



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
DEPARTAMENTO DE ESTATÍSTICA



# Estimação em Processos com Longa Dependência e Sazonalidade

Autor: Lisiane de Souza Nunes de Moura  
Orientador: Professor Dr. Cleber Bisognin

Porto Alegre, 09 de Dezembro de 2008.

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática  
Departamento de Estatística

# Estimação em Processos com Longa Dependência e Sazonalidade

Autor: Lisiane de Souza Nunes de Moura

Monografia apresentada para obtenção  
do grau de Bacharel em Estatística.

Banca Examinadora:  
Professor Cleber Bisognin  
Professor Fernando Augusto Boeira Sabino da Silva

Porto Alegre, 09 de Dezembro de 2008.

---

*Dedico este trabalho aos meus pais, que sempre estiveram presentes e sempre me apoiaram me dando forças pra seguir em frente. Dedico também a todos os colegas e, sobre tudo, amigos que estiveram comigo nesta jornada passando por tantas dificuldades, mas também por tantos momentos de alegria e prazer. E dedico ao meu amor, que me acompanhou todo este tempo de faculdade e sempre soube a coisa certa a dizer quando eu mais precisava.*

## Resumo

Neste trabalho analisamos alguns processos com as propriedades de longa dependência e sazonalidade. Nosso estudo tem por objetivo principal estudar alguns estimadores do parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , onde  $s$  é a sazonalidade. Apresentamos diversos estimadores na classe dos semi-paramétricos para estimar o parâmetro de diferenciação sazonal  $D$ . Na classe paramétrica, apresentamos um estimador de máxima verossimilhança para o mesmo parâmetro. Comparamos nova metodologia de estimação para o parâmetro  $D$ , os chamados estimadores robustos. Realizamos a análise da série temporal dos níveis mensais do rio Nilo, em Aswan.

# Sumário

1. Introdução.....	1
2. Processos $SARFIMA(p,d,q)\times(P,D,Q)_s$ .....	3
2.1 Definição e Propriedades .....	3
3. Estimaco Clssica e Robusta .....	8
3.1 Estimador $GPH$ Clssico e Robusto .....	12
3.2 Estimador $R$ Clssico e Robusto.....	14
3.3 Estimador $SPR$ Clssico e Robusto .....	14
3.4 Estimador $SR$ Clssico e Robusto.....	16
3.5 Estimador $GPHT$ Clssico e Robusto .....	16
3.6 Estimador $W$ .....	17
4. Simulao de Monte Carlo .....	20
4.1 Resultados .....	21
5. Aplicao.....	29
5.1 Srie Temporal dos Nveis Mensais do Rio Nilo .....	29
6. Concluses.....	32
Referncias Bibliogrficas .....	33
Anexo A .....	35
A.1 Estimadores da Funo Densidade Espectral .....	35
A.2 Esperana e Varincia Condicionais.....	36
Anexo B .....	38

# 1. Introdução

Nas últimas duas décadas, tem ocorrido grande interesse no estudo de séries temporais com propriedade de *longa dependência*. *Longa dependência* pode ser definida, no domínio do tempo, como a característica na qual a função de autocorrelação não é absolutamente convergente. No domínio da frequência, a *longa dependência* aparece como a característica na qual a função densidade espectral torna-se ilimitada para frequências muito pequenas, em torno de zero.

O hidrologista Harold E. Hurst, em 1951, foi o primeiro a apresentar um estudo de séries temporais com esta característica, enquanto investigava as séries temporais dos níveis mensais do rio Nilo.

Um modelo muito utilizado para estudar séries temporais com longa dependência é o *modelo auto-regressivo fracionalmente integrado média móvel*, primeiramente proposto por Granger e Joyeux (1980) e Hosking (1981 e 1984), o chamado processo  $ARFIMA(p,d,q)$ . Diversas séries temporais, além da propriedade de longa dependência, apresentam também sazonalidade. Esta é caracterizada pela repetição que ocorre em certo período de tempo fixo. O *modelo auto-regressivo fracionalmente integrado média móvel com sazonalidade  $s$* , denotado por  $SARFIMA(p,d,q) \times (P,D,Q)_s$ , foi proposto por Porter-Hudak (1990) com o objetivo de estudar séries temporais com estas duas características.

Diversas séries temporais, além da propriedade de longa dependência, apresentam também sazonalidade. Para estudar estas séries temporais Porter-Hudak (1990) propõe o modelo  $SARFIMA(0,D,0)_s$ , aplicando o mesmo para analisar a série temporal “*US monetary aggregates*”. Para a estimação do parâmetro de diferenciação sazonal  $D$ , Porter-Hudak (1990) utiliza o estimador proposto por Gewek e Porter-Hudak (1983). Ray (1993) analisa os modelos rígidos e flexíveis para os processos  $SARFIMA(0,D,0)_s$ , com sazonalidade  $s = 4$ . O autor trabalhou com estimador da função periodograma proposto por Gewek e Porter-Hudak (1983), em torno de todas as frequências sazonais. Ooms (1995) analisa um conjunto de dados reais sazonais utilizando o estimador baseado na função periodograma, o estimador de máxima verossimilhança aproximado, proposto por Fox e Taqqu (1986) e o estimador de máxima verossimilhança exato, proposto por Sowell (1992). Montanari et al. (2000) analisa a série temporal do fluxo mensal do rio Nilo em Aswan, utilizando o estimador de máxima verossimilhança

aproximado para estimar os parâmetros do processo  $SARFIMA(p,d,q)\times(P,D,Q)_s$ . Peiris e Singh (1996) apresentam alguns resultados de previsão  $h$ -passos à frente para alguns casos particulares deste modelo. Brietzke et al. (2005) demonstram diversas propriedades dos processos  $SARFIMA(0,D,0)_s$ , incluindo estacionariedade, inversibilidade, funções de autocovariância e autocovariância parcial do processo. Neste artigo apresentam uma fórmula fechada para o algoritmo de Durbin-Levinson o qual relaciona a função de autocorrelação e de autocorrelação parcial do processo estocástico. Bisognin e Lopes (2007) estende o método de geração dos processos  $ARFIMA(p,d,q)$ , proposto por Hosking (1984), para os processos  $SARFIMA(0,D,0)_s$  e apresentam alguns resultados de previsão para os mesmos. Ademais, apresentam a média e a variância condicionais destes processos, que são importantes para geração e previsão com os processos  $SARFIMA(0,D,0)_s$ .

Nosso objetivo neste trabalho é estudar os processos  $SARFIMA(0,D,0)_s$  considerando a estimação do parâmetro de diferenciação sazonal, denotado por  $D$ , com  $s$  conhecido.

Para estimar o parâmetro  $D$ , consideramos cinco estimadores semi-paramétricos e um paramétrico. Na classe dos estimadores semi-paramétricos utilizamos os estimadores propostos por Geweke e Porter-Hudak (1983), Robinson (1995), Reisen (1994), Hurvich e Ray (1995) e Velasco (1999), utilizando as metodologias clássica  $MQ$  e robustas  $MM$  e  $MQP$ , e o estimador paramétrico proposto por Fox e Taquq (1986). Estes serão comparados entre si e avaliaremos a média, o viés, a variância e o erro quadrático médio de cada um.

Na seção 2 apresentamos as definições e propriedades do processo  $SARFIMA(0,D,0)_s$ . Na seção 3 apresentamos as duas metodologias de estimação robusta utilizadas neste trabalho, assim como os cinco estimadores semi-paramétricos na metodologia clássica e nas duas metodologias robustas e um estimador paramétrico. Na seção 4 se encontra os passos utilizados para simulação dos processos e as tabelas de resultado dos 16 estimadores apresentados neste trabalho (cinco semi-paramétricos, cada um com três metodologias, mais o estimador paramétrico). Na seção 5 utilizamos todos estes estimadores para estudar uma série real. Na seção 6, finalmente, apresentamos nossas conclusões com relação aos estimadores que estão sendo comparados.

## 2. Processos $SARFIMA(p,d,q) \times (P,D,Q)_s$

A seguir, definiremos os *processos sazonais auto-regressivos fracionariamente integrados média móvel*, denotados por  $SARFIMA(p,d,q) \times (P,D,Q)_s$ , os quais são uma extensão dos modelos  $ARFIMA(p,d,p)$ , propostos por Granger e Joyeux (1980) e Hosking (1981). Em seguida apresentamos os processos  $SARFIMA(0,D,0)_s$  juntamente com algumas das suas propriedades.

### 2.1 Definição e Propriedades

Nesta seção apresentaremos algumas definições e propriedades dos processos  $SARFIMA(p,d,q) \times (P,D,Q)_s$  e outras propriedades relacionadas. A seguir se encontra a definição dos processos ruído branco.

**Definição 2.1** O processo  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é dito ser um *ruído branco* com média zero e variância  $\sigma^2$ , denotado por  $\varepsilon_t \sim WN(0, \sigma^2)$ , se

$$E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2, \quad \gamma_\varepsilon(k) = \begin{cases} \sigma_\varepsilon^2, & k = 0 \\ 0, & k \neq 0 \end{cases}, \quad f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}.$$

A seguir apresentamos a definição da propriedade de *longa dependência*.

**Definição 2.2** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário. No domínio do tempo, se existe um número  $u \in (0,1)$  tal que:

$$\rho_X(k) \sim C_1 k^{-u}, \text{ quando } k \rightarrow \infty,$$

onde  $C_1 \neq 0$  e  $\rho_X(\cdot)$  é função de autocorrelação do processo, então  $\{X_t\}_{t \in \mathbb{Z}}$  possui a propriedade de *longa dependência*. Equivalentemente, no domínio da frequência, se existe um número real  $b \in (0,1)$  tal que:

$$f_X(\omega) \sim C_2 |\omega|^{-b}, \text{ quando } \omega \rightarrow 0,$$

onde  $C_2 > 0$  e  $f_X(\cdot)$  é a função densidade espectral do processo, então  $\{X_t\}_{t \in \mathbb{Z}}$  possui a propriedade de *longa dependência* (ver Bary, 1964).

**Observação 2.1** Em relação a Definição 2.2, observamos:

- i) A relação  $b = 1 - u$  é verdadeira;



ii) No domínio do tempo, quando  $u \in (0, 1)$ ,  $\rho_X(k)$  tende a zero tão lentamente que

$$\sum_{k \geq 0} |\rho_X(k)| \text{ diverge, onde } \rho_X(\cdot) \text{ é a função de autocorrelação do processo.}$$

**Definição 2.3** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estacionário para o qual existe um número real  $b \in (0, 1)$ , uma constante  $C_f > 0$  e uma frequência  $G \in [0, \pi]$  (ou um número finito de frequências) tal que

$$f_X(\omega) \sim C_f |\omega - G|^{-\frac{d}{2}} \text{ quando } \omega \rightarrow G.$$

Então  $\{X_t\}_{t \in \mathbb{Z}}$  é chamado processo estacionário com *longa dependência*.

A seguir definimos o operador diferença e o operador diferença sazonal.

**Definição 2.4** Para todo  $d > -1$ , definimos o *operador diferença*  $\nabla^d \equiv (1 - B)^d$ , através da expansão binomial

$$\nabla^d \equiv (1 - B)^d = \sum_{k \geq 0} \binom{d}{k} (-B)^k = 1 - dB - \frac{d(1-d)}{2!} B^2 - \dots,$$

$$\text{onde } \binom{d}{k} = \frac{\Gamma(1+d)}{\Gamma(1+k)\Gamma(1+d-k)}.$$

Para todo  $D > -1$ , definimos o *operador diferença sazonal*  $\nabla_s^D \equiv (1 - B^s)^D$ , onde  $s$  é a sazonalidade,  $s \in \mathbb{N} - \{0\}$ , através da expansão binomial

$$\nabla_s^D \equiv (1 - B^s)^D = \sum_{k \geq 0} \binom{D}{k} (-B^s)^k = 1 - DB^s - \frac{D(1-D)}{2!} B^{2s} - \dots,$$

$$\text{onde } \binom{D}{k} = \frac{\Gamma(1+D)}{\Gamma(1+k)\Gamma(1+D-k)},$$

na qual  $\Gamma(\cdot)$  é a função Gama definida por

$$(i) \quad \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

$$(ii) \quad \Gamma(x+1) = x\Gamma(x), \quad x \in \mathfrak{R}$$

Na definição a seguir apresentamos os processo  $SARFIMA(p, d, q) \times (P, D, Q)_s$ .

**Definição 2.5** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico satisfazendo a equação

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D(X_t - \mu) = \Theta(B^s)\theta(B)\varepsilon_t,$$

onde  $\mu$  é a média do processo,  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é o processo ruído branco,  $s \in \mathbb{N} - \{0\}$  é a sazonalidade,  $B$  é o operador *defasagem*, isto é,  $B^m x_t = x_{t-m}$ , para  $m \in \mathbb{N}$ ,  $\nabla^d$  e  $\nabla_s^D$  são, respectivamente, os operadores diferença e diferença sazonal,  $\phi(\cdot)$ ,  $\theta(\cdot)$ ,  $\Phi(\cdot)$  e  $\Theta(\cdot)$  são os polinômios de ordem  $p$  e  $q$ ,  $P$  e  $Q$ , respectivamente, definidos por

$$\begin{aligned}\phi(B) &= \sum_{j=0}^p -\phi_j B^j, & \theta(B) &= \sum_{i=0}^q -\theta_i B^i, \\ \Phi(B^s) &= \sum_{k=0}^P -\Phi_k B^{sk}, & \Theta(B^s) &= \sum_{l=0}^Q -\Theta_l B^{sl},\end{aligned}$$

onde  $\phi_j, 1 \leq j \leq p$ ,  $\theta_i, 1 \leq i \leq q$ ,  $\Phi_k, 1 \leq k \leq P$ , e  $\Theta_l, 1 \leq l \leq Q$  são constantes reais e  $\phi_0 = \Phi_0 = -1 = \theta_0 = \Theta_0$ . Então o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo *sazonal auto-regressivo fracionariamente integrado média móvel de ordem  $(p,d,q) \times (P,D,Q)_s$* , denotados por  $SARFIMA(p,d,q) \times (P,D,Q)_s$ , onde  $d$  e  $D$  são, respectivamente, o grau de diferenciação e o grau de diferenciação sazonal.

**Observação 2.2** Queremos estudar os processos  $SARFIMA(0,D,0)_s$ , que são um caso particular dos processos  $SARFIMA$  completo, cuja expressão é dada por:

$$(1-B^s)^D(X_t - \mu) = \varepsilon_t$$

O teorema a seguir apresenta as propriedades do processo  $SARFIMA(0,D,0)_s$  e a sua demonstração pode ser encontrada em Brietzke et al. (2005).

**Teorema 2.1** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $SARFIMA(0,D,0)_s$  dado pela equação acima, onde  $s$  é a sazonalidade,  $s$  fixo. Então seguem as seguintes afirmações

- (i) Quando  $D > -0,5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo inversível e apresenta-se como um processo auto-regressivo infinito, isto é,

$$\Pi(B^s)X_t = \sum_{k \geq 0} \pi_k X_{t-sk} = \varepsilon_t,$$

onde

$$\pi_k = \frac{-D(1-D)\cdots(k-D-1)}{k!} = \frac{\Gamma(k-D)}{\Gamma(k+1)\Gamma(-D)}.$$

Quando  $k \rightarrow \infty$ ,  $\pi_k \sim \frac{k^{-D-1}}{\Gamma(-D)}$ , isto é, são assintoticamente equivalentes.

- (ii) Quando  $D < 0,5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo estacionário e apresenta-se como um processo média móvel infinito, isto é,

$$X_t = \Psi(B^s)\varepsilon_t = \sum_{k \geq 0} \psi_k \varepsilon_{t-sk},$$

onde

$$\pi_k = \frac{D(1+D)\cdots(k+D-1)}{k!} = \frac{\Gamma(k+D)}{\Gamma(k+1)\Gamma(D)}.$$

Quando  $k \rightarrow \infty$ ,  $\psi_k \sim \frac{k^{D-1}}{\Gamma(D)}$ .

Nos itens a seguir, assumimos que  $D \in (-0,5; 0,5)$ .

- (iii) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função densidade espectral dada por

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \left[ 2 \operatorname{sen} \left( \frac{\omega s}{2} \right) \right]^{-2D}, \quad 0 < \omega \leq \pi.$$

Nas frequências sazonais, para  $\nu = 0, 1, \dots, \lfloor s/2 \rfloor$ , onde  $\lfloor x \rfloor$  denota a parte inteira de  $x$ ,

$$f_X \left( \frac{2\pi\nu}{s} + \omega \right) \sim \frac{\sigma_\varepsilon^2}{2\pi} (\omega s)^{-2D}, \quad \text{quando } \omega \rightarrow 0.$$

- (iv) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função de autocovariância e autocorrelação de ordem  $k$ ,  $k \in \mathbb{N}$ , dadas, respectivamente por

$$\gamma_X(sk) = \frac{(-1)^k \Gamma(1-2D)}{\Gamma(k-D+1)\Gamma(1-k-D)} \sigma_\varepsilon^2,$$

$$\gamma_X(sk + \zeta) = 0, \quad \text{para } \zeta = 1, \dots, s-1$$

$$\rho_X(sk) = \frac{\Gamma(k+D)\Gamma(1-D)}{\Gamma(1+k-D)\Gamma(D)},$$

e

$$\rho_X(sk + \zeta) = 0, \quad \text{para } \zeta = 1, \dots, s-1.$$

Quando  $k \rightarrow \infty$ ,  $\rho_X(sk) \sim \frac{\Gamma(1-D)}{\Gamma(D)} k^{2D-1}$ .

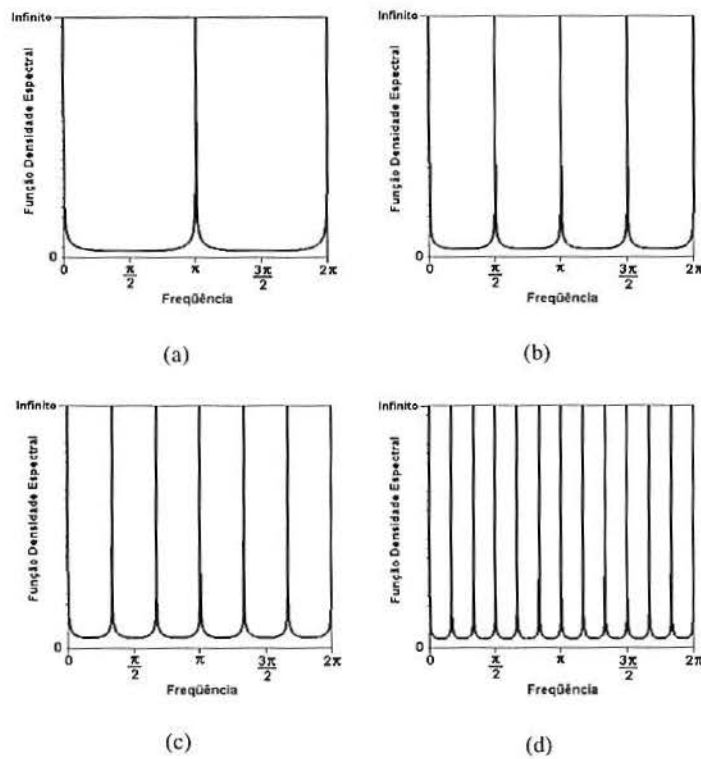
(v) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função de autocorrelação parcial dada por

$$\phi_X(sk, sk) = \frac{D}{k-D}, \text{ para todo } k \in \mathbb{N}$$

$$\phi_X(t, j) = \phi_X(t-1, j) - \phi_X(t, t)\phi_X(t-1, t-j), \quad j = 1, \dots, t-1$$

e  $\phi_X(sk + \zeta, sk + \zeta) = 0$ , para  $\zeta = 1, \dots, s-1$ .

A seguir um exemplo gráfico da função densidade espectral de um processo  $SARFIMA(0, D, 0)_s$ , onde é possível identificar que o número pico é igual à sazonalidade.



**Figura 1** Função densidade espectral do processo  $SARFIMA(0, D, 0)_s$ : (a)  $D = 0,3$ ;  $s = 2$ ; (b)  $D = 0,3$ ;  $s = 4$ ; (c)  $D = 0,3$ ;  $s = 6$ ; (d)  $D = 0,3$ ;  $s = 12$ .

### 3. Estimação Clássica e Robusta

Na literatura sobre processos estocásticos *SARFIMA* existem diversos métodos para estimar o parâmetro  $D$  de diferenciação sazonal. Estes métodos podem ser semi-paramétricos, paramétricos e não-paramétricos. Neste trabalho consideramos diversos estimadores na classe dos métodos semi-paramétricos e paramétricos. Também apresentamos uma nova metodologia de estimação para o parâmetro  $D$ , os chamados *estimadores robustos*.

A seguir, descrevemos os procedimentos de estimação através do método dos mínimos quadrados e da metodologia robusta. Inicialmente, definimos o modelo de regressão linear geral com  $l$  variáveis independentes  $X_k$ ,  $1 \leq k \leq l$ , e uma variável aleatória  $Y$  dependente, o qual possui a seguinte forma

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_l X_l + \epsilon$$

Assumindo  $g(n)$  independentes observações de  $Y$  associadas aos valores  $x_k$ , para  $k \in \{1, \dots, l\}$ , o modelo acima apresenta-se como

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \dots + \beta_l X_{1l} + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \dots + \beta_l X_{2l} + \epsilon_2 \\ &\vdots \\ Y_{g(n)} &= \beta_0 + \beta_1 X_{g(n)1} + \dots + \beta_l X_{g(n)l} + \epsilon_{g(n)}, \end{aligned}$$

onde os erros  $\epsilon_j$ , satisfazem as seguintes suposições, para todo  $j \in \{1, \dots, g(n)\}$ ,

- i)  $E(\epsilon_j) = 0$ ,
- ii)  $Var(\epsilon_j) = \sigma_\epsilon^2$ ,
- iii)  $Cov(\epsilon_l, \epsilon_k) = 0$ , se  $l \neq k$ .

O estimador de  $\beta = (\beta_0, \beta_1, \dots, \beta_l)$  pelo método dos mínimos quadrados, denotado neste trabalho por *MQ*, é o valor  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_l) \in \mathfrak{R}^{l+1}$  que minimiza a função perda

$$P_l(g(n)) = \sum_{j=1}^{g(n)} r_j^2,$$

onde os resíduos  $r_j$  são dados por

$$r_j = y_j - \hat{\beta}_0 - \hat{\beta}_1 X_{j1} - \dots - \hat{\beta}_l X_{jl}.$$

Para maiores detalhes sobre regressão linear múltipla referenciamos o leitor a Draper e Smith (1981).

Os estimadores obtidos através do processo *MQ*, sob hipótese de normalidade dos erros, são consistentes e tem mínima variância entre todos os estimadores não-viciados. A presença de *outliers* e pontos de alavanca, ou mesmo a perda da hipótese de normalidade dos erros são responsáveis por um considerável vício e ineficiência dos estimadores *MQ* (ver Huber, 1981 e Russeeuw e Leroy, 2003).

Outro procedimento utilizado para estimar o vetor de parâmetros  $\beta = (\beta_0, \beta_1, \dots, \beta_l)$  são os procedimentos de estimação robusta, os quais apresentam estimadores que não são fortemente afetados por *outliers*.

Os procedimentos de estimação robusta são indicados quando os dados contém uma certa porcentagem de *outliers*. Por isso, definimos o *ponto de ruptura* destes estimadores. Esta definição é uma versão finita do conceito de ponto de ruptura, introduzido por Donaho e Huber (1983).

Seja  $Z$  uma amostra de tamanho  $g(n)$ , dada por

$$Z = \{(1, x_{11}, \dots, x_{1l}, y_1), \dots, (1, x_{g(n)1}, \dots, x_{g(n)l}, y_{g(n)})\},$$

e  $T$  um estimador dos coeficientes de regressão  $\beta = (\beta_0, \beta_1, \dots, \beta_l)$ , isto é,  $T(Z) = \hat{\beta}$ .

Supomos que  $\tilde{m}$  valores arbitrários (aberrantes ou *outliers*) substituam  $\tilde{m}$  observações originais de  $Z$ , produzindo uma amostra contaminada  $Z'$ .

Denotamos por *vício*( $\tilde{m}; T, Z$ ) o vício máximo causado por tal contaminação, isto é,

$$vicio(\tilde{m}; T, Z) = \sup_{Z'} \|T(Z') - T(Z)\|,$$

onde o supremo é sobre todas as possíveis amostras  $Z'$  e  $\|\cdot\|$  representa a normal Euclidiana. Se o *vício*( $\tilde{m}; T, Z$ ) é infinito, isto significa que os  $\tilde{m}$  *outliers* podem ter um efeito arbitrariamente grande sobre  $T$ .

Assim, para amostras de tamanho  $g(n)$  finitas, o ponto de ruptura do estimador  $T$ , utilizando a amostra  $Z$ , é definido como

$$\epsilon^* := \epsilon^*(T, Z) = \min \left\{ \frac{m}{g(n)} ; \text{vício}(\tilde{m}; T, Z) \text{ é finito} \right\}.$$

Em outras palavras, o ponto de ruptura é a menor porcentagem de dados contaminados que pode fazer com que o estimador tome um valor alto e arbitrariamente anormal (ver Hampel, 1974). No caso dos estimadores *MQ*, Rowsseeuw e Leroy (2003) observam que um *outlier* é suficiente para que o estimador tenha um alto vício. Portanto seu *ponto de ruptura* é igual a

$$\epsilon^* = \frac{1}{g(n)},$$

o qual converge a zero, quando o tamanho da amostra cresce. Assim, podemos dizer que os estimadores obtidos pelo método *MQ* possuem ponto de ruptura de 0%. Isto novamente reflete a extrema sensibilidade destes estimadores na presença de algum ponto atípico.

Em vista deste problema, apresentamos os procedimentos de estimação robusta.

No procedimento de estimação robusta ao invés de minimizarmos a soma dos quadrados dos resíduos, minimizamos uma versão robusta da dispersão dos resíduos. Um estimador pode ser considerado robusto se:

- (i) ele é altamente eficiente sob a verdadeira distribuição dos dados e razoavelmente eficiente numa vizinhança da mesma;
- (ii) se possui alto (0,5) ponto de ruptura e portanto não fornece estimativas distorcidas sob a presença de alguns pontos atípicos;
- (iii) se ele possui uma função de influência contínua e limitada.

Um procedimento de estimação robusta é o dos *mínimos quadrados podados* (denotado por *MQP*), proposto por Rousseeuw (1984). Os estimadores obtidos através do procedimento *MQP* são os estimadores  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_l) \in \mathfrak{R}^{l+1}$  que minimizam a função perda

$$P_2(g(n)) = \sum_{j=1}^{m^*} (r^2)_{j\tilde{m}},$$

onde  $(r^2)_{l:\bar{m}} \leq \dots \leq (r^2)_{m^*:\bar{m}}$  são os resíduos ao quadrado ordenados e  $m^*$  é o número de pontos usados no procedimento de otimização. A constante  $m^*$  é responsável pelo ponto de ruptura e eficiência.

Yohai (1987) introduziu outra classe de estimadores robustos baseados na regressão. Os estimadores baseados nesse procedimento serão denotados aqui por *MM*, e possuem as seguintes propriedades:

- (i) são altamente eficientes quando os erros possuem distribuição normal;
- (ii) são qualitativamente robustos;
- (iii) possuem alto ponto de ruptura.

Os estimadores *MM* são definidos por um procedimento de três estágios e podem ser descritos da seguinte forma. No primeiro estágio, um estimador inicial é calculado o qual é qualitativamente robusto e tem alto ponto de ruptura, mas não é necessariamente eficiente. No segundo estágio, um estimador  $M$  da escala de erros é calculado utilizando os resíduos baseados na estimação inicial. Finalmente, no terceiro estágio, um estimador  $M$  dos parâmetros de regressão é calculado baseado em uma função apropriada.

Os estimadores no procedimento *MM* são definidos como a solução  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_l) \in \mathfrak{R}^{l+1}$  a qual minimiza a função perda

$$P_3(g(n)) = \sum_{j=1}^{g(n)} \rho_2\left(\frac{r_j}{k}\right)^2,$$

sujeita a restrição

$$\frac{1}{n} \sum_{j=1}^{g(n)} \rho_1\left(\frac{r_j}{k}\right) \leq b,$$

onde  $\rho_1(\cdot)$  e  $\rho_2(\cdot)$  são funções simétricas, limitadas e não decrescentes em  $[0, \infty)$ , com  $\rho_v(0) = 0$ ,  $\lim_{u \rightarrow \infty} \rho_v(u) = 1$ , para  $v = 1, 2$ ,  $k$  é um parâmetro de escala,  $b$  é definido por  $E_\phi(\rho_1(u)) = b$ , onde  $\phi$  denota a função distribuição acumulada da normal padrão e  $r_j$  são os resíduos para todo  $j = 1, \dots, g(n)$ . O ponto de ruptura dos estimadores *MM* depende somente da função  $\rho_1(\cdot)$  e é dado por  $\min\{b, 1 - b\}$ . Os estimadores baseados no procedimento *MM* são consistentes e assintoticamente normais.



### 3.1 Estimador GPH Clássico e Robusto

Nesta seção, apresentamos o estimador proposto por Geweke e Porter-Hudak (1983), o qual se baseia no método de regressão utilizando a função periodograma. A seguir descrevemos tal método de estimação.

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $SARFIMA(0, D, 0)_s$ , com  $D \in (-0,5; 0,5)$ . A função densidade espectral deste processo é dado por

$$f_X(\omega) = f_\epsilon(\omega) \left[ 2 \operatorname{sen} \left( \frac{\omega s}{2} \right) \right]^{-2D}, \quad 0 < \omega \leq \pi,$$

onde  $f_\epsilon(\cdot)$  é a função densidade espectral de um ruído branco.

Aplicando a função logarítmica a ambos os lados da equação acima, temos

$$\begin{aligned} \ln f_X(\omega) &= \ln \left[ 2 \operatorname{sen} \left( \frac{\omega s}{2} \right) \right]^{-2D} + \ln f_\epsilon(\omega) \\ &= \ln f_\epsilon(\omega) - D \ln \left[ 2 \operatorname{sen} \left( \frac{\omega s}{2} \right) \right]^2 \end{aligned}$$

Adicionando  $\ln f_\epsilon(0)$  e  $\ln I(\omega)$ , a ambos os lados da equação, onde  $I(\omega)$  é a função periodograma (ver equação A.1.1 no Anexo A.1) e fazendo uso das propriedades da função logarítmica, obtemos

$$\ln I(\omega) = \ln f_\epsilon(0) - D \ln \left[ 2 \operatorname{sen} \left( \frac{\omega s}{2} \right) \right]^2 + \ln \left[ \frac{f_\epsilon(\omega)}{f_\epsilon(0)} \right] + \ln \left[ \frac{I(\omega)}{f_X(\omega)} \right]$$

Substituindo  $\omega$  pelas frequências de Fourier  $\omega_j = \frac{2\pi j}{n}$ ,  $j \in \{0, 1, \dots, \frac{n}{2} \mid j \neq \frac{n}{2}\}$  na equação, obtemos

$$\ln I(\omega_j) = \ln f_\epsilon(0) - D \ln \left[ 2 \operatorname{sen} \left( \frac{\omega_j s}{2} \right) \right]^2 + \ln \left[ \frac{f_\epsilon(\omega_j)}{f_\epsilon(0)} \right] + \ln \left[ \frac{I(\omega_j)}{f_X(\omega_j)} \right]$$

Considerando o limite máximo de  $j$  igual a  $g(n)$ , o qual é escolhido de tal forma que  $g(n)/n \rightarrow 0$  quando  $n \rightarrow \infty$  e  $\omega_j \leq \omega_{g(n)}$ , onde  $\omega_{g(n)}$  é pequeno, o termo  $\ln \left[ \frac{f_\epsilon(\omega_j)}{f_\epsilon(0)} \right]$  é desprezível se comparado com os outros termos da equação. Assim obtemos uma forma aproximada, dada por

$$\ln I(\omega_j) \approx \ln f_\epsilon(0) - D \ln \left[ 2 \operatorname{sen} \left( \frac{\omega_j s}{2} \right) \right]^2 + \ln \left[ \frac{I(\omega_j)}{f_X(\omega_j)} \right].$$

Podemos observar que a equação é uma forma aproximada da equação de regressão linear simples dada por

$$y_j \approx a + bx_j + \epsilon_j, \text{ para todo } j = 1, 2, \dots, g(n),$$

onde

$$y_j = \ln I(\omega_j), \quad x_j = \ln \left[ 2 \operatorname{sen} \left( \frac{\omega_j s}{2} \right) \right]^2, \quad \epsilon_j = \ln \left[ \frac{I(\omega_j)}{f_X(\omega_j)} \right] + c$$

$$b = -D, \quad a = \ln f_\epsilon(0) - c, \quad c = E \left[ \ln \left[ \frac{I(\omega_j)}{f_X(\omega_j)} \right] \right].$$

Utilizando a regressão de mínimos quadrados, podemos encontrar o estimador para o parâmetro  $D$  pelo método proposto por Geweke e Porter-Hudak (1983), denotado por *GPH*, através da função periodograma.

Assim,

$$\hat{D}_{GPH.MQ} = - \frac{\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2}$$

com

$$E(\hat{D}_{GPH.MQ}) \approx D, \quad \operatorname{Var}(\hat{D}_{GPH.MQ}) \approx \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (x_j - \bar{x})^2}$$

onde  $\bar{x}$  é a média dos  $x_j$  para  $j = 1, 2, \dots, g(n)$ .

**Observação** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo *SARFIMA*(0,  $D$ , 0)<sub>s</sub> com  $D \in (-0, 5; 0)$ . Seja  $\hat{D}_{GPH.MQ}$  o estimador de  $D$ , onde  $I(\cdot)$  é a função periodograma de  $\{X_t\}_{t \in \mathbb{Z}}$ , nas frequências de Fourier  $\omega_j = \frac{2\pi j}{n}, j \in \{0, 1, \dots, \frac{n}{2} | j \neq \frac{n}{2}\}$ . Suponha que  $g(n) = n^\alpha, \alpha \in$

$(0;1)$  satisfaça  $g(n) \rightarrow \infty$  e  $g(n)/n \rightarrow 0$ , quando  $n \rightarrow \infty$ . Então  $\lim_{n \rightarrow \infty} E(\hat{D}_{GPH.MQ}) = D$ . Se

$\lim_{n \rightarrow \infty} \frac{[\ln(n)]^2}{n} = 0$  então,

$$\frac{\hat{D}_{GPH.MQ} - D}{\sqrt{\text{Var}(\hat{D}_{GPH.MQ})}} \xrightarrow{\mathcal{D}} Z$$

onde  $Z \sim N(0,1)$  e  $\mathcal{D}$  significa convergência em distribuição. Este resultado foi estendido para  $D \in (-0,5;0,5)$  por Robinson (1995).

Desta forma encontramos o estimador para o parâmetro  $D$  pelo método proposto por Geweke e Porter-Hudak (1983), denotado por  $\hat{D}_{GPH.MQ}$ .

Os estimadores para  $D$  baseados no procedimento de estimação robusta são denotados por  $\hat{D}_{GPH.MQP}$  e  $\hat{D}_{GPH.MM}$ .

### 3.2 Estimador $R$ Clássico e Robusto

Proposto por Robinson (1995), este estimador semiparamétrico é baseado na função periodograma, de forma similar ao estimador proposto por Geweke e Porter-Hudak (1983). Este estimador é obtido aplicando o método dos mínimos quadrados, mas considerando as frequências de Fourier  $\omega_j$ , para  $j = \ell, \ell + 1, \dots, g(n)$ , onde  $\ell > 1$  é o valor de corte (“trimming”), que tende para o infinito de forma mais lenta do que  $g(n)$ . Denotamos este estimador de  $D$  por  $\hat{D}_{R.MQ}$ .

Analogamente ao estimador anterior, utilizamos a metodologia robusta para obtermos os estimadores robustos para o parâmetro  $D$  denotados por  $\hat{D}_{R.MQP}$  e  $\hat{D}_{R.MM}$ .

### 3.3 Estimador $SPR$ Clássico e Robusto

Nesta seção apresentamos o estimador proposto por Reisen (1994). Este estimador se baseia no uso da função periodograma suavizado, onde  $\lambda(\cdot)$  é a janela de Parzen, em vez da função periodograma no estimador proposto por Geweke e Porter-Hudak (1983). Esta mudança se deve ao fato de que a função periodograma é um estimador não

viciado para a função densidade espectral, porém a função periodograma suavizado é um estimador não viciado e consistente para ela.

Assim, substituindo-se a função periodograma pela função periodograma suavizado (ver equação A.1.2 no Anexo A.1), obtemos

$$\ln I_s(\omega_j) \approx \ln f_\varepsilon(0) - D \ln \left[ 2 \operatorname{sen} \left( \frac{\omega_j s}{2} \right) \right]^2 + \ln \left[ \frac{I_s(\omega_j)}{f_X(\omega_j)} \right].$$

Da mesma forma que no estimador  $\hat{D}_{GPH.MQ}$ , a equação acima é uma forma aproximada da equação de regressão linear simples, isto é

$$y_j \approx a + bx_j + \varepsilon_j, \text{ para todo } j = 1, 2, \dots, g(n),$$

onde

$$y_j = \ln I_s(\omega_j), \quad x_j = \ln \left[ 2 \operatorname{sen} \left( \frac{\omega_j s}{2} \right) \right]^2, \quad \varepsilon_j = \ln \left[ \frac{I_s(\omega_j)}{f_X(\omega_j)} \right] + c$$

$$b = -D, \quad a = \ln f_\varepsilon(0) - c, \quad c = E \left( \ln \left[ \frac{I_s(\omega_j)}{f_X(\omega_j)} \right] \right).$$

Portanto, o estimador de  $D$ , denotado por  $\hat{D}_{SPR.MQ}$ , através do método de regressão de mínimos quadrados, utilizando a função periodograma suavizado é dado por

$$\hat{D}_{SPR.MQ} = - \frac{\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2}$$

com

$$E(\hat{D}_{SPR.MQ}) \approx D, \quad \operatorname{Var}(\hat{D}_{SPR.MQ}) \approx 0,53928 \frac{m}{n \sum_{j=1}^{g(n)} (x_j - \bar{x})^2}$$

onde  $\bar{x}$  é a média dos  $x_j$  para  $j = 1, 2, \dots, g(n)$  e  $0,53928 \frac{m}{n}$  é a variância assintótica de  $\ln \left[ \frac{I_s(\omega_j)}{f_X(\omega_j)} \right]$ .

O estimador  $\hat{D}_{SPR.MQ}$  é, sob algumas condições de regularidade, aproximadamente normal com  $\lim_{n \rightarrow \infty} E(\hat{D}_{SPR.MQ}) = D$  e a variância assintótica é dada pela expressão acima.

Maiores detalhes podem ser encontrados em Reisen (1994).

O estimador para o parâmetro  $D$ , pelo método de mínimos quadrados, utilizando a função periodograma suavizado de covariâncias é denotado por  $\hat{D}_{SPR.MQ}$ . Este estimador é não viciado. Da mesma forma que o estimador proposto por Geweke e Porter-Hudak (1983), aplicando a metodologia robusta *MQP* e *MM*, obtemos os estimadores para  $D$  baseados no procedimento de estimação robusta que são denotados por  $\hat{D}_{SPR.MQP}$  e  $\hat{D}_{SPR.MM}$ .

### 3.4 Estimador *SR* Clássico e Robusto

Este estimador é obtido substituindo-se a função periodograma pela função periodograma suavizado no estimador proposto por Robinson (1995). Denotaremos este estimador de  $D$  por  $\hat{D}_{SR.MQ}$ .

Analogamente aos estimadores semi-paramétricos anteriores, utilizamos a metodologia robusta para obtermos os estimadores robustos para o parâmetro  $D$  denotados por  $\hat{D}_{SR.MQP}$  e  $\hat{D}_{SR.MM}$ .

### 3.5 Estimador *GPHT* Clássico e Robusto

Considerando o estimador proposto por Hurvich e Ray (1995) e Velasco (1999) que é um estimador semiparamétrico baseado na função periodograma modificado pela expressão

$$I_{\text{mod}}(\omega) = \frac{1}{\sum_{t=0}^{n-1} g(t)^2} \left| \sum_{t=0}^{n-1} g(t) X_t e^{-it\omega} \right|^2, \text{ para todo } \omega \in (0, \pi],$$

onde  $g(\cdot)$  é a janela espectral dada pela expressão

$$g(t) = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi(t+5,0)}{n} \right) \right].$$

Neste caso, a janela espectral  $g(\cdot)$  é aplicada diretamente ao processo estocástico e o estimador é denotado por  $\hat{D}_{GPHT}$ .

O estimador  $\hat{D}_{GPHT.MQ}$  é obtido aplicando a metodologia *MQ* à equação de regressão com  $I_{mod}(\cdot)$ , a função periodograma modificado. Analogamente, consideramos os estimadores  $\hat{D}_{GPHT.MQP}$  e  $\hat{D}_{GPHT.MM}$  através das metodologias robustas *MQP* e *MM*.

### 3.6 Estimador *W*

Nesta seção apresentamos o estimador de máxima verossimilhança. Para modelos *ARFIMA*( $p, d, q$ ), Fox e Taqqu (1986), utilizando a aproximação sugerida por Whittle (1951), apresentam condições que permitem que este estimador, para seqüências com forte dependência, seja consistente e tenha distribuição normal assintótica. Sowell (1992) apresenta resultados para o estimador de máxima verossimilhança exata. Ooms (1995) apresenta alguns resultados para modelos com longa dependência sazonais fazendo uma comparação entre os estimadores de máxima verossimilhança exata proposto por Sowell (1992), com a de máxima verossimilhança aproximada proposto por Fox e Taqqu (1986) e ainda com o estimador baseado na regressão da função periodograma proposto por Geweke e Porter-Hudak (1983).

A seguir, descrevemos o procedimento para encontrar o estimador de máxima verossimilhança aproximada.

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo Gaussiano com média  $\mu > 0$  e função densidade espectral dada por

$$f_X(\omega, \eta, \sigma^2) = \sigma^2 f(\omega, \eta), \quad \omega \in (-\pi, \pi),$$

onde  $\sigma^2 > 0$  (não precisa ser a variância do processo  $X_t$ ) e  $f(\cdot, \eta)$  é uma função da forma

$$f(\omega, \eta) \approx |\omega|^{\alpha(\eta)} G_\eta(\omega), \quad \text{quando } \omega \rightarrow 0,$$

com  $0 < \alpha(\eta) < 1$  e  $G_\eta(\omega)$  uma função que varia suavemente quando  $\omega \rightarrow 0$  e  $\eta$  é o vetor de parâmetros desconhecidos dos quais desejamos estimar.

A estimativa dos parâmetros do vetor  $\eta$  e  $\sigma^2$  é obtida através da série temporal  $X_1, X_2, \dots, X_n$ , cuja função distribuição conjunta, ou a de verossimilhança, é dada por

$$h(x, \eta, \sigma^2) = (2\pi)^{-\frac{n}{2}} |Q_n(\eta)|^{-\frac{1}{2}} \exp\left\{-\frac{z' Q_n^{-1}(\eta) z}{2}\right\},$$

onde  $x = (x_1, x_2, \dots, x_n)' \in \mathfrak{R}^n$  e  $Q_n(\eta)$  é a matriz de Teoplitz contendo os valores da função de autocovariância do processo  $\{X_t\}_{t \in \mathbb{Z}}$  cujos elementos são dados por  $\{Q_n(\eta)\}_{i,j} = \gamma_x(|i-j|)$ ,  $1 \leq i, j \leq n$ .

Aplicando a função logarítmica na equação acima obtemos a seguinte expressão

$$L(x, \eta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|Q_n(\eta)| - \frac{1}{2} z' Q_n^{-1}(\eta) z.$$

O estimador de máxima verossimilhança de  $\eta$  e de  $\sigma^2$  é obtido maximizando-se a função  $L(x, \eta, \sigma^2)$  com respeito ao vetor de parâmetros  $\eta$  e  $\sigma^2$ .

Computacionalmente, é complicado obter a inversa da matriz de autocovariâncias  $Q_n(\eta)$ . Fox e Taqqu (1986), fazendo uso de uma aproximação, aplicam o método de máxima verossimilhança aproximada para estimar  $\eta$  e  $\sigma^2$  maximizando a função

$$h(x, \eta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{Z' A_n(\eta) Z}{2n\sigma^2}\right\},$$

onde  $Z = (X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$  com  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  e  $A_n(\eta)$  é uma matriz de ordem  $n \times n$ , proposta por Whittle, para aproximar  $Q_n^{-1}(\eta)$ , onde o  $(i, j)$ -ésimo elemento é dado por

$$\{A_n(\eta)\}_{i,j} = a_{i-j}(\eta) = a_t(\eta) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} e^{i\omega t} [f_x(\omega, n)]^{-1} d\omega.$$

Os estimadores de máxima verossimilhança aproximada de  $\eta$  e  $\sigma^2$  que maximizam a equação acima, são denotados por  $\hat{\eta}_n$  e  $\hat{\sigma}_n^2$ .

Maximizar a função  $h(x, \eta, \sigma^2)$  é equivalente a encontrar  $\hat{\eta}_n$  que minimize

$$\hat{\sigma}_n^2(\eta) = \frac{Z' A_n(\eta) Z}{n}.$$

Computacionalmente, o estimador de máxima verossimilhança é obtido minimizando a função, na forma discreta,

$$L(x, \eta, \sigma^2) = \frac{1}{2n} \sum_{j=1}^T \left( \ln f_x(\omega_j, \eta) + \frac{I(\omega_j)}{f_x(\omega_j, \eta)} \right),$$

onde  $\eta$  denota o vetor de parâmetros desconhecidos,  $\omega_j = \frac{2\pi j}{n}$ ,  $j \in \{0, 1, \dots, T \mid j \neq \frac{in}{s}\}$ , são as frequências de Fourier e  $T = \lfloor \frac{n-1}{2} \rfloor$ .

**Observação** Denotado por  $\hat{D}_w$ , o estimador de máxima verossimilhança aproximada proposto por Fox e Taquq (1986), sob certas condições de regularidade, é consistente e tem distribuição assintótica normal. Um estudo completo deste estimador pode ser encontrado em Fox e Taquq (1986) e Beran (1994).



## 4. Simulação de Monte Carlo

Para a simulação do processo  $SARFIMA(0,D,0)_s$ , foi utilizado o programa Splus na versão 6.2. A simulação foi feita utilizando o método de simulação descrito por Hosking (1984). A seguir descrevemos as etapas para a simulação de amostras de tamanho  $n$  de um processo  $SARFIMA(0,D,0)_s$ .

**Etapa I:** Calcular a função de autocorrelação parcial.

**Etapa II:** Gerar uma variável aleatória  $N(0,1)$  para simular um processo  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ , Ruído Branco Gaussiano.

**Etapa III:** Calcular a média e a variância da variável  $X_t$  (ver Anexo A.2).

**Etapa III:** Gerar uma variável aleatória  $X_t$ , com distribuição Normal com média  $m_t$  e variância  $v_t$ , para  $t = 1, 2, \dots, n$ .

Para as simulações utilizamos tamanho da amostra  $n = 1000$ , o número de replicações  $re = 1000$ ,  $\alpha = \{0,5; 0,6; 0,7; 0,8\}$ ,  $s = \{2; 3; 6; 12\}$ ,  $\ell = \{2; 3\}$  e  $D = \{0,1; 0,2; 0,3; 0,4\}$ . Os resultados apresentados a seguir serão apenas utilizando  $\alpha = \{0,6; 0,7\}$ ,  $\ell = 2$ ,  $D = \{0,2; 0,4\}$  e para todas as sazonalidades citadas acima.

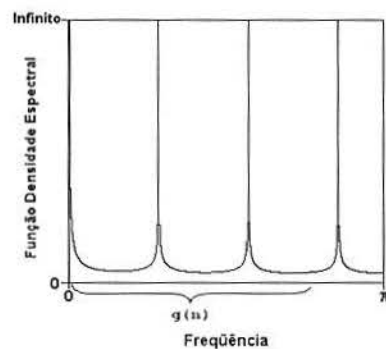


Figura 2 Representação de  $g(n) = n^\alpha$

O programa Splus também foi utilizado para calcular as estimativas para o parâmetro  $D$  para cada um dos estimadores especificados na seção 3. São eles: os estimadores semi-paramétricos  $GPH$ ,  $R$ ,  $SPR$ ,  $SR$  e  $GPHT$  nas metodologias clássica ( $MQ$ ) e robustas ( $MM$  e  $MQP$ ) e o estimador paramétrico  $W$ .

Para efeito de comparação entre os estimadores utilizamos as seguintes relações:

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i; \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2;$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (D_i - D_0)^2; \quad \text{Viés} = \bar{D} - D_0,$$

onde  $D_0$  é o valor nominal de  $D$ , isto é, o valor utilizado para as gerações do processo.

## 4.1 Resultados

Para cada estimador e cada valor atribuído a  $s$  e  $\alpha$ , foram calculados a média, o viés, a variância e o erro quadrático médio a fim de fazermos comparações entre os 16 estimadores estudados neste trabalho. Os valores destas medidas estão apresentados nas tabelas, da **Tabela 1** à **Tabela 8**.

O número de replicações utilizado ( $re = 1000$ ) é suficiente para avaliar estes estimadores, não é necessário aumentar este valor para conseguir melhores resultados. O tamanho amostral ( $n = 1000$ ) influencia fortemente nos resultados, principalmente no estimador paramétrico  $W$  (de máxima verossimilhança), quanto maior o valor de  $n$  melhor serão as estimações. A variabilidade e o erro quadrático médio se comportam de forma muito semelhante, já que os vieses são pequenos, logo as observações feitas para um valem para o outro.

**Tabela 1** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

Estimador. Metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2062	0,0062	0,0079	0,0078	0,4140	0,0140	0,0089	0,0087
GPH.MM	0,2026	0,0026	0,0104	0,0104	0,4091	0,0091	0,0106	0,0105
GPH.MQP	0,2020	0,0020	0,0100	0,0100	0,4106	0,0106	0,0102	0,0101
R.MQ	0,2090	0,0090	0,0095	0,0094	0,4147	0,0147	0,0106	0,0104
R.MM	0,2037	0,0037	0,0123	0,0123	0,4101	0,0101	0,0114	0,0113
R.MQP	0,2045	0,0045	0,0123	0,0123	0,4130	0,0130	0,0117	0,0115
SPR.MQ	0,1893	-0,0107	0,0058	0,0057	0,3956	-0,0044	0,0059	0,0059
SPR.MM	0,2006	0,0006	0,0103	0,0103	0,4027	0,0027	0,0114	0,0114
SPR.MQP	0,1994	-0,0006	0,0083	0,0083	0,4061	0,0061	0,0086	0,0086
SR.MQ	0,2063	0,0063	0,0064	0,0063	0,4168	0,0168	0,0066	0,0063
SR.MM	0,2097	0,0097	0,0124	0,0123	0,4109	0,0109	0,0123	0,0122
SR.MQP	0,2073	0,0073	0,0089	0,0088	0,4133	0,0133	0,0089	0,0088
GPHT.MQ	0,2072	0,0072	0,0129	0,0129	0,4374	0,0374	0,0146	0,0132
GPHT.MM	0,1996	-0,0004	0,0161	0,0161	0,4199	0,0199	0,0203	0,0199
GPHT.MQP	0,2007	0,0007	0,0147	0,0147	0,4238	0,0238	0,0176	0,0171
W	0,1909	-0,0091	0,0015	0,0014	0,3952	-0,0048	0,0015	0,0015

**Tabela 2** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

Estimador. Metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2221	0,0221	0,0044	0,0039	0,4255	0,0255	0,0050	0,0044
GPH.MM	0,2149	0,0149	0,0072	0,0070	0,4179	0,0179	0,0068	0,0065
GPH.MQP	0,2193	0,0193	0,0048	0,0045	0,4227	0,0227	0,0048	0,0043
R.MQ	0,2255	0,0255	0,0052	0,0045	0,4272	0,0272	0,0057	0,0049
R.MM	0,2189	0,0189	0,0082	0,0078	0,4187	0,0187	0,0080	0,0076
R.MQP	0,2234	0,0234	0,0057	0,0052	0,4248	0,0248	0,0054	0,0048
SPR.MQ	0,2121	0,0121	0,0029	0,0028	0,4153	0,0153	0,0031	0,0029
SPR.MM	0,2202	0,0202	0,0057	0,0053	0,4239	0,0239	0,0054	0,0049
SPR.MQP	0,2192	0,0192	0,0044	0,0040	0,4227	0,0227	0,0042	0,0037
SR.MQ	0,2237	0,0237	0,0036	0,0030	0,4289	0,0289	0,0038	0,0030
SR.MM	0,2248	0,0248	0,0060	0,0054	0,4281	0,0281	0,0058	0,0050
SR.MQP	0,2244	0,0244	0,0047	0,0041	0,4276	0,0276	0,0045	0,0037
GPHT.MQ	0,2222	0,0222	0,0069	0,0065	0,4382	0,0382	0,0079	0,0064
GPHT.MM	0,2141	0,0141	0,0105	0,0103	0,4250	0,0250	0,0117	0,0111
GPHT.MQP	0,2189	0,0189	0,0070	0,0066	0,4309	0,0309	0,0085	0,0076
W	0,1909	-0,0091	0,0015	0,0014	0,3952	-0,0048	0,0015	0,0015

**Tabela 3** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2009	0,0009	0,0080	0,0080	0,4109	0,0109	0,0074	0,0073
GPH.MM	0,1980	-0,0020	0,0102	0,0102	0,4011	0,0011	0,0092	0,0092
GPH.MQP	0,1954	-0,0046	0,0102	0,0102	0,4008	0,0008	0,0099	0,0100
R.MQ	0,2022	0,0022	0,0098	0,0099	0,4104	0,0104	0,0093	0,0092
R.MM	0,1966	-0,0034	0,0128	0,0128	0,4005	0,0005	0,0124	0,0125
R.MQP	0,1954	-0,0046	0,0122	0,0122	0,4014	0,0014	0,0123	0,0123
SPR.MQ	0,2009	0,0009	0,0080	0,0080	0,4109	0,0109	0,0074	0,0073
SPR.MM	0,1980	-0,0020	0,0102	0,0102	0,4011	0,0011	0,0092	0,0092
SPR.MQP	0,1954	-0,0046	0,0102	0,0102	0,4008	0,0008	0,0099	0,0100
SR.MQ	0,2022	0,0022	0,0098	0,0099	0,4104	0,0104	0,0093	0,0092
SR.MM	0,1966	-0,0034	0,0128	0,0128	0,4005	0,0005	0,0124	0,0125
SR.MQP	0,1954	-0,0046	0,0122	0,0122	0,4014	0,0014	0,0123	0,0123
GPHT.MQ	0,2139	0,0139	0,0141	0,0140	0,4313	0,0313	0,0124	0,0115
GPHT.MM	0,2061	0,0061	0,0156	0,0156	0,4089	0,0089	0,0174	0,0173
GPHT.MQP	0,2074	0,0074	0,0144	0,0143	0,4095	0,0095	0,0157	0,0156
W	0,1871	-0,0129	0,0021	0,0019	0,3898	-0,0102	0,0022	0,0021

**Tabela 4** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2027	0,0027	0,0040	0,0040	0,4096	0,0096	0,0039	0,0038
GPH.MM	0,1942	-0,0058	0,0068	0,0068	0,3957	-0,0043	0,0081	0,0081
GPH.MQP	0,1992	-0,0008	0,0045	0,0045	0,4014	0,0014	0,0047	0,0047
R.MQ	0,2036	0,0036	0,0044	0,0044	0,4092	0,0092	0,0045	0,0045
R.MM	0,1964	-0,0036	0,0075	0,0075	0,3960	-0,0040	0,0089	0,0089
R.MQP	0,1994	-0,0006	0,0051	0,0051	0,4027	0,0027	0,0056	0,0056
SPR.MQ	0,2027	0,0027	0,0040	0,0040	0,4096	0,0096	0,0039	0,0038
SPR.MM	0,1942	-0,0058	0,0068	0,0068	0,3957	-0,0043	0,0081	0,0081
SPR.MQP	0,1992	-0,0008	0,0045	0,0045	0,4014	0,0014	0,0047	0,0047
SR.MQ	0,2036	0,0036	0,0044	0,0044	0,4092	0,0092	0,0045	0,0045
SR.MM	0,1964	-0,0036	0,0075	0,0075	0,3960	-0,0040	0,0089	0,0089
SR.MQP	0,1994	-0,0006	0,0051	0,0051	0,4027	0,0027	0,0056	0,0056
GPHT.MQ	0,2087	0,0087	0,0072	0,0072	0,4201	0,0201	0,0066	0,0062
GPHT.MM	0,1987	-0,0013	0,0098	0,0098	0,4034	0,0034	0,0107	0,0107
GPHT.MQP	0,2040	0,0040	0,0070	0,0070	0,4091	0,0091	0,0072	0,0071
W	0,1871	-0,0129	0,0021	0,0019	0,3898	-0,0102	0,0022	0,0021

**Tabela 5** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2048	0,0048	0,0085	0,0085	0,4175	0,0175	0,0091	0,0088
GPH.MM	0,1951	-0,0049	0,0118	0,0118	0,4043	0,0043	0,0123	0,0122
GPH.MQP	0,1947	-0,0053	0,0122	0,0121	0,4066	0,0066	0,0115	0,0115
R.MQ	0,2037	0,0037	0,0108	0,0108	0,4177	0,0177	0,0116	0,0113
R.MM	0,1935	-0,0065	0,0156	0,0156	0,4071	0,0071	0,0146	0,0146
R.MQP	0,1926	-0,0074	0,0158	0,0157	0,4098	0,0098	0,0155	0,0155
SPR.MQ	0,2048	0,0048	0,0085	0,0085	0,4175	0,0175	0,0091	0,0088
SPR.MM	0,1951	-0,0049	0,0118	0,0118	0,4043	0,0043	0,0123	0,0122
SPR.MQP	0,1947	-0,0053	0,0122	0,0121	0,4066	0,0066	0,0115	0,0115
SR.MQ	0,2037	0,0037	0,0108	0,0108	0,4177	0,0177	0,0116	0,0113
SR.MM	0,1935	-0,0065	0,0156	0,0156	0,4071	0,0071	0,0146	0,0146
SR.MQP	0,1926	-0,0074	0,0158	0,0157	0,4098	0,0098	0,0155	0,0155
GPHT.MQ	0,2061	0,0061	0,0137	0,0136	0,4412	0,0412	0,0147	0,0130
GPHT.MM	0,1964	-0,0036	0,0167	0,0167	0,4225	0,0225	0,0203	0,0198
GPHT.MQP	0,1968	-0,0032	0,0159	0,0159	0,4231	0,0231	0,0176	0,0171
W	0,1763	-0,0237	0,0051	0,0046	0,3802	-0,0198	0,0049	0,0045

**Tabela 6** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2064	0,0064	0,0071	0,0070	0,4147	0,0147	0,0071	0,0069
GPH.MM	0,1827	-0,0173	0,0155	0,0152	0,3886	-0,0114	0,0163	0,0162
GPH.MQP	0,2020	0,0020	0,0088	0,0089	0,4106	0,0106	0,0090	0,0089
R.MQ	0,2057	0,0057	0,0090	0,0090	0,4143	0,0143	0,0089	0,0087
R.MM	0,1811	-0,0189	0,0185	0,0181	0,3862	-0,0138	0,0199	0,0197
R.MQP	0,1985	-0,0015	0,0113	0,0113	0,4116	0,0116	0,0116	0,0115
SPR.MQ	0,2064	0,0064	0,0071	0,0070	0,4147	0,0147	0,0071	0,0069
SPR.MM	0,1827	-0,0173	0,0155	0,0152	0,3886	-0,0114	0,0163	0,0162
SPR.MQP	0,2020	0,0020	0,0088	0,0089	0,4106	0,0106	0,0090	0,0089
SR.MQ	0,2057	0,0057	0,0090	0,0090	0,4143	0,0143	0,0089	0,0087
SR.MM	0,1811	-0,0189	0,0185	0,0181	0,3862	-0,0138	0,0199	0,0197
SR.MQP	0,1985	-0,0015	0,0113	0,0113	0,4116	0,0116	0,0116	0,0115
GPHT.MQ	0,2084	0,0084	0,0115	0,0114	0,4356	0,0356	0,0122	0,0109
GPHT.MM	0,1846	-0,0154	0,0230	0,0227	0,4017	0,0017	0,0266	0,0267
GPHT.MQP	0,2016	0,0016	0,0132	0,0133	0,4244	0,0244	0,0138	0,0132
W	0,1763	-0,0237	0,0051	0,0046	0,3802	-0,0198	0,0049	0,0045

**Tabela 7** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2091	0,0091	0,0145	0,0144	0,4184	0,0184	0,0139	0,0136
GPH.MM	0,1933	-0,0067	0,0268	0,0268	0,3989	-0,0011	0,0283	0,0283
GPH.MQP	0,2014	0,0014	0,0226	0,0227	0,4073	0,0073	0,0212	0,0211
R.MQ	0,2096	0,0096	0,0214	0,0214	0,4189	0,0189	0,0202	0,0199
R.MM	0,1937	-0,0063	0,0345	0,0345	0,3992	-0,0008	0,0388	0,0389
R.MQP	0,2017	0,0017	0,0309	0,0309	0,4096	0,0096	0,0302	0,0301
SPR.MQ	0,2091	0,0091	0,0145	0,0144	0,4184	0,0184	0,0139	0,0136
SPR.MM	0,1933	-0,0067	0,0268	0,0268	0,3989	-0,0011	0,0283	0,0283
SPR.MQP	0,2014	0,0014	0,0226	0,0227	0,4073	0,0073	0,0212	0,0211
SR.MQ	0,2096	0,0096	0,0214	0,0214	0,4189	0,0189	0,0202	0,0199
SR.MM	0,1937	-0,0063	0,0345	0,0345	0,3992	-0,0008	0,0388	0,0389
SR.MQP	0,2017	0,0017	0,0309	0,0309	0,4096	0,0096	0,0302	0,0301
GPHT.MQ	0,2244	0,0244	0,0280	0,0274	0,4602	0,0602	0,0265	0,0229
GPHT.MM	0,1980	-0,0020	0,0516	0,0516	0,4195	0,0195	0,0583	0,0580
GPHT.MQP	0,2069	0,0069	0,0317	0,0317	0,4340	0,0340	0,0371	0,0360
W	0,1499	-0,0501	0,0114	0,0089	0,3513	-0,0487	0,0121	0,0097

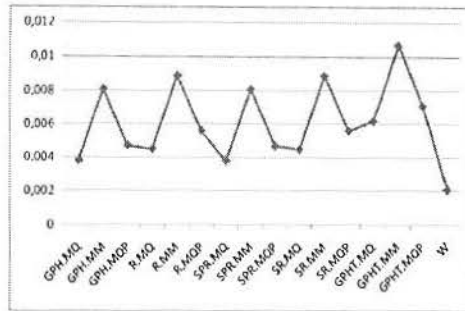
**Tabela 8** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $s = 12$ ,  $n = 1000$ ,  $rc = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

Estimador. metodologia	D = 0,2				D = 0,4			
	média	viés	eqm	variância	média	viés	eqm	variância
GPH.MQ	0,2095	0,0095	0,0141	0,0140	0,4196	0,0196	0,0137	0,0133
GPH.MM	0,1881	-0,0119	0,0382	0,0381	0,3972	-0,0028	0,0311	0,0311
GPH.MQP	0,1978	-0,0022	0,0204	0,0204	0,4048	0,0048	0,0186	0,0186
R.MQ	0,2096	0,0096	0,0146	0,0146	0,4198	0,0198	0,0142	0,0138
R.MM	0,1846	-0,0154	0,0413	0,0411	0,3984	-0,0016	0,0348	0,0348
R.MQP	0,1979	-0,0021	0,0219	0,0219	0,4053	0,0053	0,0198	0,0198
SPR.MQ	0,2095	0,0095	0,0141	0,0140	0,4196	0,0196	0,0137	0,0133
SPR.MM	0,1881	-0,0119	0,0382	0,0381	0,3972	-0,0028	0,0311	0,0311
SPR.MQP	0,1978	-0,0022	0,0204	0,0204	0,4048	0,0048	0,0186	0,0186
SR.MQ	0,2096	0,0096	0,0146	0,0146	0,4198	0,0198	0,0142	0,0138
SR.MM	0,1846	-0,0154	0,0413	0,0411	0,3984	-0,0016	0,0348	0,0348
SR.MQP	0,1979	-0,0021	0,0219	0,0219	0,4053	0,0053	0,0198	0,0198
GPHT.MQ	0,2241	0,0241	0,0270	0,0265	0,4661	0,0661	0,0254	0,0211
GPHT.MM	0,1898	-0,0102	0,0489	0,0488	0,4094	0,0094	0,0506	0,0505
GPHT.MQP	0,2053	0,0053	0,0301	0,0301	0,4377	0,0377	0,0306	0,0292
W	0,1499	-0,0501	0,0114	0,0089	0,3513	-0,0487	0,0121	0,0097

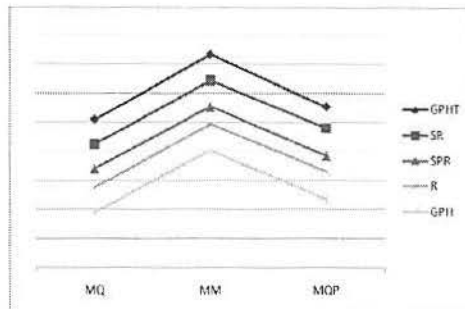
O estimador  $W$  apresenta, em todos os casos, a menor variabilidade dentre todos estimadores estudados, como é possível observar na **Figura 3**. A **Figura 4** mostra o comportamento da variabilidade dentro de cada estimador semi-paramétrico, nas diferentes metodologias. Conclui-se que em um mesmo estimador, a metodologia clássica ( $MQ$ ) apresenta sempre a menor variabilidade, seguida das metodologias robustas  $MQP$  e  $MM$ , que é a que apresenta maior variabilidade. Esse comportamento não se altera de estimador para estimador.

Nas figuras em anexo (da **Figura 13** até a **Figura 28**) estão apresentados o boxplot dos estimadores para cada variação de  $D$ ,  $s$  e  $\alpha$ , onde é possível perceber também este comportamento da variabilidade.



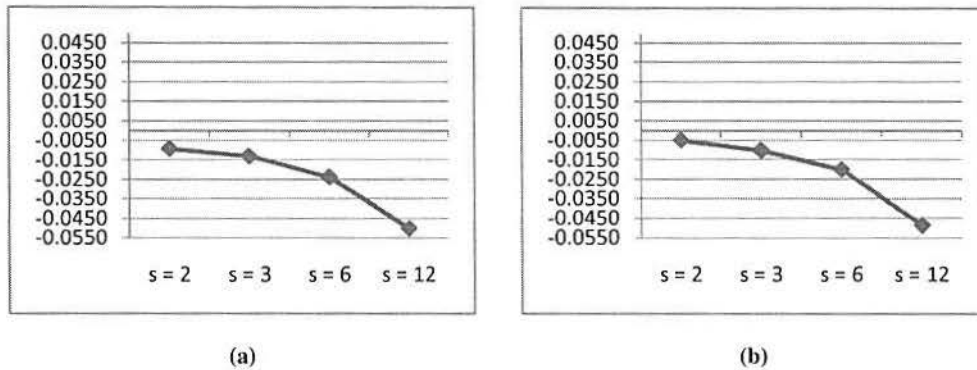


**Figura 3** Variância dos estimadores do parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,4$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

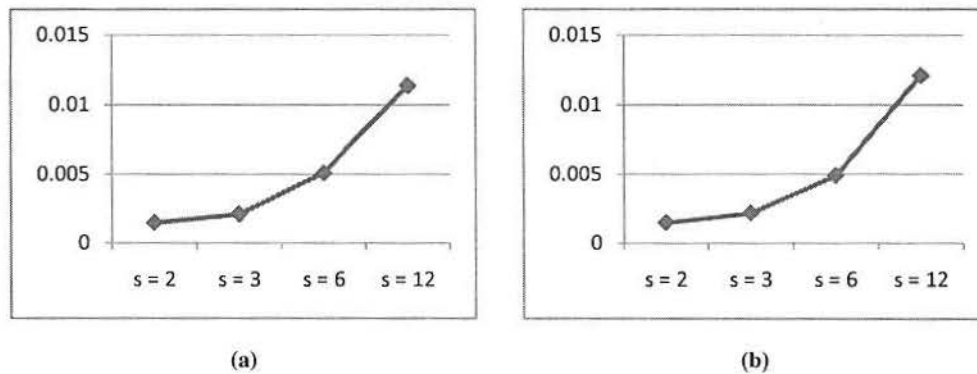


**Figura 4** Comportamento das variâncias dentro dos estimadores semi-paramétricos, em cada metodologia utilizada, nos estimadores do parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,4$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

O estimador paramétrico  $W$ , de máxima verossimilhança, subestimou o parâmetro  $D$  em todos os casos estudados. Na análise do viés isto é observado facilmente, já que o viés deste estimador é sempre negativo. Percebe-se também que este estimador se comporta de forma semelhante para os diferentes valores de  $D$ , mas sua característica de subestimação se agrava para valores maiores de  $s$ , como se observa na **Figura 5**. Isso implica num crescimento no erro quadrático médio para maiores valores de  $s$ , como é possível observar na **Figura 6**. Para  $s = 2$ , o viés deste estimador é muito pequeno, tornando  $W$  um estimador bom. Aumentando o tamanho amostral o viés tende a diminuir para todos os valores de  $s$ , pois quanto maior é o valor de  $s$  maior deve ser o valor de  $n$  para que  $W$  forneça boas estimativas, assim como aconteceu neste caso com  $s = 2$ .



**Figura 5** Comportamento do viés do estimador  $W$  para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , em cada sazonalidade estudada: (a)  $D = 0,2$ ; (b)  $D = 0,4$ .



**Figura 6** Comportamento do erro quadrático médio do estimador  $W$  para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , em cada sazonalidade estudada: (a)  $D = 0,2$ ; (b)  $D = 0,4$ .

Os estimadores semi-paramétricos são muito semelhantes, apenas o estimador *GPHT*, que aplica a janela espectral na série ao invés de aplicar na função periodograma, se comporta um pouco diferente dos demais. As características descritas a seguir são encontradas em todos os estimadores semi-paramétricos: o viés dos estimadores, para os diferentes  $\alpha$  e  $s$ , são sempre menores (não necessariamente menores em módulo) para  $D = 0,2$  do que para  $D = 0,4$ ; as três metodologias utilizadas parecem influenciar o estimador de forma semelhante. É possível perceber isso, por exemplo, com o viés que se apresenta de forma bem semelhante dentro de uma mesma metodologia e de forma diferente entre as metodologias ou mesmo no comportamento das variâncias, como já foi visto anteriormente; a estimação pontual (média das estimativas) apresenta formato bem parecido dentro de cada estimador, para as metodologias, as figura em anexo (da **Figura 13** até a **Figura 28**) e a **Figura 7** mostram este comportamento.



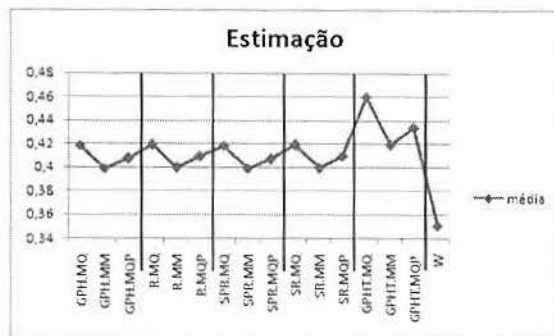
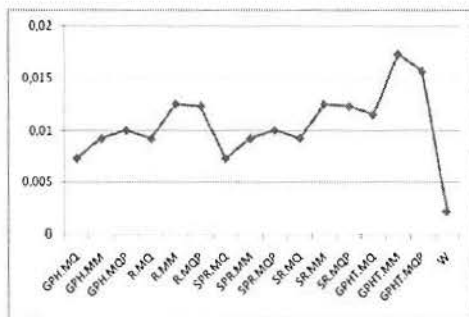
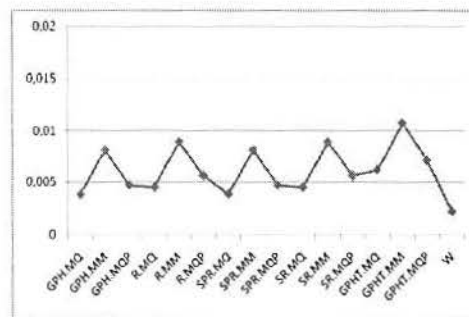


Figura 7 Comportamento da média das estimativas para cada estimador do parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,4$ ,  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

Os diferentes  $\alpha = \{0,6; 0,7\}$  não parecem influenciar nas médias nem nos vieses das estimativas, apenas, não muito fortemente, na variabilidade, mas apenas para  $s = \{2; 3\}$ , para os demais  $s$  isso não é verificado. Obviamente para o parâmetro  $W$  não ocorre diferença alguma, pois não é utilizado  $\alpha$  para sua estimação.



(a)



(b)

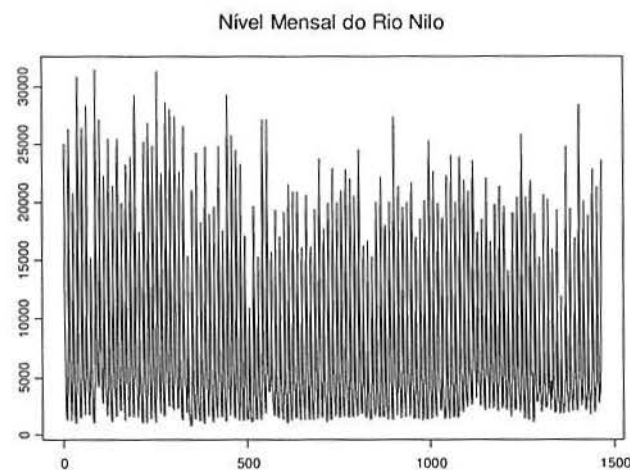
Figura 8 Variância dos estimadores do parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,4$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$  e  $\ell = 2$ : (a)  $\alpha = 0,6$ ; (b)  $\alpha = 0,7$ .

## 5. Aplicação

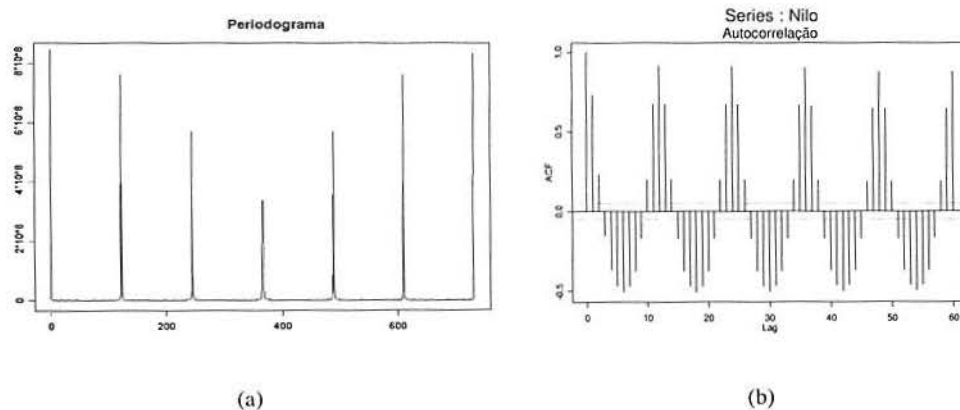
Nesta seção analisamos uma série temporal real utilizando a metodologia desenvolvida ao longo deste trabalho.

### 5.1 Série Temporal dos Níveis Mensais do Rio Nilo

Consideramos a série temporal dos níveis mensais do Rio Nilo, em Aswan, cordialmente cedida pelo Professor Dr. A. Montanari (Università di Bologna, Itália). Esta série temporal consiste de 1466 observações, durante o período de agosto de 1872 a setembro de 1994. A análise desta série pode ser encontrada em Montanari et al. (2000) e Bisognin e Lopes (2007).



**Figura 9** Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan. (Período: Agosto de 1872 a Setembro de 1994;  $n = 1466$ )



**Figura 10** Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan: (a) função periodograma; (b) função de autocorrelação amostral.

A **Figura 10(a)** e a **Figura 10(b)** mostram que a série temporal apresenta um comportamento de longa dependência, já que a função de autocorrelação amostral

possui um decaimento hiperbólico lento, e sua função periodograma apresenta um comportamento periódico causado pelo seu ciclo anual. Este comportamento já tinha sido observado por Montanari et al. (2000) e Bisognin e Lopes (2007). Aplicamos o teste R/S, criado por Hurst (1951) e modificado por Lo (1991), cuja hipótese nula é de que a série não possui longa dependência. O resultado deste teste foi significativo ao nível de 5% (Estatística do teste R/S = 2,0766, p-valor < 0,05), confirmando estatisticamente a conclusão que tiramos da análise visual.

Aplicamos os 16 estimadores estudados na série e obtivemos as seguintes estimativas.

**Tabela 9** Estimativas para o parâmetro  $D$  do modelo  $SARFIMA(0,D,0)_s$ .

Estimador.metodologia	Estimativa para $D$
GPH.MQ	0.3678887
GPH.MM	0.3861863
GPH.MQP	0.2544854
R.MQ	0.3690743
R.MM	0.3890693
R.MQP	0.2725735
SPR.MQ	0.3678887
SPR.MM	0.3861863
SPR.MQP	0.2544854
SR.MQ	0.3690743
SR.MM	0.3890693
SR.MQP	0.2725735
GPHT.MQ	0.317106
GPHT.MM	-0.0372659
GPHT.MQP	0.1459409
W	0.1985201

Para ajustar o modelo utilizamos a estimativa dada pelo estimador  $W$ . O melhor modelo ajustado à esta série temporal, segundo os critérios de seleção AIC e BIC, é um processo  $SARFIMA(p,d,q) \times (P,D,Q)_s$ , com  $P = p = l = q = Q$ ,  $\hat{D} = \hat{D}_w$ ,  $\hat{d} = 0,0$  e  $s = 12$ .

**Tabela 10** Modelo ajustado para a Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan.

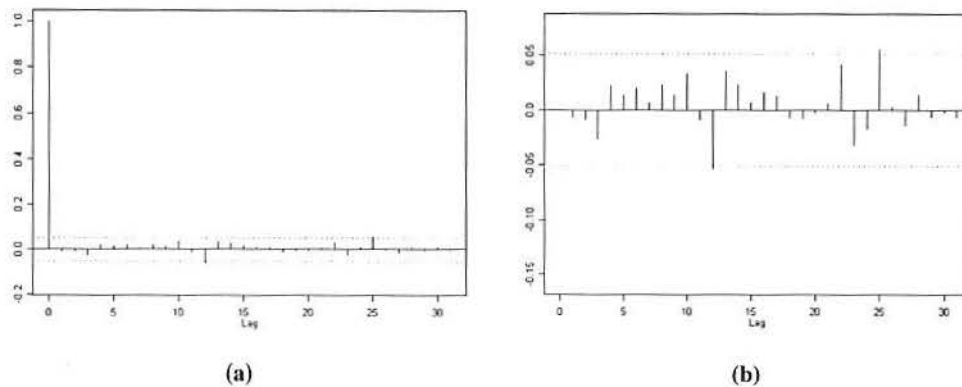
Parâmetro	$\phi$	$\Phi$	$D_w$	$d_w$	$\Theta$	$\theta$
Estimador	0.6137	0.9845	0.1985	0.0045	-0.2145	0.9348
Desvio Padrão	0.0287	0.0289	0.0237	0.0201	0.0334	0.0135

A variância estimada dos resíduos é  $\hat{\sigma}_\epsilon^2 = 1,0007$ .

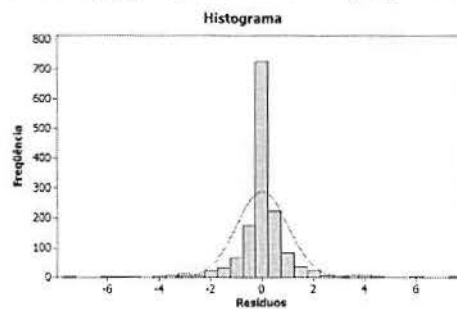
A **Tabela 10** apresenta os resultados para os estimadores dos parâmetros do modelo ajustado para esta série temporal com sazonalidade  $s = 12$ . Consideramos apenas o estimador proposto por Fox e Taquq (1986). O parâmetro de diferenciação  $d$  é igual a zero (na verdade,  $\hat{d} = 0,0045$ ). Este modelo foi selecionada pelos critérios de seleção de modelo AIC e BIC, cujos valores são apresentados na **Tabela 11**.

**Tabela 11** Valores dos Critérios de Seleção do Modelo para a Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan.

Critério de Seleção	AIC	BIC
Valor	15.7354	34.6598



**Figura 11** Resíduos do Modelo Ajustado à Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan: (a) função de autocorrelação amostral; (b) função de autocorrelação parcial amostral.



**Figura 12** Histograma dos Resíduos do Modelo Ajustado à Série Temporal dos Níveis Mensais do Rio Nilo, em Aswan.

A **Figura 12(a)** e a **Figura 12(b)** apresentam as funções de autocorrelação e autocorrelação parcial amostrais dos resíduos do modelo. Analisando estas figuras, observamos que o modelo está bem ajustado. Observando o histograma na **Figura 12**, percebemos que os resíduos do modelo ajustado à série temporal possuem distribuição aproximadamente normal padrão com uma variância muito pequena com relação a 1,0. O teste de Shapiro-Wilks não rejeitou a hipótese nula de normalidade dos resíduos ( $W = 0,999$ ;  $p$ -valor = 0,8746).

## 6. Conclusões

Analisamos as estimativas do parâmetro de diferenciação sazonal  $D$ , para o processo  $SARFIMA(0,D,0)_s$ , quando  $D \in (-0,5;0,5)$  e quando  $s \in \{2,3,6,12\}$ , concluímos que o erro quadrático médio dos estimadores diminui quando o tamanho amostral aumenta. Para o estimador  $W$  o erro quadrático médio é menor quanto menor for a sazonalidade  $s$ .

Entre os estimadores estudados,  $W$  possui o menor erro quadrático médio, em todas as variações de  $D$  e  $s$ . Mesmo para  $s$  grande, onde o erro quadrático médio de  $W$  cresce muito, este ainda assim é o menor.

Os estimadores semi-paramétricos tiveram um comportamento muito semelhante, para todos os casos estudados. As metodologias robustas apresentaram, em grande parte dos casos, menor viés do que a metodologia clássica. Entretanto a metodologia clássica apresentou menores variância e erro quadrático médio em praticamente todas as situações.

Comparando as duas metodologias robustas empregadas, a metodologia  $MQP$  se comporta melhor do que a  $MM$ , quando se trata do viés, da variância e do erro quadrático médio.

Nos diferentes valores de  $\alpha$  analisados, pode-se perceber apenas uma pequena diferença no comportamento da variabilidade dos estimadores. Quando  $\alpha = 0,7$  a variabilidade é um pouco menor do que quando  $\alpha = 0,6$ . Este comportamento só foi observado quando  $s = \{2,3\}$ .

## Referências Bibliográficas

- Beran, J. (1994). "Statistics for Long-Memory Processes". *New York: Chapman & Hall*.
- Bisognin, C. e S.R.C. Lopes (2007). "Estimating and Forecasting the Long Memory Parameter in the Presence of Periodicity". *Journal of Forecasting*, 26(6), 405-427.
- Brietzke, E. H. M. et al (2005). "A Closed Formula For The Durbin-Levinson's Algorithm In Seasonal Fractionally Integrated Processes". *Mathematical and Computer Modelling*, 42, 11-12.
- Chan, N., & Palma, W. (1998). "State space modeling os long-memory processes". *The Annals os Statistics*, 26(2), 719-740.
- Draper, N. e H. Smith (1981). *Applied Regression Analysis*. *New York: John Wiley*.
- Fox, R. e M.S. Taqqu (1986). "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, 14, 517-532.
- Geweke, J. e S. Porter-Hudak (1983). "The Estimation and Application of Long Memory Time Series Model". *Journal of Time Series Analysis*, 4(4), 221-238.
- Granger, C.W.J. e R. Joyeux (1980). "An Introduction to Long Memory Time Series Models and Fractional Differencing". *Journal of Time Series Analysis*, 1(1), 15-29.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, 68(1), 165-176.
- Huber, P.J. (1981). *Robust Statistics*. *New York: Wiley*.
- Hurst, H.E. (1951). "Long-term Storage Capacity of Reservoirs". *Transactions of American Society in Civil Engineers*, 116, 165-167.
- Hurvich, C.M. e B.K. Ray (1995). "Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes". *Journal of Time Series Analysis*, 16(1), 17-042.
- Lo, A. W. (1991). "Long Term Memory in Stock Market Prices", *Econometrica*, 59, 1279-1313.
- Montanari, A., R. Rosso e M.S. Taqqu (2000). "A Seasonal Fractional ARIMA Model Applied to the Nile River Monthly Flows at Aswan". *Water Resources Research*}, 36(5), 1249-1259.

- Ooms, M (1995). "Flexible Seasonal Long Memory and Economic Time Series". *Preprint of the Econometric Institute*, Erasmus University, Rotterdam.
- Peiris, M.S. e N. Singh (1996). "Predictors for Seasonal and Nonseasonal Fractionally Integrated Arima Models". *Biometrika*, 38(6), 741-752.
- Porter-Hudak, S. (1990). "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates". *Journal of the American Statistical Association*, 85(410), 338-344.
- Ray, B.K. (1993). "Long-Range Forecasting of IBM Product Revenues Using a Seasonal Fractionally Differenced ARMA Model". *International Journal of Forecasting*, 9, 255-269.
- Reisen, V.A. (1994). "Estimation of the Fractional Difference Parameter in the ARIMA(p,d,q) Model using the Smoothed Periodogram". *Journal of Time Series Analysis*, 15(3), 335-350.
- Robinson, P.M. (1995). "Log-Periodogram Regression of Time Series with Long Range Dependence". *Annals of Statistics*, 23(3), 1048-1072.
- Rousseeuw, P.J. (1984). "Least Median of Square Regression". *Journal of the American Statistical Association*, 79, 871-880.
- Rousseeuw, P.J. e A.M. Leroy (2003). *Rubust Regression and Outlier Detection*. New Jersey: John Wiley.
- Sowell, F. (1992). Maximum likelihood estimation os stationary univariate fractionally integrated time series models. *Journal os Econometrics*, 53, 165-188.
- Velasco C. (1999). "Gaussian Semiparametric Estimation of Nonstationary Time Series". *Journal of Time Series Analysis*, 20(1), 87-127.
- Whittle, P. (1951). "Hypothesis Testing in Time Series Analysis". *New York: Hafner*.
- Yohai, V. J. (1987). "High breakdown point and high efficiency robust estimates for regression". *Annals of Statistics*, 15, 642-656.

## Anexo A

### A.1 Estimadores da Função Densidade Espectral

Nesta seção, apresentamos o estimador para a função densidade espectral de um processo estocástico. Estes resultados são muito úteis na seção 3, onde descrevemos alguns métodos de estimação para estimar o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ .

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário com média  $\mu$  e função de autocovariância, denotada por  $\gamma_X(\cdot)$ , absolutamente convergente. Sob estas condições, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função densidade espectral dada por

$$f_X(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) e^{-ik\omega}, \text{ para } \omega \in (0, \pi].$$

Seja  $\{X_t\}_{t=1}^n$  uma série temporal gerada a partir de um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$ , com função de autocovariância não necessariamente absolutamente convergente. Então, a função periodograma da série temporal  $\{X_t\}_{t=1}^n$  é definida por

$$I(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-it\omega} \right|^2, \text{ para todo } \omega \in (0, \pi]. \quad (\text{A.1})$$

Em vista disso, Tukey (1949) propõe um estimador para a função densidade espectral baseado na função de autocovariância ponderada ou suavizada o qual é não-viciado e consistente. Neste caso, o processo de suavização ocorre no domínio do tempo e para a obtenção do estimador passa-se para o domínio da frequência.

Seja  $\{X_t\}_{t=1}^n$  uma série temporal gerada a partir de um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$ . Então, o estimador da função densidade espectral, chamado estimador suavizado de covariâncias, denotado por  $I_s(\cdot)$ , é dado por

$$I_s(\omega) = \frac{1}{2\pi} \sum_{|k| \leq m} \Lambda\left(\frac{k}{m}\right) \hat{\gamma}_X(k) e^{-ik\omega}, \text{ para todo } \omega \in (0, \pi] \quad (\text{A.2})$$

onde  $\hat{\gamma}_X(\cdot)$  é a função de autocovariância amostral dada pela expressão

$$\hat{\gamma}_X(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X})$$



$m = n^\beta$ , para  $0 \leq \beta \leq 1$  e  $\Lambda(x)$  é uma função par, contínua por intervalos de  $x$  satisfazendo as condições

A.i)  $0 \leq \Lambda(x) \leq \Lambda(0) = 1$ ,

A.ii)  $\Lambda(-x) = \Lambda(x)$ , para todo  $x$ ,

A.iii)  $\Lambda(x) = 0$ , para  $|x| > 1$ .

A função  $\Lambda(\cdot)$  é chamada de função de ponderação, função peso ou núcleo.

Note que, pela propriedade (A.iii) da função de ponderação, o produto  $\Lambda(\frac{k}{m})\hat{\gamma}_X(k) = 0$ , para  $|k| > m$ .

## A.2 Esperança e Variância Condicionais

O teorema, a seguir, apresenta para  $\{X_t\}_{t \in \mathbb{Z}}$ , de um processo  $SARFIMA(0,D,0)_s$ , com sazonalidade  $s$ , média zero e  $D \in (-0,5; 0,5)$ . A esperança e a variância condicionais que dependem apenas dos valores passados distantes por múltiplos da sazonalidade  $s$ . Este teorema é importante no método de geração dos processos  $\{X_t\}_{t \in \mathbb{Z}}$ .

**Teorema** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $SARFIMA(0,D,0)_s$  com sazonalidade  $s \in \mathbb{N} - \{0\}$  e  $D \in (-0,5; 0,5)$ . A esperança condicional e a variância condicional de  $X_t$ , dado  $X_l, l < t$ , denotadas respectivamente por  $m_t = E(X_t | X_l, l < t)$  e  $v_t = Var(X_t | X_l, l < t)$ , são dadas por

$$\begin{cases} m_\zeta = 0, \text{ para } \zeta = 1, \dots, s-1, \\ m_{sk} = \sum_{j=1}^k \phi_X(sk, sj) X_{s(k-j)}, \text{ para } k \in \mathbb{N}, \\ m_{sk+\zeta} = \sum_{j=1}^k \phi_X(sk+\zeta, sj) X_{sk+\zeta-sj}, \end{cases}$$

e

$$\begin{cases} v_\zeta = \sigma_\varepsilon^2, \text{ para } \zeta = 1, \dots, s-1, \\ v_{sk} = \sigma_\varepsilon^2 \prod_{j=1}^k (1 - \phi_X^2(sj, sj)), \text{ para } k \in \mathbb{N}, \\ v_{sk+\zeta} = v_{sk}, \end{cases}$$

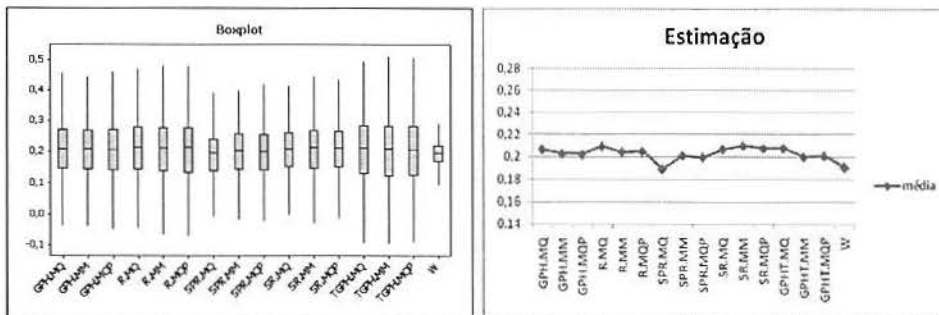
onde  $t = \zeta$  determina a média e a variância em defasagens menores que  $s$ ,  $t = sk$  para defasagens múltiplas de  $s$ ,  $\phi_X(\cdot, \cdot)$  é a função de autocorrelação parcial do processo  $\{X_t\}_{t \in \mathbb{Z}}$  e  $\sigma_\varepsilon^2$  é a variância do processo ruído branco.

**Prova** Ver Bisognin e Lopes (2007).

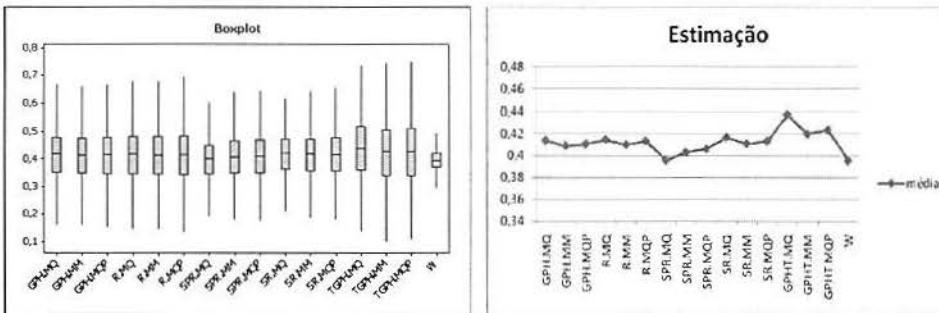
## Anexo B

As **Figura 13** à **Figura 28**, a seguir, apresentam os resultados da estimação do parâmetro  $D$  dos processos  $SARFIMA(0,D,0)_s$ , quando  $\alpha \in \{0,6;0,7\}$ ,  $D \in \{0,2;0,4\}$ ,  $n = 1000$ ,  $\ell = 2$  e  $re = 1000$ .

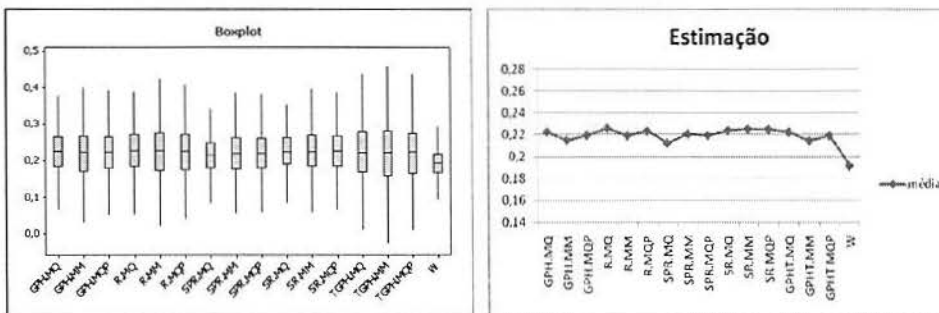
Para melhor visualização das comparações dos estimadores para cada  $D$ ,  $s$  e  $\alpha$  do processo  $SARFIMA(0,D,0)_s$ , apresentamos os *boxplot* e gráficos de média dos estimadores.



**Figura 13** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,2$ ,  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .



**Figura 14** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,4$ ,  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .



**Figura 15** Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)_s$ , quando  $D = 0,2$ ,  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

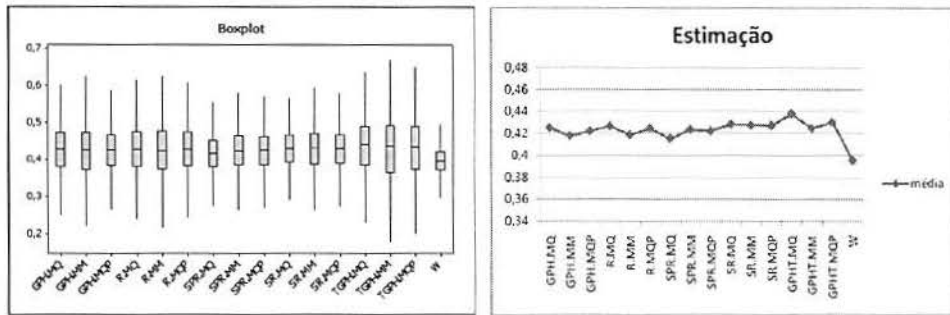


Figura 16 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 2$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

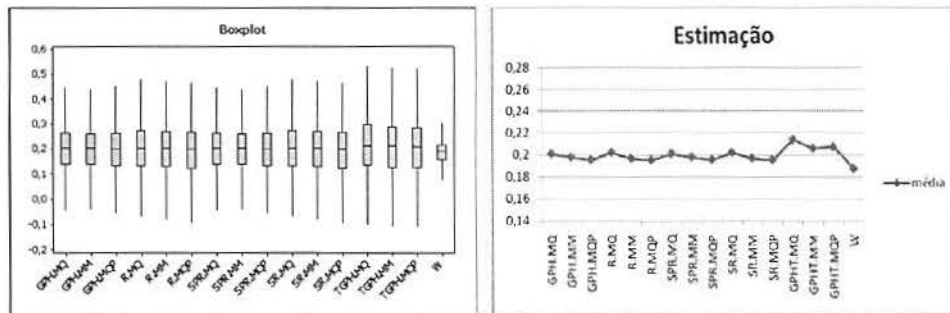


Figura 17 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

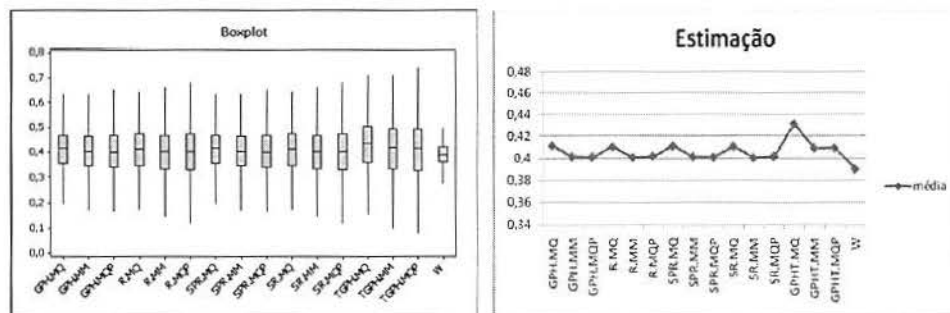


Figura 18 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

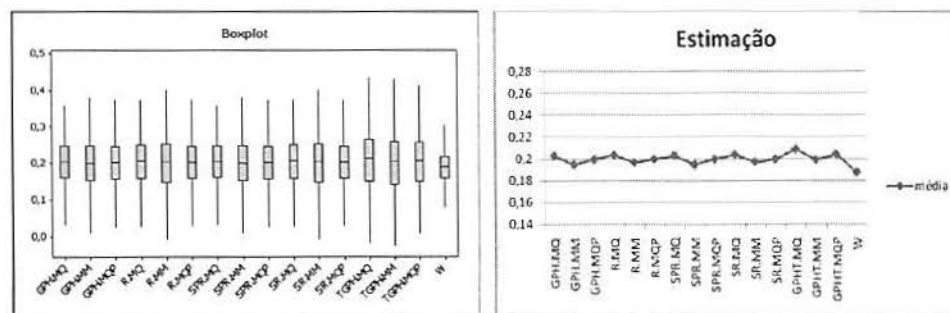


Figura 19 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

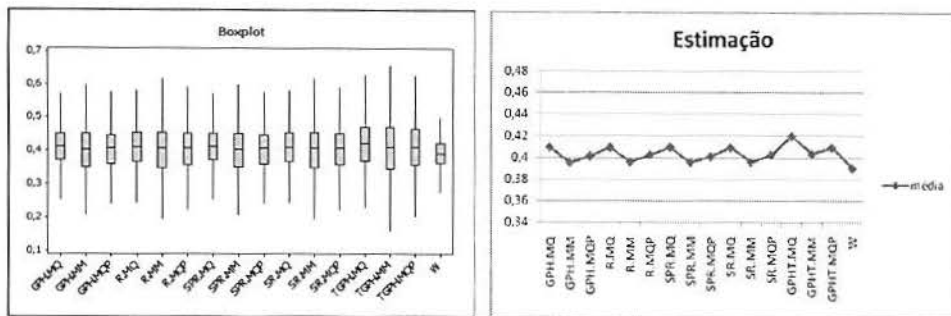


Figura 20 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 3$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

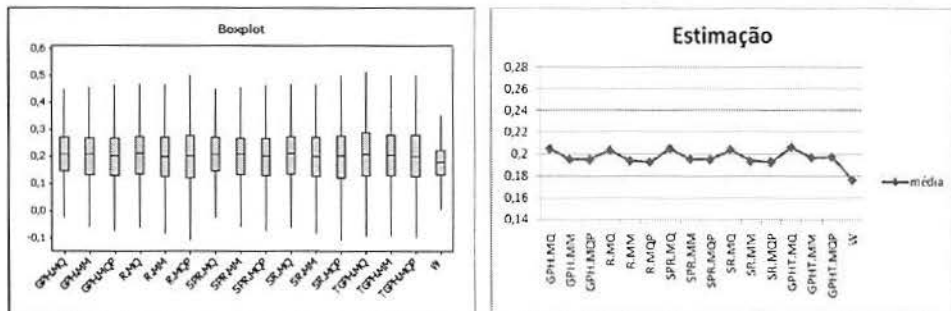


Figura 21 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

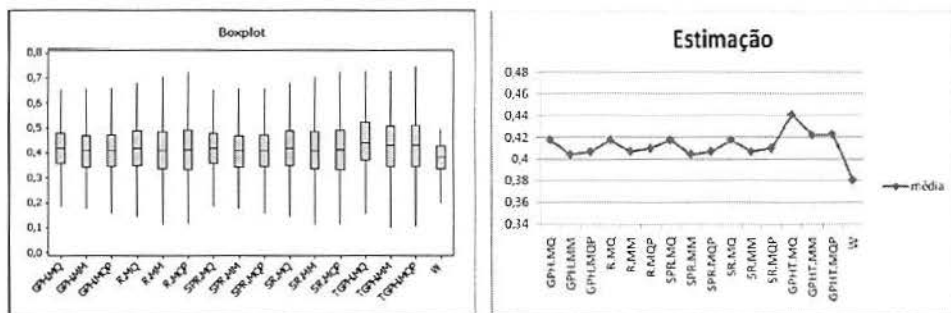


Figura 22 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

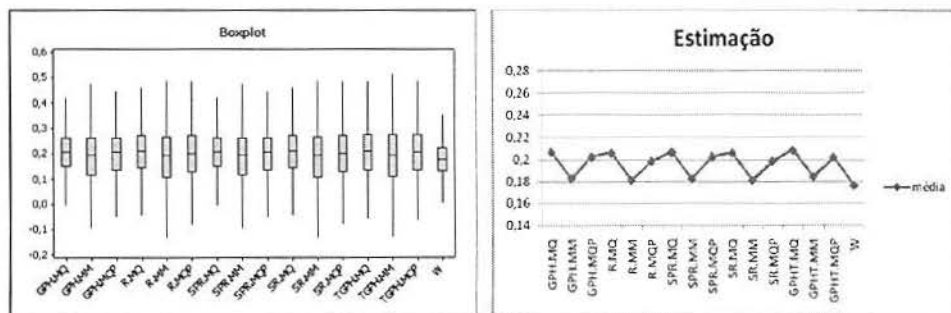


Figura 23 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

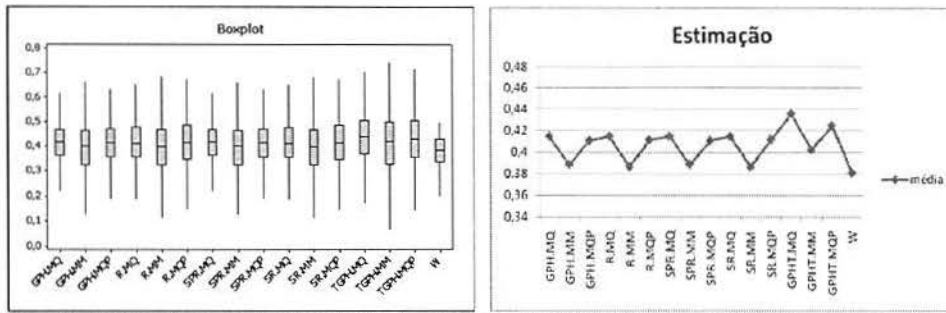


Figura 24 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 6$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

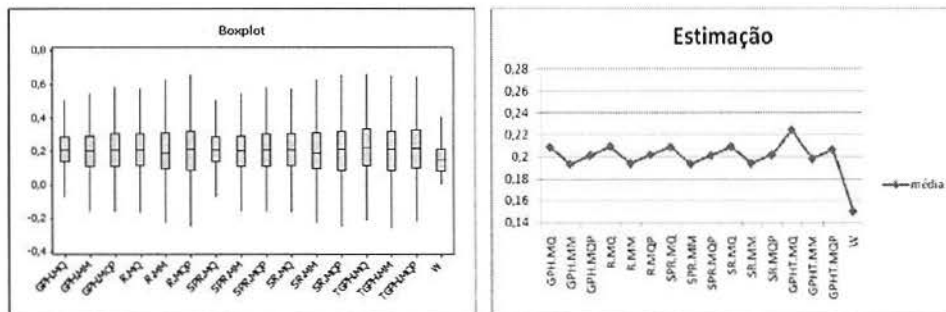


Figura 25 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

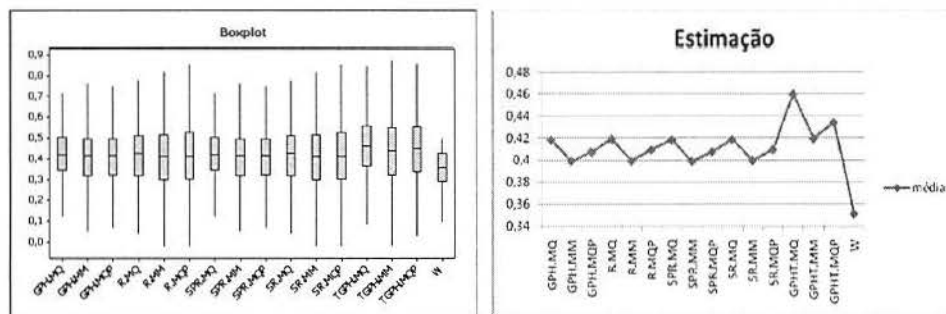


Figura 26 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,6$ .

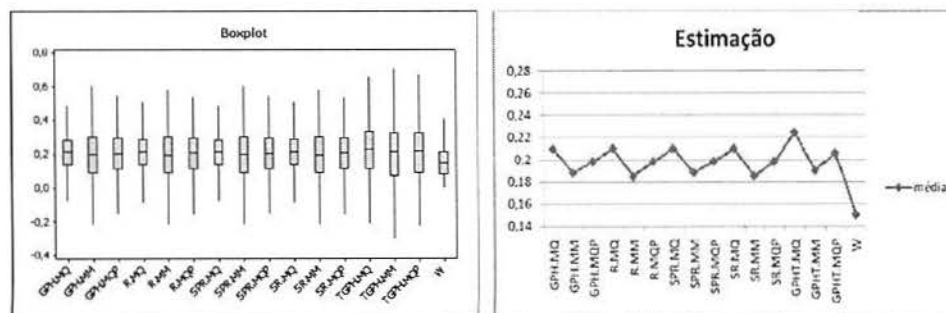


Figura 27 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,2$ ,  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .

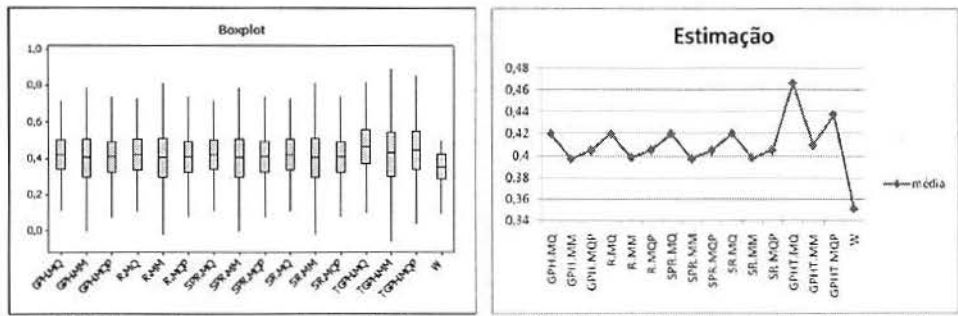


Figura 28 Estimação para o parâmetro  $D$  do processo  $SARFIMA(0,D,0)$ , quando  $D = 0,4$ ,  $s = 12$ ,  $n = 1000$ ,  $re = 1000$ ,  $\ell = 2$  e  $\alpha = 0,7$ .