0.35	0.12	2.5

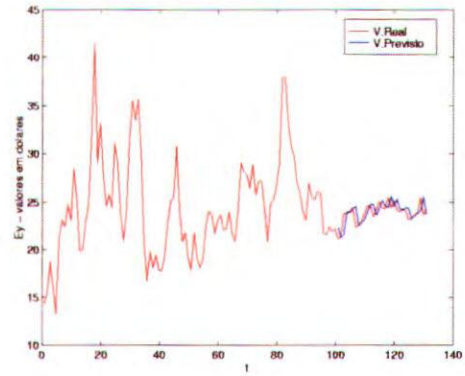
$m$  = média de  $\lambda$   
 $S$  = desvio padrão de  $\lambda$   
 $\gamma$  = desvio padrão de  $\omega_t$   
 $\tau$  = desvio padrão de  $\epsilon_t$

A partir das distribuições *a posteriori* condicionais, segundo o algoritmo Gibbs, geramos amostras de tamanho 131 para cada um dos parâmetros de interesse e, assim, fizemos a previsão para os últimos 30 dias observados do período.

A figura 5.5 ilustra a evolução do preço real da arroba do boi gordo e a figura 5.6 visualiza a evolução do preço previsto um passo (dia) à frente e preço real nos últimos 30 dias do período considerado.



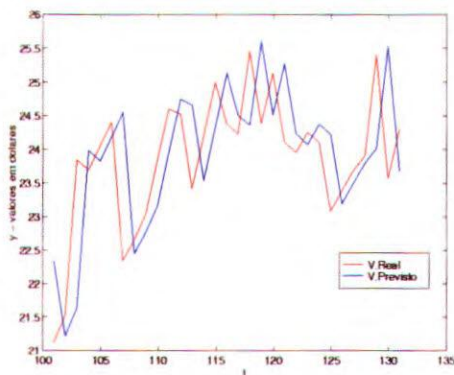
**Fig. 5.5: Evolução diária do preço da arroba do boi gordo.**



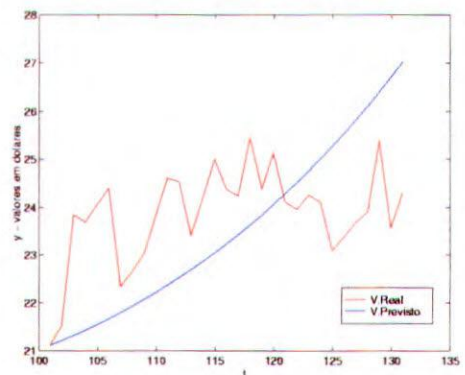
**Fig. 5.6: Previsão do preço em 30 dias, da arroba do boi gordo.**

Pelas figuras 5.5 e 5.6, podemos observar que tanto os preços reais quanto os preços previstos apresentam não-estacionariedade.

Comparamos graficamente a evolução dos preços reais com os preços previstos. Na figura 5.7, apresentamos os resultados obtidos na previsão de 30 dias, considerando um passo (dia) à frente, enquanto que, na figura 5.8, apresentamos a tendência da previsão dos 30 dias considerando todos os dias anteriores, deixando de lado aqui também os últimos 30 dias da série dos dados usados para o ajuste do modelo.



**Fig. 5.7 : Comparação diária de preços reais e previstos da arroba do boi.**



**Fig. 5.8: Tendência de preços da arroba do boi.**

Conforme podemos observar nas figuras 5.7 e 5.8, os valores previstos possuem, basicamente, a mesma tendência que os valores reais.

Com o intuito de visualizar esses resultados, apresentamos o erro percentual de previsão na tabela 5.6.

**Tabela 5.4: Valores Previstos, um passo à frente para a série de preço da arroba do boi gordo, em dólares, no período de 01/1998 a 08/1999.**

$k$	$y_t$	$E(\theta_{t+k} y^{(t)})$	$E(y_{t+k} y^{(t)})$	$e_t(\%)$
1	21.1100	1.0045	22.0450	4.42
2	21.5100	1.0046	21.1076	1.40
3	23.8400	1.0048	21.3110	10.61
4	23.6900	1.0057	21.4203	9.58
5	24.0500	1.0056	21.5355	10.45
6	24.4000	1.0057	21.6567	11.24
7	22.3300	1.0049	21.7841	2.44
8	22.6500	1.0049	21.9177	3.23
9	23.0300	1.0051	22.0576	4.17
10	23.8400	1.0053	22.2040	6.68
11	24.6000	1.0056	22.3569	9.11
12	24.5200	1.0055	22.5164	8.17
13	23.4100	1.0050	22.6827	3.10
14	24.2000	1.0053	22.8560	5.55
15	24.9900	1.0055	23.0362	7.81
16	24.3700	1.0052	23.2237	4.70
17	24.2300	1.0051	23.4185	3.34
18	25.4500	1.0055	23.6208	7.18
19	24.3800	1.0051	23.8307	2.25
20	25.1300	1.0053	24.0485	1.19
21	24.1100	1.0049	24.2742	0.68
22	23.9500	1.0048	24.5081	2.33
23	24.2500	1.0049	24.7504	2.06
24	24.1000	1.0048	25.0013	3.73
25	23.0800	1.0044	25.2609	9.44
26	23.3800	1.0045	25.5295	9.19
27	23.6800	1.0046	25.8073	8.98
28	23.9000	1.0046	26.0946	9.18
29	25.4000	1.0050	26.3915	0.43
30	23.5600	1.0044	26.6984	13.32

Fonte: Valores de  $y_t$  fornecidos por Tortuga Cia. Z. Agrária.

Temos, também, que o erro percentual de previsão é dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|y_{obs} - \hat{y}|}{y_{obs}} 100 = 5.80\%$$

Notamos um erro médio percentual superior a 5% , esse valor pode ter ocorrido devido à grande mudança no comportamento da série nos últimos 30 dias, quando comparados com as observações anteriores.

## 5.2 Modelo Hierárquico com $\tau^2$ Desconhecida

Agora, analisamos os dados das séries, considerando o modelo (3.1), onde a variância do ruído ( $\tau^2$ ) é considerada desconhecida.

Na tabela 5.5, apresentamos os valores utilizados considerando *a priori* informativa normal e gama inversa para lambda e tau, respectivamente, para gerar amostras desses parâmetros a partir das distribuições condicionais *a posteriori*, considerando 5 cadeias com 3000 iterações cada uma.

**Tabela 5.5: Valores das densidades *a Priori*.**

$m$	$S$	$\gamma$	$a$	$b$
1.0	0.5	1	50	10

$m =$  média de  $\lambda$

$S =$  desvio padrão de  $\lambda$

$\gamma =$  desvio padrão de  $\omega_t$

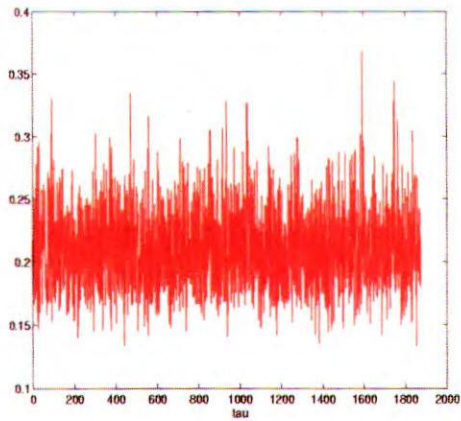
$a$  e  $b$  parâmetros da distribuição *a priori* de  $\tau$

Devido ao fato de as distribuições consideradas não apresentarem uma forma padrão, utilizamos, então, o algoritmo Gibbs combinado com Métropolis-Hasting, para gerar amostras considerando a série de preços das ações da Vale do Rio Doce .

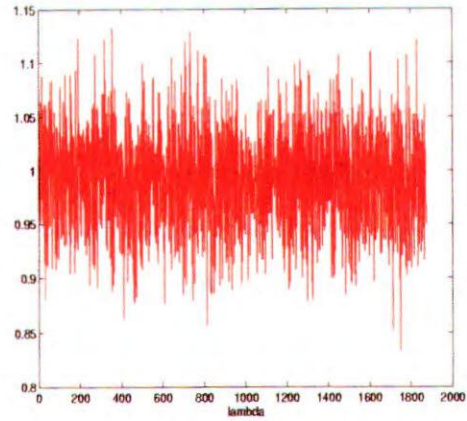
Para obtermos uma independência dos pontos amostrais, desprezamos 50% dos valores iniciais e selecionamos um valor a cada 5, totalizando assim, uma amostra de 300 valores em cada cadeia, obtendo uma amostra final com 1500 dados.

As figuras 5.9, 5.10 e 5.11 mostram a convergência para os dados da série simulada, partindo de uma condição inicial arbitrária, demonstrando, de forma gráfica, a análise de convergência para  $\lambda$ , para  $\tau$  e para  $\lambda$  e  $\tau$  e as figuras 5.12 e 5.13 apresentam os histogramas dessas distribuições *a posteriori* estimados por simulação para os parâmetros  $\tau$  e  $\lambda$ , uma vez que não é possível deduzir explicitamente a expressão dessas distribuições.

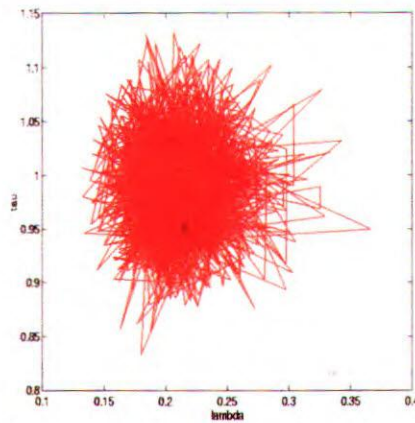




**Fig. 5.9: Convergência para  $\tau$ .**

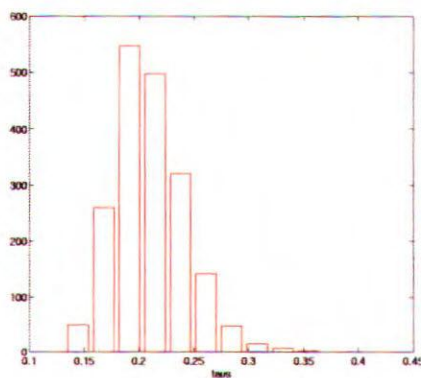


**Fig. 5.10: Convergência para  $\lambda$ .**

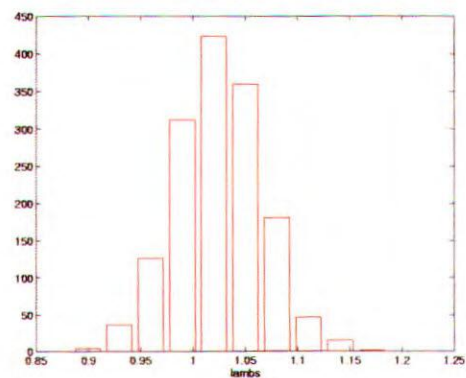


**Fig. 5.11: Convergência para  $\lambda$  e  $\tau$ .**

A convergência das amostras geradas pelo algoritmo Gibbs com Metropolis-Hasting foi monitorada utilizando o método proposto por Gelman e Rubin (1992), o qual se baseia na análise de variância (ver anexo A).



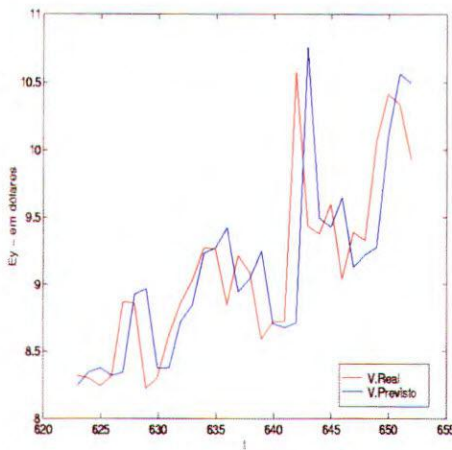
**Fig. 5.12: Dist. a Posteriori para  $\tau$ .**



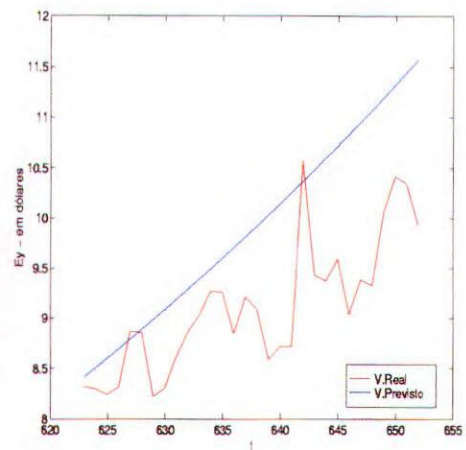
**Fig. 5.13: Dist. a Posteriori para  $\lambda$ .**

As figuras 5.12 e 5.13, mostram que as distribuições *a posteriori* são assimétricas.

Na figura 5.14, visualizamos a evolução de preços reais e de preços previstos um passo (dia) à frente ( $y_{t+k}|y^{(t+k-1)}$  ;  $k = 1, \dots, 30$ ), enquanto que na figura 5.15, visualizamos a tendência dos preços das ações da Siderúrgica Vale do Rio Doce, negociadas nos últimos 30 dias do período analisado , sendo que os valores de  $y_{t+k}$  são calculados considerando todos os dados anteriores, ou seja, ( $y_{t+k}|y^{(t)}$ ), observando que nas duas figuras os últimos 30 dias foram deixados fora dos dados usados para o ajuste do modelo.



**Fig. 5.14 : Comparação diária de preços de ações da Vale do Rio Doce.**



**Fig. 5.15: Tendência de preços de ações da Vale do Rio Doce.**

Na figura 5.14, é possível observar que, pelo fato de acrescentarmos uma incerteza na variância, ocorre uma discrepância nos valores previstos a partir da amostra gerada, em relação aos valores reais dos dados.

Na tabela 5.6, mostramos os valores reais dos últimos 30 dias do período, os valores de  $\theta_{t+k}$ , os valores previstos  $y_{t+k}$  correspondentes e o erro de previsão, como intuito de fazer a comparação dos valores previstos e os valores reais.

**Tabela 5.6: Valores Previstos, um passo à frente para a série de preços de ações da Vale do Rio Doce, no Período de 02/01/96 a 01/02/99.**

$k$	$y_t$	$E(\theta_{t+k} y^{(t)})$	$E(y_{t+k} y^{(t+k-1)})$	$e_t(\%)$
1	8.3154	1.0084	8.2351	0.96
2	8.2968	0.9967	8.3372	0.48
3	8.2393	0.9958	8.3711	1.59
4	8.3123	0.9959	8.1323	2.16
5	8.8657	0.9961	8.3367	5.96
6	8.8541	0.9957	8.9186	0.72
7	8.2184	0.9959	8.9616	9.04
8	8.3013	0.9956	8.3650	0.76
9	8.6175	0.9960	8.3702	2.86
10	8.8554	0.9972	8.7158	1.57
11	9.0247	0.9970	8.8360	5.24
12	9.2677	0.9957	9.2256	0.45
13	9.2593	0.9956	9.2665	0.07
14	8.8430	0.9944	9.4157	6.47
15	9.2118	0.9979	8.9353	3.00
16	9.0842	0.9939	9.0385	0.50
17	8.5851	0.9945	9.2424	7.65
18	8.7167	0.9958	8.7016	0.17
19	8.7161	0.9947	8.6706	0.52
20	10.5737	0.9931	8.7068	17.65
21	9.4319	0.9961	10.7540	14.01
22	9.3710	0.9964	9.4875	1.13
23	9.5964	0.9960	9.6433	6.73
24	9.0351	0.9963	9.1229	2.78
25	9.3847	0.9959	9.2149	1.16
26	9.3234	0.9960	9.2699	7.89
27	10.0604	0.9963	10.0953	3.05
28	10.4134	0.9952	10.5008	0.83
29	10.3368	0.9960	10.5593	2.15
30	9.9297	0.9955	10.4883	5.62

Fonte: Valores de  $y_t$  fornecidos pela BOVESPA.

Nesse caso, temos um erro médio de previsão dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|y_{obs} - \hat{y}|}{y_{obs}} 100 = 3.75\%$$

Comparando o erro médio de previsão nos dois casos, a variância conhecida e a variância desconhecida, observamos que, no segundo caso quando acrescentamos uma incerteza na variância, temos um grande aumento no valor do erro médio. Esse resultado é esperado devido ao aumento da incerteza nos parâmetros do modelo.

Agora, analisamos os dados da série de preço da arroba do boi gordo com o modelo (3.1), onde a variância do ruído ( $\tau^2$ ) é considerada desconhecida.

Na tabela 5.7, apresentamos os valores utilizados considerando *as priors* informativas normal e gama inversa para lambda e tau, respectivamente, para gerar amostras desses parâmetros a partir das distribuições condicionais *a posteriori*, considerando 5 cadeias com 3000 iterações cada uma.

Devido ao fato de as distribuições consideradas não apresentarem uma forma padrão, utilizamos, então, o algoritmo Gibbs combinado com Metropolis-Hasting para gerar essas amostras e, para obtermos uma independência dos pontos, desprezamos 50% dos valores iniciais e selecionamos um valor a cada 5, totalizando, assim, 5 amostras de 300 valores em cada cadeia, obtendo uma amostra final de 1500 valores.

**Tabela 5.7: Valores das Densidades *a Priori*.**

$m$	$S$	$\gamma$	$a$	$b$
1.0	0.75	0.3	40	10

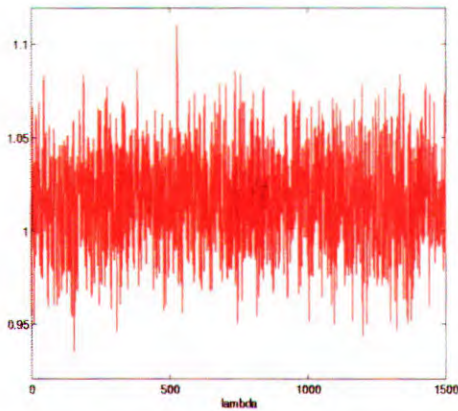
$m =$  média de  $\lambda$

$S =$  desvio padrão de  $\lambda$

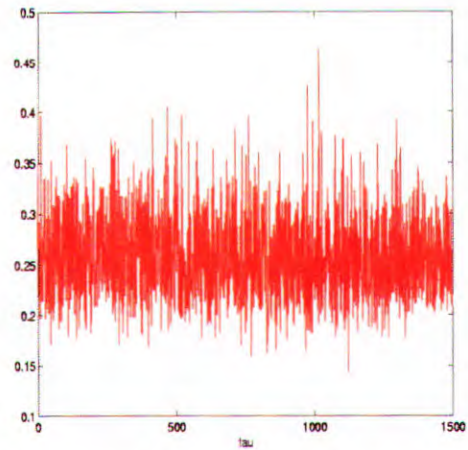
$\gamma =$  desvio padrão de  $\omega_t$

$a$  e  $b$  parâmetros da distribuição *a priori* de  $\tau$

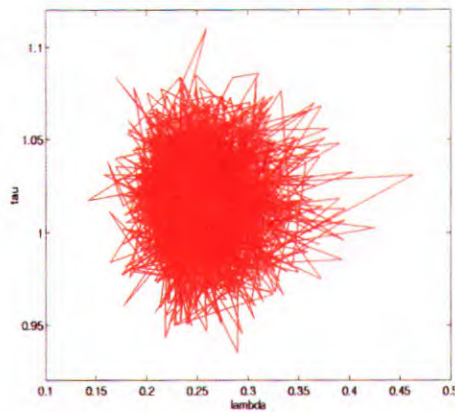
As figuras 5.16, 5.17 e 5.18 mostram a convergência para os dados da série simulada, partindo de uma condição inicial arbitrária, demonstrando, de forma gráfica, a análise de convergência para  $\lambda$ , para  $\tau$  e para  $\lambda$  e  $\tau$ .



**Fig. 5.16: Convergência para  $\tau$ .**



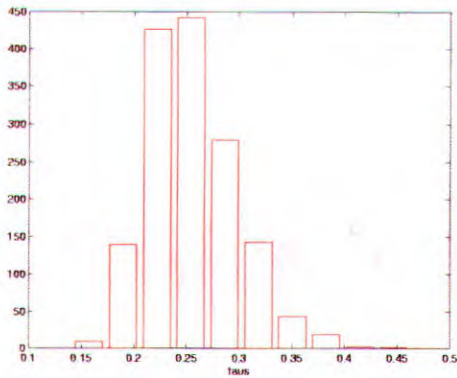
**Fig. 5.17: Convergência para  $\lambda$ .**



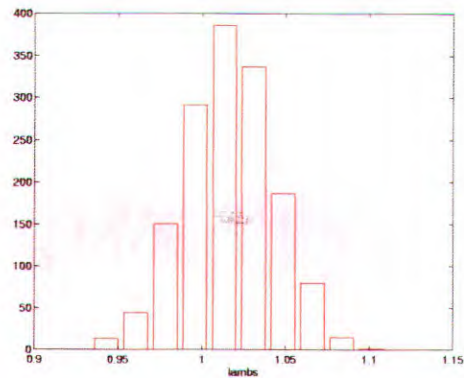
**Fig. 5.18: Convergência para  $\lambda$  e  $\tau$ .**

A convergência das amostras geradas pelo algoritmo Gibbs com Metropolis-Hasting foi monitorada utilizando o método proposto por Gelman e Rubin (1992) (ver anexo A).

Agora, nas figuras 5.19 e 5.20, apresentamos os histogramas das distribuições *a posteriori* estimados por simulação para os parâmetros  $\tau$  e  $\lambda$ , uma vez que não é possível deduzir explicitamente a expressão dessas distribuições.



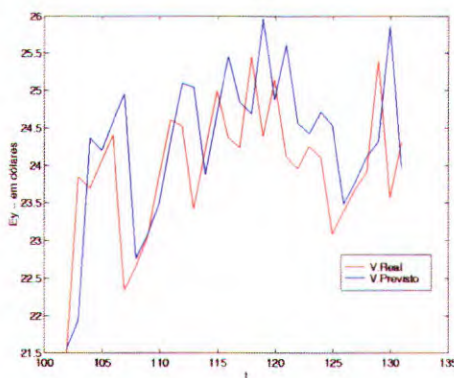
**Fig. 5.19: Dist. a Posteriori. para  $\tau$**



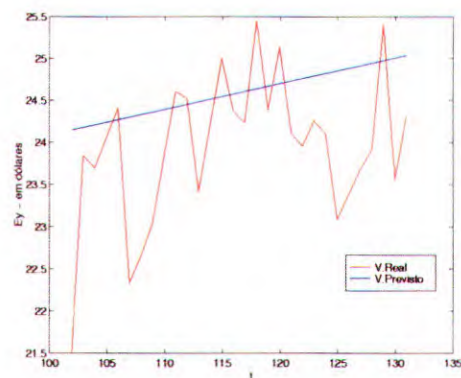
**Fig. 5.20: Dist. a Posteriori. para  $\lambda$**

A figura 5.19 mostra que a distribuição *a posteriori* de  $\tau$  é assimétrica.

A figura 5.21 ilustra a evolução de preços reais e de preços previstos um passo (dia) à frente, enquanto que a figura 5.22 mostra a tendência de preços da arroba do boi gordo nos últimos 30 dias do período considerado, sendo que, aqui também, os mesmos foram deixados fora do conjunto de dados utilizados no ajuste do modelo e, para se obterem os valores previstos, foram considerados todos os outros dados anteriores



**Fig. 5.21: Comparação diária de preços da arroba do boi gordo.**



**Fig.5.22: Tendência de preços da arroba do boi gordo.**

Na figura 5.22, é possível observar que, pelo fato de acrescentarmos uma incerteza na variância, ocorre uma discrepância nos valores previstos a partir da amostra gerada, em relação aos valores dos dados reais.

Na tabela 5.8, comparamos os valores reais dos últimos 30 dias do período com os valores previstos  $y_{t+k}$  correspondentes.

Tabela 5.8: Valores Previstos, um passo à frente para a série de preços da arroba do boi gordo no Período de 01/1998 a 08/1999.

$k$	$y_t$	$E(\theta_{t+k} y^{(t)})$	$E(y_{t+k} y^{(t)})$	$e_t(\%)$
1	21.5100	1.0210	21.5535	0.21
2	23.8400	1.0195	21.9289	8.01
3	23.6900	1.0218	24.3600	1.40
4	24.0500	1.0209	24.1856	0.56
5	24.4000	1.0220	24.5795	0.73
6	22.3300	1.0225	24.9492	11.72
7	22.6500	1.0189	22.7525	0.45
8	23.0300	1.0183	23.0641	0.14
9	23.8400	1.0192	23.4733	1.53
10	24.6000	1.021	24.3189	1.14
11	24.5200	1.0198	24.086	2.30
12	23.4100	1.0208	25.0303	6.93
13	24.2000	1.0194	23.8639	1.38
14	24.9900	1.0200	24.6844	2.77
15	24.3700	1.0184	25.4497	4.43
16	24.2300	1.0191	24.8359	2.50
17	25.4500	1.0185	24.6794	3.02
18	24.3800	1.0197	25.9505	6.44
19	25.1300	1.0199	24.8658	1.05
20	24.1100	1.0189	25.6039	6.19
21	23.9500	1.0184	24.5546	2.52
22	24.2500	1.0194	24.4157	0.68
23	24.1000	1.0187	24.7034	2.50
24	23.0800	1.0175	24.5217	6.24
25	23.3800	1.0172	23.4764	0.41
26	23.6800	1.0167	23.7701	0.38
27	23.9000	1.0180	24.1068	0.86
28	25.4000	1.0171	24.3084	4.28
29	23.5600	1.0181	25.8594	9.75
30	24.3000	1.0169	23.9582	1.40

Fonte: Valores de  $y_t$  fornecidos por Tortuga Cia Z. Agrária.

Nesse caso, temos um erro médio de previsão dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|y_{obs} - \hat{y}|}{y_{obs}} 100 = 3.06\%$$

Aqui, mesmo acrescentando uma incerteza na variância, o erro médio de previsão foi menor que no caso anterior; portanto podemos afirmar que este modelo é adequado para ajustar dados financeiros bem como para fazer previsões de observações futuras.

### 5.3 Modelo Dinâmico com Variância Conhecida

A ilustração para o processo dado pelo modelo proposto em (4.16) é análoga às ilustrações anteriores. Consideramos os mesmos conjuntos de dados; primeiramente, analisamos os preços de ações da Vale do Rio Doce.

Na tabela 5.9, apresentamos os valores aplicados a partir de distribuições *a priori* normal para  $\lambda$  e  $\theta_t$  e geramos amostras a partir das distribuições condicionais *a posteriori* para os parâmetros, considerando 5 cadeias com 3000 iterações cada uma.

**Tabela 5.9: Valores *a Priori*.**

$m$	$S$	$\gamma$	$\sigma$	$y_0$
1.0	0.5	0.65	0.3	10

$m$  = média de  $\lambda$

$S$  = desvio padrão de  $\lambda$

$\gamma$  = desvio padrão do ruído da eq. de estados

$\sigma$  desvio padrão do ruído da eq. observações

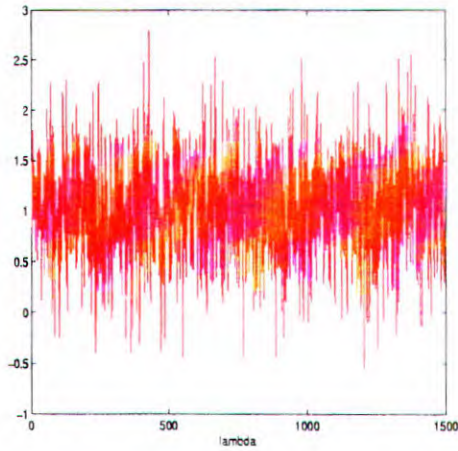
$y_0$  = valor inicial de  $y_t$ , arbitrário.

Pelo fato de as distribuições *a posteriori* condicionais não apresentarem uma forma padrão, simulamos amostras aplicando o algoritmo Gibbs com Metropolis-Hasting, sendo que, para assegurarmos a independência entre os valores simulados para cada parâmetro, consideramos 50% das 3000 iterações, ou seja, desprezamos 1500 valores e, desses, damos um salto de 5 para cada valor, selecionando, assim, uma amostra de tamanho 300 em cada cadeia, totalizando uma amostra de 1500 valores.

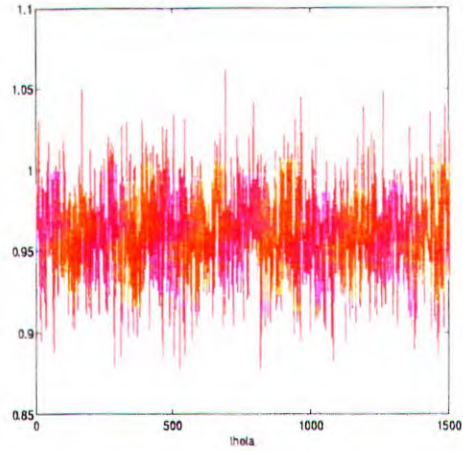
Os resultados obtidos dessa amostra são apresentados de forma gráfica, conforme as figuras a seguir.

As figuras 5.23, 5.24 e 5.25 mostram a análise de convergência para os parâmetros de interesse, partindo de condições iniciais arbitrárias dos dados da série.

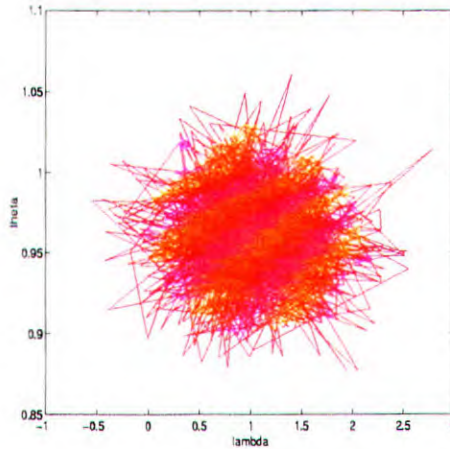




**Fig. 5.23: Convergência para  $\lambda$ .**



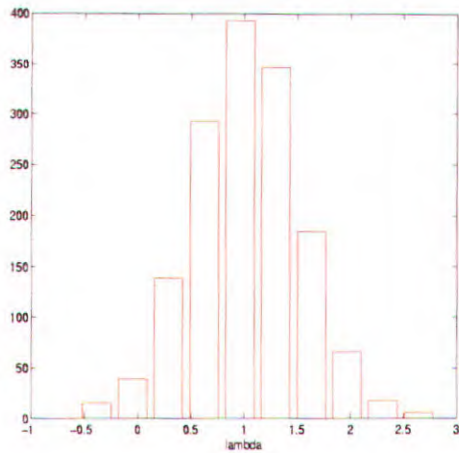
**Fig. 5.24: Convergência para  $\theta_t$ .**



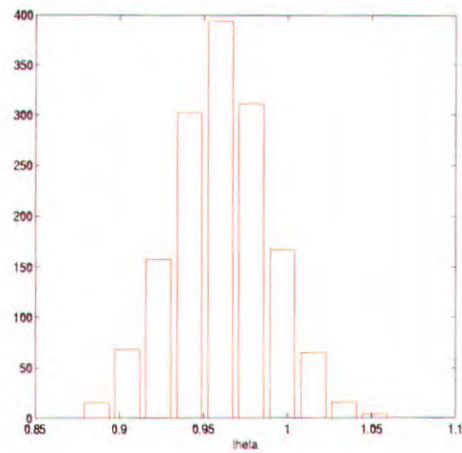
**Fig.5.25: Convergência para  $\lambda$  e  $\theta_t$ .**

Aqui, também, a convergência das amostras geradas pelo algoritmo Gibbs com Metropolis-Hasting foi monitorada utilizando-se o método proposto por Gelman e Rubin (1992) (ver anexo A).

Agora, apresentamos graficamente a distribuição *a posteriori* de  $\lambda$  e  $\theta_t$ , onde os gráficos 5.26 e 5.27 mostram os histogramas estimados por simulação, pois as mesmas não apresentam uma forma padrão.



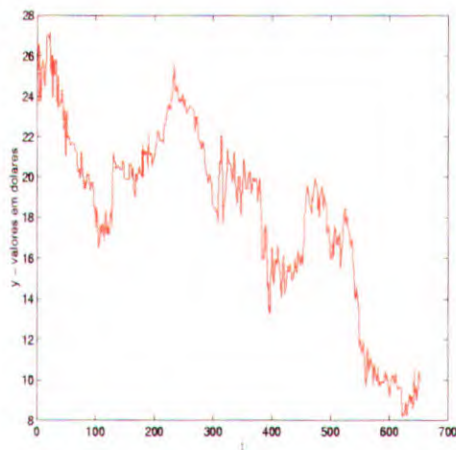
**Fig. 5.26: Dist. a Posteriori para  $\lambda$ .**



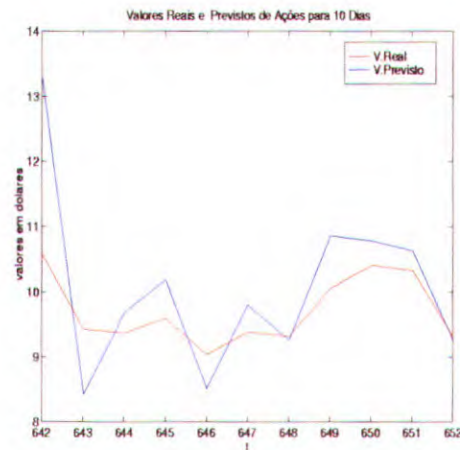
**Fig. 5.27: Dist. a Posteriori para  $\theta_t$ .**

Através das figuras, podemos verificar que as distribuição *a posteriori* são simétricas, no caso dos dois parâmetros.

Na figura 5.28, visualizamos a evolução de preços reais, e na figura 5.29, os preços reais e previstos das ações da Siderúrgica Vale do Rio Doce, negociadas em dólares, pela BOVESPA, nos últimos 10 dias do período de 02 de janeiro de 1996 a 01 de fevereiro de 1999, considerando um passo (dia) à frente.



**Fig. 5.28: Evolução diária de preços de ações da Vale do Rio Doce.**



**Fig. 5.29: Evolução de preços em 10 dias, de ações da Vale do Rio Doce.**

Na figura 5.29, é possível observar que , os valores previstos estão bem próximos dos valores reais com pequeno erro de previsão conforme podemos observar na tabela 5.10.

**Tabela 5.10: Valores Previstos, um passo à frente para a série de preços de ações da Vale do Rio Doce, no período de 02/01/96 a 01/02/99.**

$t + 1$	$y_t$	$E(\theta_{t+1} y_t)$	$E(y_{t+1} y_t)$	$e_t(\%)$
643	9.4319	0.8932	8.4249	10.67
644	9.3710	1.0333	9.6831	3.33
645	9.5964	1.0611	10.1832	6.11
646	9.0351	0.9423	8.5138	5.76
647	9.3847	1.0443	9.8005	4.43
648	9.3234	0.9931	9.2591	0.68
649	10.0604	1.0794	10.8594	7.94
650	10.4134	1.0357	10.7853	3.57
651	10.3368	1.0288	10.6345	2.88
652	9.9297	0.9994	9.2400	6.94

Fonte: Valores de  $y_t$  fornecidos pela BOVESPA.

Temos também que o erro percentual de previsão é dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|(y_{obs} - \hat{y})|}{y_{obs}} 100 = 5.23\%$$

Mediante os resultados obtidos neste exemplo, vale ressaltar que a dependência das variáveis (coeficiente) do modelo auto-regressivo não apresenta mudanças muito acentuadas na previsão das observações futuras.

A ilustração para o processo dado pelo modelo proposto em (4.16) é análoga a ilustrações anteriores. Consideramos a série que representa o preço da arroba do boi gordo.

Na tabela 5.11, apresentamos os valores aplicados a partir de distribuições *a priori* normal para  $\lambda$  e  $\theta_t$  e geramos amostras a partir das distribuições condicionais *a posteriori* para os parâmetros, considerando 5 cadeias com 3000 iterações, cada uma.

**Tabela 5.11: Valores *a Priori*.**

$m$	$S$	$\gamma$	$\sigma$	$y_0$
1.0	0.3	0.12	2.5	10

$m$  = média de  $\lambda$

$S$  = desvio padrão de  $\lambda$

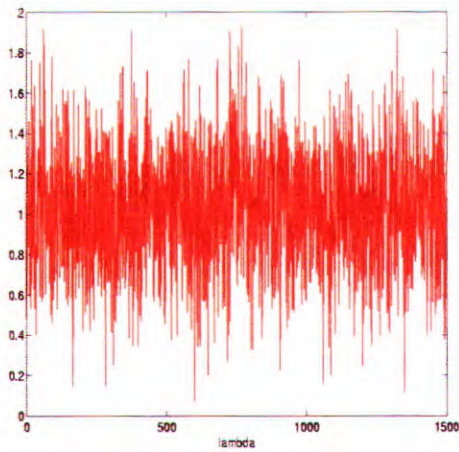
$\gamma$  = desvio padrão do ruído da eq. de estados

$\sigma$  = desvio padrão do ruído da eq. observações

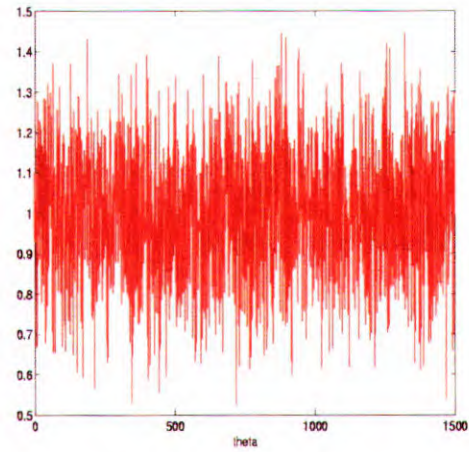
$y_0$  = valor inicial de  $y_t$  arbitrário.

Pelo fato de as distribuições *a posteriori* condicionais não apresentarem uma forma padrão, simulamos amostras aplicando o algoritmo Gibbs com Metropolis-Hasting, sendo que, para assegurarmos a independência entre os valores simulados para cada parâmetro, consideramos 50% das 3000 iterações, ou seja, desprezamos 1500 valores e desses demos um salto de 5 para cada valor, selecionando 5 amostras de 300 valores, totalizando uma amostra de tamanho 1500 . Os resultados obtidos por essa amostra são apresentados de forma gráfica.

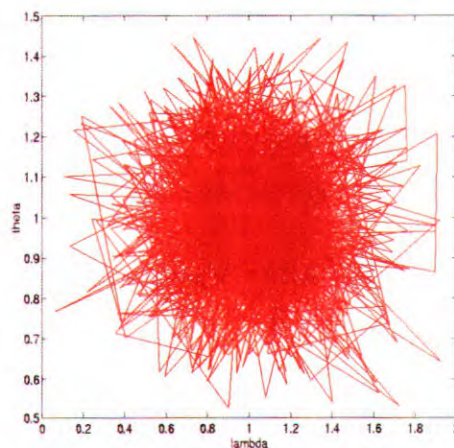
As figuras 5.30, 5.31 e 5.32 mostram, graficamente, a análise de convergência para os parâmetros de interesse, partindo de condições iniciais arbitrárias dos dados da série.



**Fig. 5.30: Convergência para  $\lambda$ .**



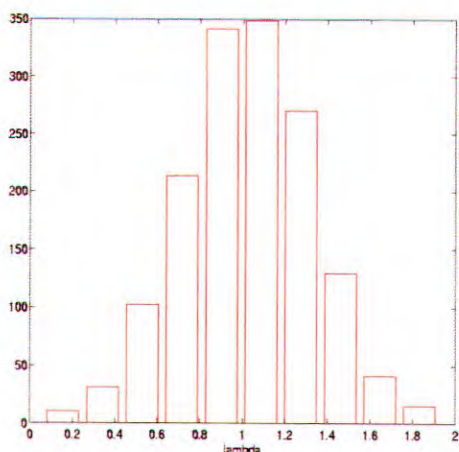
**Fig. 5.31: Convergência para  $\theta_t$ .**



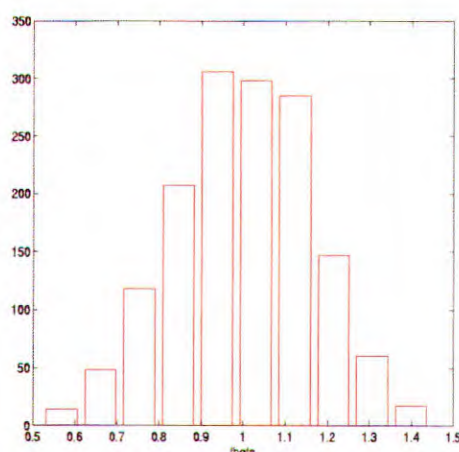
**Fig. 5.32: Convergência para  $\lambda$  e  $\theta_t$ .**

Aqui, também, a convergência das amostras geradas pelo algoritmo Gibbs com Metropolis-Hasting foi monitorada pelo método de Gelman e Rubin (1992) (ver anexo A).

Agora, apresentamos graficamente a distribuição *a posteriori* de  $\lambda$  e  $\theta_t$  onde os gráficos 5.33 e 5.34 mostram os histogramas estimados por simulação, pois as mesmas não apresentam uma forma padrão.



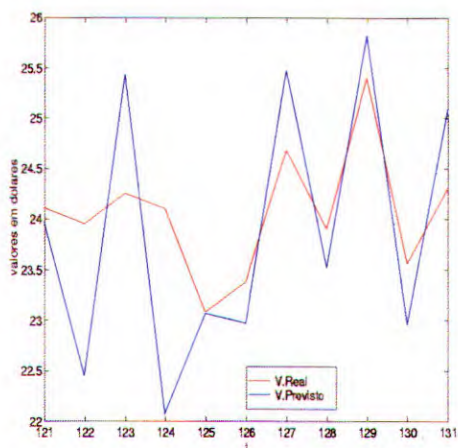
**Fig. 5.33: Dist.a Posteriori para  $\lambda$ .**



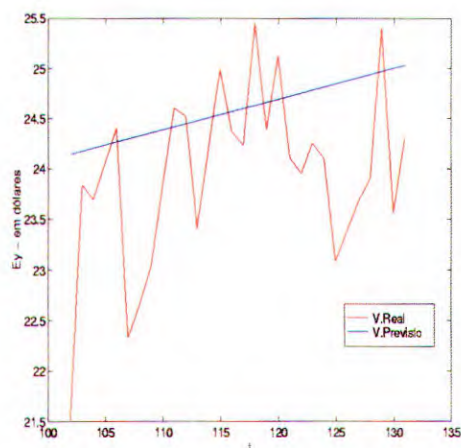
**Fig. 5.34: Dist. a Posteriori para  $\theta_t$ .**

Através das figuras, podemos verificar que as distribuições *a posteriori* são simétricas, no caso dos dois parâmetros.

Na figura 5.35, visualizamos a evolução de preços reais e de preços previstos da arroba do boi gordo, enquanto que, na figura 5.36, visualizamos a comparação entre o preço real e o preço previsto, considerando um passo (dia) à frente, nos últimos 10 dias do período de agosto de 1999, sendo que esses não foram computados no conjunto dos dados usados para ajustar o modelo.



**Fig. 5.35: Evolução diária do preço da arroba do boi gordo**



**Fig. 5.36: Evolução do preço da arroba de boi gordo, em 10 dias**

Na figura 5.36, é possível observar que os valores previstos, dado o preço real do dia anterior, estão bem próximos dos valores reais com pequeno erro de previsão. Isso é comprovado pelos valores da tabela 5.12.

**Tabela 5.12: Valores Previstos, um passo (dia) à frente, para a série de preço da arroba de boi gordo, no período de 01/1998 a 08/1999.**

$t + 1$	$y_t$	$E(\theta_{t+1} y_t)$	$E(y_{t+1} y_t)$	$e_t(\%)$
121	24.1100	0.9941	23.9665	0.59
122	23.9500	0.9372	22.4461	6.27
123	24.2500	1.0486	25.4284	4.85
124	24.1000	0.9157	22.0683	8.43
125	23.0800	0.9950	23.0676	0.05
126	23.3800	0.9824	22.9676	1.70
127	23.6800	0.9830	23.2780	2.60
128	23.9000	0.9840	23.5183	1.59
129	25.4000	1.0165	25.8195	1.65
130	23.5000	0.9743	22.9550	2.56
131	24.3000	1.0323	25.0852	3.23

Fonte: Valores de  $y_t$  fornecidos por Tortuga Cia Z. Agrária.

Temos também que o erro percentual de previsão é dado por:

$$\bar{e}_t = \frac{1}{N} \sum_{i=1}^N \frac{|(y_{obs} - \hat{y})|}{y_{obs}} 100 = 3.53\%$$

Dentre os resultados obtidos neste estudo de caso, podemos concluir que o modelo auto-regressivo de primeira ordem com coeficiente aleatório, é adequado para ajustar dados financeiros bem como para fazer previsões de observações futuras, com ressalva do caso em que consideramos a variância do ruído ( $\tau^2$ ) desconhecida.

# Capítulo 6

## Considerações Finais

### 6.1 Conclusão

O uso de modelos hierárquico e dinâmico para modelagem de processos auto-regressivos com parâmetros aleatórios são alternativas eficientes para analisar e para fazer previsões futuras da evolução de valores econômicos.

Nesse caso, a análise bayesiana, usando método de simulação de Monte Carlo em Cadeia de Markov, tem implementação razoavelmente fácil, não exigindo desenvolvimento de aproximações numéricas, e as inferências obtidas são bastante precisas. Entretanto, conforme resultados gráficos obtidos, podemos observar que o modelo hierárquico, onde consideramos a variância do ruído ( $\tau^2$ ) conhecida, apresenta os valores previstos mais próximos dos valores reais, no período considerado. Porém a abordagem bayesiana é estendida para modelos mais realistas, onde temos incertezas sobre as variâncias dos ruídos.

Nesses modelos mais gerais, os erros de previsão crescem como era de se esperar; mesmo assim, os modelos mostram-se mais adequados para previsão de séries não-estacionárias.

### 6.2 Proposta Futura

Este trabalho pode ser aplicado para dados em outras situações que envolvam modelos com parâmetros aleatórios, variantes no tempo; esses modelos encontram aplicação em áreas como economia, engenharia.

As técnicas utilizadas aqui podem ser aplicadas, também, considerando-se as extensões dos modelos. O modelo hierárquico pode ser estendido considerando a variância do ruído ( $\gamma^2$ ) desconhecida, ou seja:

$$\begin{aligned}
 y_t &= \theta_t y_{t-1} + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \tau^2) \quad \tau^2 \text{ desconhecido} \\
 \theta_t &= \lambda + \omega_t \quad ; \quad \omega_t \sim N(0, \gamma^2) \quad \gamma^2 \text{ desconhecido} \\
 \lambda &\sim N(m, S^2) \quad ; \quad m \text{ e } S^2 \text{ conhecidos}
 \end{aligned}
 \tag{6.1}$$

Pode-se, também, estender o procedimento do Modelo Dinâmico, descrito no capítulo 4, considerando-se desconhecida a variância do ruído ( $\epsilon_t$ ) da equação de observações, isto é:

$$\begin{aligned}
 y_t &= \theta_t y_{t-1} + \epsilon_t \\
 \theta_t &= \lambda \theta_{t-1} + \omega_t,
 \end{aligned}
 \tag{6.2}$$

onde 
$$\begin{pmatrix} \epsilon_t \\ \omega_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} \right], \quad \sigma^2 \text{ e } \gamma^2 \text{ desconhecidos}$$

Outra extensão seria considerar o processo auto-regressivo que modela a variação dos coeficientes aleatórios como um modelo AR(p),  $p \geq 2$ .



## Anexo A

### Cr terio de Converg ncia de Gelman e Rubin

Para verificarmos se as amostras geradas pelos algoritmos amostrador de Gibbs com Metropolis-Hasting est o realmente convergindo para uma distribui o estacion ria, utilizamos o algoritmo baseado na t cnica proposta por Gelman e Rubin (1992), o qual considera pelo menos 2 cadeias paralelas com valores iniciais amostrados de uma distribui o bem comportada. Ap s as cadeias atingirem estacionariedade, digamos na  $n$ - sima itera o, consideramos, ent o, uma sele o  $(\theta_j, \theta_{j+h}, \dots, \theta_{j+Nh})$ , como uma amostra aleat ria da distribui o desejada. Devemos assumir  $h$  razoavelmente grande, de tal forma que dois valores sucessivos de  $\theta_t$  sejam independentes e, assim, teremos uma amostra independente identicamente distribu da (iid).

A converg ncia da simula o   monitorada estimando o fator em que a escala da distribui o deva ser reduzida se as simula es s o feitas at  o limite quando o n mero de itera es tendem para o infinito ( $N \rightarrow \infty$ ).

Se cada seq ncia tem comprimento  $M + N$ , descartamos as  $M$  primeiras amostras e consideramos as  $N$   ltimas, da  calculamos:

$$U = \frac{\sum_{i=1}^k (\bar{\theta}_i - \bar{\theta}_{..})^2}{k - 1}$$

onde  $\bar{\theta}_i$  s o as  $k$  m dias baseadas nas  $N$   ltimas itera es da seq ncia.

$$W = \frac{\sum_{i=1}^k S_i^2}{k}$$

onde

$$S_i^2 = \frac{\sum_{i=1}^k (\bar{\theta}_i - \bar{\theta}_{..})^2}{N - 1}$$

Podemos observar que  $U$  representa a variabilidade entre as  $k$  seq ncias e  $W$  representa a m dia das  $k$  vari ncias dentro das seq ncias. Portanto, a m dia e a vari ncia da distribui o estacion ria podem ser estimadas por:

$$\hat{\mu} = \hat{\theta}_{..} \quad ; \quad \sigma^2 = \frac{N-1}{N}W + \frac{1}{N}U$$

Portanto, temos que a distribuição de  $\theta_i$  condicionada aos dados tem aproximadamente uma distribuição t-Student com média  $\hat{\mu}$  e desvio padrão dado por :

$$S = \sqrt{\hat{\nu}} = \sqrt{\sigma^2 + \frac{U}{kN}}$$

e graus de liberdade ( $g.l$ )

$$g.l = \frac{2\hat{\nu}^2}{Var(\hat{\nu})}$$

onde

$$Var(\hat{\nu}) = \left(\frac{N-1}{N}\right)^2 \left(\frac{1}{k}\right) Var(\sigma^2) + \left(\frac{k+1}{kN}\right)^2 + \left(\frac{2}{k-1}\right) U^2 +$$

$$\left(\frac{2(k+1)(N-1)}{kN}\right) \frac{N}{k} \left[ cov(\sigma^2, \bar{\theta}_i^2) - 2\bar{\theta}_i \cdot cov(\sigma^2, \bar{\theta}_i) \right]$$

Com as variâncias e as covariâncias estimadas, podemos estimar o fator de redução de escala potencial por:

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{\hat{\nu}}{\bar{W}}\right) \left(\frac{g.l+3}{g.l+1}\right)}$$

Esperamos que esse valor convirja à medida que o número de iterações tende para o infinito ( $N \rightarrow \infty$ ); caso não ocorra, devemos considerar mais simulações para melhorar a convergência sobre a distribuição de interesse, ou seja, quando  $\sqrt{\hat{R}} \approx 1$ , Gelman e Rubin (1992) argumentam que as amostras selecionadas convergem em distribuição para a distribuição condicionada aos dados, e os pontos amostrais são iid.

## Anexo B - Programa I- Matlab

### Algoritmo Gibbs para o Modelo Hierárquico com Variância Conhecida

```
%      Leitura dos dados
clear
load ibovespa.m;
dados=ibovespa;
y=dados(:,2);
T=length(y);
T0=T-30;
%      Condições iniciais
m=1;
S=0.5;
g=0.65;
tau=0.33;
mt(1)=m/S^2;
s(1)=1/S^2;
r(1)=tau^2;
    s(2)=(1/S^2+y(1)^2/( g^2*y(1)^2+tau^2))^(-1);
    mt(2)=(m/S^2+y(2)*y(1)/((g^2)*(y(1)^2)+tau^2))*s(2);
    r(2)=g^2*y(1)^2+tau^2;
for j=3:T
    r(j)=g^2*y(j-1)^2 + tau^2;
    s(j)=(s(j-1)*r(j))/(s(j-1)*y(j-1)^2+r(j));
mt(j)=(s(j-1)*y(j)*y(j-1)+mt(j-1)*r(j))/(s(j-1)y(j-1)^2+r(j) );
    Eth(j)=mt(j-1);
    Vth(j)=(s(j-1)+g^2);
    Ey(j)=mt(j-1)*y(j-1);
    Vy(j)=s(j)*y(j-1)^2 + (g^2*y(j-1)^2 + tau^2);
end
    Eyk(1:T0)=y(1:T0);
for k=1:(T-T0)
```

```

Eyk(T0+k)=mt(T0)*Eyk(T0+k-1)+(k-1)*s(T0)*Eyk(T0+k-2);
end
%           Intervalo de credibilidade
eth=1.96*sqrt(Vth);
ey=1.96*sqrt(Vy);
disp('Valores de y, Eyk(T0+k)');
disp([y(T0:T), Eyk(T0:T)' ] )
%           Gráficos
x=[1:1:T];
whitebg('w')
plot(y)
print -depsc fig1
figure(2)
plot(x,y,'r',x(T0:T),Ey(T0:T),'b')
%title('Valores Reais e Valores Previstos das Ações y')
legend('V.Real','V.Previsto')
print -depsc fig2
figure(3)
plot(x(T0:T),y(T0:T),'r',x(T0:T),Ey(T0:T),'b')
%title('Valores Das Ações Previstos Um Dia')
legend('V.Real','V.Previsto')
print -depsc fig3
figure(4)
plot(x(T0:T),y(T0:T),'r',x(T0:T),Eyk(T0:T),'b')
title('Valores Das Ações Previstos para 30 Dias')
legend('V.Real','V.Previsto')
print -depsc fig4

```

## Anexo C - Programa II - Matlab

### Algoritmo Gibbs com Metropolis-Hasting para o Modelo Hierárquico com Variância Desconhecida

```
clear
% Leitura do arquivo de dados para execução do programa
load ibovespa.m;
YY=ibovespa(:,2);
% Entrada dos dados (condições iniciais)
kk=30;
for k=1:kk
    clear y
    T = 652-31+k;
    y = YY(1:T);
    m = 5 ;           % n ° de cadeias
    t = 3000;        % n° de interações p/ cadeia
    M=1;
    S= 0.5;
    a=50;
    b=10;
    g=1;
    tau=linspace(0.1,0.3,m); % tau inicial
    lambi=linspace(0.1,0.5,m); % lambda inicial
% valor inicial da j-ésima cadeia
for j=1:m
    tau(1,j)=tau(j);
    lamb(1,j)=lambi(j);
    c(j)=0;
    for i=2:t
% gerando o valor para lambda
        rtv=g^2*y(1:T-1).^2+tau(i-1,j)^2;
        Stv= (sum (y(1:T-1).^2./rtv )+1/S^2 )^(-1);
        Mtv=(m/S^2+sum((y(2:T).*y(1:T-1))./rtv))*Stv;
```

```

        lamb(i,j)=normrnd(Mtv, (Stv)^(1/2));
%           gerando valor para tau
AB=gamrnd(a,1/b);
tau(i,j)=1/AB;
rtn=g^2* y(1:T-1).^2+tau(i,j)^2 ;
Stn= ( sum ( y(1:T-1).^2./rtn )+1/S^2 )^(-1);
Mtn= (m/S^2+sum((y(2:T).*y(1:T-1))./rtn))*Stn;
% Cálculo da probabilidade de Aceitação
F1 = (-1/2*Stn)*(lamb(i,j) - Mtn)^2;
F2 = (-1/2*sum( (y(1:T-1).^2./rtn).*( y(2:T)./
y(1:T-1)-m ).^2));
r1 = F1+F2;
F3 = (-1/2*Stv)*(lamb(i,j) - Mtv)^2;
F4 = (-1/2*sum( (y(1:T-1).^2./rtv).*( y(2:T)
./y(1:T-1) - m ).^2));
r2 = F3+F4;

Vold = tau(i-1,j);      % Observação anterior
Vnew = AB;              % Observação gerada
prior1 = gampdf(Vnew,a,1/b);
prior2 = gampdf(Vold,a,1/b);
r = exp(r1-r2) * (rtv(T-1)^0.5/rtn(T-1)^0.5);
f = min(1, (prior1/prior2)*r);
%           gerando u de Uniforme
u=rand;
    if u > f,
        tau(i,j) = tau(i-1,j);
        c(j) = c(j)+1;          % rejeição
    end
end
end
end
IR=mean(c/t);

```

```

% Critério de convergência de Gelman-Rubin
% desprezar 50% e selecionar de s em s
s=5;
taus=tau(t/2+1:s:t,:);
lambs=lamb(t/2+1:s:t,:);
R_tau=gr(taus)
R_lamb=gr(lambs)
Mlambs5=mean(lambs);           % média das cadeias
Mtaus5=mean(taus);

    % Estimando o valor de theta
rt=(g* y(T-1))^2+taus(:).^2 ;
thc(k)=1/length(taus)*(sum( (lambs(:).*taus(:)
.^2+g^2*y(T)*y(T-1) ) ./rt) );
V(k)=1/length(taus)*sum( (lambs(:).*taus(:)
.^2+g^2*y(T)*y(T-1) ) ./rt).^2-thc(k)^2;
sigma(k)=1/length(taus)*sum( taus(:).^2*g^2./rt)+V(k);
%
    % Estatísticas
% Valor esperado e desvio padrão de lambda
lambs=[lambs(:)];
Mlamb=mean(lambs);
Slamb=std(lambs);
Linflamb=prctile(lambs,2.5);
Lsuplamb=prctile(lambs,97.5);
% Valor esperado e desvio padrão de tau
taus=[taus(:)];
Mtau=mean(taus);
Stau=std(taus);
Linftau=prctile(taus,2.5);
Lsuptau=prctile(taus,97.5);
% Valor estimado para média e variância de theta;
ths=[thc(:)];
Mths=mean(ths);           % p/ c\{a}lculo IC

```

```

Mths;
Vths=var(th);
Sth=[std(th)];
%           Estimação das preditivas
Etheta(k)=Mlamb;
Vtheta(k)=var(Mlamb)+g^2;
Eyt(k)=mean(Mlamb)*y(T); % previsão para um passo a frente
end           %final do loop para kk=30;
disp('Valores de y   Etheta(T+1)   Eyt(T+1)');
disp([YY(623:622+k),   Etheta',   Eyt' ])
%           Gráficos
x=[1:1:622+k];
whitebg('w')
plot(taus(:))
%title('Gráfico de Convergência de tau')
xlabel ('tau')
print -depsc fig5
figure(6)
hist(taus)
%title('Histograma de tau selecionados')
xlabel('taus')
ylabel('frequência')
print -depsc fig6
figure(7)
plot(lambs(:))
%title('Gráfico de Convergência de lambda')
xlabel('lambda')
ylabel('frequência de lambda')
print -depsc fig7
figure(8)
hist(lambs)
title('Histograma de lamb selecionados')

```



```

xlabel('lambs'),
ylabel('frequência')
print -depsc fig8
figure(9)
plot(taus,lambs)
%title('convergencia (lambda, tau)')
xlabel('lambda')
ylabel('tau')
print -depsc fig9
figure(10)
plot(YY)
%title('Preço De Ações Da Vale do Rio Doce')
xlabel('t')
ylabel('yt- valores em dólares')
print -depsc fig10
figure(11)
plot(x(623:622+k),YY(623:622+k),'r',x(623:622+k),Eyt,'b')
%title('Valores Reais e Valores Previstos das Ações y')
legend('V.Real','V.Previsto')
xlabel('t')
ylabel('Ey - em dólares')
print -depsc fig11

```

## Anexo D - Programa III - Matlab

### Algoritmo de Gibbs com Metropolis-Hanstings para o Modelo Dinâmico - Variância Conhecida

```
clear
% Leitura do arquivo de dados para execução do programa
load boi.m;
dados=boi;
y=dados(:,2);
T0=10;
T=length(y)-T0;
% Entrada dos dados (condições iniciais)
    m=5 ;           % n ° de cadeias
    ni=3000;       % n° de interações p/ cadeia
% geração dos thetas estimados até t-1
y0= 10;
M=1;
S=0.3;
g=0.12;
sig=2.5;
r_i=linspace(0,0.3,m);
thc_i=linspace(0.5,1.5,m);
lamb_i=linspace(2,5,m);
c=0;
% valor inicial da j-enésima cadeia
for j=1:m
    r(1,j)=r_i(j);
    thc(1,j)=thc_i(j);
    lamb(1,j)=lamb_i(j);
    for i=2:ni
        si(1,j)=r(1,j)*sig^2/(sig^2+y(1)^2*r(1,j));
    for t= 2:T
        r(t,j)=lamb(i-1,j)^2*si(t-1,j)+g^2;
```

```

    si(t,j)=r(t,j)*sig^2/(sig^2+y(t-1)^2*r(t,j));
    thc(t,j)=lamb(i-1,j)*thc(t-1,j)+(r(t,j)*...
    y(t-1)/(sig^2+y(t-1)^2*r(t,j)))*(y(t)-y(t-1)*...
    lamb(i-1,j)*thc(t-1,j));
end
    Sigma(t,j)=(r(t,j)*sig^2)/(sig^2+y(T-1)^2*r(t,j));
%   geração de theta(t) para gerar lambda
    th(i,j)= normrnd(thc(t,j),Sigma(t,j)^(1/2));
%           gerando valor para lambda
for t=2:T
    l1v(t-1)=(y(t-1)^2*r(t,j)+sig^2)^(-1/2);
    l2v(t-1)=exp((-1/2*(y(t)^2*r(t,j)+sig^2))*...
    (y(t)-lamb(i-1,j)*thc(t-1,j)*y(t-1))^2);
    lv(t-1)=l1v(t-1)*l2v(t-1);
end
    lvv=prod(lv);
    lamb(i,j)=normrnd(M,S);
for t=2:T
    rn(t,j)=lamb(i,j)^2*si(t-1,j)+g^2;
    sin(t,j)=rn(t,j)*sig^2/(sig^2+y(t-1)^2*rn(t,j));
    l1n(t-1)=(y(t-1)^2*rn(t,j)+sig^2)^(-1/2);
    l2n(t-1)=exp((-1/2*(y(t)^2*rn(t,j)+sig^2))*...
    (y(t)-lamb(i,j)*thc(t-1,j)*y(t-1))^2);
    ln(t-1)=l1n(t-1)*l2n(t-1);
end
    lnn=prod(ln);
%   Cálculo da probabilidade de aceitação
    if(lnn~=0)
        L=lvv/lnn;
    else
        L=1;
    end
end

```

```

    f= min(1,L);
%           gerando u de Uniforme
    u=rand;
    if u > f
        lamb(i,j)= lamb(i-1,j);
        c = c+1;           % rejeição
    end
end
    IR(j)=c/ni;
end
%           Critério de convergência de Gelman-Rubin
%           desprezar 50% e selecionar de s em s
    s=5;
    ths=th(ni/2+1:s:ni,:);
    lambs=lamb(ni/2+1:s:ni,:);
    R_th=CC_GR(ths)
    R_lamb=CC_GR(lambs)
    Mlambs5=mean(lambs);   % média das cadeias
    Mths5=mean(ths);
%           Estatísticas
%           Valor esperado e desvio padrão de lambda
    lambs=[lambs(:)];
    Mlamb=mean(lambs);
    Slamb=std(lambs);
    Linflamb=prctile(lambs,2.5);
    Lsuplamb=prctile(lambs,97.5);
%           Valor esperado e desvio padrão de theta
    thcs=[ths(:)];
    Mthc=mean(thcs);
    Sthc=std(thcs);
    Linfthc=prctile(thcs,2.5);
    Lsupthc=prctile(thcs,97.5);

```

```

%                Cálculo das preditivas
    Eth=mean(lambs*thc(T));
    Vth=rn(t,j)*sig^2/sig^2+y(T-1)^2*rn(t,j);
    Ey=mean(lambs*y(T)*thc(T));
disp('T    y    Eth    Ey');
disp([ T,    y(T),    Eth' ,    Ey' ])
%                Gráficos
whitebg('w')
plot(lambs(:))
%title('Gráfico de Convergência de lambda')
xlabel('lambda')
print -depsc fig30
figure(3)
plot(thc(:))
%title('Gráfico de Convergência de theta')
xlabel('theta')
print -depsc fig31
figure(4)
plot(lambs(:),thc(:))
%title('convergencia (lambda, theta)')
xlabel('lambda')
ylabel('theta')
print -depsc fig32
figure(5)
hist(lambs(:))
%title('Histograma de lambda selecionados')
xlabel('lambda'),
%ylabel('frequência')
print -depsc fig33
figure(6)
hist(thc(:))
%title('Histograma de theta selecionados')

```

```

xlabel('theta'),
ylabel('frequência')
print -depsc fig34
figure(7)
plot(y)
%title('Preço da Arroba de Boi ')
print -depsc fig35
T2= [121 122 123 124 125 126 127 128 129 130 131];
y2= [24.1100 23.9500 24.2500 24.1000 23.0800 ...
     23.3800 24.6800 23.9000 25.4000 23.5600 24.3000];
Ey2=[23.9665 22.4461 25.4284 22.0683 23.0676 ...
     22.9676 25.4727 23.5183 25.8195 22.9550 25.0852];
figure(8)
plot(T2,y2,'r',T2,Ey2,'b')
%title('Valores Reais e Previstos da Arroba de Bois para 10 Dias')
xlabel('t')
ylabel(' valores em dolares')
%legend('V.Real','V.Previsto')
print -depsc fig36

```

# Referências Bibliográficas

- [1] Box,G.E.P.; Jenkins,G.M.; Reinsel,G. *Time Series Analysis:Forecasting and Control.*, 3<sup>th</sup> edition, Holden-Day, 1994.
- [2] Broemeling,D.L.; Cook,P. *Bayesian estimation of the mean of an autoregressive process.* Journal of Applied Statistics, vol.20, N° 1, 1993
- [3] Nicholls,D.f.; Quinn,B.G. *The Estimation of Random coefficient Auto-regressive Models I*, J. Times Series Anal.1, p.37 – 46, 1980
- [4] Liu,S.I. *Comparison of Forecastes for ARMA Models Between a Random Coeficent Approach and a Bayesian Approach.* Commu.Statist Theory Meth.24(2), p.319-333, 1995.
- [5] Soyer,R. *Random Coefficient Autoregressive Process and Their Ramifications: Applications to Reliability Growth Assessment.* D.Sc. Dissertation, School Eng. Appl. Sci., George Washington Univ., Washington D.C., 1985
- [6] Diaz,J. *Bayesian Forecasting for AR(1) Models With Normal Coefficients.* Commu. Statst.-Theory Meth.,196, p.2229 – 2246, 1990.
- [7] Singpurwalla,M.D.; Soyer,R. *Assessing (Software) Realiability Growth Using a Random Coefficient Autoregressive Process and Its Ramifications.* IEEE Transactions on Software Engineering, v.seII, n°12, December,1985.
- [8] Guyton,D.A.; Zhang,N.; Foutz,R.V. *A Random Parameter Process for Modeling and Forecastting Time Series*, Journal of Time Series Analysis, v.7, n°2, p.105 - 115, 1986.
- [9] Meinhold,R.J.; Singpurwalla,N.D. *Understanging the Kalman Filter*, The American Statistician. vol.37, n°2, p.229 – 246, 1983.

- [10] Morris,C. *Parametric Empirical Bayes Inference: Teory and Applications*. Journal of American Statistical Association, vol 78, p. 47-65, 1983.
- [11] Anderson,T.W. *Repeated Measurements on Autoregressive Process*. Journal of American Atatistical Association, vol.73, p 371 – 378, 1978.
- [12] Lindely,D.V. *Aproximate Bayesian Methods*. Trabajos Estatistica, Vol. 31, p.223 - 237, 1980.
- [13] Casella,G.; George,E.J. *Explaining the Gibbs Sampler*. Amer.Statis. Assoc., vol. 46, N°3, p.167 – 174, 1992.
- [14] Box,G.E.; Tiao,G.C. *Bayesian Inference in Statistical Analysis*. New York: Addison Wesley, p.74, 1973
- [15] Gamerman,D. *Simulação Estocástica Via Cadeia de Markov*. ABE, 1996.
- [16] Chib,S.; Greenberg,E.J. *Understanding the Metropolis-Hasting Algorithm*. Amer. Statis. Assoc., vol. 49, N°4, 1995.
- [17] Johnson,R.A.; Wechern,D.W. *Applied Multivariate Statistical Analysis*. 3<sup>th</sup> Edition, Prentice Hall, 1992.
- [18] Gelmam,A.E.; Rubin,D. *Inference From Iterative Using Multiple Sequences*. Statistical Science,7, p.457 – 511, 1992.