



UFRJ

MODELOS BAYESIANOS DE LONGA  
DEPENDÊNCIA COM ERROS  
HIPERBÓLICOS GENERALIZADOS

Ralph dos Santos Silva

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

Orientador: Helio S. Migon

Rio de Janeiro

Março de 2003

# MODELOS BAYESIANOS DE LONGA DEPENDÊNCIA COM ERROS HIPERBÓLICOS GENERALIZADOS

Ralph dos Santos Silva

Orientador: Prof. Helio S. Migon

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

Aprovada por :

---

Presidente, Prof. Helio S. Migon

---

Prof. Marco Antonio R. Ferreira

---

Prof<sup>a</sup>. Ela M. M. de Toscano

Rio de Janeiro

Março de 2003

Silva, Ralph dos Santos

Modelos bayesianos de longa dependência com erros hiperbólicos/ Ralph dos Santos Silva.- Rio de Janeiro: UFRJ/IM, 2003.

ix, 79f.: il.; 31cm.

Orientador: Helio S. Migon

Dissertação (mestrado) - UFRJ/IM/ Programa de Pós-graduação em Estatística, 2003.

Referências Bibliográficas: f.62-64.

1. Longa dependência. 2. Erros hiperbólicos generalizados.  
3. MCMC. 4. Teste de longa dependência I. Migon, Helio S. II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. III. Título.

## Resumo

# MODELOS BAYESIANOS DE LONGA DEPENDÊNCIA COM ERROS HIPERBÓLICOS GENERALIZADOS

Ralph dos Santos Silva

Orientador: Prof. Helio S. Migon

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

Nos últimos anos tem crescido o interesse em modelos estatísticos de séries temporais que possuam a propriedade de longa dependência e quais as consequências de se omitir tal propriedade. Nesta dissertação apresenta-se um estudo sob o enfoque Bayesiano dos processos ARFIMA(0,d,0) utilizando a função de verossimilhança aproximada considerando que a distribuição dos erros do processo gerador das séries temporais possam ser hiperbólicos generalizados e como casos particulares erros normais e t-Student, além de considerar uma classe de prioris que permitam testar a presença de longa dependência. Utilizando os métodos MCMC, as principais características da distribuição a posteriori dos parâmetros de interesse são obtidas através de uma amostra desta distribuição. Duas aplicações são tratadas: uma sobre o nível mínimo do rio Nilo e outra sobre os índices diários de mercado do Chile calculados pela MSCI.

Palavras-chave: longa dependência, erros hiperbólicos generalizados e MCMC.

# Abstract

## BAYESIAN MODELS OF LONG DEPENDENCE WITH GENERALIZED HYPERBOLIC ERRORS

Ralph dos Santos Silva

Orientador: Prof. Helio S. Migon

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

Recently, it has been growth the interest on statistical models of time series with long dependence properties. The consequences of not including these components in the model have also been intensively studied. In this dissertation a study, under the Bayesian perspective, of the ARFIMA(0,d,0) process using an approximate likelihood function is presented. The data generation process is assumed to be a generalized hyperbolic distribution, a broad class including, as a particular case, the normal and t-Student distributions. A class of prior distribution that permits to test the presence of long dependence is also presented. The main characteristics of the posterior distribution of the parameter of the model are obtained via MCMC, making all the inference ease. The methodology presented here is applied to two set of data: the Nile River minima and daily indexes of market of the Chile calculated by MSCI.

Key-words: long dependence, generalized hyperbolic errors and MCMC.

## Agradecimentos

Em especial ao professor e meu orientador Helio S. Migon;

Ao suporte dado pelo CNPq;

Ao professor Hedibert Lopes e

Aos amigos Luis Medrano, Mariane Alves e Rafael Martins

por colaborarem com o desenvolvimento deste trabalho.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Os modelos ARFIMA(<math>p, d, q</math>)</b>	<b>5</b>
2.1	Definição . . . . .	5
2.2	O operador de diferença fracionária . . . . .	6
2.3	A função de verossimilhança . . . . .	6
2.4	O processo ARFIMA( $0, d, 0$ ) . . . . .	7
2.5	O processo ARFIMA( $p, d, q$ ) . . . . .	8
2.5.1	O processo auto-regressivo . . . . .	11
2.5.2	O processo médias-móveis . . . . .	11
<b>3</b>	<b>Análise Bayesiana do processo ARFIMA(<math>0, d, 0</math>)</b>	<b>18</b>
3.1	Modelo ARFIMA( $0, d, 0$ ) com erros normais . . . . .	19
3.1.1	Função de verossimilhança . . . . .	19
3.1.2	Distribuições a priori . . . . .	20
3.1.3	Distribuição a posteriori . . . . .	21
3.1.4	Condicionais completas . . . . .	21
3.1.5	Simulação . . . . .	23
3.1.6	Comentários . . . . .	27
3.2	Modelo ARFIMA( $0, d, 0$ ) com erros t-Student . . . . .	29
3.2.1	Função de verossimilhança . . . . .	30

3.2.2	Distribuições a priori . . . . .	31
3.2.3	Distribuição a posteriori . . . . .	32
3.2.4	Condicionais completas . . . . .	33
3.2.5	Comentários . . . . .	36
3.3	Modelo ARFIMA(0,d,0) com erros hiperbólicos generalizados . . .	36
3.3.1	Função de verossimilhança . . . . .	37
3.3.2	Distribuições a priori . . . . .	38
3.3.3	Distribuição a posteriori . . . . .	39
3.3.4	Condicionais completas . . . . .	40
3.3.5	Comentários . . . . .	43
<b>4</b>	<b>Identificando a longa dependência</b>	<b>44</b>
4.1	Modelo ARFIMA(0,d,0) com erros normais . . . . .	44
4.1.1	Função de verossimilhança . . . . .	44
4.1.2	Distribuições a priori . . . . .	45
4.1.3	Distribuição a posteriori . . . . .	45
4.1.4	Condicionais completas . . . . .	46
4.1.5	Comentários . . . . .	46
<b>5</b>	<b>Aspectos computacionais</b>	<b>47</b>
5.1	Amostrador de Gibbs . . . . .	47
5.2	Algoritmo de Metropolis-Hastings . . . . .	48
5.2.1	Proposta uniforme não simétrica . . . . .	49
5.2.2	Proposta Normal-Logito . . . . .	49
5.3	Amostrador da fatia . . . . .	50
5.4	Amostrador para mistura de distribuições . . . . .	52
<b>6</b>	<b>Aplicação</b>	<b>54</b>
6.1	Nível mínimo do rio Nilo . . . . .	54



6.2	Índice de preço do Chile . . . . .	57
<b>7</b>	<b>Conclusões e extensões</b>	<b>60</b>
	<b>Referências Bibliográficas</b>	<b>62</b>
<b>A</b>	<b>Distribuições Generalizadas</b>	<b>65</b>
A.1	Gaussiana Inversa Generalizada (GIG) . . . . .	65
A.2	Hiperbólica Generalizada (HG) . . . . .	67
<b>B</b>	<b>Função de Bessel do Terceiro Tipo</b>	<b>69</b>
<b>C</b>	<b>Distribuições de Mistura</b>	<b>71</b>
C.1	Distribuição t-Student como uma mistura Normal-Gamma Inversa	71
C.2	Distribuição Hiperbólica Generalizada como uma mistura Normal-Gaussiana Inversa Generalizada . . . . .	72
<b>D</b>	<b>Critério de informação do desvio - DIC</b>	<b>74</b>
<b>E</b>	<b>Resultados de simulação</b>	<b>75</b>

# Capítulo 1

## Introdução

As séries temporais estacionárias com memória curta ou memória longa caracterizada pelos modelos auto-regressivos fracionalmente integrados médias-móveis (ARFIMA) têm sido estudada a pelo menos duas décadas (Granger, 1980; Hosking, 1981) e nos últimos anos despertou-se um maior interesse em se trabalhar com estes modelos (Sowell, 1992; Pai & Ravishanker, 1996; Koop, Ley, Osiewalski, & Steel, 1997; Pai & Ravishanker, 1998; Chan & Palma, 1998), os quais podem ser encontrado em várias aplicações de diversas áreas como hidrologia, economia, astronomia e ciências da computação. Acreditava-se que observações distantes entre si poderiam ser assumidas como independentes, mas isso não se mostrou adequado. Os modelos ARFIMA foram motivados para captar esta forte dependência existente entre as observações distantes. Granger e Joyeux (1980) foram os pioneiros nos estudos deste modelo, enquanto que Hosking (1981) apresentou o modelo formal pela primeira vez, generalizando o então conhecido modelo  $ARIMA(p, d, q)$  de Box e Jenkins (1976). O processo de longa dependência ou persistência é caracterizado pelo parâmetro  $d \in (-0,5; 0,5)$ . Quando este parâmetro  $d$  é diferente de zero, pode-se então dizer que o processo é de memória curta se  $d \in (-0,5; 0)$  e de memória longa se  $d \in (0; 0,5)$ , caso contrário o processo ARFIMA

passa a ser simplesmente o processo  $ARMA(p, q)$ . Quando o parâmetro  $d$  está fora deste intervalo, basta aplicar sucessivas diferenças na série temporal até que esta fique no intervalo desejado. Em Crato (2001), Beran (1994) e Reisen e Toscano (2000) encontram-se uma boa revisão do processo de longa dependência. O extenso uso destes modelos em trabalhos empíricos têm mostrado sua relevância na busca de procedimentos de estimação que sejam eficientes na inferência frequentista e Bayesiana. O procedimento sugerido em Li e McLeod (1986) é truncar a soma infinita que define a longa dependência e a partir daí utilizar a estimação padrão de séries temporais. De acordo com Sowell (1992) a estimação de máxima verossimilhança parece ser o procedimento mais eficiente no caso de modelos  $ARFIMA(0,d,0)$ , por isso, utilizando a hipótese de normalidade para os erros do processo, ele desenvolveu o método da máxima verossimilhança exata de todos os parâmetros do modelo, mas o esforço computacional, em termos de memória e CPU, são altos neste caso e além disso, segundo alguns autores, as estimativas obtidas podem ser inconsistentes dependendo se os parâmetros têm erros de especificação ou não. Uma das alternativas são os métodos semi-paramétricos, dentre eles o (GPH) devido à Geweke e Porter-Hudak (1983). Os métodos que utilizam o método da regressão através da função periodograma, como o (GPH), são os mais populares. A modelagem foi tratada através do filtro de Kalman por (Chan & Palma, 1998), onde uma aproximação da função de verossimilhança baseada na equação de estado de espaço truncada foi utilizada. Assumindo uma distribuição Beta para os quadrados dos coeficientes AR, Granger (1980) mostrou que a soma de um grande número de processos  $AR(1)$  estacionários com parâmetros aleatórios podem possuir a característica de longa dependência. Em Koop et al. (1997), Pai e Ravishanker (1996, 1998) encontram-se abordagens Bayesianas do assunto. O primeiro destes trabalha com a função impulso e resposta, além de prioris com ponto de massa para identificar a longa

dependência. Nos outros dois, o esforço computacional é grande, pois se trabalha com a verossimilhança exata. A complexidade dos modelos ARFIMA é maior se comparado aos modelos ARMA, devido a estrutura da matriz de covariâncias, portanto a função de verossimilhança e a distribuição a posteriori não possuem uma forma simples comparada aos modelos ARMA.

A proposta desta dissertação é apresentar uma abordagem de estimação Bayesiana dos processos de longa dependência baseado nos modelos ARFIMA(0,  $d$ , 0), seguindo o trabalho de Li e McLeod (1986), incluindo aspectos de seleção de modelo - entenda-se ordem de aproximação - e previsão que são importantes para quem trabalha com séries temporais. Esta abordagem é atrativa porque fornece ao usuário características completas dos parâmetros, baseadas numa amostra gerada da distribuição conjunta a posteriori dos parâmetros do modelo gerada pelos métodos MCMC - abreviação do inglês "Markov chain Monte Carlo".

A dissertação está dividida da seguinte maneira: no Capítulo 2 apresenta-se o modelo ARFIMA( $p, d, q$ ), a expansão do operador de diferença fracionária  $d$  em termos AR e MA, e faz-se uma comparação entre as funções de auto-correlações: exata dada por Hosking (1981), amostral e utilizando aproximações AR e MA, para o processo ARFIMA(0,  $d$ , 0). No Capítulo 3 apresenta-se a análise Bayesiana dos processos ARFIMA(0,  $d$ , 0) sob a hipótese que os erros possam ser normais, t-student e hiperbólicos generalizados. Descreve-se a função de verossimilhança, a distribuição a priori e a posteriori considerando as aproximações em termos AR ou MA. Tem-se também um estudo de simulação quando os erros do processo têm distribuição normal. O Capítulo 4 aborda a possibilidade de não existir longa dependência numa série temporal através da probabilidade a posteriori do parâmetro que determina essa longa dependência. No Capítulo 5 apresenta os métodos computacionais utilizados neste trabalho. Duas aplicações são tratadas no Capítulo 6: o volume mínimo do rio Nilo entre 622-1284 D.C e o índice de

mercado do Chile calculados pela MSCI. E no Capítulo 7 faz-se uma conclusão dos assuntos abordados e as possíveis extensões a serem feitas a partir deste trabalho.

# Capítulo 2

## Os modelos ARFIMA( $p, d, q$ )

### 2.1 Definição

Uma série temporal  $\{y_t\}$  gerada por um processo auto-regressivo fracionalmente integrado médias móveis, ARFIMA( $p, d, q$ ) com média  $\mu$  é descrita por

$$\Phi(B)(1 - B)^d(y_t - \mu) = \Theta(B)\epsilon_t \quad (2.1)$$

onde

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p$$

e

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \dots + \theta_q B^q$$

são polinômios em  $B$ , onde  $B$  é o operador de defasagem ( $B^m y_t = y_{t-m}$ , para  $m = 1, 2, \dots$ ),  $p$  e  $q$  são números inteiros e  $d$  um número real entre  $(-0,5; 0,5)$ , além disso as raízes do polinômio  $\Phi(z) = 0$  e  $\Theta(z) = 0$  devem estar todas fora do círculo unitário para garantir que o processo seja estacionário e inversível.  $\epsilon_t$  são variáveis aleatórias independentes e identicamente distribuídas com média 0 e variância  $\sigma^2$ . Os parâmetros deste modelo são  $\mu, \sigma^2, \Phi$  e  $\Theta$ , onde  $\Phi = (\phi_1, \phi_2, \dots, \phi_p)'$  e  $\Theta = (\theta_1, \theta_2, \dots, \theta_q)'$ .

## 2.2 O operador de diferença fracionária

O operador de diferença fracionária  $d$ ,  $(1 - B)^d$  em (2.1) é definido como a expansão Binomial,

$$(1 - B)^d = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^j (-B)^j = \sum_{j=1}^{\infty} \pi_j(d) B^j \quad (2.2)$$

onde

$$\pi_j(d) = \frac{\Gamma(j - d)}{\Gamma(j + d)\Gamma(-d)}$$

## 2.3 A função de verossimilhança

Suponha que uma série temporal de tamanho  $T$  do processo  $\{y_t\}$  - ARFIMA( $p, d, q$ ) - foi observada, denotando-a como  $\mathbf{Y}_T = (y_1, y_2, \dots, y_T)$ . Além disso, assuma que os erros do processo são Normais, então a função de verossimilhança (Sowell, 1992) é dada por:

$$p(\mathbf{Y}_T | \Psi) = (2\pi)^{-\frac{T}{2}} |\Omega_T|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_T - \mu \mathbf{1}_T)' \Omega_T^{-1} (\mathbf{Y}_T - \mu \mathbf{1}_T) \right\} \quad (2.3)$$

onde  $\Psi = (\Phi, \Theta, d, \mu, \sigma^2)$  é um vetor de parâmetros de dimensão  $(p + q + 3)$ ,  $\mathbf{1}_T = (1, \dots, 1)'_T$  e  $\Omega_T$  é matriz de covariâncias. Essa matriz depende do número de observações da série temporal, isto é, a matriz é de dimensão  $T \times T$ . Assim, fica difícil se trabalhar com verossimilhança exata. Qualquer função de verossimilhança pode ser obtida através de suas distribuições condicionais, em particular tem-se

$$p(\mathbf{Y}_T | \Psi) = p(y_T | \mathbf{Y}_{T-1}, \Psi) \dots p(y_1 | \mathbf{Y}_0, \Psi) \quad (2.4)$$

Essa forma da função de verossimilhança conduz a duas alternativas possíveis de aproxima-las: utilizando a soma infinita em termos AR ou MA, truncando qualquer uma das duas num certo número inteiro  $m \leq T$  (Li & McLeod, 1986),

como será visto a seguir. Esse procedimento permite a avaliação da matriz de covariâncias sem o problema de considerá-la diretamente. As primeiras indagações que surgem estão relacionadas a qualidade das aproximações, qual dos dois processos utilizar, AR ou MA, e qual o valor de  $m$  satisfatório. Esta dissertação procura respostas para estas questões.

## 2.4 O processo ARFIMA(0, $d$ , 0)

Considerando somente o processo ARFIMA(0,  $d$ , 0), Hosking (1981) obteve algumas de suas propriedades, as quais não serão provadas neste trabalho.

**Teorema 2.1** *Seja  $\{y_t\}$  um modelo ARFIMA(0,  $d$ , 0) tal que  $d \in (-0,5; 0,5)$ , então*

**a.1)** *Se  $d > -0,5$ ,  $\{y_t\}$  é inversível e possui representação auto-regressiva (AR) infinita,*

$$(1 - B)^d y_t = \sum_{j=0}^{\infty} \varphi_j(d) y_{t-j} = \epsilon_t \quad (2.5)$$

onde

$$\varphi_0(d) = 1, \varphi_1(d) = -d, \varphi_2(d) = \frac{d(d-1)}{2!}, \varphi_3(d) = -\frac{d(d-1)(d-2)}{3!}, \dots$$

**a.2)** *Se  $d < 0,5$ ,  $\{y_t\}$  é um processo estacionário e tem representação médias-móveis (MA) infinita,*

$$y_t = (1 - B)^{-d} \epsilon_t = \sum_{j=0}^{\infty} \psi_j(d) \epsilon_{t-j} \quad (2.6)$$

onde

$$\psi_0(d) = 1, \psi_1(d) = d, \psi_2(d) = \frac{d(d+1)}{2!}, \psi_3(d) = \frac{d(d+1)(d+2)}{3!}, \dots$$



**b.1)** A função de auto-covariância de ordem  $k$  de  $\{y_t\}$  é

$$\gamma(k) = E(y_t y_{t-k} | d) = \frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!}, \quad k \geq 1 \quad (2.7)$$

**b.2)** A função de auto-correlação de ordem  $k$  de  $\{y_t\}$  é

$$\rho(k) = \frac{(-d)!(k+d-1)!}{(k-d)!(d-1)!}, \quad k = 0, \pm 1, \dots \quad (2.8)$$

ou

$$\rho(k) = \frac{d(1+d) \dots (k-1+d)}{(1-d)(2-d) \dots (k-d)}, \quad k = 1, 2, \dots \quad (2.9)$$

**b.3)** A função de auto-correlação parcial é dado por:

$$\rho_{kk} = \frac{d}{k-d}, \quad k = 1, 2, \dots \quad (2.10)$$

## 2.5 O processo ARFIMA( $p, d, q$ )

Novamente de Hosking (1981) tem-se algumas propriedades para o processo ARFIMA( $p, d, q$ ) em geral, as quais não serão provadas neste trabalho.

**Teorema 2.1** *Seja  $\{y_t\}$  um modelo ARFIMA( $p, d, q$ ) definido em (2.1), então*

**a.1)** *Se  $d > -0,5$ ,  $\{y_t\}$  é inversível e possui representação auto-regressiva (AR) infinita,*

$$\sum_{j=0}^{\infty} \varphi_j(d) y_{t-j} = \epsilon_t \quad (2.11)$$

onde  $\varphi_j$  são os coeficientes de  $B^j$  na expansão

$$\varphi(B) = \frac{\Phi(B)}{\Theta(B)} (1-B)^d.$$

**a.2)** *Se  $d < 0,5$ ,  $\{y_t\}$  é um processo estacionário e tem representação médias-móveis (MA) infinita,*

$$y_t = \sum_{j=0}^{\infty} \psi_j(d) \epsilon_{t-j} \quad (2.12)$$

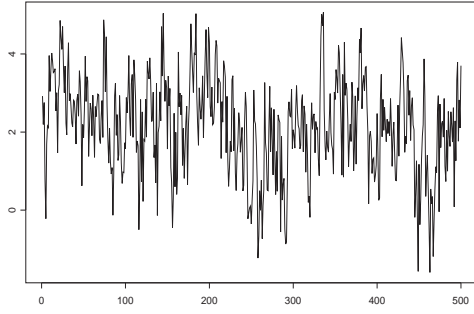
onde  $\psi_j$  são os coeficientes de  $B^j$  na expansão

$$\psi(B) = \frac{\Theta(B)}{\Phi(B)}(1 - B)^{-d}.$$

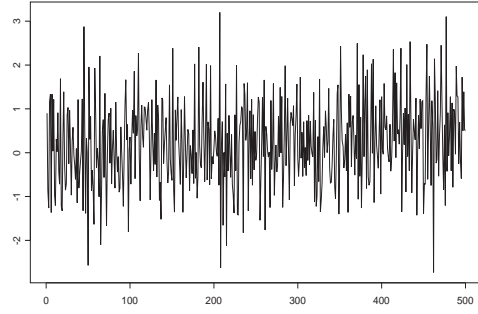
A Figura (2.1) mostra 6 séries simuladas do processo ARFIMA( $p, d, q$ ) utilizando a função `arima.fracdiff.sim` do pacote estatístico S-plus. Os modelos simulados foram

$$\begin{aligned} (1 - B)^{0,45} y_t &= \epsilon_t \\ (1 - 0,1B)(1 - B)^{0,45} y_t &= (1 - 0,8B) \epsilon_t \\ (1 - 0,3B)(1 - B)^{0,45} y_t &= (1 - 0,4B) \epsilon_t \\ (1 - B)^{0,49} y_t &= \epsilon_t \\ (1 - 0,1B)(1 - B)^{0,49} y_t &= (1 - 0,8B) \epsilon_t \\ (1 - 0,3B)(1 - B)^{0,49} y_t &= (1 - 0,4B) \epsilon_t \end{aligned}$$

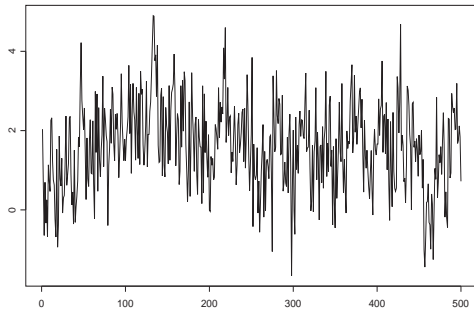
Uma das características marcante do processo ARFIMA é que sua função de auto-correlação decai a uma taxa hiperbólica, isto é,  $\rho(k) \simeq k^{-d}$ , ao contrário das auto-correlações produzidas pelo modelo ARMA( $p, q$ ) (Box & Jenkins, 1976) as quais têm um decaimento exponencial, isto é,  $\rho(k) \simeq c^k$ ,  $0 < c < 1$ . O primeiro estudo conduzido aqui foi no intuito de verificar como se comportava a estrutura de correlação utilizando aproximações do processo ARFIMA em termos AR e MA feitas em relação a correlação exata em (2.9). Para isso, apresenta-se os processos particulares AR( $p$ ) e MA( $q$ ) do ARFIMA( $p, d, q$ ), e suas respectivas funções de auto-correlação.



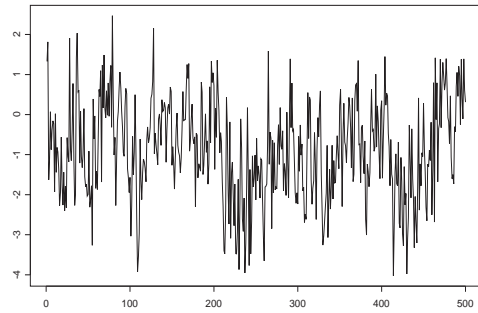
(a) ARFIMA(0,d,0)  $d = 0,45$



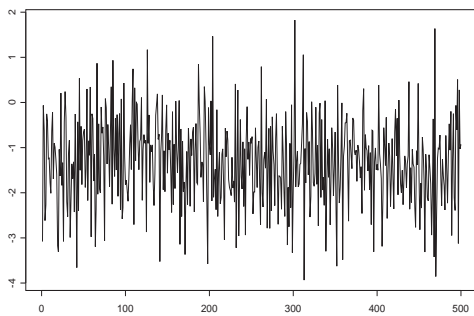
(b) ARFIMA(1,d,1)  $\phi = 0,1; d = 0,45$



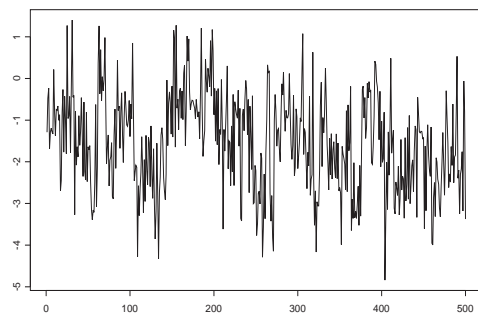
(c) ARFIMA(1,d,1)  $\phi = 0,3; d = 0,45$



(d) ARFIMA(0,d,0)  $d = 0,49$



(e) ARFIMA(1,d,1)  $\phi = 0,1; d = 0,49$



(f) ARFIMA(1,d,1)  $\phi = 0,3; d = 0,49$

Figura 2.1: Séries temporais de tamanho 500 simuladas do processo ARFIMA( $p, d, q$ ) utilizando a rotina `arima.fracdiff.sim` do pacote estatístico S-plus

### 2.5.1 O processo auto-regressivo

Uma série temporal  $\{y_t\}$  que segue um processo auto-regressivo de ordem  $p$ , AR(p), é dado pelo seguinte modelo:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)y_t = \epsilon_t$$

ou

$$(2.13)$$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

onde  $\epsilon_t \sim RB(0, \sigma^2)$ . Sua função de auto-covariância é dada por

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \dots + \phi_p \gamma(k-p), \quad k > 0 \quad (2.14)$$

e a função de auto-correlação é

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \dots + \phi_p \rho(k-p), \quad k > 0 \quad (2.15)$$

### 2.5.2 O processo médias-móveis

Uma série temporal  $\{y_t\}$  que segue um processo médias-móveis de ordem  $q$ , MA(q), é dada pelo seguinte modelo:

$$y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)\epsilon_t$$

ou

$$(2.16)$$

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

onde  $\epsilon_t \sim RB(0, \sigma^2)$  independentes. Sua função de auto-covariância é dada por

$$\gamma(k) = \begin{cases} \sigma^2(\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{q-k} \theta_q), & k = 0, 1, \dots, q \\ 0, & k > q \end{cases} \quad (2.17)$$

e a função de auto-correlação é

$$\rho(k) = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & k = 0, 1, \dots, q \\ 0, & k > q \end{cases} \quad (2.18)$$

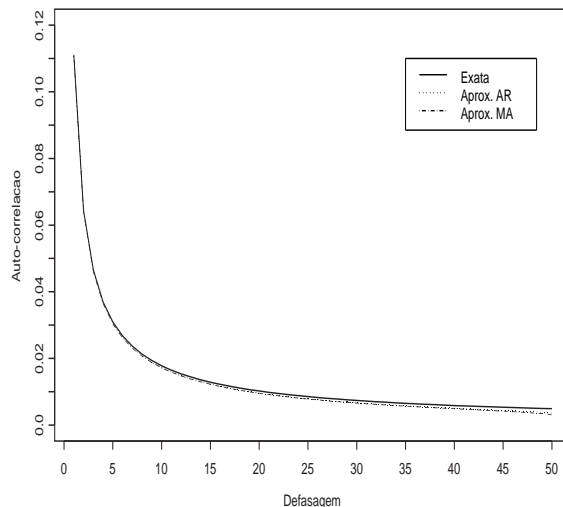


Figura 2.2: Funções de auto-correlação exata, aproximadas AR e MA do processo ARFIMA(0,d,0) para  $d=0,10$

Assim, utilizando-se o ruído fracionário - ARFIMA(0,  $d$ , 0) - calculou-se as correlações: exatas dadas por (2.9) e as aproximadas dadas por (2.15) e (2.18), quando se aproxima o processo ARFIMA(0,  $d$ , 0) em termos AR por (2.5) ou MA em (2.6), respectivamente, como somas finitas de ordem  $k < \infty$ . Foram utilizados três valores diferentes de  $d$ ,  $d = \{0, 1; 0, 45; 0, 49\}$  e a aproximação foi da ordem  $k = 50$ . Na Figura (2.2) exibe as três funções de auto-correlação: exata, AR(50) e MA(50) para  $d = 0, 10$ . Nota-se que as correlações aproximadas pelos processos AR e MA diferem pouco da correlação exata, resultado que não é surpreendente, visto que para este valor -  $d = 0, 10$  - a estrutura de auto-correlação do processo ARFIMA não é tão visível, exibindo apenas correlações menores que 0,12. Para os outros dois valores,  $d = 0, 45$  e  $d = 0, 49$ , as Figuras (2.3) e (2.4), mostram claramente a noção de forte dependência, pois para a defasagem  $k = 50$  a correlação ainda está acima de 0,88. Portanto, devido as características dos processos AR e MA, as aproximações já não conseguem acompanhar a taxa de

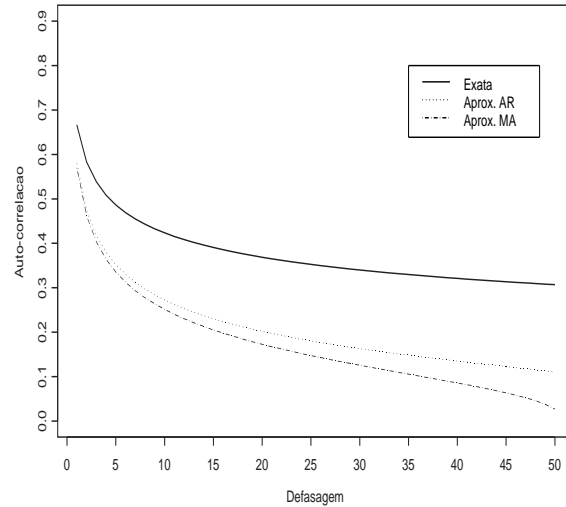


Figura 2.3: Funções de auto-correlação exata, aproximadas AR e MA do processo ARFIMA(0,d,0) para  $d=0,45$

decréscimo da correlação do processo ARFIMA(0,  $d$ , 0), como era de se esperar pois os decaimentos são exponenciais num caso e hiperbólico em outro. Para exemplificar ainda mais, gerou-se um processo ARFIMA(0,d,0) de tamanho 50.000 com valor  $d$  fracionário igual a 0,49, através da Figura 2.5 verifica-se que a auto-correlação amostral também não está próxima da correlação teórica dada por Hosking (1981).

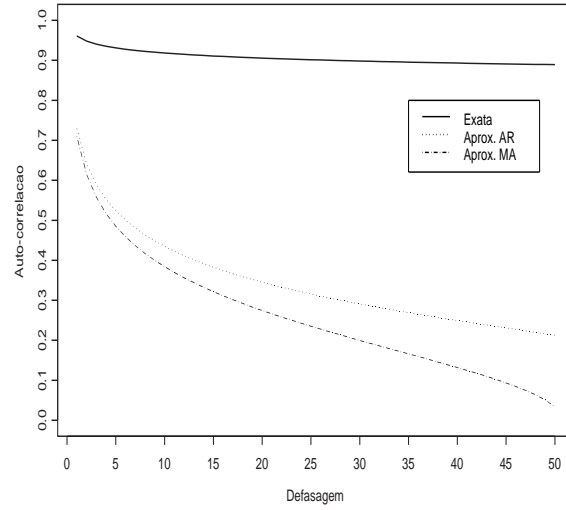


Figura 2.4: Funções de auto-correlação exata, aproximadas AR e MA do processo ARFIMA(0,d,0) para  $d=0,49$

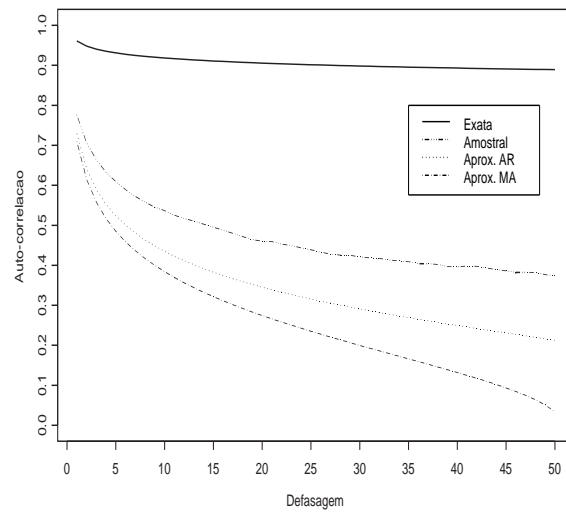


Figura 2.5: Funções de auto-correlação exata, amostral, aproximadas AR e MA do processo ARFIMA(0,d,0) para  $d=0,49$

Para auxiliar e ficar mais claro como as aproximações se desenvolvem em relação as diferenças  $k$ , utilizou-se a função de perda percentual

$$Erro(\%) = \frac{|\rho(k) - \rho_{Ap}(k)|}{\rho(x)} \times 100$$

onde  $\rho_{Ap}(k)$  é a auto-correlação na aproximação. As Figuras (2.6), (2.7) e (2.8) mostram como esses erros aumentam a medida que as observações do processo ARFIMA se tornam mais afastadas. A primeira não necessita maiores comentários, pois já foi visto que para o valor  $d = 0,10$  as correlações na aproximação estão próximas da exata com erros menores que 20%, sendo que o processo AR tem menor erro comparado ao MA. Entretanto, quando  $d = 0,45$  e  $0,49$  esses erros crescem mais rápido, sendo maior que 50% já na diferença  $k = 15$  para ambos os casos. Novamente, o processo AR possui um melhor desempenho se comparado ao MA. Mas essas aproximações, do ponto de vista de suas correlações, não são satisfatórias se o analista for exigente e quiser trabalhar com erros percentuais baixos, por exemplo menores que 10%.

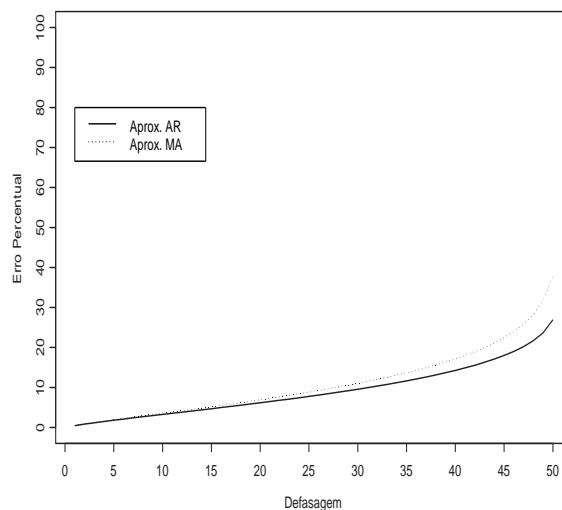


Figura 2.6: Erro percentual das funções de auto-correlação aproximadas AR e MA em relação a exata do processo ARFIMA(0,d,0) para  $d=0,10$



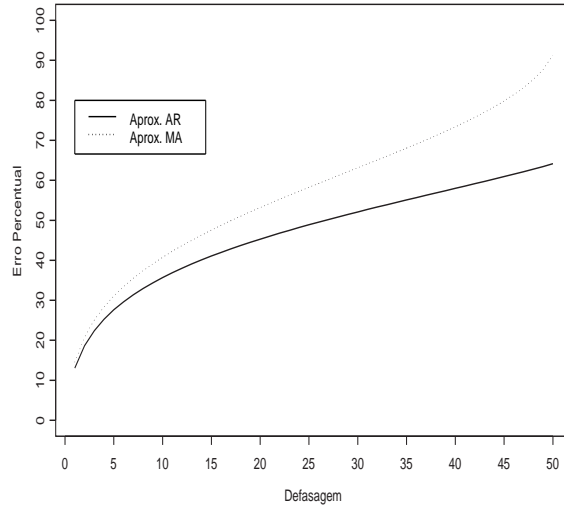


Figura 2.7: Erro percentual das funções de auto-correlação aproximadas AR e MA em relação a exata do processo ARFIMA(0,d,0) para  $d=0,45$

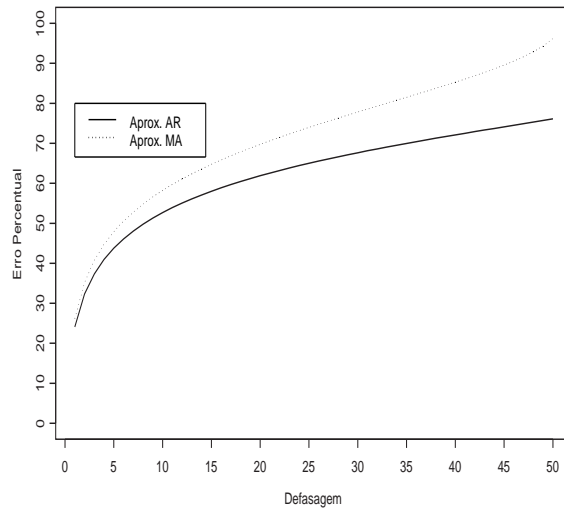


Figura 2.8: Erro percentual das funções de auto-correlação aproximadas AR e MA em relação a exata do processo ARFIMA(0,d,0) para  $d=0,49$

Conclui-se que do ponto de vista das correlações nenhuma das duas expansões - AR e MA - truncadas poderiam gerar bons resultados, ou seja, estimar de forma adequada e precisa os parâmetros do processo ARFIMA( $p, d, q$ ) em geral. Os resultados foram apenas mostrados para o ARFIMA( $0, d, 0$ ), mas o resultado para o modelo considerando que  $p \neq 0$  e  $q \neq 0$  valem, porque o processo não perde a característica de decrescimento hiperbólico da auto-correlação, enquanto que qualquer uma das duas aproximações decrescem mais rápida.

## Capítulo 3

# Análise Bayesiana do processo

## ARFIMA(0,d,0)

A análise desenvolvida neste capítulo para o modelo ARFIMA(0,  $d$ , 0) está baseada na verossimilhança aproximada gerada pelo truncamento de (2.5) e (2.6). Para a análise Bayesiana, cujo interesse é a distribuição conjunta a posteriori de todos os parâmetros, é necessário definir a função de verossimilhança, a distribuição a priori e conseqüentemente as distribuição a posteriori. Entretanto as distribuições a posteriori que serão apresentadas não possuem uma forma fechada conhecida, assim a inferência será através de uma amostra da distribuição conjunta a posteriori gerada pelo amostrador de Gibbs (Gelfand & Smith, 1990), necessitando assim definir as condicionais completas, que em alguns casos também não possuem forma fechada conhecida ou não existe ainda métodos diretos para amostrar e portanto utilizará-se o amostrador da fatia (Neal, 1997) - que se mostrou muito eficiente em estudos preliminares - e o amostrador para mistura de distribuição. O algoritmo de Metropolis-Hasting (Metropolis, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) pode também ser utilizado. Tais algoritmos são descritos no Capítulo 5 e foram todos implementados em linguagem de programação FOR-

TRAN. Neste trabalho a função de perda a ser minimizada é a quadrática e pelos resultados da inferência Bayesiana tem-se as estimativas dadas pela média da distribuição a posteriori.

### 3.1 Modelo ARFIMA(0,d,0) com erros normais

A análise é feita nesta seção sob a hipótese de que os erros  $\epsilon_t$  geradores do processo ARFIMA(0,  $d$ , 0) são normais independentes e identicamente distribuídas com média 0 e variância constante  $\sigma^2$ . Praticamente toda literatura em modelos ARFIMA trabalha sob a hipótese de que os erros do processo são normais, devido a suas características simples. Esse modelo mostrou-se adequado em várias aplicações existentes na literatura até o presente momento, portanto é de fundamental importância que se conheça profundamente suas características do ponto de vista da aproximação dada por Li e McLeod (1986) e do enfoque Bayesiano.

#### 3.1.1 Função de verossimilhança

A função de verossimilhança será definida de acordo com as aproximações de ordem  $m$  utilizadas, considerando que existem  $m$  variáveis auxiliares no processo e que a variância do erro é  $\sigma^2$ .

##### Em termos AR(m)

Considere o vetor de variáveis auxiliares do processo  $Y_0 = (y_0, y_{-1}, \dots, y_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.1)$$

onde

$$\mu_t = \sum_{j=1}^m \varphi_j(d) y_{t-j} \text{ e } \Psi = (d, \sigma^2, Y_0)$$

### Em termos MA(m)

Considere o vetor de variáveis auxiliares do processo  $E_0 = (e_0, e_{-1}, \dots, e_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.2)$$

onde

$$\mu_t = \sum_{j=1}^m \psi_j(d) e_{t-j}, \quad e_t = (y_t - \mu_t) \text{ e } \Psi = (d, \sigma^2, E_0)$$

### 3.1.2 Distribuições a priori

As prioris para os parâmetros do processo ARFIMA(0,  $d$ , 0) são vagas, respeitando as respectivas regiões onde estão definidas. Com isso toda fonte de informação está concentrada na verossimilhança.

#### Em termos AR(m)

- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ,  
onde  $\mathbf{1}_{(-0,5;0,5)}(d)$  é igual a 1 se  $d \in (-0,5; 0,5)$  e 0 caso contrário;
- $(Y_0 | \sigma^2) \sim N_m(\underline{0}; \sigma^2 I_m)$ ; e
- $(\sigma^2) \sim GI(\alpha, \beta)$ , ou seja,  $p(\sigma^2) = \frac{(\sigma^2)^{-\alpha-1} \beta^\alpha}{\Gamma(\alpha)} \exp \left[ -\frac{\beta}{\sigma^2} \right]$

### Em termos MA(m)

- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ,  
onde  $\mathbf{1}_{(-0,5;0,5)}(d)$  é igual a 1 se  $d \in (-0,5; 0,5)$  e 0 caso contrário;
- $(E_0 | \sigma^2) \sim N_m(\mathbf{0}; \sigma^2 I_m)$ ; e
- $(\sigma^2) \sim GI(\alpha, \beta)$ , ou seja,  $p(\sigma^2) = \frac{(\sigma^2)^{-\alpha-1} \beta^\alpha}{\Gamma(\alpha)} \exp\left[-\frac{\beta}{\sigma^2}\right]$

Outras prioris também são permitidas, assim como as prioris informativas, como por exemplo a priori (Beta( $\alpha, \beta$ ) – 0.5) para o parâmetro  $d$ .

### 3.1.3 Distribuição a posteriori

A distribuição a posteriori é proporcional ao produto da função de verossimilhança pela a distribuição a priori, ou seja,

$$p(\Psi | Y_T) \propto p(Y_T | \Psi)p(\Psi)$$

### Em termos AR(m)

$$p(\Psi | Y_T) \propto (\sigma^2)^{-\frac{T+m}{2}-\alpha-1} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2\right) - \frac{\beta}{\sigma^2}\right] \quad (3.3)$$

### Em termos MA(m)

$$p(\Psi | Y_T) \propto (\sigma^2)^{-\frac{T+m}{2}-\alpha-1} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2\right) - \frac{\beta}{\sigma^2}\right] \quad (3.4)$$

### 3.1.4 Condicionais completas

A partir de um vetor de parâmetros  $\Psi$  a posteriori, define-se a condicional completa de um sub-vetor paramétrico genérico  $\theta$  como a distribuição deste dado todos os outros parâmetros e os dados, que será denotado por  $p(\theta | (\Psi/\theta), Y_T)$ .

**Em termos AR(m)**

$$p(d \mid (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.5)$$

$$(y_j \mid (\Psi/y_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1 - m), (2 - m), \dots, 0 \quad (3.6)$$

$$\text{onde } \sigma_j^2 = \frac{\sigma^2}{\left(1 + \sum_{i=j:-1}^{1-m} [\varphi_{1-i}(d)]^2\right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\sigma^2} \sum_{i=1:-1}^{1-m} \varphi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \varphi_k(d) y_{1-i-k} \right)$$

$$(\sigma^2 \mid (\Psi/\sigma^2), Y_T) \sim GI(\alpha^*, \beta^*) \quad (3.7)$$

$$\text{onde } \alpha^* = \left( \alpha + \frac{T + m}{2} \right) \text{ e } \beta^* = \beta + \frac{1}{2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 \right)$$

As condicionais completas para as variáveis auxiliares foram obtidas como distribuições univariadas, pois seu cálculo foi mais rápido, além disso gera-se destas distribuições facilmente e resultados preliminares mostraram ser eficientes, mas nada impede de se gerar das variáveis auxiliares em bloco.

**Em termos MA(m)**

$$p(d | (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.8)$$

$$(e_j | (\Psi/e_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1 - m), (2 - m), \dots, 0 \quad (3.9)$$

$$\text{onde } \sigma_j^2 = \frac{\sigma^2}{\left(1 + \sum_{i=j:-1}^{1-m} [\psi_{1-i}(d)]^2\right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\sigma^2} \sum_{i=1:-1}^{1-m} \psi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \psi_k(d) e_{1-i-k} \right)$$

$$(\sigma^2 | (\Psi/\sigma^2), Y_T) \sim GI(\alpha^*, \beta^*) \quad (3.10)$$

$$\text{onde } \alpha^* = \left( \alpha + \frac{T + m}{2} \right) \text{ e } \beta^* = \beta + \frac{1}{2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 \right)$$

Novamente, as condicionais completas para as variáveis auxiliares foram obtidas como distribuições univariadas, pois seu cálculo foi mais rápido, além disso gera-se destas distribuições facilmente e resultados preliminares mostraram ser eficientes, mas nada impede de se gerar das variáveis auxiliares em bloco.

### 3.1.5 Simulação

Um estudo de simulação foi conduzido para inferir sobre a qualidade das aproximações AR e MA, além de justificar a utilização de uma das duas. O critério de seleção de modelo reportado aqui é o DIC devido a Spiegelhalter, Best, Car-



lin, e Linde (2001) que se encontra no Apêndice D. Foram considerados quatro diferentes valores para o parâmetro  $d$ ,  $d = \{0, 10; 0, 35; 0, 40; 0, 49\}$ , donde para cada um desses valores gerou-se 200 séries temporais de tamanho 1000. As ordens de aproximações foram  $m = \{5, 10, \dots, 50\}$  em termos AR e MA. A rotina `arima.fracdiff.sim` do pacote estatístico S-plus foi utilizado para gerar as séries para os quatro valores de  $d$ , além disso para  $d = 0, 49$  utilizou-se a estrutura de aproximação MA de ordem 10.000 para gerar do processo em questão. Todas as séries foram geradas com variância  $\sigma^2 = 1$ . Os resultados reportados a seguir referem-se as propriedades da inferência clássica, ou seja, a média das estimativas dos parâmetros, neste caso a média das 200 estimativas a posteriori do parâmetro  $d$  e  $\sigma^2$ . Também a taxa de cobertura (C) do intervalo de credibilidade de 0,95 para estas séries foi calculada. O DIC foi calculado para cada uma das 200 séries e verificou-se a quantidade de vezes em que cada ordem de aproximação foi superior. Estes resultados estão todos baseados em amostras de tamanho 2.000 da distribuição conjunta a posteriori após uma convergência de 1.000 iterações do MCMC.

Os resultados completos da simulação se encontram no Apêndice E. Na Tabela 3.1 tem-se os resultados para  $d = 0, 49$  utilizando-se a expansão em termos MA(m) e AR(m) para  $m = 10, 30, 50$ . Observa-se que para a aproximação em termos MA as médias da distribuição a posteriori do parâmetro  $d$  estão em média perto do valor gerado, entretanto a cobertura pode ser considerada baixa, além disso para o parâmetro  $\sigma^2$  os resultados não apresentam boa performance. Os resultados referentes ao mesmo parâmetro  $d = 0, 49$ , mas agora utilizando-se a aproximação em termos AR(m), mostram que em média as estimativas do parâmetro  $d$  estão próximas do valor gerado e com uma taxa de cobertura um pouco maior se comparado ao da aproximação MA. Quanto a variância do processo, vê-se claramente que a taxa de cobertura aumentou em relação ao modelo com expansão em termos

MA. O DIC calculado para todas as séries indicam que a ordem de aproximação AR é em média melhor que a MA e além disso indica ordens de aproximação perto de 50. A estimativa dada pelo S-plus para  $d$ , que está baseada na verossimilhança aproximada, foi em média igual a 0,470 (0,016) com aproximação igual a 100.

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,489 (0,014)	58	1,24 (0,36)	45	0,491 (0,010)	63	1,10 (0,17)	73
30	0,490 (0,014)	48	1,37 (0,57)	35	0,493 (0,010)	55	1,16 (0,26)	59
10	0,493 (0,013)	36	2,03 (1,57)	14	0,496 (0,006)	37	1,43 (0,69)	34

Tabela 3.1: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,49$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

A Tabela 3.2 exhibe os resultados para as séries geradas com  $d = 0,40$ . As estimativas do parâmetro  $d$  estão em média acima do valor gerado, apesar da taxa de cobertura ter melhorado em relação ao valor de  $d = 0,49$ . Por outro lado, a estimação da variância está em média perto do valor gerado e com uma boa taxa de cobertura. A estimação dada pela método da verossimilhança aproximada para  $d$  foi em média 0,395 (0,02). Novamente, o DIC indica que a aproximação AR é melhor que a MA.

Resultados similares a estes são dados pelo valor de  $d = 0,35$ , de acordo com a Tabela 3.3. Para estas séries o valor médio de  $d$  foi igual a 0,345 (0,02) produzidas pelo método da máxima verossimilhança aproximada. O DIC continua indicando que a aproximação AR é melhor que a MA. Para um valor mais baixo do parâmetro  $d$ , por exemplo  $d = 0,10$ , o ajuste dos dois parâmetros do modelo -  $d$  e  $\sigma^2$  - são bons, pois estão em média no valor gerado e possui uma boa taxa

	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,415 (0,033)	77	1,02 (0,05)	90	0,426 (0,029)	71	1,01 (0,05)	99
30	0,420 (0,036)	68	1,03 (0,05)	88	0,432 (0,029)	64	1,01 (0,05)	96
10	0,431 (0,041)	60	1,08 (0,08)	68	0,452 (0,032)	37	1,03 (0,08)	89

Tabela 3.2: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,40$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,360 (0,030)	85	1,01 (0,05)	92	0,370 (0,028)	80	1,01 (0,05)	99
30	0,364 (0,033)	78	1,01 (0,05)	91	0,376 (0,031)	73	1,01 (0,05)	97
10	0,373 (0,040)	69	1,04 (0,06)	83	0,396 (0,037)	56	1,02 (0,05)	93

Tabela 3.3: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,35$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

de cobertura. Tais resultados podem ser visto na Tabela 3.4. Para este valor de  $d$  tem-se os mesmos resultados em relação ao DIC, ou seja, a aproximação AR é em média melhor que a MA.

Utilizando a aproximação MA(10.000) para o processo ARFIMA(0,d,0), gerou-se 200 séries com parâmetro  $d = 0,49$  e  $\sigma^2 = 1$ , para avaliar um pouco do efeito do gerador das séries dada pelo S-plus e por esta aproximação MA, além de avaliar a possibilidade de se usar este procedimento para gerar séries com outras distribuições para o erro do processo, haja visto que este seria um método mais

	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,100 (0,024)	97	1,00 (0,05)	94	0,106 (0,025)	95	1,00 (0,05)	99
30	0,100 (0,024)	98	1,00 (0,05)	94	0,106 (0,025)	93	1,00 (0,05)	97
10	0,101 (0,025)	97	1,00 (0,05)	92	0,108 (0,026)	93	1,00 (0,05)	93

Tabela 3.4: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,10$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

geral. A escolha de se gerar somente para o valor de  $d=0,49$  é que, em geral, os valores mais perto da fronteira de estacionariedade é que produzem maiores dificuldades na estimação. A estimação pelo método da máxima verossimilhança aproximada produziu em média estimativas igual a 0,474 (0,016). Os resultados são apresentados na Tabela 3.5, donde nota-se claramente que a taxa de cobertura é maior se comparada as séries geradas pelo S-plus, para ambos  $d$  e  $\sigma^2$  nas duas abordagens: AR e MA. A melhora foi mais significativa para  $\sigma^2$ , onde em média o valor está próximo do gerado. A aproximação AR foi 187 vezes superior a MA pelo critério de seleção de modelo DIC, indicando assim que aproximação em termos AR é preferível a MA.

### 3.1.6 Comentários

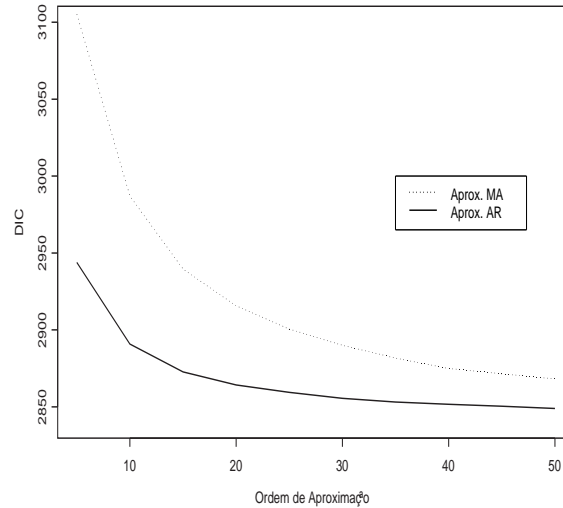
Os resultados apresentados na simulação indicam que a aproximação AR é preferível a MA, pois produz melhores resultados tanto na estimação de  $d$  quanto na de  $\sigma^2$ . A estimação Bayesiana considerando a perda quadrática, priori vaga para o parâmetro  $d$  e utilizando a função de verossimilhança aproximada dada por (Li & McLeod, 1986) produz bons resultados em todos os casos, mesmos para

	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,479 (0,015)	93	1,04 (0,06)	81	0,484 (0,010)	98	1,01 (0,05)	97
30	0,481 (0,016)	86	1,06 (0,07)	72	0,486 (0,010)	94	1,02 (0,05)	94
10	0,486 (0,016)	76	1,17 (0,16)	33	0,492 (0,007)	80	1,06 (0,08)	74

Tabela 3.5: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,49$  pela estrutura MA(10000), onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

as séries geradas pelo S-plus, que não se mostraram adequadas, principalmente para os valores de  $d$  perto da fronteira de estacionariedade. Quando comparada a estimação pelo método da máxima verossimilhança aproximada, os resultados apresentados tiveram em média um melhor desempenho. A ordem de aproximação quanto maior melhor, pois o aumento da mesma conduz a verossimilhança exata. Por outro lado o ganho que se tem com o aumento dessa ordem pode não ser significativo e não valer o esforço computacional envolvido no problema. Para exemplificar, utilizou-se o DIC calculado para as 200 séries gerados em termos MA(10.000) e calculou-se o seu valor médio. A Figura 3.1 confirma um resultado anterior que a aproximação do processo ARFIMA em termos AR é em média melhor que MA, além disso a partir da ordem 40 o DIC já não se reduz tanto quando se aumenta a ordem, o que indica que as aproximações nesta faixa de valores (40-50) já é satisfatório.

Figura 3.1: DIC médio para as 200 séries com  $d=0,49$



### 3.2 Modelo ARFIMA(0,d,0) com erros t-Student

A análise é feita nesta seção sob a hipótese de que os erros  $\epsilon_t$  geradores do processo ARFIMA(0,  $d$ , 0) são t-Student independentes e identicamente distribuídas com  $\nu > 2$  graus de liberdade, parâmetro de locação 0 e parâmetro de escala  $\sigma^2$ , sendo a variância dada por  $\nu\sigma^2/(\nu - 2)$ . Esses erros podem ser obtidos a partir de uma mistura de escala normal-gama inversa, de acordo com o Apêndice C.1. Assim, os erros  $\epsilon_t$  são normais independentes e identicamente distribuídos com média 0 e variância  $\delta^2$  e  $\delta^2$  condicional a  $\nu$  e  $\sigma^2$  tem distribuição  $GI(\eta, \eta\sigma^2)$  onde  $\eta = \nu/2$ . A distribuição t-Student é motivada por possuir a característica de caudas pesadas, ou seja, existe uma probabilidade maior de serem gerado valores nas duas caudas da distribuição se comparada ao modelo com erro normal, mesmo considerando graus de liberdade maiores que 2. O grau de liberdade que define a t-Student é tratado como um número inteiro positivo maior que 2 e menor que 30, visto que para valores acima de 30 a distribuição se aproxima bastante da normal. Uma das áreas de aplicação de modelos com erros t-Student é a economia. Os modelos

econômicos geralmente produzem dados que considerando os erros normais seriam tratados como espúrios, mas que não o seriam com erros t-Student. Este trabalho segue mostrando a aproximação dada pela expansão em termos AR ou MA.

### 3.2.1 Função de verossimilhança

A função de verossimilhança será definida de acordo com as aproximações de ordem  $m$  utilizadas, considerando que existem  $m$  variáveis auxiliares no processo.

#### Em termos AR(m)

Considere o vetor de variáveis auxiliares do processo  $Y_0 = (y_0, y_{-1}, \dots, y_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = \left( \frac{\Gamma((\nu + 1)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\sqrt{\pi\sigma^2}} \right)^T \prod_{t=1}^T \left[ \nu + \frac{(y_t - \mu_t)^2}{\sigma^2} \right]^{-(\nu+1)/2} \quad (3.11)$$

$$\mu_t = \sum_{j=1}^m \varphi_j(d)y_{t-j} \quad \text{e} \quad \Psi = (d, \sigma^2, \nu, Y_0)$$

ou através da mistura normal-gama inversa

$$p(\mathbf{Y}_T | \Psi) = (2\pi\delta^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\delta^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.12)$$

onde

$$\mu_t = \sum_{j=1}^m \varphi_j(d)y_{t-j} \quad \text{e} \quad \Psi = (d, \sigma^2, \nu, \delta^2, Y_0)$$

com

$$(\delta^2 | \nu, \sigma^2) \sim GI(\nu/2, \sigma^2\nu/2)$$

### Em termos MA(m)

Considere o vetor de variáveis auxiliares do processo  $E_0 = (e_0, e_{-1}, \dots, e_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = \left( \frac{\Gamma((\nu + 1)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\sqrt{\pi\sigma^2}} \right)^T \prod_{t=1}^T \left[ \nu + \frac{(y_t - \mu_t)^2}{\sigma^2} \right]^{-(\nu+1)/2} \quad (3.13)$$
$$\mu_t = \sum_{j=1}^m \psi_j(d)e_{t-j} \quad e_t = (y_t - \mu_t) \quad \text{e} \quad \Psi = (d, \sigma^2, \nu, Y_0)$$

ou através da mistura normal-gama inversa

$$p(\mathbf{Y}_T | \Psi) = (2\pi\delta^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\delta^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.14)$$

onde

$$\mu_t = \sum_{j=1}^m \psi_j(d)e_{t-j}, \quad e_t = (y_t - \mu_t) \quad \text{e} \quad \Psi = (d, \sigma^2, \nu, \delta^2, Y_0)$$

com

$$(\delta^2 | \nu, \sigma^2) \sim GI(\nu/2, \sigma^2\nu/2)$$

.

### 3.2.2 Distribuições a priori

O análise é feita considerando a mistura de escala normal-gama inversa para se obter erros com distribuição t-Student e portanto as prioris para os parâmetros do processo ARFIMA(0,  $d$ , 0) incluem o parâmetro  $\delta^2$  e são todas vagas, respeitando as respectivas regiões onde estão definidas. Com isso toda fonte de informação está concentrada na verossimilhança.



### Em termos AR(m)

- $(\delta^2 | \nu, \sigma^2) \sim GI(\eta, \eta\sigma^2)$  com  $\eta = \nu/2$ ;
- $p(\nu) = \frac{1}{M-2}$  para  $\nu = 3, \dots, k_{max}$
- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ;
- $(Y_0 | \sigma^2) \sim N_m(\underline{0}; \sigma^2 I_m)$ ; e
- $(\sigma^2) \sim GI(\alpha, \beta)$ , ou seja,  $p(\sigma^2) = \frac{(\sigma^2)^{-\alpha-1} \beta^\alpha}{\Gamma(\alpha)} \exp\left[-\frac{\beta}{\sigma^2}\right]$ .

### Em termos MA(m)

- $(\delta^2 | \nu, \sigma^2) \sim GI(\eta, \eta\sigma^2)$  com  $\eta = \nu/2$ ;
- $p(\nu) = \frac{1}{M-2}$  para  $\nu = 3, \dots, k_{max}$
- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ;
- $(E_0 | \sigma^2) \sim N_m(\underline{0}; \sigma^2 I_m)$ ; e
- $(\sigma^2) \sim GI(\alpha, \beta)$ , ou seja,  $p(\sigma^2) = \frac{(\sigma^2)^{-\alpha-1} \beta^\alpha}{\Gamma(\alpha)} \exp\left[-\frac{\beta}{\sigma^2}\right]$

Outras prioris também são permitidas, assim como as prioris informativas, como por exemplo a priori  $(\text{Beta}(\alpha, \beta) - 0.5)$  para o parâmetro  $d$ .

### 3.2.3 Distribuição a posteriori

A distribuição a posteriori é proporcional ao produto da função de verossimilhança pela distribuição a priori, ou seja,

$$p(\Psi | Y_T) \propto p(Y_T | \Psi)p(\Psi)$$

**Em termos AR(m)**

$$\begin{aligned}
p(\Psi | Y_T) &\propto (\delta^2)^{-\frac{T+m+\nu}{2}-1} (\sigma^2)^{-\alpha-1} \left( \frac{\nu\sigma^2}{2} \right)^{\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} \mathbf{1}_{(-0,5;0,5)}(d) \\
&\times \exp \left[ -\frac{1}{2\delta^2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 \right) - \frac{\nu\sigma^2}{2\delta^2} - \frac{\beta}{\sigma^2} \right] \quad (3.15)
\end{aligned}$$

**Em termos MA(m)**

$$\begin{aligned}
p(\Psi | Y_T) &\propto (\delta^2)^{-\frac{T+m+\nu}{2}-1} (\sigma^2)^{-\alpha-1} \left( \frac{\nu\sigma^2}{2} \right)^{\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} \mathbf{1}_{(-0,5;0,5)}(d) \\
&\times \exp \left[ -\frac{1}{2\delta^2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 \right) - \frac{\nu\sigma^2}{2\delta^2} - \frac{\beta}{\sigma^2} \right] \quad (3.16)
\end{aligned}$$

onde  $\tau_{\nu_0}(k) = k$  se  $\nu = \nu_0$  e  $\tau_{\nu_0}(k) = 1$  caso contrário.

### 3.2.4 Condicionais completas

A partir de um vetor de parâmetros  $\Psi$  a posteriori, define-se a condicional completa de um sub-vetor paramétrico genérico  $\theta$  como a distribuição deste dado todos os outros parâmetros e os dados, que será denotado por  $p(\theta | (\Psi/\theta), Y_T)$ .

**Em termos AR(m)**

$$(\delta^2 \mid (\Psi/\delta^2), Y_T) \sim GI(\alpha^*, \beta^*) \quad (3.17)$$

$$\text{onde } \alpha^* = \left( \frac{T + m + \nu}{2} \right)$$

$$\text{e } \beta^* = \beta + \frac{1}{2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 + \nu\sigma^2 \right)$$

$$p(\nu \mid (\Psi/\nu), Y_T) \propto \left( \frac{\nu\sigma^2}{2} \right)^{\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} (\delta^2)^{-\frac{\nu}{2}} \exp \left[ -\frac{\nu\sigma^2}{2\delta^2} \right] \quad (3.18)$$

$$p(d \mid (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\delta^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.19)$$

$$(y_j \mid (\Psi/y_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1 - m), (2 - m), \dots, 0 \quad (3.20)$$

$$\text{onde } \sigma_j^2 = \frac{\delta^2}{\left( 1 + \sum_{i=j:-1}^{1-m} [\varphi_{1-i}(d)]^2 \right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\delta^2} \sum_{i=1:-1}^{1-m} \varphi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \varphi_k(d) y_{1-i-k} \right)$$

$$(\sigma^2 \mid (\Psi/\sigma^2), Y_T) \sim GIG \left( \frac{\nu}{2} - \alpha, \sqrt{2\beta}, \frac{\sqrt{\nu}}{\delta} \right) \quad (3.21)$$

Em termos MA(m)

$$(\delta^2 \mid (\Psi/\delta^2), Y_T) \sim GI(\alpha^*, \beta^*) \quad (3.22)$$

$$\text{onde } \alpha^* = \left( \frac{T + m + \nu}{2} \right)$$

$$\text{e } \beta^* = \beta + \frac{1}{2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 + \nu\sigma^2 \right)$$

$$p(\nu \mid (\Psi/\nu), Y_T) \propto \left( \frac{\nu\sigma^2}{2} \right)^{\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} (\delta^2)^{-\frac{\nu}{2}} \exp \left[ -\frac{\nu\sigma^2}{2\delta^2} \right] \quad (3.23)$$

$$p(d \mid (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\delta^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.24)$$

$$(e_j \mid (\Psi/e_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1 - m), (2 - m), \dots, 0 \quad (3.25)$$

$$\text{onde } \sigma_j^2 = \frac{\delta^2}{\left( 1 + \sum_{i=j:-1}^{1-m} [\psi_{1-i}(d)]^2 \right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\delta^2} \sum_{i=1:-1}^{1-m} \psi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \psi_k(d) e_{1-i-k} \right)$$

$$(\sigma^2 \mid (\Psi/\sigma^2), Y_T) \sim GIG \left( \frac{\nu}{2} - \alpha, \sqrt{2\beta}, \frac{\sqrt{\nu}}{\delta} \right) \quad (3.26)$$

### 3.2.5 Comentários

Um estudo de simulação também poderia ser conduzido aqui para avaliar as propostas de aproximação e a qualidade das mesmas como se fez para os erros normais. As séries podem ser geradas utilizando-se uma aproximação  $MA(q)$  com  $q$  grande como se fez no caso dos erros terem distribuição normais.

## 3.3 Modelo ARFIMA(0,d,0) com erros hiperbólicos generalizados

A classe de distribuições hiperbólicas generalizadas tem sido usada para ajustar dados financeiros. O desenvolvimento desta distribuição se deve a Barndorff-Nielsen (1977), mas foi em Eberlein e Keller (1995) que se deu pela primeira vez o uso desta distribuição em finanças. O nome desta distribuição se deve ao fato do logaritmo de sua densidade formar uma hipérbole e não uma parábola como no caso da distribuição normal. Uma justificativa para seu uso é que sua distribuição pode acomodar caudas pesadas, mas não tanto como no modelo com erros t-Student, e sua variância entretanto ser finita, além disso ela pode ser assimétrica, uma característica relevante neste tipo de modelagem. A análise nesta seção considera que os erros  $\epsilon_t$  geradores do processo ARFIMA  $(0, d, 0)$  são hiperbólicos generalizados, ou seja,  $(\epsilon_t | \lambda, \gamma, \delta) \sim HG(\lambda, \gamma, 0, \delta, 0)$  ou  $(\epsilon_t | \lambda, \gamma, \delta) \sim HG(\lambda, \gamma, \delta)$ . Os erros do processo, por hipótese, são independentes e identicamente distribuídos. Esses erros podem ser obtidos a partir de uma mistura de média-escala normal-gaussiana inversa generalizada, de acordo com o Apêndice C.2. Assume-se que os erros  $\epsilon_t$  são normais independentes e identicamente distribuídos com média 0 e variância  $\sigma^2$ , e que  $\sigma^2$  condicional a  $\lambda$ ,  $\delta$  e  $\gamma$  tem distribuição  $GIG(\lambda, \delta, \gamma)$ . Poucos trabalhos tem sido feito do ponto de vista Bayesiano com esta distribuição, o que motiva o seu estudo. Novamente,

segue-se apresentando a verossimilhança aproximada em termos AR e MA.

### 3.3.1 Função de verossimilhança

A função de verossimilhança será definida de acordo com as aproximações de ordem  $m$  utilizadas, considerando que existem  $m$  variáveis auxiliares no processo.

#### Em termos AR(m)

Considere o vetor de variáveis auxiliares do processo  $Y_0 = (y_0, y_{-1}, \dots, y_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = \prod_{t=1}^T \frac{\gamma^{\frac{1}{2}}}{\sqrt{2\pi\delta^\lambda} K_\lambda(\delta\gamma)} [(y_t - \mu_t)^2 + \delta^2]^{(\lambda - \frac{1}{2})/2} \times K_{\lambda - \frac{1}{2}}(\gamma\sqrt{(y_t - \mu_t)^2 + \delta^2}) \quad (3.27)$$

$$\mu_t = \sum_{j=1}^m \varphi_j(d)y_{t-j} \quad \text{e} \quad \Psi = (d, \lambda, \gamma, \delta, Y_0)$$

ou através da mistura normal-inversa gaussiana generalizada

$$p(\mathbf{Y}_T | \Psi) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.28)$$

onde

$$\mu_t = \sum_{j=1}^m \varphi_j(d)y_{t-j} \quad \text{e} \quad \Psi = (d, \sigma^2, \lambda, \gamma, \delta, Y_0)$$

com

$$(\sigma^2 | \lambda, \gamma, \delta) \sim GIG(\lambda, \gamma, \delta)$$

### Em termos MA(m)

Considere o vetor de variáveis auxiliares do processo  $E_0 = (e_0, e_{-1}, \dots, e_{1-m})$ . A função de verossimilhança do processo, baseada em uma série temporal de tamanho  $T$  de  $\{y_t\}$  -  $Y_T = (y_1, \dots, y_T)$  - é

$$p(\mathbf{Y}_T | \Psi) = \prod_{t=1}^T \frac{\gamma^{\frac{1}{2}}}{\sqrt{2\pi\delta^\lambda} K_\lambda(\delta\gamma)} [(y_t - \mu_t)^2 + \delta^2]^{(\lambda - \frac{1}{2})/2} \times K_{\lambda - \frac{1}{2}}(\gamma\sqrt{(y_t - \mu_t)^2 + \delta^2}) \quad (3.29)$$

$$\mu_t = \sum_{j=1}^m \psi_j(d) e_{t-j}, \quad e_t = (y_t - \mu_t) \quad \text{e} \quad \Psi = (d, \lambda, \gamma, \delta, Y_0)$$

ou através da mistura normal-inversa gaussiana generalizada

$$p(\mathbf{Y}_T | \Psi) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] \quad (3.30)$$

onde

$$\mu_t = \sum_{j=1}^m \psi_j(d) e_{t-j}, \quad e_t = (y_t - \mu_t) \quad \text{e} \quad \Psi = (d, \sigma^2, \lambda, \gamma, \delta, Y_0)$$

com

$$(\sigma^2 | \lambda, \gamma, \delta) \sim GIG(\lambda, \gamma, \delta)$$

### 3.3.2 Distribuições a priori

O análise é feita considerando a mistura de média-escala normal-inversa gaussiana generalizada para se obter erros com distribuição hiperbólica generalizada e portanto as prioris para os parâmetros do processo ARFIMA(0,  $d$ , 0) incluem o parâmetro  $\sigma^2$  e são todas vagas, respeitando as respectivas regiões onde estão definidas. Com isso toda fonte de informação está concentrada na verossimilhança.

### Em termos AR(m)

- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ;
- $(Y_0 | \sigma^2) \sim N_m(\underline{0}; \sigma^2 I_m)$ ;
- $(\sigma^2 | \lambda, \delta, \gamma) \sim GIG(\lambda, \delta, \gamma)$
- $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$ ;  $\delta \sim G(\alpha_\delta, \beta_\delta)$ ; e  $\gamma \sim G(\alpha_\gamma, \beta_\gamma)$

### Em termos MA(m)

- $p(d) \propto \mathbf{1}_{(-0,5;0,5)}(d)$ ;
- $(E_0 | \sigma^2) \sim N_m(\underline{0}; \sigma^2 I_m)$ ;
- $(\sigma^2 | \lambda, \delta, \gamma) \sim GIG(\lambda, \delta, \gamma)$
- $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$ ;  $\delta \sim G(\alpha_\delta, \beta_\delta)$ ; e  $\gamma \sim G(\alpha_\gamma, \beta_\gamma)$ ;

Outras prioris também são permitidas, assim como as prioris informativas, como por exemplo a priori (Beta( $\alpha, \beta$ ) - 0.5) para o parâmetro  $d$ .

### 3.3.3 Distribuição a posteriori

A distribuição a posteriori é proporcional ao produto da função de verossimilhança pela a distribuição a priori, ou seja,

$$p(\Psi | Y_T) \propto p(Y_T | \Psi)p(\Psi)$$

### Em termos AR(m)

$$p(\Psi | Y_T) \propto (\sigma^2)^{-\frac{T+m}{2} + \lambda - 1} \gamma^{\alpha_\gamma + \lambda - 1} \delta^{\alpha_\delta - \lambda - 1} \frac{1}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{(\lambda - \mu_\lambda)^2}{2\sigma_\lambda^2} - \beta_\delta \delta - \beta_\gamma \gamma \right] \\ \times \exp \left[ -\frac{1}{2} \left( \frac{\sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 + \delta^2}{\sigma^2} + \gamma^2 \sigma^2 \right) \right] \mathbf{1}_{(-0,5;0,5)}(d) \quad (3.31)$$



**Em termos MA(m)**

$$\begin{aligned}
p(\Psi | Y_T) &\propto (\sigma^2)^{-\frac{T+m}{2}+\lambda-1} \gamma^{\alpha_\gamma+\lambda-1} \delta^{\alpha_\delta-\lambda-1} \frac{1}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{(\lambda - \mu_\lambda)^2}{2\sigma_\lambda^2} - \beta_\delta\delta - \beta_\gamma\gamma \right] \\
&\times \exp \left[ -\frac{1}{2} \left( \frac{\sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 + \delta^2}{\sigma^2} + \gamma^2\sigma^2 \right) \right] \mathbf{1}_{(-0,5;0,5)}(d) \quad (3.32)
\end{aligned}$$

### 3.3.4 Condicionais completas

A partir de um vetor de parâmetros  $\Psi$  a posteriori, define-se a condicional completa de um sub-vetor paramétrico genérico  $\theta$  como a distribuição deste dado todos os outros parâmetros e os dados, que será denotado por  $p(\theta | (\Psi/\theta), Y_T)$ .

**Em termos AR(m)**

$$p(d | (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.33)$$

$$(y_j | (\Psi/y_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1-m), (2-m), \dots, 0 \quad (3.34)$$

$$\text{onde } \sigma_j^2 = \frac{\sigma^2}{\left(1 + \sum_{i=j:-1}^{1-m} [\varphi_{1-i}(d)]^2\right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\sigma^2} \sum_{i=1:-1}^{1-m} \varphi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \varphi_k(d) y_{1-i-k} \right)$$

$$(\sigma^2 | (\Psi/\sigma^2), Y_T) \sim GIG \left( \lambda - \frac{T+m}{2}, \left( \delta^2 + \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 \right)^{\frac{1}{2}}, \gamma \right) \quad (3.35)$$

$$p(\lambda | (\Psi/\lambda), Y_T) \propto \left( \frac{\gamma}{\delta} \right)^\lambda \frac{(\sigma^2)^\lambda}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{(\lambda - \mu_\lambda)^2}{2\sigma^2\lambda} \right] \quad (3.36)$$

$$p(\gamma | (\Psi/\gamma), Y_T) \propto \frac{\gamma^{\alpha_\gamma + \lambda - 1}}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{\gamma^2 \sigma^2}{2} - \beta_\gamma \gamma \right] \quad (3.37)$$

$$p(\delta | (\Psi/\delta), Y_T) \propto \frac{\delta^{\alpha_\delta - \lambda - 1}}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{\delta^2}{2\sigma^2} - \beta_\delta \delta \right] \quad (3.38)$$

Em termos MA(m)

$$p(d \mid (\Psi/d), Y_T) \propto \exp \left[ -\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2 \right] I_{(-0,5;0,5)}(d) \quad (3.39)$$

$$(e_j \mid (\Psi/e_j), Y_T) \sim N(\mu_j, \sigma_j^2) \text{ para } j = (1-m), (2-m), \dots, 0 \quad (3.40)$$

$$\text{onde } \sigma_j^2 = \frac{\sigma^2}{\left(1 + \sum_{i=j:-1}^{1-m} [\psi_{1-i}(d)]^2\right)}$$

$$\text{e } \mu_j = \frac{\sigma_j^2}{\sigma^2} \sum_{i=1:-1}^{1-m} \psi_{1-i}(d) \left( y_{1-i} - \sum_{\substack{k=1 \\ k \neq (1-i)}}^m \psi_k(d) e_{1-i-k} \right)$$

$$(\sigma^2 \mid (\Psi/\sigma^2), Y_T) \sim GIG \left( \lambda - \frac{T+m}{2}, \left( \delta^2 + \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 \right)^{\frac{1}{2}}, \gamma \right) \quad (3.41)$$

$$p(\lambda \mid (\Psi/\lambda), Y_T) \propto \left( \frac{\gamma}{\delta} \right)^\lambda \frac{(\sigma^2)^\lambda}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{(\lambda - \mu_\lambda)^2}{2\sigma^2\lambda} \right] \quad (3.42)$$

$$p(\gamma \mid (\Psi/\gamma), Y_T) \propto \frac{\gamma^{\alpha_\gamma + \lambda - 1}}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{\gamma^2 \sigma^2}{2} - \beta_\gamma \gamma \right] \quad (3.43)$$

$$p(\delta \mid (\Psi/\delta), Y_T) \propto \frac{\delta^{\alpha_\delta - \lambda - 1}}{K_\lambda(\delta\gamma)} \exp \left[ -\frac{\delta^2}{2\sigma^2} - \beta_\delta \delta \right] \quad (3.44)$$

### 3.3.5 Comentários

Outra vez, poderia ser conduzido um estudo de simulação para se entender melhor o processo de estimação dos parâmetros desta distribuição e dos parâmetros do processo ARFIMA no enfoque Bayesiano. Gerar erros desta distribuição não é uma tarefa fácil, mas pode ser feita utilizando o amostrador da fatia com devidos ajustes para que possa produzir auto-correlações que sejam desprezíveis e possa se considerar os valores gerados como independentes. A tarefa mais difícil é estimar os parâmetros da distribuição HG, pois alguns dos parâmetros possuem correlações altas, portanto gerando problemas na convergência do algoritmo de Gibbs.

# Capítulo 4

## Identificando a longa dependência

A característica marcante do processo ARFIMA é sua capacidade de reconhecer a longa dependência através do parâmetro  $d$ , como já foi visto nas seções anteriores. Mas se o valor deste parâmetro  $d$  for igual a 0 com alta probabilidade, então o processo ARFIMA passa a ser simplesmente um processo ARMA, ou seja, ele já não possui a característica de longa dependência. Portanto, ao se introduzir priori com ponto de massa (Petris, 1997; West & Huerta, 1999) no domínio de  $d$  -  $d \in [-0,5; 0,5]$  - procura-se identificar qual a probabilidade a posteriori desse ponto de massa  $d_0$ . Em particular, o ponto de massa de interesse é o  $d_0 = 0$ , que permitirá um teste de hipótese Bayesiano de  $H_0 : d = 0$  contra  $H_1 : d \neq 0$ .

### 4.1 Modelo ARFIMA(0,d,0) com erros normais

#### 4.1.1 Função de verossimilhança

A função de verossimilhança não muda em nenhum dos dois casos anteriores - expansão em AR( $m$ ) ou MA( $m$ ) e condicional a  $d$ , ela será denotada como  $\ell_d(d)$ .

### 4.1.2 Distribuições a priori

As distribuições a priori para  $\sigma^2$ ,  $(Y_0 | \sigma^2)$  e  $(E_0 | \sigma^2)$  são as mesmas dadas anteriormente no caso dos erros normais. E a distribuição a priori para o parâmetro  $d$  e sua respectiva probabilidade do ponto de massa são:

- $p(d | \omega) := \omega \mathbf{1}_{d_0}(d) + (1 - \omega)g(d)$ ,

onde  $g(d)$  é uma densidade absolutamente contínua no domínio de  $d$ . Assume-se que  $g(d)$  é uma densidade  $(\text{Beta}(\alpha_d, \beta_d)-0,5)$ , pois esta densidade pode acomodar uma diversidade de opiniões a priori do parâmetro de longa dependência  $d$  (Petris, 1997).

- $\omega \sim \text{Beta}(\alpha_\omega, \beta_\omega)$ ,

que se justifica novamente pelo fato de poder acomodar diferentes opiniões a priori sobre a probabilidade do ponto de massa  $d_0$ .

### 4.1.3 Distribuição a posteriori

A distribuição a posteriori é proporcional ao produto da função de verossimilhança vezes a distribuição a priori, ou seja,

$$p(\Psi | Y_T) \propto p(Y_T | \Psi)p(\Psi)$$

**Em termos AR(m)**

$$p(\Psi | Y_T) \propto (\sigma^2)^{-\frac{T+m}{2}-\alpha-1} [\omega + (1 - \omega)g(d)] \omega^{\alpha-1} (1 - \omega)^{\beta-1} \\ \times \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m y_{1-j}^2 \right) - \frac{\beta}{\sigma^2} \right] \quad (4.1)$$

### Em termos MA(m)

$$p(\Psi | Y_T) \propto (\sigma^2)^{-\frac{T+m}{2}-\alpha-1} [\omega + (1-\omega)g(d)] \omega^{\alpha\omega-1} (1-\omega)^{\beta\omega-1} \\ \times \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{t=1}^T (y_t - \mu_t)^2 + \sum_{j=1}^m e_{1-j}^2 \right) - \frac{\beta}{\sigma^2} \right] \quad (4.2)$$

#### 4.1.4 Condicionais completas

A condicionais completas de  $(\sigma^2 | (\Psi/\sigma^2), Y_T)$ ,  $(Y_0 | (\Psi/Y_0), Y_T)$  e  $(E_0 | (\Psi/E_0), Y_T)$  são as mesmas dadas por (3.5) e (3.8). As outras duas condicionais completas  $(d | (\Psi/d), Y_T)$  e  $(\omega | (\Psi/\omega), Y_T)$  em ambas aproximações, AR(m) e MA(m), são:

- $p(d | (\Psi/d), Y_T) \propto [\omega \ell_d(d_0) + (1-\omega)g^*(d)\ell_d(d)] \mathbf{1}_{[-0,5;0,5]}(d)$   
onde  $g^*(d) = (0,5+d)^{\alpha_d-1} (0,5-d)^{\beta_d-1}$ .
- $p(\omega | (\Psi/\omega), Y_T) \propto [\omega \ell_d(d_0) + (1-\omega)g^*(d)\ell_d(d)] \omega^{\alpha\omega-1} (1-\omega)^{\beta\omega-1}$   
onde  $g^*(d) = (0,5+d)^{\alpha_d-1} (0,5-d)^{\beta_d-1}$ .

#### 4.1.5 Comentários

Apesar de existir uma distribuição a priori para a probabilidade de  $d = 0$  e conseqüentemente a distribuição marginal a posteriori para esta probabilidade, ela não foi utilizada, sendo portanto fixada a priori. A identificação da longa dependência através da probabilidade a posteriori do parâmetro  $d$  ser igual a 0 é uma importante ferramenta para quem trabalha com séries temporais e portanto poder-se-ia realizar simulações para se obter alguns resultados conclusivos. Além disso, a idéia de se introduzir prioris com pontos de massa para o parâmetro  $d$  aplica-se aos erros t-Student e hiperbólico generalizado ficando evidente, é claro, que a complexidade do modelo aumenta.

# Capítulo 5

## Aspectos computacionais

### 5.1 Amostrador de Gibbs

A inferência Bayesiana procura obter a densidade conjunta dos parâmetros a posteriori, assim como suas distribuições marginais e resumindo suas características principais. O amostrador de Gibbs introduzidos por Gelfand e Smith (1990) permite obter amostras das distribuições conjunta e marginais a posteriori, obtendo-se desta maneira um resumo estatísticos de tais distribuições, utilizando a estrutura Markoviana. Suponha que a distribuição de  $p(\Theta | D) = p(\theta_1, \theta_2, \dots, \theta_n | D)$ , ou seja,  $(\Theta | Y_T)$  tem distribuição  $n$ -dimensional, e  $D$  é toda a informação disponível. A distribuição de  $p_j(\theta_j | \{\theta_i\}_{j \neq i}, D)$  é chamada de condicional completa de  $\theta_j$ . Tais condicionais completas são sempre proporcionais ao produto da função de verossimilhança pela distribuição a priori, isto é,  $p_j(\theta_j | \{\theta_i\}_{j \neq i}, D) \propto p(Y_T | \Theta)p(\Theta) \forall j$ . Dado o conjunto de valores iniciais  $(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)})$ , as amostras são obtidas da seguinte maneira:



- 1)  $\theta_1^{(k)} \sim p_1(\theta_1 | \theta_2^{(k-1)}, \dots, \theta_n^{(k-1)}, D)$
- 2)  $\theta_2^{(k)} \sim p_2(\theta_2 | \theta_1^{(k)}, \theta_3^{(k-1)} \dots \theta_n^{(k-1)}, D)$
- $\vdots$
- n)  $\theta_n^{(k)} \sim p_n(\theta_n | \theta_1^{(k)}, \dots, \theta_{n-1}^{(k)}, D)$

- Repita passos 1, 2, ..., n para  $k = 1, 2, 3, \dots$

Sob certas condições de regularidade (Gelfand & Smith, 1990; Tierney, 1994) a cadeia de Markov acima converge para uma distribuição estacionária após  $K$  iterações do esquema acima e esta é a distribuição a posteriori de  $p(\theta_1, \theta_2, \dots, \theta_n | D)$ .

## 5.2 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings foi desenvolvido por Metropolis et al. (1953) e depois generalizado por Hastings (1970). Em Tierney (1994) encontra-se uma boa exposição teórica deste algoritmo. O algoritmo depende de uma densidade proposta  $q(\theta, \xi)$  onde

$$\int_{R_\xi} q(\theta, \xi) d\xi = 1.$$

Denotando-se  $U(0, 1)$  como uma distribuição uniforme no intervalo  $(0, 1)$ , tem-se uma versão bem geral do algoritmo de Metropolis-Hastings para amostrar da distribuição a posteriori de  $p(\theta | D)$  é:

- 0) Escolha um valor inicial  $\theta_0$  e faça  $i = 0$
- 1) Gere um valor candidato  $\theta^*$  da proposta  $q(\theta_i, \cdot)$  e  $u$  de  $U(0, 1)$

- 2) Faça 
$$\begin{cases} \theta_{i+1} = \theta^* & , \text{ se } u \leq a(\theta_i, \theta^*) \\ \theta_{i+1} = \theta_i & , \text{ caso contrário} \end{cases}$$

onde a probabilidade de aceitação é  $a(\theta, \xi) = \min \left\{ 1, \frac{p(\xi | D)q(\xi, \theta)}{p(\theta | D)q(\theta, \xi)} \right\}$

3) Faça  $i=i+1$  e volte para o passo 1.

Aconselha-se calcular a probabilidade de aceitação através de seu logaritmo para evitar problemas de precisão numérica. Para se obter uma amostra de uma distribuição cujo domínio é um intervalo fechado e limitado, pode-se utilizar uma das duas densidades propostas a seguir para o Metropolis-Hastings descrito acima.

### 5.2.1 Proposta uniforme não simétrica

Suponha que se deseja obter uma amostra da densidade  $p(\theta | D)$  onde  $\theta \in [a, b]$ ,  $a < b$ . Seja  $\Delta$  um valor real positivo tal que  $\Delta \leq (b - a)$ , então a densidade proposta para o algoritmo acima é

$$q(\theta, \xi) = \frac{1}{b_\theta - a_\theta}$$

onde  $a_\theta = \max(\xi - \Delta; a)$  e  $b_\theta = \min(\xi + \Delta; b)$ .

Assim, tem-se que

$$q(\xi, \theta) = \frac{1}{b_\xi - a_\xi}$$

onde  $a_\xi = \max(\theta - \Delta; a)$  e  $b_\xi = \min(\theta + \Delta; b)$ .

Portanto, a razão das proposta no cálculo da probabilidade de aceitação pode ser diferente de 1, pois a proposta já não é simplesmente uma uniforme em todo o intervalo, e sim num subintervalo contido em  $[a, b]$ , ou em outras palavras, a distribuição pode não ser simétrica. A quantidade positiva  $\Delta$  é controlada de tal forma que a taxa de aceitação fique em torno de 30%.

### 5.2.2 Proposta Normal-Logito

Novamente, suponha que se deseja obter uma amostra da densidade  $p(\theta | D)$  onde  $\theta \in [a, b]$ ,  $a < b$ . A idéia aqui é transformar esse intervalo fechado na reta dos  $\Re$  através de uma transformação logarítmica, utilizando-se como proposta a

distribuição normal. Seja a transformação dada por

$$\delta = \ln \left( \frac{\theta - a}{b - \theta} \right)$$

cujo Jacobiano é

$$J = \frac{b - a}{(\theta - a)(b - \theta)}.$$

Então

$$q_\theta(\theta, \xi) = q_\delta \left( \ln \left( \frac{\xi - a}{b - \xi}, \xi \right) \right) \frac{b - a}{(\xi - a)(b - \xi)}$$

onde  $q_\delta(\cdot, \xi)$  é uma densidade normal com média  $\gamma^{(i-1)}$  e variância  $V$ . Essa variância é determinada de tal forma que a taxa de aceitação do algoritmo de Metropolis-Hastings fique em torno de 30%. Ao avaliar-se a razão das propostas para o cálculo da probabilidade de aceitação, o que realmente importa é a razão dos Jacobianos, pois a densidade Normal é simétrica. É recomendável avaliar-se a probabilidade de aceitação através de seu logaritmo, porque fica claro que nos extremos do intervalo pode-se ter problemas de precisão numérica.

### 5.3 Amostrador da fatia

O amostrador da fatia - traduzido do inglês *slice sampler* - é um método que serve para gerar valores de uma variedade de distribuições, em especial densidades que estejam num intervalo fechado e limitado. A idéia básica é gerar valores da distribuição uniforme definida pela região abaixo da função de densidade ao qual se deseja amostrar, e considerar apenas as coordenadas horizontais. Uma cadeia de Markov que converge para esta distribuição uniforme pode ser construída alternando-se valores na direção vertical e na horizontal definindo a fatia pela posição vertical atual. Para isso se introduz variáveis auxiliares e então usa-se o amostrador de Gibbs para a área abaixo da densidade. Suponha que se deseja gerar valores de uma densidade  $p(x)$ ,  $x \in A \subseteq \mathfrak{R}$ . Para isso considera-se a região bi-dimensional abaixo de  $p(x)$  ou de  $g(x) = cp(x)$ , e então:

- Introduz-se uma variável auxiliar  $z$  com  $z | x \sim U(0, g(x))$ ;
- Segue-se que a distribuição conjunta de  $(z, x)$  é uniforme na região  $\{(z, x) : 0 \leq z \leq g(x)\}$  com densidade

$$f(z, x) = \begin{cases} \frac{1}{c} & , \text{ se } 0 \leq z \leq g(x) \\ 0 & , \text{ caso contrário.} \end{cases}$$

- A distribuição condicional de  $x | z$  é  $f(x | z) \propto f(z, x) = \begin{cases} \frac{1}{c} & , \text{ se } 0 \leq z \leq g(x) \\ 0 & , \text{ caso contrário.} \end{cases}$ ,  
ou seja,  $(x | z) \sim U(S(z))$ , onde  $S(z) = \{x : g(x) \geq z\}$ .

Assim,  $S(x)$  é a união dos intervalos que constituem a fatia através da densidade definida por  $z$ . A obtenção de  $S(x)$  na maioria das vezes não é simples. A estrutura do modelo acima nos conduz a simular tais valores usando o amostrador de Gibbs, onde as condicionais completas são uniformes.

- Gere  $z^{(i)} \sim U(0, g(x^{(i-1)}))$
- Gere  $x^{(i)} \sim U(S(z^{(i)}))$  onde  $S(z^{(i)}) = \{x : g(x) \geq z^{(i)}\}$

Dentre as vantagens desse método estão: aplica-se a várias distribuições, não necessita especificar uma densidade proposta como no Metropolis-Hastings e só utiliza a distribuição uniforme para gerar os valores. A principal desvantagem é que a determinação de  $S(z)$  pode ser difícil. O método é detalhado em Neal (1997), onde se propõe uma maneira de solucionar a determinação de  $S(z)$  e a generalização do algoritmo.

## 5.4 Amostrador para mistura de distribuições

Em muitas ocasiões é de fundamental importância determinar-se prioris com pontos de massa para a avaliação de certas características de interesse do modelo em questão. Suponha que uma priori deste tipo seja dada por

$$p(\theta) = \omega \mathbf{1}_{\theta_0}(\theta) + (1 - \omega)g(\theta) \quad (5.1)$$

onde  $\mathbf{1}_{\theta_0}(\theta) = 1$  se  $\theta = \theta_0$  e 0 caso contrário;  $g(\theta)$  é uma densidade absolutamente contínua; e  $\omega \in [0, 1]$  é a probabilidade a priori de  $\{\theta = \theta_0\}$ . Além disso, suponha que a verossimilhança seja  $\ell(\theta)$ , então de (5.1), tem-se

$$p(\theta | D) \propto \omega \ell(\theta_0) \mathbf{1}_{\theta_0}(\theta) + (1 - \omega) \ell(\theta) g(\theta) \quad (5.2)$$

ou de maneira equivalente tem-se

$$p(\theta | D) \propto \frac{\omega}{1 - \omega} \ell(\theta_0) \mathbf{1}_{\theta_0}(\theta) + \ell(\theta) g(\theta) \quad (5.3)$$

Assim, calcula-se o inverso da constante de normalização da densidade  $p(\theta | D)$  integrando-se (5.3), ou seja,

$$c^{-1} = \frac{\omega}{(1 - \omega)} \ell \theta_0 + \int_{\Theta} \ell \theta g(\theta) d\theta$$

e portanto,

$$p(\theta | D) = c \frac{\omega}{1 - \omega} \ell(\theta_0) \mathbf{1}_{\theta_0}(\theta) + c \ell(\theta) g(\theta) \quad (5.4)$$

O cálculo da constante de normalização, mesmo por métodos numéricos, na maioria das vezes não é simples de se obter. A idéia é distribuir a massa de probabilidade num intervalo uniforme e transformar a densidade em (5.4) numa densidade absolutamente contínua.

Seja  $\xi$  uma variável aleatória auxiliar cuja densidade é

$$h_{\xi}(\xi) = \begin{cases} c \ell(\xi + z) g(\xi + z) & , \text{ se } \xi < \theta_0 - z \\ c \frac{\omega}{1 - \omega} \ell(\theta_0) & , \text{ se } |\xi - \theta_0| < z \\ c \ell(\xi - z) g(\xi - z) & , \text{ se } \xi > \theta_0 + z \end{cases} \quad (5.5)$$

Assim, a massa de probabilidade em  $\theta_0$  foi transferida para o intervalo  $(\theta_0 - z, \theta_0 + z)$ . É fácil verificar que  $2z = \omega/(1-\omega)$ . Gera-se valores da distribuição da variável auxiliar  $\xi$  com algum dos métodos de simulação para densidades absolutamente contínuas, por exemplo o Metropolis-Hastings ou amostrador da fatia. Ao obter o valor gerado de  $\xi$  e usando a transformação inversa

$$\theta = \begin{cases} \xi + z & , \text{ se } \xi < \theta_0 - z \\ \theta_0 & , \text{ se } |\xi - \theta_0| < z \\ \xi - z & , \text{ se } \xi > \theta_0 + z \end{cases} \quad (5.6)$$

obtém-se os valores gerados da distribuição  $p(\theta | D)$ . Prova-se facilmente que a transformação acima leva a distribuição desejada  $p(\theta | D)$ .

# Capítulo 6

## Aplicação

### 6.1 Nível mínimo do rio Nilo

O volume anual mínimo do rio Nilo tem aparecido em diversos trabalhos, pois o logaritmo desta série apresenta uma longa dependência. De acordo com estes trabalhos o parâmetro  $d$  encontra-se no intervalo  $(0,35;0,40)$ , veja por exemplo (Beran & Terrin, 1996). Na Figura 6.1 tem-se a série temporal do volume mínimo do Rio Nilo baseado nas medidas feitas em Roda Gauge perto do Cairo entre 622 e 1284 D.C ( $T=663$ ). A modelagem foi feita com o modelo ARFIMA(0, $d$ ,0) sob a hipótese de erros normais e média do processo  $\mu$ . As distribuições a priori foram vagas para os três parâmetros  $d, \sigma^2, \mu$ . A estimação foi feita utilizando os procedimentos descritos até agora: amostrador de Gibbs e amostrador da fatia para a condicional completa do parâmetro  $d$ . As densidades marginais a posteriori estão caracterizadas pela Tabela 6.1, baseadas numa amostra a posteriori de tamanho 2.000 logo após convergência em 1.000 iterações, que exibem a média dos parâmetros e os desvios-padrões associados a elas, e a densidade a posteriori pode ser vista na Figura 6.2. O resultado para o parâmetro  $d$  de interesse está dentro do esperado e em conformidade com a literatura presente até o momento

com esta série temporal. O intervalo de credibilidade de 0,95 para  $d$  é  $(0,33;0,45)$ . Além disso temos as médias a posteriori de  $\sigma^2$  e  $\mu$  e seus respectivos desvios-padrões.

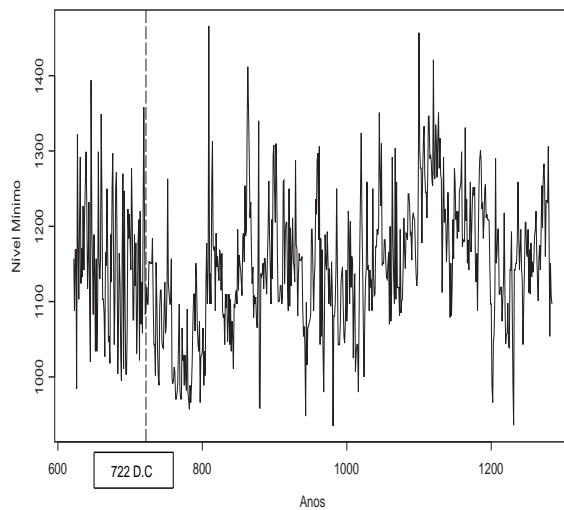


Figura 6.1: Nível mínimo do rio Nilo medido em Roda Gauche entre os anos 622 a 1284 D.C

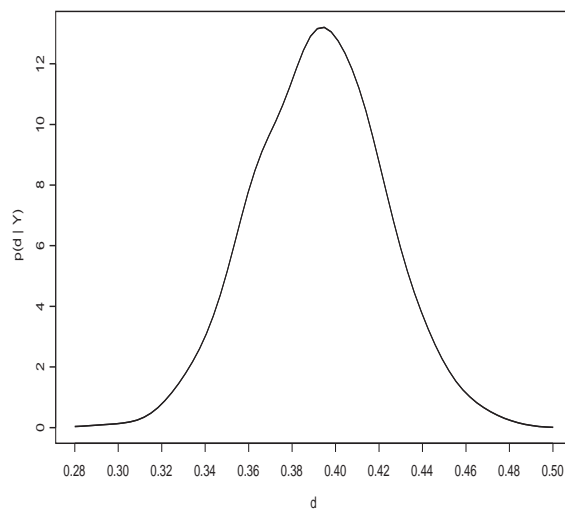


Figura 6.2: Distribuição a posteriori do parâmetro  $d$  referentes aos dados do rio Nilo.



	$d$	$\sigma^2$	$\mu$
Posteriori	0,391 (0,0293)	4929 (267,8)	1147,6 (2,8)
Whittle	0,399 (0,0498)	4840 (-)	-
GPH	0,500 (0,1570)	4899 (-)	-

Tabela 6.1: Estimação dos parâmetros referentes a série temporal do volume mínimo do rio Nilo entre os anos 622-1284 D.C. Os resultados dados pelos estimadores de Whittle e GPH estão reportados em Reisen, Abraham e Toscano (2000)

Alguns autores que trabalharam com esta série, veja por exemplo (Beran, 1994), detectaram uma mudança no valor do parâmetro  $d$  no ano de 722 D.C., que passava de 0,05 para os primeiros 100 anos para 0,45 no restante. Seguindo esses autores, a estimação foi feita para os anos antes e depois da mudança e o resultado está resumido na Tabela 6.2 que trazem a média a posteriori e os respectivos desvios-padrões. Os resultados condizem com o esperado. O valor para os primeiros 100 anos indicam um valor baixo para o parâmetro de interesse  $d$ , e além disso o intervalo de credibilidade de 0,95 é  $(-0,13;0,23)$ , indicando a possibilidade de existir apenas um passeio aleatório, tal resultado pode ser visto na Figura 6.3. Para os outros 563 anos o intervalo de credibilidade de 0,95 é  $(0,3916;0,4954)$ , confirmando a mudança dada pela média a posteriori e a condição de que a série possui uma longa dependência. Para os primeiros 100 anos se faz necessário estimar a probabilidade a posteriori do parâmetro  $d$  ser igual a 0 pela priori com ponto de massa no 0. A probabilidade a posteriori do parâmetro  $d$  ser igual a zero é 1 mesmo quando a probabilidade a priori é pequena (0,02), portanto pode-se considerar os primeiros 100 anos como um passeio aleatório. A Figura 6.3 já induzia tal resultado.

<b>662-721</b>		
$d$	$\sigma^2$	$\mu$
0,0391 (0,0933)	8373,92 (1222,93)	1150,83 (9,12)
<b>722-1284</b>		
$d$	$\sigma^2$	$\mu$
0,4480 (0,0277)	4142,58 (248,5)	1147,88 (2,71)

Tabela 6.2: Distribuição a posteriori na mudança de  $d$  para os dados do rio Nilo

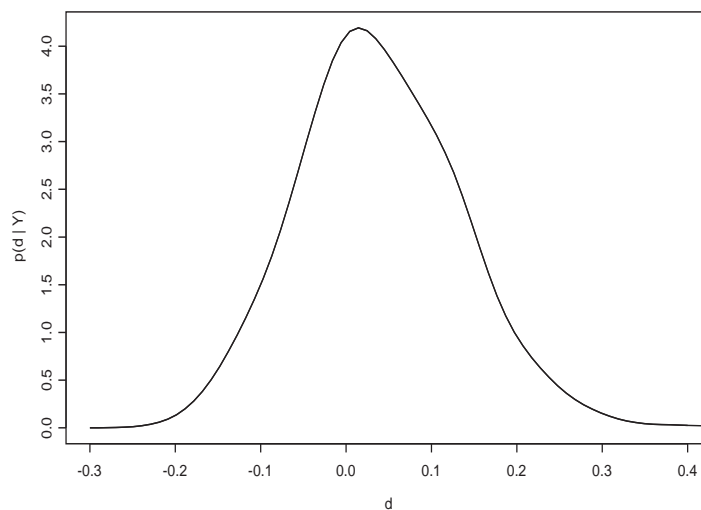
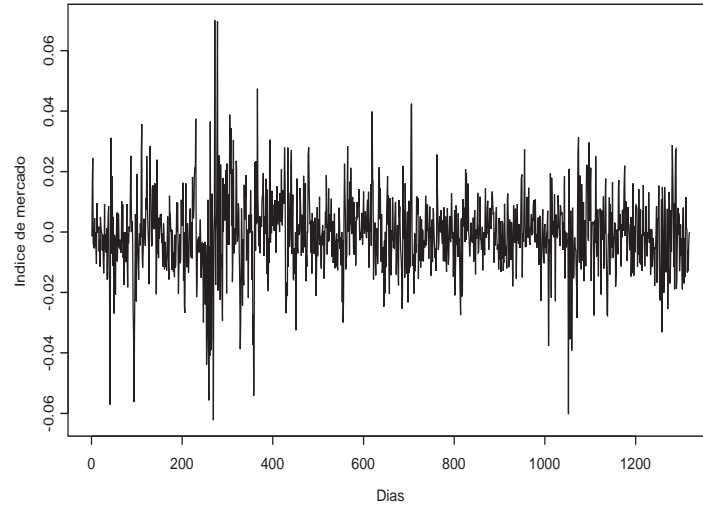


Figura 6.3: Distribuição a posteriori de  $d$  para os 100 anos iniciais

## 6.2 Índice de preço do Chile

A série refere-se ao índice diário de mercado do Chile na primeira diferença de seu logaritmo, calculados para 5 anos - de 29/08/1997 a 19/09/2002 ( $T=1320$ ) - pela "MSCI data" e disponível na rede mundial de computadores. A série é utilizada para verificar a existência de longa dependência ou não, e qual o efeito

Figura 6.4: Série do índice de preço do Chile



de se utilizar erros normais e t-Student em dados financeiros. Esta série temporal pode ser vista na Figura 6.4. As Tabelas 6.3 e 6.4 trazem as médias a posteriori dos parâmetros e seus respectivos desvios-padrões, sob a hipótese que os erros do processo são normais e t-Student, respectivamente. Nas tabelas pode-se ver que a série exibe uma longa dependência sob a hipótese de erro normal com  $d = 0,20$ , o que já não se pode dizer quando os erros do processo são t-Student, pois o desvio-padrão de  $d$  é alto, incluindo assim o 0. Isto pode ser visto pelos intervalos de credibilidade de 0,95 são, respectivamente,  $(0,15;0,25)$  sob a hipótese de normalidade e  $(-0,03;0,48)$  no caso da t-Student, portanto indicando a presença significativa de longa dependência no caso de erro normal e uma certa probabilidade a posteriori de não existir longa dependência quando se está sob a hipótese de erros t-Student. A probabilidade a posteriori para os graus de liberdade é 0,95 para  $\nu = 3$  e 0,05 para  $\nu = 4$ , sendo nula para os demais valores. Utilizando o DIC como critério de seleção de modelos, vê-se desta maneira que a hipótese de erros t-Student exibe melhores resultados em relação ao erro normal. Isso se deve as características de uma série financeira como esta que exibem pontos

mais afastados de suas médias, usualmente chamado de distribuição com caldas pesadas.

<b>Erro Normal</b>			
$d$	$\sigma^2$	$\mu$	DIC
0,196 (0,024)	0,0002 ( $6 \times 10^{-6}$ )	0,0 (0,0003)	7864

Tabela 6.3: Distribuição a posteriori sob a hipótese de erros normais

<b>Erro t-Student</b>			
$d$	$\sigma^2$	$\mu$	DIC
0,249 (0,140)	0,0583 (0,010)	0,0 (0,0030)	235

Tabela 6.4: Distribuição a posteriori sob a hipótese de erros t-Student

# Capítulo 7

## Conclusões e extensões

Foi visto neste trabalho que os modelos ARFIMA são importantes para quem trabalha com séries temporais, pois este modelo consegue captar a forte dependência mesmo para observações distantes entre si através do parâmetro  $d$ . Esse parâmetro é fundamental na identificação da estrutura de auto-correlação das séries temporais, tendo que ser bem estimado. Os procedimentos de estimação utilizados neste trabalho através da função de verossimilhança aproximada sugeridas por Li e McLeod (1986) produzem bons resultados e evitam o problema de se trabalhar com matrizes de covariâncias da dimensão dos dados o qual demanda muito esforço computacional em termos de memória e CPU. Ao tentar compreender o comportamento dos processos ARFIMA e suas aproximações do ponto de vista de suas correlações, vê-se que a auto-correlação amostral mesmo para uma série enorme (50.000) não chega perto da correlação exata, mas a aproximação produziu bons resultado na estimação dos parâmetros. A simulação considerando os erros normais mostrou duas maneiras alternativas de se gerar dados de um processo ARFIMA. Ao utilizar-se o pacote estatístico S-plus o analista deve ficar precavido de que este não possui um bom gerador, enquanto que ao se utilizar uma aproximação  $MA(q)$  para gerar deste mesmo processo obtém-se bons resul-

tados, além disso este método pode ser utilizado para gerar processos ARFIMA com erros t-Student e hiperbólico generalizado. As duas aplicações deste trabalho mostraram que a estimação está de acordo com o encontrado na literatura sobre o assunto. As prioris com ponto de massa para o parâmetro  $d$  tem um importante papel na identificação da longa dependência como foi visto pela probabilidade a posteriori deste parâmetro. O modelo com erros t-Student mostraram melhor ajuste que os com erro normal para os dados financeiros apresentados neste trabalho.

Um estudo mais detalhado do comportamento da aproximação em cada um dos casos descritos neste trabalho devem ser melhor avaliados, assim como o procedimento para gerar séries temporais que sigam um processo ARFIMA com hipótese de diferentes erros. A estimação sob a hipótese de erro hiperbólico generalizado deve ser melhor trabalhada, pois esta distribuição possui um parâmetro que caracteriza assimetria da distribuição e isto é uma ferramenta relevante para uma distribuição. Os procedimentos aplicados nesta dissertação podem ser estendidos para os modelos de volatilidade estocástica e modelos que detectam mudanças no parâmetro de longa dependência  $d$ . Uma outra possibilidade é utilizar o filtro de Kalman para estimar os parâmetros do modelo como em (Chan & Palma, 1998).

# Referências Bibliográficas

- Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London, Series A*, 353, 401–419.
- Beran, J. (1994). *Statistics for long memory process*. New York: Chapman and Hall.
- Beran, J., & Terrin, N. (1996). Testing for a change of the long-memory parameter. *Biometrika*, 83, 627–638.
- Box, G., & Jenkins, G. (1976). *Time series analysis forecasting and control*. San Francisco: Holden-Day.
- Chan, N., & Palma, W. (1998). State space modeling of long-memory processes. *The Annals of Statistics*, 26(2), 719–740.
- Crato, N. (2001). *Aplicações de modelos de memória longa*. Minas Gerais: ABE.
- Eberlein, E., & Hammerstein, E. (2002, September). *Generalized hyperbolic and inverse gaussian distributions: limiting cases and approximation of processes*.
- Eberlein, E., & Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1(3), 281–299.

- Gelfand, A., & Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.
- Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series model. *Journal of Time Series Analysis*, *4*(4), 221–238.
- Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, *14*, 227–238.
- Granger, C., & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*, 97–109.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, *68*(1), 165–176.
- Jørgensen, S. (1982). *Statistical properties of the generalized inverse gaussian distribution*. New York: Springer-Verlag.
- Koop, G., Ley, E., Osiewalski, J., & Steel, M. (1997). Bayesian analysis of long memory and persistence using arfima models. *Journal of Econometrics*, *76*(2), 149–169.
- Li, W., & McLeod, A. (1986). Fractional time series modelling. *Biometrika*, *73*(1), 217–221.
- Metropolis, N., Rosenbluth, A., Teller, A., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1092.



- Neal, R. (1997). *Markov chain monte carlo method based on 'slicing' the density function* (Tech. Rep.). Toronto, Canada. (<http://www.cs.utoronto.ca/~radford>)
- Pai, J., & Ravishanker, N. (1996). Bayesian modelling of arfima processes by markov chain monte carlo methods. *Journal of Forecasting*, *15*, 63–82.
- Pai, J., & Ravishanker, N. (1998). Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journal of Time Series Analysis*, *19*(1), 99–112.
- Petris, G. (1997). *Bayesian analysis of long memory time series*. Unpublished doctoral dissertation, Duke University.
- Reisen, V., & Toscano, E. (2000). Parametric and semi-parametric estimation of stationary univariate arfima models. *Brazilian Journal of Probability and Statistics*, *14*(2), 167–206.
- Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, *53*, 165–188.
- Spiegelhalter, D., Best, N. G., Carlin, B. P., & Linde, A. (2001). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Serie B*, *64*, 583–639.
- Tierney. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*, 1701–1786.
- West, M., & Huerta, G. (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, Series B*, *61*(4), 881–899.

# Apêndice A

## Distribuições Generalizadas

### A.1 Gaussiana Inversa Generalizada (GIG)

A distribuição  $GIG(\lambda, \delta, \gamma)$  tem a seguinte função de densidade de probabilidade

$$p(x | \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\gamma\delta)} x^{\lambda-1} \exp\left[-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right] \mathbf{I}_{\{x>0\}}, \quad (\text{A.1})$$

onde  $K_\lambda(\cdot)$  é a função de Bessel modificada do terceiro tipo com índice  $\lambda$ . O domínio de variação dos parâmetros em (A.1) são dados por

$$\delta \geq 0, \quad \gamma > 0, \quad \text{se } \lambda > 0,$$

$$\delta > 0, \quad \gamma > 0, \quad \text{se } \lambda = 0,$$

$$\delta > 0, \quad \gamma \geq 0, \quad \text{se } \lambda < 0.$$

Quando  $\lambda > 0$  e  $\delta = 0$ , (A.1) se reduz a uma densidade Gama e se  $\lambda < 0$  e  $\gamma = 0$  então (A.1) se reduz a uma densidade Gama Inversa.

A distribuição (A.1) é unimodal e com ponto de moda

$$x^* = \begin{cases} \frac{\lambda - 1 + \sqrt{(\lambda - 1)^2 + \delta^2 \gamma^2}}{\gamma^2} & , \text{ se } \gamma > 0, \\ \frac{\delta^2}{2(1 - \lambda)} & , \text{ se } \gamma = 0. \end{cases}$$

e além disso se  $\lambda \leq 1$  a distribuição é log-côncava.

Por causa da forma exponencial da densidade (A.1), a função geradora de momentos da distribuição  $\text{GIG}(\lambda, \delta, \gamma)$  com  $\delta > 0$  e  $\gamma > 0$  é simplesmente a razão das constantes de normalização correspondentes aos parâmetros  $(\lambda, \delta, \sqrt{\gamma^2 - 2u})$  e  $(\lambda, \delta, \gamma)$ :

$$M_x(t) = \int_0^\infty e^{tx} p(x | \lambda, \delta, \gamma) dx = \left( \frac{\gamma^2}{\gamma^2 - 2t} \right)^{\frac{\lambda}{2}} \frac{K_\lambda(\delta \sqrt{\gamma^2 - 2t})}{K_\lambda(\delta \gamma)}$$

com a restrição que  $2t < \gamma^2$ . Assim, os momentos  $\mu^k = \text{E}X^k$  da distribuição (A.1) são

$$\mu^k = \frac{K_{\lambda+k}(\delta \gamma)}{K_\lambda(\delta \gamma)} \left( \frac{\delta}{\gamma} \right)^k, \quad k \in \mathfrak{R}.$$

Se  $\delta = 0$  e  $\lambda > 0$  então

$$\mu^k = \begin{cases} \frac{\Gamma(\lambda + k)}{\Gamma(\lambda)} \left( \frac{2}{\gamma^2} \right)^k, & \text{se } k > -\lambda \\ \infty & \text{se } k \leq -\lambda \end{cases}$$

e se  $\gamma = 0$  e  $\lambda < 0$  então

$$\mu^k = \begin{cases} \frac{\Gamma(-\lambda - k)}{\Gamma(-\lambda)} \left( \frac{\delta^2}{2} \right)^k, & \text{se } k < -\lambda \\ \infty & \text{se } k \geq -\lambda \end{cases}$$

Portanto, verifica-se com certa facilidade que a variância de (A.1) é

$$\text{Var}(X) = \left( \frac{\delta}{\gamma} \right)^2 \left[ \frac{K_{\lambda+2}(\delta \gamma)}{K_\lambda(\delta \gamma)} - \left( \frac{K_{\lambda+1}(\delta \gamma)}{K_\lambda(\delta \gamma)} \right)^2 \right].$$

Entre algumas formulas de convoluções e transformadas da GIG de (A.1), temos:

- $X^{-1} \sim \text{GIG}(-\lambda, \gamma, \delta)$
- $cX \sim \text{GIG}(\lambda, c\delta, c^{-1}\gamma)$ , se  $c > 0$
- $\text{GIG}(-\frac{1}{2}, \delta_1, \gamma) + \text{GIG}(-\frac{1}{2}, \delta_2, \gamma) = \text{GIG}(-\frac{1}{2}, (\delta_1 + \delta_2)^2, \gamma)$
- $\text{GIG}(-\lambda, \delta, \gamma) + \text{GIG}(\lambda, 0, \gamma) = \text{GIG}(\lambda, \delta, \gamma)$ , se  $\lambda > 0$

- $\text{GIG}(-\frac{1}{2}, \delta_1, \gamma) + \text{GIG}(\frac{1}{2}, \delta_2, \gamma) = \text{GIG}(\frac{1}{2}, (\delta_1 + \delta_2)^2, \gamma)$
- $\text{GIG}(\lambda_1, 0, \gamma) + \text{GIG}(\lambda_2, 0, \gamma) = \text{GIG}(\lambda_1 + \lambda_2, 0, \gamma)$

Maiores detalhes sobre a distribuição GIG e suas propriedades estatísticas podem ser encontradas em Jørgensen (1982).

## A.2 Hiperbólica Generalizada (HG)

A distribuição  $\text{HG}(\lambda, \alpha, \beta, \delta, \mu)$  tem a seguinte função de densidade de probabilidade

$$p(x | \lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}} \exp[\beta(x - \mu)]}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^\lambda K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})} [(x - \mu)^2 + \delta^2]^{(\lambda - \frac{1}{2})/2} \times K_{\lambda - \frac{1}{2}}(\alpha \sqrt{(x - \mu)^2 + \delta^2}), \quad (\text{A.2})$$

onde  $K_\lambda(\cdot)$  é a função de Bessel modificada do terceiro tipo com índice  $\lambda$ . Os parâmetros são tais que:

- $\alpha > 0$  e determina a forma;
- $0 \leq |\beta| < \alpha$  a assimetria;
- $\mu \in \mathfrak{R}$  é parâmetro de locação;
- $\delta > 0$  que serve como escala comparado a  $\sigma$  em modelos Gaussianos; e
- $\lambda \in \mathfrak{R}$  caracteriza certas subclasses e influência consideravelmente o tamanho da massa contida na cauda da distribuição.

Dois possíveis parametrizações alternativas podem ser:

$$\rho = \frac{\beta}{\alpha}, \quad \zeta = \delta \sqrt{\alpha^2 - \beta^2} \quad \text{ou} \quad \chi = \rho \xi, \quad \xi = (1 + \zeta)^{-\frac{1}{2}}$$

A função geradora de momentos da distribuição  $HG(\lambda, \alpha, \beta, \delta, \mu)$  com a restrição que  $|\beta - t| < \alpha$  é

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} p(x | \lambda, \alpha, \beta, \delta, \mu) dx = e^{t\mu} \left( \frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + t)^2} \right)^{\frac{\lambda}{2}} \frac{K_{\lambda}(\delta\sqrt{\alpha^2 - (\beta + t)^2})}{K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})}$$

Assim, a média  $\theta = EX$  e a variância  $\tau^2 = \text{Var}(X)$  da distribuição (A.2) são, respectivamente,

$$\theta = \mu + \frac{\beta\delta^2}{\delta\sqrt{\alpha^2 - \beta^2}} \frac{K_{\lambda+1}(\delta\sqrt{\alpha^2 - \beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})}$$

e

$$\tau^2 = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}} \frac{K_{\lambda+1}(\delta\sqrt{\alpha^2 - \beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})} + \frac{\beta^2\delta^2}{(\alpha^2 - \beta^2)} \left[ \frac{K_{\lambda+2}(\delta\sqrt{\alpha^2 - \beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2 - \beta^2})} - \frac{K_{\lambda+1}^2(\delta\sqrt{\alpha^2 - \beta^2})}{K_{\lambda}^2(\delta\sqrt{\alpha^2 - \beta^2})} \right]$$

Maiores detalhes sobre a distribuição HG e seus casos limites podem ser encontrados em Eberlein e Hammerstein (2002).

# Apêndice B

## Função de Bessel do Terceiro

### Tipo

A função de Bessel do terceiro tipo com índice  $\lambda \in \Re$  é denotada por  $K_\lambda(\cdot)$ , é uma função integral, que dentre as várias representações, tem-se

$$K_\lambda(u) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp \left[ -\frac{1}{2} u (y + y^{-1}) \right] dy$$

Algumas propriedades desta função são:

- $K_\lambda(u) = K_{-\lambda}(u)$
- $K_{\lambda+1}(u) = \frac{2\lambda}{u} K_\lambda(u) + K_{\lambda-1}(u)$
- $K'_{\lambda-1}(u) = -\frac{1}{2} [K_{\lambda-1}(u) + K_{\lambda+1}(u)]$
- $K_{n+\frac{1}{2}}(u) = \sqrt{\frac{\pi}{2u}} e^{-u} \left[ 1 + \sum_{i=1}^n \frac{(n+i)!}{(n-i)!i!} (2u)^{-i} \right]$ , se  $\lambda = n + \frac{1}{2}$  e  $n = 0, 1, 2, \dots$

Em casos limites onde  $u \rightarrow 0$  tem-se

- $K_\lambda(u) \simeq \frac{1}{2} \Gamma(\lambda) \left( \frac{u}{2} \right)^{-\lambda}$ , se  $\lambda > 0$

- $K_\lambda(u) \simeq \frac{1}{2}\Gamma(-\lambda) \left(\frac{u}{2}\right)^\lambda$ , se  $\lambda < 0$
- $K_0(u) \simeq -\ln(u)$ .

No caso que  $u \rightarrow \infty$  tem-se

- $K_\lambda(u) \simeq \sqrt{\frac{\pi}{2u}} e^{-u}$ .

Para  $\lambda \rightarrow \infty$  tem-se a seguinte relação assintótica

- $K_\lambda(u) \simeq 2^\lambda \lambda^{\lambda-\frac{1}{2}} e^{-\lambda} u^{-\lambda\sqrt{\frac{\pi}{2}}}$  que pode ser vista como análoga a fórmula de Stirling para a função gama.

# Apêndice C

## Distribuições de Mistura

### C.1 Distribuição t-Student como uma mistura Normal-Gamma Inversa

A distribuição t-Student pode ser obtida como uma mistura de escala entre uma distribuição Normal (N) e uma Gamma Inversa (GI).

**Lema:** Considere que a distribuição da variável aleatória  $X$  condicionada em  $\mu$  e  $\sigma^2$  seja

$$(X | \mu, \delta^2) \sim N(\mu, \delta^2)$$

e que a distribuição da variável aleatória  $\delta^2$  condicionada em  $\nu$  e  $\sigma^2$  seja

$$(\delta^2 | \nu, \sigma^2) \sim GI\left(\frac{\nu}{2}, \frac{\nu\sigma^2}{2}\right)$$

então,  $(X | \mu, \sigma^2, \nu) \sim t_\nu(\mu, \sigma^2)$

**Prova:** Temos que

$$p(x | \mu, \delta^2) = (2\pi\delta^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\delta^2}(x - \mu)^2\right]$$



e

$$p(\delta^2 | \nu, \sigma^2) = \frac{\left(\frac{\nu\sigma^2}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} (\delta^2)^{-\frac{\nu}{2}-1} \exp\left[-\frac{\nu\sigma^2}{2\delta^2}\right], \quad \delta^2 > 0.$$

Além disso,

$$p(x | \mu, \sigma^2, \nu) = \int_0^\infty p(x, \delta^2 | \mu, \sigma^2, \nu) d\delta^2 = \int_0^\infty p(x | \mu, \delta^2) p(\delta^2 | \nu, \sigma^2) d\delta^2.$$

Logo,

$$\begin{aligned} p(x | \mu, \sigma^2, \nu) &\propto \int_0^\infty (\delta^2)^{-\left(\frac{\nu+1}{2}\right)-1} \exp\left[-\frac{1}{\delta^2} \left[\frac{(x-\mu)^2 + \nu\sigma^2}{2}\right]\right] d\delta^2 \\ &\propto \left(\nu + \frac{(x-\mu)^2}{\sigma^2}\right)^{-\left(\frac{\nu+1}{2}\right)} \end{aligned}$$

Portanto, tem-se que  $(x | \mu, \sigma^2, \nu) \sim t_\nu(\mu, \sigma^2)$

## C.2 Distribuição Hiperbólica Generalizada como uma mistura Normal-Gaussiana Inversa Generalizada

A distribuição Hiperbólica Generalizada (HG) pode ser obtida como uma mistura média-variância entre uma distribuição Normal (N) e uma Gaussiana Inversa Generalizada (GIG).

**Lema:** Considere que a distribuição da variável aleatória  $X$  condicionada em  $\mu$ ,  $\beta$  e  $\sigma^2$  seja

$$(X | \mu, \beta, \sigma^2) \sim N(\mu + \beta\sigma^2, \sigma^2)$$

e que a distribuição da variável aleatória  $\sigma^2$  condicionada em  $\lambda$ ,  $\delta$  e  $\alpha$  seja

$$(\sigma^2 | \lambda, \delta, \alpha, \beta) \sim GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})$$

então,  $(X | \lambda, \alpha, \beta, \delta, \mu) \sim GH(\lambda, \alpha, \beta, \delta, \mu)$

**Prova:** Temos que

$$p(x | \mu, \beta, \sigma^2) = (2\pi\sigma^2)^{\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma^2}(x - \mu - \beta\sigma^2)^2 \right]$$

e

$$p(\sigma^2 | \lambda, \delta, \alpha, \beta) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} (\sigma^2)^{\lambda-1} \exp \left[ -\frac{1}{2}(\delta^2\sigma^{-2} + (\alpha^2 - \beta^2)\sigma^2) \right],$$

onde  $K_\lambda(\cdot)$  é a função de Bessel modificada do terceiro tipo com índice  $\lambda$ . Além disso,

$$p(x | \lambda, \alpha, \beta, \delta, \mu) = \int_0^\infty p(x | \mu, \beta, \sigma^2) p(\sigma^2 | \lambda, \delta, \alpha, \beta) d\sigma^2$$

Logo,

$$p(x | \lambda, \alpha, \beta, \delta, \mu) = c \int_0^\infty (\sigma^2)^{(\lambda-\frac{1}{2})-1} \exp \left[ -\frac{1}{2} \left( \frac{1}{\sigma^2} ((x - \mu)^2 + \delta^2) + \alpha^2 \sigma^2 \right) \right] d\sigma^2 \quad (\text{C.1})$$

onde

$$c = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}} \exp[\beta(x - \mu)]}{2\sqrt{2\pi}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}$$

O resultado da integral em (C.1) é o inverso da constante de normalização de uma  $GIG(\lambda - \frac{1}{2}, \sqrt{((x - \mu)^2 + \delta^2)}, \alpha)$ . Portanto,

$$p(x | \lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}} \exp[\beta(x - \mu)]}{\sqrt{2\pi}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} \times \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{(x - \mu)^2 + \delta^2}) ((x - \mu)^2 + \delta^2)^{(\lambda-\frac{1}{2})/2}}{\alpha^{\lambda-\frac{1}{2}}},$$

isto é,  $(X | \lambda, \alpha, \beta, \delta, \mu) \sim HG(\lambda, \alpha, \beta, \delta, \mu)$

# Apêndice D

## Critério de informação do desvio - DIC

O DIC - abreviação do inglês *deviance information criterion* - serve para avaliar a complexidade de um modelo e compara-lo a outros diferentes.

O DIC é dado por  $DIC = \bar{D} + p_D = D(\bar{\xi}) + 2p_D$ , onde

- $\bar{D}$  é a média a posteriori do desvio. O desvio é definido como  $D = -2\log(p(Y_T | \xi))$
- $D(\bar{\xi})$  é uma estimativa pontual do desvio obtido ao substituir as médias dos parâmetros a posteriori ( $\xi$ ), assim  $D(\bar{\xi}) = -2\log(p(Y_T | \bar{\xi}))$
- $p_D$  é o número efetivo de parâmetros dado por  $p_D = \bar{D} - D(\bar{\xi})$

Quanto menor o DIC melhor será o modelo, no mesmo sentido que o AIC. Vale resaltar que o DIC só pode ser comparado quando se trabalha com um mesmo conjunto de dados.

Maiores detalhes podem ser encontrados em (Spiegelhalter et al., 2001)

# Apêndice E

## Resultados de simulação

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,489 (0,014)	58	1,24 (0,36)	45	0,491 (0,010)	63	1,10 (0,17)	73
45	0,489 (0,015)	55	1,26 (0,40)	42	0,492 (0,010)	62	1,11 (0,19)	68
40	0,490 (0,014)	53	1,29 (0,45)	39	0,492 (0,010)	59	1,12 (0,21)	66
35	0,490 (0,015)	51	1,32 (0,50)	35	0,492 (0,010)	57	1,14 (0,23)	62
30	0,490 (0,014)	48	1,37 (0,57)	35	0,493 (0,010)	55	1,16 (0,26)	59
25	0,491 (0,014)	46	1,44 (0,68)	32	0,493 (0,010)	51	1,19 (0,30)	54
20	0,492 (0,013)	45	1,54 (0,83)	29	0,494 (0,090)	49	1,23 (0,37)	50
15	0,492 (0,014)	42	1,71 (1,08)	23	0,495 (0,080)	44	1,30 (0,47)	42
10	0,493 (0,013)	36	2,03 (1,57)	14	0,496 (0,006)	37	1,43 (0,69)	34
5	0,494 (0,012)	31	2,94 (2,93)	4	0,497 (0,004)	27	1,84 (1,31)	24

Tabela E.1: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,49$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,415 (0,033)	77	1,02 (0,05)	90	0,426 (0,029)	71	1,01 (0,05)	99
45	0,416 (0,034)	75	1,02 (0,05)	90	0,427 (0,030)	68	1,01 (0,05)	99
40	0,417 (0,034)	71	1,02 (0,05)	90	0,428 (0,030)	69	1,01 (0,05)	97
35	0,418 (0,036)	70	1,03 (0,05)	90	0,430 (0,031)	66	1,01 (0,05)	97
30	0,420 (0,036)	68	1,03 (0,05)	88	0,432 (0,029)	64	1,01 (0,05)	96
25	0,421 (0,036)	68	1,03 (0,06)	84	0,434 (0,032)	62	1,01 (0,05)	93
20	0,425 (0,037)	64	1,04 (0,06)	82	0,438 (0,032)	57	1,02 (0,05)	93
15	0,426 (0,039)	61	1,05 (0,06)	79	0,443 (0,029)	51	1,02 (0,05)	91
10	0,431 (0,041)	60	1,08 (0,08)	68	0,452 (0,032)	37	1,03 (0,08)	89
5	0,438 (0,041)	51	1,15 (0,13)	39	0,468 (0,027)	20	1,05 (0,07)	79

Tabela E.2: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,40$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,360 (0,030)	85	1,01 (0,05)	92	0,370 (0,028)	80	1,01 (0,05)	99
45	0,360 (0,031)	86	1,01 (0,05)	92	0,371 (0,029)	77	1,01 (0,05)	98
40	0,361 (0,031)	83	1,01 (0,05)	92	0,372 (0,029)	77	1,01 (0,05)	99
35	0,362 (0,032)	81	1,01 (0,05)	93	0,374 (0,030)	76	1,01 (0,05)	97
30	0,364 (0,033)	78	1,01 (0,05)	91	0,376 (0,031)	73	1,01 (0,05)	97
25	0,365 (0,033)	79	1,02 (0,05)	91	0,378 (0,032)	70	1,01 (0,05)	95
20	0,368 (0,034)	76	1,02 (0,05)	91	0,381 (0,033)	67	1,01 (0,05)	94
15	0,369 (0,037)	73	1,03 (0,05)	89	0,387 (0,035)	64	1,01 (0,05)	93
10	0,373 (0,040)	69	1,04 (0,06)	83	0,396 (0,037)	56	1,02 (0,05)	93
5	0,379 (0,041)	65	1,08 (0,07)	64	0,418 (0,040)	30	1,03 (0,05)	87

Tabela E.3: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,35$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,100 (0,024)	97	1,00 (0,05)	94	0,106 (0,025)	95	1,00 (0,05)	99
45	0,100 (0,024)	97	1,00 (0,05)	93	0,106 (0,025)	94	1,00 (0,05)	99
40	0,100 (0,024)	98	1,00 (0,05)	93	0,106 (0,025)	95	1,00 (0,05)	99
35	0,100 (0,024)	98	1,00 (0,05)	93	0,106 (0,025)	95	1,00 (0,05)	98
30	0,100 (0,024)	98	1,00 (0,05)	94	0,106 (0,025)	93	1,00 (0,05)	97
25	0,100 (0,024)	98	1,00 (0,05)	94	0,107 (0,025)	94	1,00 (0,05)	96
20	0,101 (0,024)	97	1,00 (0,05)	93	0,107 (0,025)	93	1,00 (0,05)	95
15	0,100 (0,025)	98	1,00 (0,05)	93	0,107 (0,025)	93	1,00 (0,05)	94
10	0,101 (0,025)	97	1,00 (0,05)	92	0,108 (0,026)	93	1,00 (0,05)	93
5	0,100 (0,025)	96	1,00 (0,05)	92	0,110 (0,027)	90	1,00 (0,05)	92

Tabela E.4: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,10$  no pacote estatístico S-Plus, onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95

m	MA				AR			
	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%	$\tilde{d}$	C%	$\tilde{\sigma}^2$	C%
50	0,479 (0,015)	93	1,04 (0,06)	81	0,484 (0,010)	98	1,01 (0,05)	97
45	0,479 (0,016)	90	1,04 (0,06)	81	0,484 (0,010)	98	1,02 (0,05)	97
40	0,481 (0,015)	92	1,05 (0,06)	78	0,485 (0,010)	96	1,02 (0,05)	96
35	0,480 (0,016)	88	1,05 (0,07)	74	0,486 (0,010)	96	1,02 (0,05)	95
30	0,481 (0,016)	86	1,06 (0,07)	72	0,486 (0,010)	94	1,02 (0,05)	94
25	0,482 (0,016)	86	1,07 (0,08)	66	0,487 (0,010)	93	1,03 (0,05)	91
20	0,483 (0,017)	84	1,09 (0,09)	56	0,488 (0,009)	91	1,03 (0,06)	87
15	0,484 (0,015)	84	1,12 (0,12)	48	0,490 (0,009)	89	1,04 (0,06)	83
10	0,486 (0,016)	76	1,17 (0,16)	33	0,492 (0,007)	80	1,06 (0,08)	74
5	0,488 (0,015)	72	1,34 (0,29)	10	0,495 (0,005)	64	1,12 (0,13)	48

Tabela E.5: Resultados da simulação das séries temporais geradas pelo processo ARFIMA(0,d,0) com  $d = 0,49$  pela estrutura MA(10000), onde  $\tilde{d} = \frac{1}{200} \sum_1^{200} E(d | Y_T)$ ,  $\tilde{\sigma}^2 = \frac{1}{200} \sum_1^{200} E(\sigma^2 | Y_T)$  e C% é a taxa de cobertura em porcentagem do intervalo de credibilidade de 0,95