Clementina Maria Silva Marques Oliveira

PROCESSOS DE MEMÓRIA LONGA E MÉTODOS DE ESTIMAÇÃO DO PARÂMETRO QUE CONTROLA A MEMÓRIA LONGA



Departamento de Matemática Aplicada
Faculdade de Ciências da Universidade do Porto
Janeiro de 2004

Clementina Maria Silva Marques Oliveira

PROCESSOS DE MEMÓRIA LONGA E MÉTODOS DE ESTIMAÇÃO DO PARÂMETRO QUE CONTROLA A MEMÓRIA LONGA

Tese submetida à Faculdade de Ciências da Universidade do Porto
para a obtenção do grau de Mestre em Estatística

Departamento de Matemática Aplicada
Faculdade de Ciências da Universidade do Porto
Janeiro de 2004

Dissertação orientada por

Maria Eduarda da Rocha Pinto Augusto da Silva

Professora Auxiliar do
Departamento de Matemática Aplicada da
Faculdade de Ciências da Universidade de Porto

Ao Luís e ao Rafael

Agradecimentos

Quero agradecer à Professora Doutora Maria Eduarda Silva toda a disponibilidade e dedicação que sempre mostrou orientando-me na elaboração deste trabalho, e à Susana Barbosa que me ajudou no esclarecimento de algumas dúvidas.

Também quero agradecer à minha família e ao meu marido toda a ajuda, apoio, incentivo e carinho em todos os momentos, principalmente nos mais difíceis.

Resumo

Em certas áreas de investigação tais como hidrologia e economia, constatou-se que a dependência entre observações distantes no tempo, embora seja pequena não é desprezável. As séries que apresentam esta característica são designadas por séries temporais de memória longa. Desde então foram surgindo diversos estimadores que permitem estimar o parâmetro que caracteriza a memória longa, sendo as suas propriedades, ao longo do tempo, sucessivamente generalizadas de forma a justificar a aplicação daqueles estimadores a uma maior variabilidade de séries temporais.

O objectivo deste trabalho é investigar os diversos métodos de simulação de processos de memória longa e apresentar os principais estimadores para séries temporais de memória longa estacionárias e com variância infinita.

Posteriormente são aplicados alguns estimadores a uma série hidrológica com o objectivo de se inferir quanto à possibilidade desta possuir memória longa.

Abstract

It has been verified in certain areas of researching such as hydrology and economy, the dependence among some observations, distant in time, is small but not small enough to disregard. Series that present such characteristic are called as long-memory time series. Since then, several estimators have been developed to estimate the long-memory parameter. Its properties along with the time scale are consequently generalized to justify the application of those estimators to a major variability of time series.

The objective of this work is to study various simulation methods for long-memory processes and to present well-known estimators for the long-memory parameter of a stationary time series and with infinite variance.

Some estimators were applied to a hydrological series with the purpose to infer about the presence of a long-memory behaviour.

Índice

Abreviaturas **9**

1. Introdução **10**

2. Processos de memória longa **13**

2.1. Definição de memória longa 13

2.2. Modelos de memória longa 14

2.2.1. Processos FGN 15

2.2.2. Processos ARFIMA 17

2.2.3. Processos FEXP 21

2.2.4. Outros processos de memória longa 22

2.3. Definição de memória longa em processos com variância infinita 24

2.4. Processos ARFIMA com inovações com variância infinita (inovações não Gausseanas) 24

2.5. Caracterização da memória longa 30

2.6. Métodos de simulação de processos de memória longa 31

3. Estimação dos parâmetros dos processos de memória longa **41**

3.1. Estimação semi-paramétrica 41

3.1.1. Métodos locais 42

3.1.1.1. Estimadores do tipo GPH 42

3.1.1.2. GSE ou LWE 61

3.1.2. Métodos globais 64

3.1.2.1. Estimador FEXP 64

3.1.2.2. Estimador FEXP generalizado 70

3.2. Estimação paramétrica 75

3.2.1 Estimador de máxima verosimilhança exacto 75

3.2.2. Estimador de Whittle 82

3.2.3. Estimador dos parâmetros do processo ARFIMA com inovações SαS 84

4. Aplicação de alguns métodos de estimação a uma série temporal hidrológica **88**

4.1. Análise da série em estudo 88

4.2. Estimador de Geweke e Porter-Hudak 92

4.3. Estimador de Reisen 94

4.4. Estimador de Robinson 96
4.5. Estimador de Velasco 99
4.6. Estimação pelo método de máxima verosimilhança 101
4.7. Comentários finais 102

Apêndice A 103

Apêndice B 106

Apêndice C 110

Referências 115

Abreviaturas

ACF	função de auto-correlação ("autocorrelation function")
AR	auto-regressivo ("autoregressive")
ARFIMA	"autoregressive fractional integrated moving average"
ARFISMA	processo ARFIMA sazonal ("autoregressive fractional integrated seasonal moving average")
ARIMA	auto-regressico integrado e de médias móveis ("autoregressive integrated moving average")
EQM	erro quadrático médio
et al.	e outros ("et alii")
EXP	exponencial ("exponential")
FBM	movimento Browniano fraccionário ("fractional Brownian motion")
FEXP	"fractional exponential"
FGN	ruído Gausseano fraccionário ("fractional Gaussian noise")
i. i. d.	independentes e identicamente distribuídas
k-factor GARMA	factor-k auto-regressivo e de médias móveis Gegenbauer (" k-factor autoregressive moving average Gegenbauer")
MA	médias móveis ("moving average")
NPL	"normal profile log likelihood"
PACF	função de auto-correlação parcial ("partial autocorrelation function")
SαS	α-estável simétrica ("symmetric α stable")
SCALM	memória longa sazonal/ cíclico assimétrico ("seasonal/ cyclical asymmetric long memory")
SPP	processos sazonais persistentes ("seasonal persistent process")
STL	"seasonal-trend decomposition procedure on loess"
TSS	"typical spectral shape"

1. Introdução

Tem-se verificado em certas séries temporais ligadas a diversas áreas de investigação tais como astronomia, biologia, meteorologia, hidrologia e economia que a dependência entre observações distantes apesar de ser pequena não é desprezável, exibindo uma propriedade que se designa por memória longa (ou persistência).

Na literatura têm sido propostos vários modelos para modelar séries com memória longa. Destes destacam-se os modelos ARFIMA ("Autoregressive Fractional Integrated Moving Average") e FEXP ("Fractional Exponential"), os quais permitem modelar séries temporais que exibem simultaneamente memória longa e curta.

Foram sugeridos na literatura uma grande variedade de estimadores do parâmetro de memória longa. Existem duas abordagens principais, no domínio da frequência, que permitem a estimação do parâmetro que controla a estrutura das correlações das observações distantes: abordagens paramétrica e semi-paramétrica. Na abordagem paramétrica é necessário especificar totalmente o modelo, enquanto que na abordagem semi-paramétrica não é preciso definir correctamente a totalidade do modelo paramétrico.

Os estimadores paramétricos do parâmetro que controla a estrutura das correlações das observações distantes são consistentes, possuem normalidade assimptótica e com base no pressuposto da Gausseanidade são assimptoticamente eficientes. O EQM (erro quadrático médio) destes estimadores, baseado numa amostra de dimensão n, é tipicamente $O(1/n)$ se o modelo paramétrico estiver correctamente especificado, mas o estimador pode ser inconsistente se o modelo estiver mal especificado (Agiakloglou et al. (1993), Cheung & Diebold (1994), Reisen (1994), Robinson (1994a, 1994b, 1995a, 1995b), Taqqu & Teverovski (1996, 1997) e Smith et al. (1997)).

Este revés na estimação paramétrica motiva o uso de estimadores semi-paramétricos.

Dentro da abordagem semi-paramétrica existem dois métodos de estimação principais: método de estimação local e método de estimação global. No primeiro método usa-se um número reduzido de frequências de Fourier na vizinhança de zero na regressão do logaritmo do periodograma, enquanto que no método de estimação global usam-se todas as frequências de Fourier na referida regressão.

Do ponto de vista teórico, o maior inconveniente dos estimadores semi-paramétricos locais é que estes estimadores têm tipicamente EQM assimptótico proporcional a $n^{-\alpha}$ com $\alpha \leq 4/5$ independentemente da regularidade da função que caracteriza a estrutura das correlações das observações próximas (Giraitis et al. (1997), Hurvich et al. (1998)).

No entanto, usando métodos de estimação semi-paramétricos globais é possível obter estimadores cuja convergência para zero do EQM aumenta à medida que a função que governa a estrutura das correlações das observações próximas se torna mais "suave", podendo aquela convergência ser tão rápida quanto (log (n))/n se o logaritmo da função referida for infinitamente diferenciável.

Surgem frequentemente na literatura (Bollerslev & Jubinski (1999)) notas dizendo que é difícil estimar os modelos ARFIMA pelo método de máxima verosimilhança estando esta dificuldade, em grande parte, relacionada com problemas computacionais.

Também é difícil aplicar os estimadores semi-paramétricos locais do parâmetro que caracteriza a memória longa a uma série temporal concreta. Dependendo da estrutura da componente de memória curta do processo, o número de observações usadas na regressão do logaritmo do periodograma pode influenciar fortemente o viés, a variância e o EQM dos estomadores obtidos através de métodos semi-paramétricos locais, o que pode levar à obtenção de uma grande variedade de estimativas consoante o número de observações usadas na regressão (Robinson (1995a, 1995b), Hurvich et al. (1998)).

Apesar de ter sido deduzido o número óptimo de observações a serem usadas, no caso do estimador de Geweke & Porter-Hudak (1983), este resultado não pode ser usado na prática porque pressupõe o conhecimento da função de densidade espectral da componente de memória curta do processo, o que geralmente se desconhece no estudo de uma série.

Andrews & Guggenberger (2003) propuseram um estimador semi-paramétrico local que apresenta um viés inferior ao viés do estimador de Geweke & Porter-Hudak (1983) tendo também deduzido o número óptimo de observações a serem usadas na regressão do logaritmo do periodograma. No entanto, o valor óptimo depende de um número inteiro não negativo r, o qual estabelece o número de potências pares das frequências de Fourier a serem usadas na regressão do logaritmo do periodograma, ω_j^2, ω_j^4, ..., ω_j^{2r}. Como, na prática, não se sabe qual o valor de r que se deve usar, não se pode calcular o número óptimo de observações que se devem usar na referida regressão.

Apesar da grande quantidade de estimadores que actualmente existem para estimar o parâmetro que caracteriza a memória longa, apenas alguns se adequam ao caso das séries não Gausseanas (Robinson(1995b), Kokoszka & Taqqu (1996), Velasco (2000)).

O objectivo deste trabalho é examinar os principais estimadores do parâmetro que caracteriza a memória longa e suas respectivas propriedades. Também se pretende fazer um estudo dos diferentes métodos de simulação de processos de memória longa e do "software" que permite simular os processos. Finalmente, pretende-se aplicar diferentes métodos de estimação a uma

série real para se inferir quanto à existência ou não da característica de memória longa na série.

No capítulo 2 define-se processo de memória longa estacionário e processo com variância infinita e são apresentados vários processos relativos a estes dois casos. Neste capítulo também são apresentados vários métodos de simulação de processos de memória longa sendo comparados em termos de rapidez computacional e de precisão dos resultados pretendidos. Também são abordadas algumas funções disponíveis para a simulação dos processos ARFIMA com inovações Gausseanas e não Gausseanas em alguns programas computacionais de estatística.

No capítulo 3 são apresentados vários estimadores, no domínio da frequência, dos processos de memória longa e suas respectivas propriedades principais.

No capítulo 4 vão-se aplicar diversos estimadores a uma série hidrológica, a qual corresponde às medições diárias do fluxo médio (m^3/s) do rio Shiel, Reino Unido.

Mediante as características dos diversos estimadores vai-se procurar explicar os resultados obtidos para cada caso e comparar as estimativas do parâmetro que controla a estrutura das correlações das observações distantes obtidas por diferentes estimadores. Finalmente são feitos alguns comentários ao trabalho realizado.

2. Processos de memória longa

Neste capítulo vai-se definir processo de memória longa, quer no caso dos processos estacionários (secção 2.1.), quer no caso dos processos com variância infinita (secção 2.3.)
Nas secções 2.2. e 2.4. são descritos vários processos estacionários de memória longa e os processos de memória longa com variância infinita.
Na secção 2.5., é discutida a possibilidade de uma definição geral de dependência de memória longa. Finalmente, na secção 2.6. vão ser apresentados diversos métodos de simulação de processos de memória longa. É feita uma breve discussão quanto à sua rapidez em termos computacionais e quanto à sua fiabilidade em termos de implementação de forma a se obterem simulações precisas. Também são abordadas algumas funções que estão disponíveis para a simulação de processos ARFIMA com inovações Gausseanas ou não Gausseanas.

2.1. Definição de memória longa

Um processo estacionário de memória longa pode ser definido através da sua função de auto-correlação ou da sua função de densidade espectral.

Definição 2.1.1: *Seja $\{X_t\}$ um processo para o qual existe um número real $\alpha \in]0, 1[$ e uma constante $c_\rho > 0$ tal que*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_\rho k^{-\alpha}} = 1$$

onde $\rho(k)$ representa a função de auto-correlação de $\{X_t\}$. Então $\{X_t\}$ é um processo estacionário de memória longa.

Definição 2.1.2: *Seja $\{X_t\}$ um processo para o qual existe um número real $\beta \in]0, 1[$ e uma constante $c_f > 0$ tal que*

$$\lim_{\omega\to 0} \frac{f(\omega)}{c_f |\omega|^{-\beta}} = 1 \qquad \textbf{(2. 1)}$$

onde $f(\omega)$ é a função de densidade espectral normalizada de $\{X_t\}$. Então $\{X_t\}$ é um processo estacionário de memória longa.

Portanto a persistência em séries temporais pode ser caracterizada através do decaimento hiperbólico da função de auto-correlação à medida que o "lag" aumenta ou da não sumabilidade da série das correlações e através do crescimento ilimitado da função de densidade espectral à medida que a frequência tende para zero.

Na literatura, os processos estacionários de memória longa surgem em diferentes contextos, consoante o trabalho que se pretende realizar.

Hurvich et al. (1998), Hurvich & Deo (1999), Moulines & Soulier (1999, 2000), entre outros, definem um processo estacionário, $\{X_t\}$, através da sua função de densidade espectral,

$$f(\omega) = \left|1 - e^{-i\omega}\right|^{-2d} f_u(\omega)$$

onde d é o parâmetro que governa a memória do processo. $d \in \left]-\frac{1}{2}, \frac{1}{2}\right[$, ou seja, o processo é estacionário e invertível. Quando d é positivo o processo é dito de memória longa, quando $d = 0$ o processo é de memória curta (ou seja, a função de auto-correlação decai exponencialmente à medida que o "lag" tende para infinito) e quando $d < 0$ o processo é de memória intermediária (a soma infinita do valor absoluto das correlações é sumável). A função $f_u(\omega)$ controla a estrutura das correlações das observações próximas.

É de notar que f_u além de dotar o modelo com uma estrutura de memória curta também é livre de qualquer imposição paramétrica. A função f_u pertence a uma classe ampla de funções, sendo f_u especificada pelo seu comportamento na frequência zero ou por uma dada regularidade em todo o domínio das frequências. Consoante o tipo de resultados que se pretendem deduzir, fazem-se diferentes imposições quanto às propriedades que f_u deve possuir. Por exemplo, Hurvich et al. (1998) assumem que $f_u(\omega)$ é uma função positiva, par, contínua em $[-\pi, \pi]$, limitada em $]-\pi, \pi[\setminus \{0\}$, $f_u'(0) = 0$ e as derivadas de segunda e terceira ordem são limitadas na vizinhança de zero. Moulines & Soulier (1999) impõem que $f_u(\omega)$ deve ser uma função positiva, periódica, de período 2π, contínua, e diferenciável em $[-\pi, \pi] \setminus \{0\}$ e $\exists\, c^* : \forall\, \omega \in]-\pi, \pi[\setminus \{0\}, |\, f_u'(0)| \leq c^* / |\, \omega|$.

2.2. Modelos de memória longa

Entre os diversos modelos para os quais (2. 1) se verifica, existem três classes de modelos que são de especial interesse:

(a) processos auto-semelhantes de incrementos estacionários[1], em particular, o ruído Gausseano fraccionário (FGN);

(b) os processos ARFIMA;

(c) os processos FEXP.

Os modelos mais simples que exibem memória longa (no sentido de possuirem menor número de parâmetros) são os processos auto-semelhantes.

Definição 2.2.1: *Um processo estocástico em tempo contínuo de parâmetro t, Y_t, é um processo auto-semelhante com parâmetro $H = d + 0.5$ se para qualquer factor de escala positivo c, o processo transformado (após uma transformação de escala do processo original Y_t) com parâmetro contínuo ct, $c^{-H}Y_{ct}$, é igual em distribuição ao processo Y_t.*

Os processos auto-semelhantes são caracterizados pelo decaimento hiperbólico das funções de auto-correlação. Os modelos auto-semelhantes e assimptóticamente auto-semelhantes são especialmente atractivos porque a memória longa pode ser caracterizada por um único parâmetro $H \in]0.5, 1[$. O grau de auto-semelhança e de persistência de uma série aumenta à medida que H tende para um. O parâmetro H é designado por coeficiente de Hurst ou por parâmetro de auto-semelhança (Hurst (1951)).

A auto-semelhança pode ser caracterizada de duas formas: a densidade espectral diverge na origem ($f(\omega) \sim \frac{1}{\omega^{\alpha}}$, $0 < \alpha < 1$) e a função de auto-correlação não é sumável.

2.2.1. Processos FGN

O primeiro processo de memória longa que surgiu na literatura foi o movimento Browniano fraccionário (FBM) (ver Mandelbrot (1965), Mandelbrot & Van Ness (1968)) em tempo contínuo, sendo o ruído Gausseano fraccionário (FGN) o correspondente em tempo discreto (ver Mandelbrot & Van Ness (1968), Mandelbrot & Wallis (1969)).

O movimento Browniano fraccionário $\{B_t, t \geq 0\}$ é um processo auto-semelhante Gausseano com incrementos estacionários, média zero, variância

---

[1] B_t é um processo com incrementos estacionários se para qualquer $k \geq 1$ e quaisquer k constantes $t_1, \ldots, t_k$, a distribuição de $\left(B_{t_1+c} - B_{t_1+c-1}, \ldots, B_{t_k+c} - B_{t_k+c-1}\right)$ não depende de $c \in \mathbb{R}$.

$$\mathrm{Var}(B_t) = t^{2H}$$

e covariância

$$\mathrm{E}(B_s B_t)) = \tfrac{1}{2}\,\sigma^2\{s^{2H} + t^{2H} - |s - t|^{2H}\}$$

onde σ^2 representa a variância dos incrementos do processo.

O FGN $\{X_t, t \geq 1\}$ corresponde aos incrementos do FBM, ou seja $X_t = B_{t+1} - B_t$.

X_t é um processo estacionário, Gausseano com média zero e com função de auto-correlação dada por

$$\rho(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]$$

com $k > 0$ e $-\frac{1}{2} < d < \frac{1}{2}$, sendo portanto, um processo persistente quando $H \in \left]\frac{1}{2}, 1\right[$. À medida que k tende para infinito

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \to 1.$$

A figura seguinte mostra o decaimento lento das correlações (teóricas) de um processo FGN com parâmetro auto-semelhante $H = 0.9$.

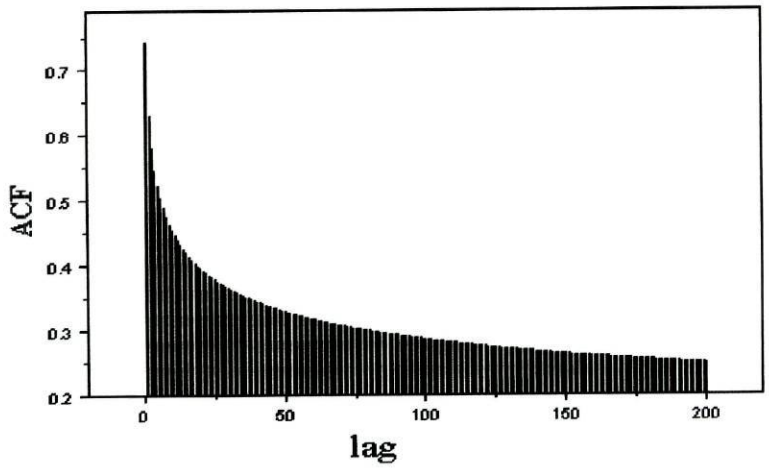


Figura 2.1: Função de auto-correlação teórica (ACF) de um processo FGN com $H = 0.9$.

A função de densidade espectral $f(\omega)$ quando $\omega \to 0$ é dada pela expressão

$$f(\omega) = 2c(1-\cos\omega)\sum_{j=-\infty}^{\infty}|2\pi j + \omega|^{-2H-1} \sim c\,|\omega|^{1-2H}, \quad \omega \in [-\pi, \pi]$$

e

$$c = \frac{\sigma^2}{2\pi}\sin(\pi H)\Gamma(2H+1)$$

onde $\sigma^2 = \text{var}(X_t)$ e $\Gamma(.)$ é a função Gama.

Quando uma série temporal é assimptóticamente auto-semelhante, a sua função de auto-correlação é da forma $\rho(k) \sim k^{2H-2}.L(k), k \to \infty$ para $H \in \left]\frac{1}{2}, 1\right[$ sendo L uma função de variação lenta[2] .

O processo assimptóticamente auto-semelhante que tem sido mais estudado é o ARFIMA (Beran (1992), Samorodnitsky & Taqqu (1994), Garrett & Willinger (1994), Houser (1999) e Kokoszka & Taqqu (2001)).

2.2.2. Processos ARFIMA

Os modelos ARFIMA foram introduzidos por Granger & Joyeux (1980) e Hosking (1981) e surgiram como uma extensão natural dos modelos ARIMA ("Autoregressive Integrated Moving Average") propostos por Box & Jenkins (1976), na medida em que permitem que d seja real em vez de ser inteiro, como considerado no modelo ARIMA. Este tipo de modelo permite modelar séries temporais que apresentam simultaneamente características de memória longa e intermediária.

Definição 2.2.2: (Hosking (1981)) *$\{X_t\}_{t\geq 1}$ é um processo ARFIMA(0, d, 0) com média zero quando pode ser representado na forma* $(1-B)^d X_t = \varepsilon_t$ *onde, $\{\varepsilon_t\}_{t\geq 1}$ é um processo de ruído branco que consiste numa sequência de variáveis aleatórias independentes e identicamente distribuídas (i. i. d.) com média zero, variância* σ_ε^2. $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$ *com* $BX_t = X_{t-1}$.

O processo ARFIMA(0, d, 0) é equivalente ao processo FGN, estando os parâmetros dos dois processos relacionados através da igualdade $H = d + ½$ (ver Reisen (1994)).

Para os resultados seguintes vai-se assumir que $\sigma_\varepsilon^2 = 1$.

Teorema 2.2.1: (Hosking (1981)) *Seja $\{X_t\}$ um processo ARFIMA(0, d, 0).*

---

[2] $L(x)$ é uma função de variação lenta se e só se $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ para qualquer $t > 0$.

a) *Quando* $d < \frac{1}{2}$, $\{X_t\}$ *é um processo estacionário e pode ser representado na forma de um processo de médias móveis de ordem infinita (MA(*∞*)):*

$$X_t = \Psi(B)\varepsilon_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$$

onde

$$\psi_k = \frac{(k+d-1)!}{k!(d-1)!} \sim \frac{k^{d-1}}{(d-1)!}, \quad k \to \infty.$$

b) *Quando* $d > -\frac{1}{2}$, $\{X_t\}$ *é um processo invertível e pode ser representado na forma de um processo auto-regressivo de ordem infinita (AR(*∞*)):*

$$\Pi(B)X_t = \sum_{k=0}^{\infty} \pi_k X_{t-k} = \varepsilon_t$$

onde

$$\pi_k = \frac{(k-d-1)!}{k!(-d-1)!} \sim \frac{k^{-d-1}}{(-d-1)!}, \quad k \to \infty$$

Para os resultados seguintes assuma-se que $-\frac{1}{2} < d < \frac{1}{2}$, *condição para a invertibilidade e estacionaridade.*

c) *A função de densidade espectral de* $\{X_t\}$ *é*

$$f_{ARFIMA}(\omega) = \frac{1}{2\pi}\left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d} \sim \omega^{-2d}, \quad \omega \to 0$$

para $0 < \omega \leq \pi$.

d) *A função de auto-covariância de* $\{X_t\}$ *é*

$$\gamma(k) = E(X_t X_{t-k}) = \frac{(-1)^k(-2d)!}{(k-d)!(-k-d)!}.$$

A função de auto-correlação de $\{X_t\}$ *é*

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!} \sim \frac{(-d)!}{(d-1)!}k^{2d-1}, \quad k \to \infty \qquad (k = 0, \pm 1, \ldots).$$

Em particular,

$$\gamma(0) = \frac{(-2d)!}{[(-d)!]^2} \quad e \quad \rho(1) = \frac{d}{1-d}.$$

e) *As auto-correlações inversas de* $\{X_t\}$ *são*

$$\rho_{inv}(k) = \frac{d!(k-d-1)!}{(-d-1)!(k+d)!} \sim \frac{d!}{(-d-1)!} k^{-1-2d}, \quad k \to \infty.$$

f) *As correlações parciais de* $\{X_t\}$ *são* $\phi_{kk} = \frac{d}{k-d}$ *(k = 1, 2, ...).*

Definição 2.2.3: (Hosking (1981)) *Um processo ARFIMA(p, d, q) onde p e q são inteiros e d é real é um processo* $\{X_t\}$ *que pode ser representado na forma*

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \qquad \textbf{(2. 2)}$$

onde $\Phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ *e* $\Theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ *são polinómios e* $\{\varepsilon_t\}$ *é um ruído branco.*

Re-escrendo (2. 2) da forma

$$X_t = \Phi^{-1}(B)\Theta(B)X_t^*$$

sendo

$$X_t^* = (1-B)^{-d}\varepsilon_t$$

um processo ARFIMA(0, *d*, 0). O processo ARFIMA(*p*, *d*, *q*) corresponde assim a passar um processo ARFIMA(0, *d*, 0) por um filtro ARMA(*p*, *q*).

Teorema 2.2.2: (Hosking (1981)) *Seja* $\{X_t\}$ *um processo ARFIMA(p, d, q). Então:*

a) $\{X_t\}$ *é estacionário se* $d < \frac{1}{2}$ *e todas as raízes da equação* $\Phi(z) = 0$ *estão fora do círculo unitário;*

b) $\{X_t\}$ *é invertível se* $d > -\frac{1}{2}$ *e todas as raízes da equação* $\Theta(z) = 0$ *estão fora do círculo unitário.*

Se $\{X_t\}$ *é um processo estacionário e invertível, com função de densidade espectral* $f(\omega)$ *e função de auto-correlação* $\rho(k)$*, então:*

c) $\lim_{\omega \to 0} \omega^{2d} f_{ARFIMA}(\omega)$ *existe e é finito;*

d) $\lim_{k \to \infty} k^{1-2d} \rho(k)$ *existe e é finito.*

A função de densidade espectral de um processo ARFIMA(*p*, *d*, *q*) é dada por

$$f_{ARFIMA}(\omega) = \left|1 - e^{-i\omega}\right|^{-2d} f_{ARMA}(\omega) = \left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d} f_{ARMA}(\omega)$$

onde $f_{ARMA}(\omega)$ é a função de densidade espectral de um processo ARMA(p, q). Então,

$$f_{ARFIMA}(\omega) = \frac{1}{2\pi}\frac{\left|\Theta\left(e^{-i\omega}\right)\right|^2}{\left|\Phi\left(e^{-i\omega}\right)\right|^2}\left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d}$$

(Beran(1994)).

O decaimento hiperbólico da função de auto-correlação e o crescimento ilimitado da função de densidade espectral nas frequências próximas de zero para $p = q = 0$ são ilustrados através das figuras 2.2 e 2.3, reapectivamente.

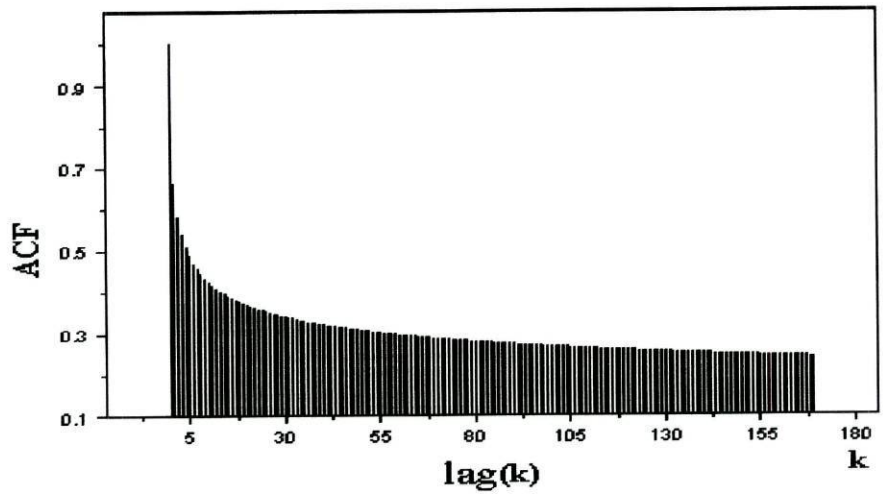


Figura 2.2: Função de auto-correlação teórica (ACF) de um processo ARFIMA(0, 0.4, 0)

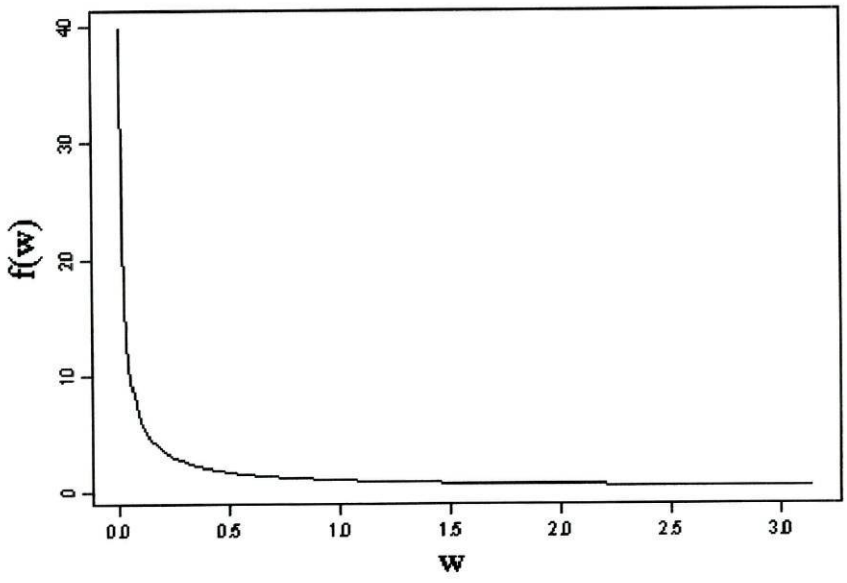


Figura 2.3: Função de densidade espectral de um processo ARFIMA(0, 0.4, 0)

Na prática, os modelos ARFIMA(p, d, q) de maior interesse são aqueles que apresentam valores pequenos para p e q sendo no apêndice A apresentados, de uma forma mais detalhada os processos ARFIMA(1, d, 0) e ARFIMA(0, d, 1).

2.2.3 Processos FEXP

Os modelos FEXP são uma extensão dos modelos EXP ("exponential") de Bloomfield (1973) e foram introduzidos no contexto dos processos de memória longa por Janacek (1982) para modelarem processos estacionários de memória longa.

Os modelos EXP são definidos através da sua função de densidade espectral

$$f(\omega) = \exp\left(\sum_{j=0}^{h} \eta_j \cos(\omega j) \right).$$

A estimação dos parâmetros dos modelos FEXP foi discutida por Robinson (1990) e por Beran (1993). Tal como os modelos ARFIMA, os processos FEXP também permitem modelar séries temporais que apresentam simultaneamente, características de memória longa e curta.

Definição 2.2.4: (Beran (1993)) *Seja* $g : [-\pi, \pi] \rightarrow \mathbb{R}^+$ *uma função positiva de tal forma que* $\lim_{x \to 0} \frac{g(x)}{x} = 1$ *e* $g(x) = g(-x)$. *Defina-se* $h_0 \equiv 1$, *e sejam* $v_1, v_2, \ldots, v_h$ *funções continuas por pedaços no intervalo* $[-\pi, \pi]$. *Assuma-se também que* $v_k(x) = v_k(-x)$ *e, para qualquer n, a matriz V,* $n^* \times (h+1)$ *com vectores coluna*

$$\left\{ v_k\left(\frac{2\pi}{n}\right), v_k\left(\frac{2\pi 2}{n}\right), v_k\left(\frac{2\pi 3}{n}\right), \ldots, v_k\left(\frac{2\pi n^*}{n}\right) \right\}^T \qquad (k = 0, 1, \ldots, h)$$

é não singular. Além disso, seja $\theta = (\eta_0, H, \eta_1, \ldots, \eta_h)$ *um vector real com* $\frac{1}{2} \leq H < 1$.

Designamos Y_t *por processo FEXP com componentes de memória curta* $v_1, v_2, \ldots, v_h$ *e com componente de memória longa g, se a função de densidade espectral é dada por*

$$f(\omega;\theta) = g(x)^{1-2H} \exp\left\{ \sum_{j=0}^{h} \eta_j v_j(\omega) \right\}$$

É de notar que n^* é a parte inteira de $\frac{1}{2}(n-1)$.

As duas classes seguintes de modelos FEXP são especialmente úteis.

(a) $f(\omega)=\left|1-e^{-i\omega}\right|^{1-2H}\exp\left(\sum_{j=0}^{h}\eta_j\cos(\omega j)\right)$

Esta classe de modelos FEXP é útil porque qualquer função "suave" pode ser expressa como uma série de Fourier (finita) $\sum_{j=0}^{h}\eta_j\cos(\omega j)$ e portanto, estes modelos podem modelar séries cuja função de densidade espectral seja suficientemente "suave".

(b) $f(\omega)=\left|1-e^{-i\omega}\right|^{1-2H}\exp\left(\sum_{j=0}^{h}\eta_j\omega^j\right)$

O logaritmo da função de densidade espectral é a soma da componente de memória longa $(1-2H)\log\left|1-e^{-i\omega}\right|$ e um polinómio em ω, permitindo uma grande flexibilidade em termos de modelação.

Ao contrário dos modelos ARFIMA, nos quais é preciso impor restrições nos parâmetros auto-regressivos e de médias móveis para assegurar a estacionaridade e a invertibilidade, não são impostas restrições a v_k nos modelos FEXP.

A vantagem desta classe de modelos é que se pode estimar os parâmetros pelos bem conhecidos modelos lineares generalizados.

Os modelos FEXP são usados para modelar a função de densidade espectral $f(\omega)$ de um processo estacionário quando $f(\omega)$ pode ser decomposta da forma $f(\omega)=\omega^{-2d}f^*(\omega)$ onde f é uma função limitada em todo o seu domínio, excepto, na frequência zero, e contínua. Portanto, estes modelos podem-se aproximar da densidade espectral de um processo qualquer de memória longa com componentes de memória curta se h for suficientemente grande.

2.2.4 Outros processos de memória longa

A característica de memória longa tem sido observada numa larga variedade de séries temporais. Os modelos cuja função de densidade espectral exibe um polo na origem têm sido designados por modelos de memória longa. Uma generalização de tais modelos permite que a função de densidade espectral apresente um polo em qualquer frequência produzindo um decaimento oscilatório lento da função de auto-covariância. Este tipo de modelos são

conhecidos por modelos sazonais de memória longa e têm vindo a serem usados para modelarem séries temporais nas áreas da economia e das ciências. Em especial, tem-se vindo a aplicar este tipo de processos a séries temporais relacionadas com medições atmosféricas (por exemplo, medições de dióxido de carbono) e com medições climatéricas (por exemplo, medição de cheias e de precipitação).

Como exemplo de processos sazonais de memória longa existem o processo persistente sazonal (SPP), os processo ARFISMA ou processo ARFIMA sazonal, o processo SCALM (processo de memória longa sazonal/ cíclico assimétrico), o processo ARUMA e o processo factor-k GARMA (factor-k autorregressivo e de médias móveis Gegenbauer). Destes processos vai-se definir os dois últimos com o objectivo de exemplificar este tipo de processos.

Definição 2.2.5: *$\{U_t\}$ é um processo ARUMA com média zero quando pode ser representado na forma*

$$\Phi(B)\lambda(B)U_t = \Theta(B)\varepsilon_t$$

onde todas as raízes de $\Phi(B)$ estão fora do círculo unitário, $\lambda(B) = 1 - \lambda_1 B - \ldots - \lambda_s B^s$, e todas as raízes de $\lambda(z) = 0$ estão no círculo unitário.

O processo ARUMA não estacionário é um processo de memória longa.

Definição 2.2.6: *$\{V_t\}$ é um processo factor-k GARMA quando pode ser representado na forma*

$$\Phi(B)\prod_{i=1}^{k}\left(1 - 2u_i B + B^2\right)^{\lambda_i} V_t = \Theta(B)\varepsilon_t$$

onde u_i especifica a frequência na qual o comportamento de memória longa ocorre e λ_i indica a velocidade de decaimento das auto-correlações.

$\{V_t\}$ é um processo estacionário de memória longa se todas as raízes da equação $\Phi(z) = 0$ estão fora do círculo unitário e se para $i = 1, \ldots, k$, os u_i são distintos e $0 < \lambda_i < ½$ quando $|u_i| \neq 1$ e $0 < \lambda_i < ¼$ quando $|u_i| = 1$.

Para uma descrição detalhada dos cinco processos sazonais referidos anteriormente ver Anděl (1986), Hassler (1994), Giraitis & Leipus (1995), Woodward et al. (1998) e Arteche & Robinson (2000).

2.3. Definição de memória longa em processos com variância infinita

Os processos de variância infinita permitem modelar séries temporais "impulsivas", ou seja, séries que apresentam ocasionalmente picos muito afiados. Este tipo de processos são muito úteis em áreas como a economia e as telecomunicações.

A definição usual de auto-covariância não é válida para os processos com variância infinita. Consequentemente, a definição usual de memória longa não se aplica a este tipo de processos.

A dependência entre duas variáveis aleatórias X_1 e X_2 com variância infinita pode ser medida através da codiferença, a qual é definida através da expressão

$$\tau(X_1, X_2) = \ln E\left(e^{i(X_1 - X_2)}\right) - \ln E\left(e^{iX_1}\right) - \ln E\left(e^{-iX_2}\right).$$

As medidas I e U também permitem medir a estrutura de dependência entre variáveis com variância infinita.

$$I(X_1, X_2; \theta_1, \theta_2) = \ln\left(E\left(e^{i\theta_1 X_1}\right)\right) + \ln\left(E\left(e^{i\theta_2 X_2}\right)\right) - \ln\left(E\left(e^{i(\theta_1 X_1 + \theta_2 X_2)}\right)\right), \quad (\theta_1, \theta_2) \in IR^2$$

Notar que

$$\tau(X_1, X_2) = -I(X_1, X_2; 1, -1)$$

e quando (X_1, X_2) for um vector Gausseano,

$$I(X_1, X_2; \theta_1, \theta_2) = \theta_1 \theta_2 \operatorname{cov}(X_1, X_2)$$

A medida de dependência U é definida pela igualdade

$$U(X_1, X_2; \theta_1, \theta_2) = e^{\ln\left(E\left(e^{i\theta_1 X_1}\right)\right) + \ln\left(E\left(e^{i\theta_2 X_2}\right)\right)} \cdot \left(e^{-I(X_1, X_2; \theta_1, \theta_2)} - 1\right), \quad (\theta_1, \theta_2) \in IR^2.$$

2.4. Processos ARFIMA com inovações com variância infinita (inovações não-Gausseanas)

Diz-se que uma variável aleatória Z tem distribuição com caudas pesadas se Z é de variação regular[3] com $0 < \alpha < 2$. Nestes casos a variância de Z é infinita.

Por exemplo, a distribuição Pareto, com função de distribuição complementar

---

[3] A variável aleatória Z é de variação regular de índice α se $P(|Z| \geq z) \sim \frac{L(z)}{z^{\alpha}}$, quando $z \to \infty$ em que $L(z)$ é de variação lenta.

$$P(Z \geq z) = \begin{cases} \left(\frac{z_0}{z}\right)^{\alpha} & , z \geq z_0 \\ 1 & , z < z_0 \end{cases}$$

onde z_0 é uma constante positiva e $0 < \alpha < 2$, é uma distribuição com caudas pesadas.

Uma classe particular de distribuições com caudas pesadas são as distribuições α-estáveis simétricas ($S\alpha S$), as quais são melhor definidas pela sua função característica

$$\phi(u) = e^{i\beta u - b|u|^{\alpha}} \text{ com } \beta = 0$$

onde α é o expoente característico, β é o parâmetro de localização, b é o parâmetro de dispersão da distribuição e $b^{\frac{1}{\alpha}}$ é o parâmetro de escala. O parâmetro de dispersão, b , determina o comportamento da distribuição em torno de β enquanto que o expoente característico α determina a forma da distribuição, como se pode ver na figura 2.4.

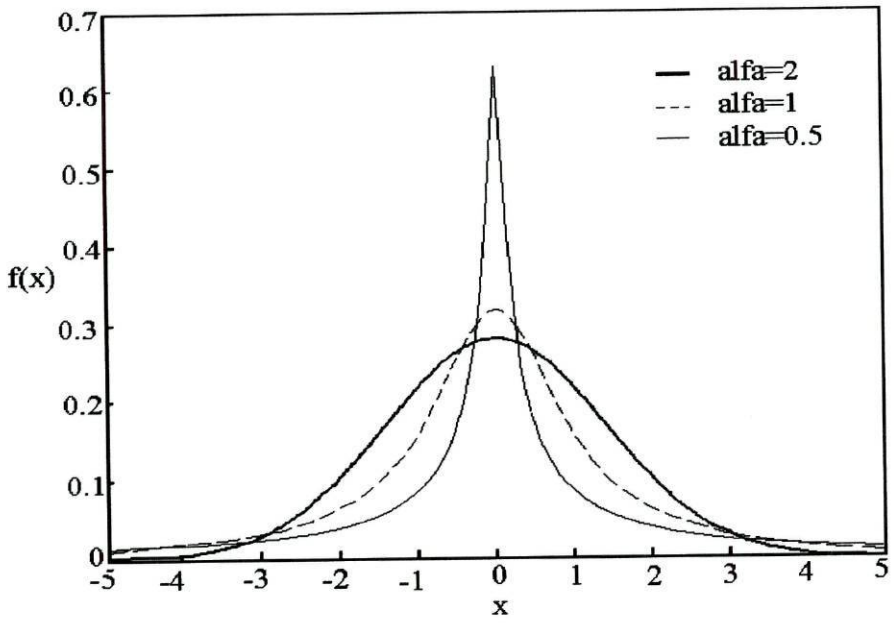


Figura 2.4: Diferentes funções de densidade de probabilidade $S\alpha S$ com $\beta = 0$ e $b = 1$.

É importante notar que existem apenas momentos de ordem inferior a α para os membros da família $S\alpha S$, $\alpha \in]0, 2[$.

Os momentos de ordem fraccionária inferior a α de uma variável aleatória Z que segue uma distribuição $S\alpha S$, com parâmetros de localização $\beta = 0$ e de dispersão b são dados por

$$E\left(|X|^{\delta}\right) = \frac{2^{\delta+1}\Gamma\left(\frac{\delta+1}{2}\right)\Gamma\left(-\frac{\delta}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(-\frac{\delta}{2}\right)} b^{\frac{\delta}{\alpha}}$$

para $0 < \delta < \alpha$.

Os modelos α-estáveis constituem uma subclasse, especialmente interessante, de modelos de variância infinita.

Definição 2.4.1: *Uma sequência de variáveis aleatórias $\{Z_t, t \in T\}$ onde T é um conjunto arbitrário de índices, é um processo estocástico $S\alpha S$ se para todas as combinações de índices distintos $t_1, \ldots, t_n$, as variáveis aleatórias $Z_{t_1}, \ldots, Z_{t_n}$ têm distribuição conjunta $S\alpha S$ com o mesmo expoente característico.*

A variância não está definida para este tipo de processos e consequentemente, neste caso, as definições 2.1.1. e 2.1.2. não são válidas. Portanto, a dependência vai ter que ser estabelecida de outra forma.

Duas medidas alternativas de dependência são a covariação e a codiferença, as quais são definidas para duas variáveis aleatórias que seguem a distribuição $S\alpha S$.

Definição 2.4.2: *A covariação de Z_{t_1} em Z_{t_2}, $[Z_{t_1}, Z_{t_2}]_\alpha$ de duas variáveis aleatórias que têm distribuição conjunta $S\alpha S$ com medida espectral Λ[4] é definida para $\alpha > 1$ pela igualdade* $[Z_{t_1}, Z_{t_2}]_\alpha = \int_{S_2} s_{t_1} (s_{t_2})^{<\alpha-1>} \Lambda(ds)$[5].

Se Z_{t_1} e Z_{t_2} são variáveis aleatórias independentes, então $[Z_{t_1}, Z_{t_2}]_\alpha = 0$, mas o inverso não é verdadeiro.

Se (Z_{t_1}, Z_{t_2}) for um vector Gausseano, então $[Z_{t_1}, Z_{t_2}]_\alpha = 1/2\text{cov}(Z_{t_1}, Z_{t_2})$.

No entanto, esta medida não está definida para $0 < \alpha \leq 1$. Além disso, em geral, $[Z_{t_1}, Z_{t_2}]_\alpha \neq [Z_{t_2}, Z_{t_1}]_\alpha$ e $[Z_t, Z_{t_1}^* + Z_{t_2}^*]_\alpha \neq [Z_t, Z_{t_1}^*]_\alpha + [Z_t, Z_{t_2}^*]_\alpha$.

Como a covariação não possui algumas das "boas" propriedades da covariância não é tão popular como a codiferença, a qual também é uma medida que permite medir a dependência entre duas variáveis aleatórias.

---

[4] A medida espectral Λ de um vector aleatório $S\alpha S$, (Z_{t_1}, Z_{t_2}), é a única medida de Borel finita, positiva e simétrica no círculo unitário $S_2 = \{(s_{t_1}, s_{t_2}) \in IR^2 : s_{t_1}^2 + s_{t_2}^2 = 1\}$ tal que

$$E\left(e^{i(a_1 Z_{t_1} + a_2 Z_{t_2})}\right) = \exp\left\{-\int_{S_2} |a_1 s_{t_1} + a_2 s_{t_2}|^\alpha \Lambda(ds)\right\}.$$

[5] $a^{<q>} = |a|^q .\text{sign}(a)$ onde $sign(a) = \frac{a}{|a|}$.

Definição 2.4.3: *Se* Z_{t_1} *e* Z_{t_2} *têm distribuição conjunta* $S\alpha S$ *com medida espectral* Λ*, então a codiferença,* τ *é definida pela igualdade*

$$\tau(Z_{t_1}, Z_{t_2}) = \ln E\left(e^{i(Z_{t_1}-Z_{t_2})}\right) - \ln E\left(e^{iZ_{t_1}}\right) - \ln E\left(e^{-iZ_{t_2}}\right) = \int_{S_2}\left[|s_{t_1}|^{\alpha} + |s_{t_2}|^{\alpha} - |s_{t_1} - s_{t_2}|^{\alpha}\right]\Lambda(ds).$$

Neste caso,

$$\tau(Z_{t_1}, Z_{t_2}) = \tau(Z_{t_2}, Z_{t_1}).$$

Se (Z_{t_1}, Z_{t_2}) for um vector Gausseano, então

$$\tau(Z_{t_1}, Z_{t_2}) = \mathrm{cov}(Z_{t_1}, Z_{t_2}).$$

Se $(Z^*_{t_1}, Z^*_{t_2})$ é outro vector aleatório $S\alpha S$ tal que $\tau(Z_{t_1}, Z_{t_2}) \leq \tau(Z^*_{t_1}, Z^*_{t_2})$, então

$$\forall c > 0,\ P(|Z_{t_1} - Z_{t_2}| > c) \geq P(|Z^*_{t_1} - Z^*_{t_2}| > c).$$

Isto significa que existe menor probabilidade de $Z^*_{t_1}$ *e* $Z^*_{t_2}$ diferirem do que Z_{t_1} e Z_{t_2} diferirem, sendo portanto "mais dependentes".

Os processos $S\alpha S$ permitem modelar sequências que seguem distribuições com caudas pesadas.

Os modelos ARFIMA com inovações simétricas α-estáveis são uma classe dos processos $S\alpha S$. Aqueles modelos são muito flexíveis porque capturam, não só as características de memória longa e curta, mas também os afastamentos da Gausseanidade.

A dependência de longo alcance surge quando $d > 0$, o que é possível apenas se $\alpha > 1$.

Seja $\{a_n\}$ uma sequência de variáveis aleatórias i. i. d. que seguem a distribuição $S\alpha S$. Sem perda de generalidade, assuma-se que o parâmetro escala dos a_n é 1 ($E\left(e^{i\theta a_n}\right) = e^{-|\theta|^{\alpha}}$). Sejam $\Phi_p(z)$ e $\Theta_q(z)$ os polinómios com coeficientes reais definidos por:

$$\Phi_p(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$$

$$\Theta_q(z) = 1 - \theta_1 z - \ldots - \theta_q z^q .$$

Vai-se assumir que $\Phi_p(z)$ e $\Theta_q(z)$ não têm raízes comuns e que $\Phi_p(z)$ não tem raízes no disco unitário $\{z: |z| \leq 1\}$. Este último pressuposto assegura que os coeficientes na expansão em série de $\Theta_q(z)/\Phi_p(z)$ tendem exponencialmente para zero e que as séries temporais ARFIMA consideradas são médias móveis causais.

Definição 2.4.4: *Seja $0 < \alpha \leq 2$ e $\alpha(d-1) < -1$. O processo ARFIMA SαS é a única solução das equações*

$$\Phi_p(B)Z_n = \Theta_q(B)(1-B)^{-d}a_n, \quad n \in \mathbb{Z} \tag{2. 3}$$

onde $(1 - B)^{-d}$ é um filtro linear definido pela igualdade

$$(1-B)^{-d}a_n = \sum_{j=0}^{\infty} b_j a_{n-j}$$

onde os b_j's são os coeficientes da expansão em série de $(1 - z)^{-d}$, $|z| < 1$, isto é

$$b_0 = 1, \qquad b_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}, \quad j \geq 1.$$

Kokoszka e Taqqu (1995) estabeleceram a existência e a invertibilidade dos processos ARFIMA com inovações α-estáveis de variância infinita.

Teorema 2.4.1: (Kokoszka & Taqqu (1995)) *Se $\alpha(d - 1) < -1$, então a sequência $\{Z_n, -\infty < n < \infty\}$ definida por $Z_n = \sum_{j=0}^{\infty} c_j a_{n-j}$, sendo*

$$c_d(z) = \frac{\Theta_q(z)}{\Phi_p(z)}(1-z)^{-d} = \sum_{j=0}^{\infty} c_j z^j, \tag{2. 4}$$

é o único processo de médias móveis causal satisfazendo (2. 3).

Teorema 2.4.2: (Kokoszka & Taqqu (1995)) *Suponhamos que*

$$Z_n = \sum_{j=0}^{\infty} c_j a_{n-j}$$

é solução de (2. 3) com coeficientes c_j definidos por (2. 4). Suponhamos que $\Theta_q(z)$ não tem raízes no círculo unitário $\{z : |z| \leq 1\}$ e que $1 < \alpha \leq 2$. Se $|d| < 1 - \frac{1}{\alpha}$, então

$$\lim_{m\to\infty} E\left|\sum_{k=0}^{m} \tilde{c}_k Z_{n-k} - a_n\right|^{\delta} = 0$$

para qualquer $0 < \delta < \alpha$. Os coeficientes $\tilde{c}_j$ são definidos por

$$\sum_{j=0}^{\infty} \tilde{c}_j z^j = (1-z)^d \frac{\Phi_p(z)}{\Theta_q(z)}, \quad |z| < 1.$$

Além disso, para $0 < d < 1 - \frac{1}{\alpha}$ as somas parciais

$$\sum_{k=0}^{m} \widetilde{c}_k Z_{n-k}$$

convergem absolutamente, quase certamente para a_n.

Este teorema diz que as sequências ARFIMA que são soluções únicas de médias móveis causais de (2. 3), também são soluções das equações AR(∞) $= \sum_{j=0}^{\infty} \widetilde{c}_j Z_{n-j}$ quase certamente, desde que $\alpha > 1$ e $|d| < 1 - \frac{1}{\alpha}$.

Se Z_n tem uma representação em médias móveis causal

$$Z_n = \sum_{j=0}^{\infty} c_j a_{n-j}\,,\ n \in \mathbb{Z}$$

com inovações a_n $S\alpha S$ com parâmetro escala 1, a codiferença é da forma

$$\tau_n = \tau(Z_n, Z_0) = \sum_{j=0}^{\infty} \left[|c_j|^{\alpha} + |c_{j+n}|^{\alpha} - |c_j - c_{j+n}|^{\alpha} \right]$$

enquanto que para $\alpha > 1$, a covariação $[Z_n, Z_0]_\alpha$ é dada por

$$[Z_n, Z_0]_\alpha = \sum_{j=0}^{\infty} c_{j+n} c_j^{<\alpha-1>}\,.$$

Teorema 2.4.3: (Kokoszka & Taqqu (1995)) *suponha-se que*

$$Z_n = \sum_{j=0}^{\infty} c_j a_{n-j}$$

é um processo ARFIMA(p, d, q) com inovações a_n $S\alpha S$. *Suponha-se também que* $0 < \alpha \leq 2$ *e que* $d \in \mathbb{R} \mid \mathbb{Z}$.

a) *se (i)* $\alpha \leq 1$ *ou (ii)* $\alpha > 1$ *e* $(\alpha - 1)(d - 1) > -1$, *então*

$$\lim_{n\to\infty} \frac{\tau_n}{n^{\alpha(d-1)+1}} = \left| \frac{\Theta_q(1)}{\Gamma(d)\Phi_p(1)} \right|^{\alpha} \int_0^{\infty} h(z)dz$$

onde

$$h(z) = z^{(d-1)\alpha} + (1+z)^{(d-1)\alpha} - \left(z^{d-1} - (1+z)^{d-1} \right)^{\alpha}.$$

b) *Se* $\alpha > 1$ *e* $(\alpha - 1)(d - 1) < -1$, *então*

$$\lim_{n\to\infty} \frac{\tau_n}{n^{d-1}} = \frac{\alpha\, \Theta_q(1)}{\Gamma(d)\Phi_q(1)} \sum_{j=0}^{\infty} c_j^{<\alpha-1>}\,.$$

(Se d é inteiro, então $\{Z_n\}$ pode ser representado na forma de médias móveis finita e consequentemente $\tau_n = 0$ para n grande.)

O teorema seguinte demonstra que para $\alpha > 1$ a covariação $[Z_n, Z_0]_\alpha$ é assimptóticamente proporcional a τ_n .

Teorema 2.4.4: (Kokoszka & Taqqu (1995)) *Suponha-se que $\{Z_n\}$ é um processo ARFIMA(p, d, q) SαS, que $1 < \alpha \leq 2$ e que $d \in \mathbb{R} \setminus \mathbb{Z}$.*

a) *Se $(\alpha - 1)(d - 1) > -1$, então*

$$\lim_{n\to\infty} \frac{[Z_n, Z_0]_\alpha}{n^{\alpha(d-1)+1}} = \left| \frac{\Theta_q(1)}{\Gamma(d)\Phi_p(1)} \right|^\alpha \int_0^\infty (y+1)^{d-1} y^{(\alpha-1)(d-1)} dy .$$

b) *Se $(\alpha - 1)(d - 1) < -1$, então*

$$\lim_{n\to\infty} \frac{[Z_n, Z_0]_\alpha}{n^{d-1}} = \frac{\Theta_q(1)}{\Gamma(d)\Phi_p(1)} \sum_{j=0}^{\infty} c_j^{<\alpha-1>} .$$

2.5. Caracterização da memória longa

De uma forma intuitiva uma série é classificada como sendo uma série de memória longa quando o passado da série influencia permanentemente o seu comportamento futuro. No entanto, não existe uma definição geral que se aplique a todos os tipos de séries.

Uma definição amplamente divulgada na literatura baseia-se na TSS ("typical spectral shape"). Esta propriedade significa que numa vizinhança de zero o espectro dá a sua maior contribuição para a variância do espectro.

No entanto, a caracterização da memória longa através da TSS só é adequada para séries lineares estacionárias. O espectro não "explica" totalmente o comportamento das séries não lineares e algumas destas séries temporais podem exibir formas de memória longa profundamente relacionadas com a distribuição de probabilidade dos seus processos subjacentes, não apresentando características estatísticas de segunda ordem. Logo a característica de memória longa não implica a TSS. Mas a TSS em séries temporais estacionárias implica a memória longa. Portanto a definição 2.1.2. é válida, embora a

implicação contrária não o seja. Esta implicação contrária só seria válida se o processo fosse linear.

Para as séries não Gausseanas e não estacionárias as medidas de memória longa usuais (função de auto-correlação e função de densidade espectral) podem não ser apropriadas ou podem não estar definidas.

A não estacionaridade pode fazer com que uma série temporal, por vezes, exiba uma tendência "lenta" e fazer com que esta apresente a TSS, apesar do processo não ter sequer memória. Portanto, para as séries temporais não estacionárias, a definição do espectro não é adequada.

No caso dos processos lineares com variância infinita a definição de auto-covariância não é válida e consequentemente a função de densidade espectral usual também não está definida. Logo, neste caso, a caracterização da memória longa através da TSS não faz sentido. Nestes casos é necessário estabelecer novas medidas de memória longa. Por exemplo, nos processos $S\alpha S$ existem, entre outras medidas de memória longa a covariação e a codiferença, pois nestes processos, para $\alpha < 2$, não existem momentos de segunda ordem.

No entanto, existem processos não Gausseanos onde a TSS fornece uma caracterização adequada da memória longa. Por exemplo, Velasco (2000) trabalhou com os processos ARFIMA não Gausseanos (ou seja, processos ARFIMA com inovações não Gausseanas, nomeadamente, inovações Exponenciais, Uniformes e t_5) e usou a TSS para caracterizar a memória longa.

Assim, uma caracterização geral adequada da dependência de memória longa é difícil e deve-se pensar nesta característica como uma "não-propriedade" em vez de uma propriedade de algumas séries.

2.6. Métodos de simulação de processos de memória longa

Nesta secção vão-se descrever diversos métodos de geração de processos de memória longa Gausseanos e não Gausseanos.

Os métodos 1 a 6 permitem simular processos Gaussenos. Só é possível implementar os métodos 1, 2, 5 e 6 se se conhecer a função de auto-correlação do processo que se pretende gerar.

Nos métodos 3 a 6 particulariza-se o método geral para o caso dos ARFIMA(p, d, q) apresentando-se os respectivos algoritmos.

O método 7 permite gerar uma aproximação do processo FGN, enquanto que os métodos 8 e 9 geram processos ARFIMA(p, d, q) Gausseanos.

O método 9 é uma extensão do algoritmo de Durbin-Levinson e permite simular sequências ARFIMA com inovações com variância infinita.

Além da descrição dos métodos referidos vão ser mencionadas algumas funções que permitem gerar amostras de processos ARFIMA Gausseanos e não Gausseanos, as quais têm por base na sua construção alguns dos métodos inicialmente descritos.

Método 1: Método baseado no algoritmo de Durbin-Levinson

Seja $\{X_t\}$ um processo Gausseano de média zero e função de auto-covariância γ. Pretende-se gerar uma realização de n observações $\{X_t\}_{t=1, \ldots, n}$.

a) Gerar uma sequência de variáveis aleatórias independentes e identicamente distribuídas N(0, 1), $\{\varepsilon_t\}_{t=1, \ldots, n}$

b) $X_1 = \sqrt{\gamma(0)}\,\varepsilon_1$ e $v_0 = \gamma(0)$

c) Para $t = 1, \ldots, n$

c1) Calcular recursivamente $\phi_{tj}, j = 1, \ldots, t$ através das equações

$$\phi_{t,t} = \left[\gamma(t) - \sum_{j=1}^{t-1} \phi_{t-1,j}\gamma(t-1)\right] v_{t-1}^{-1}$$

$$\phi_{t,j} = \phi_{t-1,j} - \phi_{t,t}\phi_{t-1,t-j} \quad , j = 1, \ldots, t\text{-}1$$

e a variância através da equação

$$v_t = v_{t-1}\left(1 - \phi_{t,t}\right)^2$$

c2) $X_{t+1} = \sum_{j=0}^{t-1} \phi_{t,j+1} X_{t-j} + \sqrt{v_t}\,\varepsilon_{t+1}$

Método 2: Este método baseia-se na decomposição de Cholesky (Beran (1994))

Este método permite gerar qualquer série Gausseana se a função de auto-covariância for conhecida.

1. Escrever a matriz das covariâncias como um produto da matriz real triangular inferior pela sua transposta: $\Sigma = LL^t$. Esta decomposição é designada decomposição de Cholesky.

2. Simular um vector de variáveis aleatórias independentes que seguem a distribuição N(0, 1), $Z = (Z_1, \ldots, Z_n)^t$.

3. Defina-se $X = (X_1, \ldots, X_n)^t = LZ$. A sequência $X_1, \ldots, X_n$ tem as propriedades desejadas.

Método 3: Método baseado numa aproximação que consiste numa soma de processos independentes AR(1)

Granger (1980) provou que a soma de processos de memória curta pode levar a um processo de memória longa.

Considere-se o ruído branco $\{\varepsilon_i\}_{i\in\mathbb{N}}$. Sejam $\{X_j\}_{j\in\mathbb{N}}$ processos auto-regressivos AR(1) com parâmetros $\{\alpha_j\}_{j\in\mathbb{N}}\in]-1, 1[^{\mathbb{N}}$ tal que $X_j = \{X_j(i), i\in\mathbb{N}\}$ e para

$$X_j(i+1) = \alpha_j X_j(i) + \varepsilon_{i+1}$$

e defina-se a soma

$$S_M(i) = \sum_{j=1}^{M} X_j(i).$$

Assuma-se que $\{\alpha_j\}_{j\in\mathbb{N}}$ são variáveis aleatórias i. i. d. com função de densidade f e independentes de $\{\varepsilon_i\}_{i\in\mathbb{N}}$.

Granger (1980) mostrou que se f é uma função de densidade em [0, 1] da forma

$$f(u) = \phi(u)(1-u)^{q-1},\ q > 0,$$

onde ϕ é contínua em $u = 1$ e $\phi(1) \neq 0$, então à medida que $M \to \infty$, S_M/M aproxima-se de um processo de memória longa cuja função de auto-covariância é da forma

$$\rho(k) \sim k^{1-2\beta}\quad, k \to \infty.$$

Por exemplo, se a função de densidade é a distribuição Beta de parâmetros a e o, obtém-se um processo ARFIMA(0, d, 0) para $0 < d < ½$, tomando $o = 1\text{-}d$ e $a = d$.

O algoritmo seguinte permite gerar uma amostra de um processo ARFIMA(p, d, q).

1. Gerar uma amostra de dimensão n de um processo ARFIMA(0, d, 0) (aproximado).

a) Gerar uma sequência i. i. d. $(\varepsilon_i)_{i=1,\ldots,n}$ com distribuição N(0, 1).

b) Gerar uma sequência i. i. d. $(\alpha_j)_{j=1,\ldots,M}$ com distribuição Beta, B(d, 1-d), $0 < d < ½$.

c) Para cada $j \in\{1, \ldots, M\}$, gerar processos AR(1), X_j, com coeficientes α_j e com ruído branco (comum) ε:

$$X_j(i+1) = \alpha_j X_j(i) + \varepsilon_{i+1}\ ,\quad i = 0, \ldots, n\text{-}1.$$

d) O processo aproximado ARFIMA(0, d, 0) é obtido da seguinte forma:

$$\frac{1}{M}\sum_{j=1}^{M} X_j(i)\ ,\quad i = 1, \ldots, \text{n}.$$

2. Gerar uma amostra de dimensão n do processo ARFIMA(p, d, q) da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + X_t - \theta_1 X_{t-1} - \ldots - \theta_q X_{t-q}.$$

Podem-se obter outros processos de memória longa usando funções de densidade diferentes da distribuição Beta.

Método 4: Método baseado no truncamento da expansão em médias móveis

Este método foi proposto por Mandelbrot & Wallis (1969) e consiste numa aproximação baseada num processo de médias móveis de ordem elevada, podendo ser usado para simular qualquer amostra de um processo linear de memória longa, $\{X_t\}$, tal que para todo o $t \in \mathbb{N}$,

$$X_t = \sum_{k=-\infty}^{\infty} a_k \varepsilon_{t-k}$$

com $\sum_{k=-\infty}^{\infty} a_k^2 < \infty$ sendo $\varepsilon = \varepsilon_k$, $k \in \mathbb{Z}$, uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância um.

O algoritmo seguinte permite gerar uma amostra de um processo ARFIMA(p, d, q).

1. Gerar uma amostra de dimensão n de um processo ARFIMA(0, d, 0) (aproximado).

a) Gerar uma sequência de variáveis aleatórias i. i. d. com distribuição N(0, 1):

$\{\varepsilon_k\}_{-M \le k \le N+M}$.

b) Para $t = 1, \ldots, n$,

$$X_t^M = \sum_{k=0}^{M} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \varepsilon_{t-k}$$

2. Gerar uma amostra de dimensão n do processo ARFIMA(p, d, q) da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + X_t - \theta_1 X_{t-1} - \ldots - \theta_q X_{t-q}.$$

Para cada t, à medida que $M \to \infty$, X_t^M converge, em L^2, quase certamente para X_t.

Método 5: Este método baseia-se na resolução das equações de Yule-Walker

Granger & Joyeux (1980) propuseram um método aproximado de gerar séries temporais fraccionariamente integradas. Aqui, vai-se aplicar aquele método aos processos ARFIMA(p, d, q).

1. Gerar uma amostra de dimensão n de um processo ARFIMA(0, d, 0) (aproximado).

a) Gerar uma sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição N(0, 1): $\{\varepsilon_t\}_{t=1,...m}$.

b) Seja

$$C_m(d) = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{m-1} \\ \rho_1 & 1 & \cdots & \rho_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m-1} & \rho_{m-2} & \cdots & 1 \end{bmatrix}$$

a matriz das auto-correlações.

Para $k = 1, ..., m$ calcular

$$\rho_k = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!}$$

c) Calcular $a_j, j = 1, ..., m$, resolvendo as equações de Yule-Walker

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = C_m^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{bmatrix}$$

d) O processo ARFIMA(0, d, 0) é obtido resolvendo a equação

$$X_t + \sum_{j=1}^{m} a_j X_{t-j} = \varepsilon_t, \quad t = 1, ..., n.$$

2. Gerar uma amostra de dimensão n do processo ARFIMA(p, d, q) da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + X_t - \theta_1 X_{t-1} - ... - \theta_q X_{t-q}.$$

Método 6: Este método baseia-se nas transformadas de Fourier rápidas

Este método foi proposto por Davies & Harte (1987) para simular séries temporais Gausseanas estacionárias de dimensão n com função de auto-covariância $\gamma(0), \gamma(1), ..., \gamma(n\text{-}1)$.

1. Gerar uma amostra de dimensão n de um processo ARFIMA(0, d, 0).

a) Defina-se

$$\lambda_k = \frac{2\pi(k-1)}{2n-2}, \quad \text{k} = 1, ..., \text{2n-2}.$$

A transformada de Fourier finita g_k da sequência $\gamma(0), \gamma(1), ..., \gamma(n\text{-}2), \gamma(n\text{-}1), \gamma(n\text{-}2), ..., \gamma(1)$ é dada por

$$g_k = \sum_{j=1}^{n-1} \gamma(j-1) e^{i(j-k)\lambda_k} + \sum_{j=n}^{2n-2} \gamma(2n-j-1) e^{i(j-1)\lambda_k}, \; k = 1, ..., 2n\text{-}2. \qquad \textbf{(2. 5)}$$

b) Verificar se $g_k > 0$ para todo o $k = 1, ..., 2n\text{-}2$.

c) Simular duas séries independentes de variáveis aleatórias normais de média zero: $U_1, U_2, ..., U_n$ e $V_2, ..., V_{n-1}$ tais que

$$\text{var}(U_1) = \text{var}(U_n) = 2$$

$$\text{var}(U_k) = \text{var}(V_k) = 1 \quad \forall\, k \neq 1, n.$$

Defina-se $V_1 = V_n = 0$ e as variáveis complexas

$$Z_k = U_k + iV_k, \qquad k = 1, ..., n$$

$$Z_k = U_{2n-k} - iV_{2n-k}, \qquad k = n+1, ..., 2n\text{-}2.$$

d) Para $t = 1, ..., n$ defina-se

$$X_t = \frac{1}{2\sqrt{n-1}} \sum_{k=1}^{2n-2} \sqrt{g_k}\, e^{i(t-1)\lambda_k} Z_k . \qquad \textbf{(2. 6)}$$

A sequência $\{X_t\}$ tem as propriedades pretendidas.

2. Gerar uma amostra de dimensão n do processo ARFIMA(p, d, q) da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + X_t - \theta_1 X_{t-1} - ... - \theta_q X_{t-q} .$$

Se se usar a função de auto-covariância

$$\gamma(k) = \frac{1}{2} \frac{\sigma^2}{n^{2d+1}} \left[(k+1)^{2d+1} - 2k^{2d+1} + (k-1)^{2d-1} \right]$$

obtém-se uma amostra de ruído Gausseano fraccionário (FGN).

Método 7: Método espectral directo

Voss (1988), Feder (1988), Peitgen & Saupe (1988) e Paxon (1997) geraram uma aproximação do FGN usando a transformada de Fourier discreta.

A estratégia que está por detrás deste método consiste em começar com a densidade espectral do processo nas frequências de Fourier para simular uma sequência aleatória complexa com estas densidades espectrais e depois usar a transformada de Fourier inversa rápida para obter uma aproximação do processo nos instantes $t = \frac{k}{n}$, $k = 0, 1, ..., n\text{-}1$, sendo n a dimensão da amostra que se pretende.

1. Escolher n de forma que seja inteiro e par. Defina-se o vector $(\omega_1, ..., \omega_{n/2})$ de frequências de Fourier onde $\omega_k = \frac{2\pi k}{n}$ e calcule-se o vector

$$F = \left(F_1(\omega_1), ..., F_{\frac{n}{2}}\left(\omega_{\frac{n}{2}}\right) \right) = \left(f_H(\omega_1), ..., f_H\left(\omega_{\frac{n}{2}}\right) \right)$$

onde

$$f_H(\omega) = (C_2(H))^{-2} 2(1-\cos\omega) \sum_{k \in Z} |2\pi k + \omega|^{-2H-1} \qquad \textbf{(2. 7)}$$

e

$$(C_2(H))^{-2} = \frac{1}{2\pi} \sin(\pi H)\Gamma(2H+1).$$

$f_H(\omega)$ é a densidade espectral do FGN.

2. Gerar n/2 variáveis aleatórias exponenciais, $\varepsilon(1)$, i. i. d., E_1, ...$E_{n/2}$ e $n/2$ variáveis aleatórias uniformes U[0, 1] i. i. d., U_1, ..., $U_{n/2}$.

3. Calcular

$$Z_k^* = \sqrt{F_k E_k}\, e^{2i\pi U_k} \text{ para } k = 1, ..., n/2.$$

4. Construir o vector de dimensão n,

$$\left(0, Z_1^*, ..., Z_{(n/2)-1}^*, Z_{n/2}^*, \overline{Z_{(n/2)-1}^*}, ..., \overline{Z_1^*}\right).$$

5. Calcular a transformada de Fourier inversa rápida para se obter a amostra pretendida.

Método 8: Método proposto por Hosking (1982)

Este método permite simular processos ARFIMA(p, d, q) com inovações Gausseanas e faz uso das correlações parciais do processo.

1. Gerar uma amostra de dimensão n de um processo ARFIMA(0, d, 0)

a) Gerar um valor inicial X_0, a partir da distribuição estacionária do processo, nomeadamente N(0, v_0) onde

$$v_0 = \frac{(-2d)!}{[(-d)!]^2}.$$

b) Para $t = 1, ..., n$

b1) Calcular $\phi_{tj}, j = 1, ..., t$ recursivamente através das equações

$$\phi_{t,t} = \frac{d}{t-d}$$

e

$$\phi_{t,j} = \phi_{t-1,j} - \phi_{t,t}\phi_{t-1,t-j}, \qquad j = 1, ..., t-1.$$

b2) Calcular

$$m_t = \sum_{j=1}^{t} \phi_{t,j} X_{t-j}$$

e

$$v_t = \left(1 - \phi_{t,t}^2\right) v_{t-1}.$$

b3) Gerar X_t a partir da distribuição $N(m_t, v_t)$.

2. Gerar uma amostra de dimensão n do processo ARFIMA(p, d, q) da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + X_t - \theta_1 X_{t-1} - \ldots - \theta_q X_{t-q}.$$

Método 9: Extensão do algoritmo de Durbin-Levinson

Kokoszka & Taqqu (2001) discutiram em que condições o algoritmo de Durbin-Levinson pode ser usado para gerar sequências ARFIMA com variância infinita. Eles mostraram que, em princípio, se pode usar aquele algoritmo com inovações α-estáveis ou com inovações que estão no domínio de atracção de uma lei α-estável para gerar séries temporais ARFIMA estáveis se se desprezar a parte inicial da série gerada.

Suponha-se que $\{Z_t, -\infty < t < \infty\}$ é uma sequência de variáveis aleatórias $S\alpha S$ i. i. d. com função característica

$$E\left(e^{i\theta Z_t}\right) = e^{-|\theta|^\alpha}.$$

Vai-se assumir que $1 < \alpha < 2$ porque para tais α, as equações

$$(1\text{-}B)^d Y_t = Z_t$$

podem ser representadas na forma

$$Y_t = \sum_{i=0}^{\infty} b_i Z_{t-i} \qquad \textbf{(2. 8)}$$

as quais permitem $d > 0$.

A série (2. 8) é convergente para $d < 1 - \frac{1}{\alpha}$.

Seja $\{W_t, t \geq 0\}$ o processo α-estável gerado pelo algoritmo de Durbin-Levinson

$$W_0 = \sqrt{v_0} Z_0$$

$$W_t = \sum_{j=1}^{t} \phi_{tj} X_{tj} + \sqrt{v_t} Z_t \qquad para \quad t \geq 1 \qquad \textbf{(2. 9)}$$

Kokoszka & Taqqu (2001) provaram que estes dois processos têm distribuições bidimensionais diferentes para qualquer $1 < \alpha < 2$.

Suponha-se que as inovações $\{Z_t, -\infty < t < \infty\}$ são independentes e identicamente distribuídas com média zero e estão no domínio de atracção de uma lei α-estável com $1 < \alpha < 2$, isto é

$$P\left(|Z_t| > x\right) = x^{-\alpha} L(x) \qquad \textbf{(2. 10)}$$

e

$$\frac{P(Z_t > x)}{P(|Z_t| > x)} \xrightarrow[x\to\infty]{} a \qquad \textbf{(2. 11)}$$

para algum $0 \le a \le 1$ e onde L é uma função de variação lenta.

Defina-se $\{K_t\}$ através da equação (2. 8) e $\{W_t,\ t \ge 0\}$ através da equação (2. 9), mas com inovações satisfazendo (2. 10) e (2. 11).

Vai-se assumir que $0 < d < 1 - \frac{1}{\alpha}$ porque as equações ARFIMA(p, d, q) com inovações satisfazendo (2. 10) e (2. 11) têm solução estacionária apenas para $d < 1 - \frac{1}{\alpha}$.

Sejam

$$\Phi(B)\widetilde{K}_t = \Theta(B)K_t$$

e

$$\widetilde{W}_t = W_t = 0 \quad \text{para } t < 0$$
$$\widetilde{W}_t = \phi_1\widetilde{W}_{t-1} + \ldots + \phi_p\widetilde{W}_{t-p} + W_t + \theta_1 W_{t-1} + \ldots + \theta_q W_{t-q} \quad \text{para } t \ge 0 \qquad \textbf{(2. 12)}$$

Teorema 2.6.1: (Kokoszka & Taqqu (2001)) *Suponha-se que as inovações Z_t satisfazem (2. 10) e (2. 11). Seja $\{\widetilde{K}_t\}$ uma sequência ARFIMA(p, d, q) e $\{\widetilde{W}_t, t \ge 0\}$ uma sequência obtida através da aplicação do algoritmo de Durbin-Levinson e das recursões (2. 12). Se $1 < \alpha < 2$ e $0 < d < 1 - \frac{1}{\alpha}$, então para qualquer $1 < \delta < \alpha$*

$$\lim_{t\to\infty} E\left(\left|\widetilde{W}_t - \widetilde{K}_t\right|^{\delta}\right) = 0.$$

Corolário 2.6.1: (Kokoszka & Taqqu (2001)) *Suponha-se que se verificam os pressupostos do teorema anterior e que K_t é definido por (2. 8) e W_t é definido por (2. 9). Então,*

$$\lim_{t\to\infty} E\left(|W_t - K_t|^{\delta}\right) = 0.$$

O método 2 só pode ser usado se se conhecer a expressão da função de auto-covariância e é um método relativamente lento.

No método 3 surge uma dificuldade: a escolha adequada de M.

O método 4 permite gerar processos lineares cuja expressão da expansão em médias móveis, MA(∞), seja conhecida.

A vantagem do método 6 é que as transformadas de Fourier rápidas podem ser usadas para calcular (2. 5) e (2. 6). Isto torna o método computacionalmente rápido.

Embora o método 7 seja rápido, a sequência por ele gerada é periódica, ou seja, a parte final da sequência está fortemente correlacionada com a parte inicial. Para atenuar este problema pode-se gerar uma sequência que é várias vezes maior do que a necessária e extrai-se a sequência de tamanho n. No entanto, na aplicação deste método surge uma dificuldade: a dimensão adequada da série a gerar para se obter a sequência pretendida de dimensão n.

A soma em (2. 7) tem que ser recalculada para cada valor de ω. No entanto, Ledesma & Liu (2000) mostraram que $\sum_{|k|\geq 3}|2\pi k+\omega|^{-2H-1}$ pode ser aproximada por uma função linear $p\lambda+q$.

Este método é mais rápido que o método 1, pois a complexidade deste algoritmo ($O(n\log(n))$) é menor que a do algoritmo do método 1 ($O(n^2)$). No entanto, este método só permite gerar uma amostra do FGN.

O método 9 é uma extensão do método 1. Este método permite gerar processos ARFIMA(p, d, q) cujas inovações são α-estáveis ou estão no domínio de atracção de uma lei α-estável.

Actualmente existem vários programas que permitem gerar amostras do processo ARFIMA(p, d, q) Gausseano como é o caso do Ox e do S-Plus. No S-Plus é a função *arima.fracdiff.sim()* que permite gerar aquelas amostras. O método de simulação baseia-se no algoritmo de Haslett & Raftery (1989).

Velasco (2000) usou uma versão modificada daquela função para gerar séries ARFIMA(0,d,0) com inovações não Gausseanas, nomeadamente, com inovações seguindo as distribuições Uniforme$\left[-\sqrt{3},\sqrt{3}\right]$, Exponencial com parâmetro 1, re-centrado com média zero e t_5. Em vez de usar as inovações originais (Gausseanas), Velasco (2000) substituiu a sub-rotina *rnorm*, a qual gerava inovações Gausseanas, pelas sub-rotinas *rt* , *runif* e *rexp*, as quais geram sequências que seguem respectivamente as distribuições t, uniforme e exponencial.

Como a função *arima.fracdiff.sim()* envolve o cálculo das covariâncias, é inapropriada para gerar séries α-estáveis, as quais têm variância infinita.

No entanto, a "package" S-Plus também tem disponível a função *rstab()* que permite simular séries temporais α-estáveis.

Na internet, no endereço http: // math.bu.edu/ people/ murad/ methods / implementations/ time_series/farima.generate, existe o código de duas funções desenvolvidas por Murad Taqqu: *farima.generate()* e *farima.generate.pareto()*. A primeira permite simular processos ARFIMA(1, d, 1) com inovações α-estáveis, enquanto que a segunda permite simular processos ARFIMA(0, d, 0) com inovações Pareto.

3. Estimação dos parâmetros dos processos de memória longa

Existem duas formas principais de estimação dos parâmetros dos modelos de memória longa no domínio das frequências: estimação semi-paramétrica e paramétrica.

No caso da estimação semi-paramétrica os estimadores podem ser construídos usando-se métodos locais ou métodos globais. No primeiro caso usam-se apenas as frequências que estão numa vizinhança de zero, enquanto que no segundo caso são usadas todas as frequências.

Neste capítulo vão ser apresentados vários estimadores dos processos de memória longa. Na secção 3.1. vão ser apresentados estimadores semi-paramétricos, estando na subsecção 3.1.1. os estimadores obtidos através dos métodos locais (estimadores do tipo GPH e o GSE), enquanto que na subsecção 3.1.2. se encontram os estimadores obtidos através da utilização de métodos globais (estimador FEXP e estimador FEXP generalizado). Na secção 3.2. vão ser apresentados vários estimadores paramétricos: estimador de máxima verosimilhança exacto (subsecção 3.2.1.), estimador de Whittle (subsecção 3.2.2.), o qual é um estimador de máxima verosimilhança aproximado, e o estimador dos parâmetros dos processos ARFIMA com inovações $S\alpha S$ (subsecção 3.2.3.).

3.1. Estimação semi-paramétrica

Seja $f(\omega)$ a função de densidade espectral de um processo de memória longa.

Na estimação semi-paramétrica $f_u(.)$ é livre de qualquer imposição paramétrica e pertence a uma ampla classe de funções, sendo caracterizada pelo seu comportamento na frequência zero ou por uma dada regularidade em todo o domínio das frequências. A primeira caracterização estimula o uso de métodos de estimação locais de d, enquanto que no segundo caso se usam métodos de estimação globais.

3.1.1. Métodos locais

Os métodos locais visam construir estimadores de d consistentes, sem impor qualquer tipo de restrição a $f_u(.)$ fora da frequência zero.

3.1.1.1. Estimadores do tipo GPH

Estimador GPH

Seja $\{X_t\}$ um processo ARFIMA(p, d, q), isto é um processo linear que satisfaz a igualdade

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t$$

para $d \in]-\frac{1}{2}, \frac{1}{2}[$ onde $\{\varepsilon_t\}$ é um ruído branco.

A função de densidade espectral deste processo é dada pela expressão:

$$f(\omega) = f_u(\omega)\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d},$$

onde f_u é a função de densidade espectral de um processo ARMA(p, q).

Sejam $X_1, ..., X_n$, n observações de $\{X_t\}$. O periodograma $I(\omega)$ é definido pela expressão

$$I(\omega) = \frac{1}{2\pi}\left[R(0) + 2\sum_{s=1}^{n-1} R(s)\cos(s\omega)\right], \quad \omega \in [-\pi, \pi]$$

sendo $R(s)$ a função de auto-covariância amostral,

$$R(s) = \frac{1}{n}\sum_{i=1}^{n-s}(X_i - \overline{X})(X_{i+s} - \overline{X}), \quad s = 0, \pm 1, ..., \pm(n-1),$$

onde $\overline{X}$ é a média amostral.

$$\log[f(\omega_j)] = \log(f_u(0)) - d\log\left[2\sin\left(\frac{\omega_j}{2}\right)\right]^2 + \log\frac{f_u(\omega_j)}{f_u(0)}$$

$$\Leftrightarrow \log I(\omega_j) = \log f_u(0) - d\log\left[2\sin\left(\frac{\omega_j}{2}\right)\right]^2 + \log\frac{f_u(\omega_j)}{f_u(0)} + \log\frac{I(\omega_j)}{f(\omega_j)}, j = 1, ..., m \quad \textbf{(3. 1)}$$

Se m for escolhido de tal forma que

$$\frac{m}{n} \xrightarrow[n\to\infty]{} 0,$$

as frequências de Fourier

$$\omega_j = \frac{2\pi j}{n} \xrightarrow[n\to\infty]{} 0 .$$

Então, o termo

$$\log \frac{f_u(\omega_j)}{f_u(0)}$$

é desprezável .

Logo,

$$\log I(\omega_j) \approx \log f_u(0) - d \log\left[2\sin\left(\frac{\omega_j}{2}\right)\right]^2 + \log \frac{I(\omega_j)}{f(\omega_j)}, \qquad j = 1, 2, ..., m \qquad \textbf{(3. 2)}$$

Considere-se a regressão linear

$$Y_j = a + bX_j + e_j, \quad j = 1, ..., m \qquad \textbf{(3. 3)}$$

onde

$$Y_j = \log I(\omega_j),$$

$$X_j = \log\left[2\sin\left(\frac{\omega_j}{2}\right)\right]^2,$$

$$a = \log f_u(0) - c,\ b = -d,\ e_j = \log\left[\frac{I(\omega_j)}{f(\omega_j)}\right] + c \text{ e } c = E\left[-\log \frac{I(\omega_j)}{f(\omega_j)}\right].$$

Aplicando o método dos mínimos quadrados à equação de regressão linear (3. 3) obtém-se

$$\hat{d}_{GPH} = -\frac{\sum_{j=1}^{m}(X_j - \bar{X})Y_j}{\sum_{j=1}^{m}(X_j - \bar{X})^2}.$$

Geweke & Porter-Hudak (1983) propuseram $m = n^{\alpha}$, $0 < \alpha < 1$.

Suponha-se que $\{X_t\}$ é um processo estacionário Gausseano com $d \in]-0.5, 0.5[$. Hurvich et al. (1998) obtiveram expressões para o viés, a variância e EQM assimptóticos de $\hat{d}_{GPH}$ em função do número de ordenadas do periodograma usadas na regressão. Além disso provaram a consistência do estimador e estabeleceram a sua normalidade assimptótica.

Sejam m o número de observações usadas na regressão linear e n a dimensão da amostra.

Pressuposto 3.1.1: $m \to \infty$, $n \to \infty$, *com* $\frac{m}{n} \to 0$ *e* $\frac{m \log m}{n} \to 0$

Pressuposto 3.1.2: $f_u'(0)=0$, $|f_u''(\omega)|<\widetilde{B}_2<\infty$ *e* $|f_u'''(\omega)|<\widetilde{B}_3<\infty$ *para todo o* ω *na vizinhança de zero.*

Teorema 3.1.1: (Hurvich et al. (1998)) *Suponha-se que os pressuposto 3.1.1 e 3.1.2 são válidos. Então,*

$$E\left(\hat{d}_{GPH}-d\right)=-\frac{2\pi^2}{9}\frac{f_u''(0)}{f_u(0)}\frac{m^2}{n^2}+o\left(\frac{m^2}{n^2}\right)+O\left(\frac{\log^3 m}{m}\right)$$

$$\operatorname{var}\left(\hat{d}_{GPH}\right)=\frac{\pi^2}{24m}+o\left(\frac{1}{m}\right)$$

$$EQM\left(\hat{d}_{GPH}\right)=E\left(\hat{d}_{GPH}-d\right)^2=\frac{4\pi^4}{81}\left\{\frac{f_u''(0)}{f_u(0)}\right\}^2\frac{m^4}{n^4}+\frac{\pi^2}{24m}+O\left\{\frac{m\left(\log^3 m\right)}{n^2}\right\}+o\left(\frac{m^4}{n^4}\right)+o\left(\frac{1}{m}\right)$$

Corolário 3.1.1: (Hurvich et al. (1998)) *Suponha-se que os pressupostos 3.1.1 e 3.1.2 são válidos. Então o estimador* $\hat{d}_{GPH}$ *é um estimador consistente de d.*

Como a escolha de m influencia o viés, a variância e o erro quadrático médio do estimador $\hat{d}_{GPH}$, convém usar um valor de m que minimize o EQM. Hurvich et al. (1998) chegaram à conclusão de que a escolha óptima para m é

$$m^{opt}=\left(\frac{27}{128\pi^2}\right)^{\frac{1}{5}}\left(\frac{f_u(0)}{f_u''(0)}\right)^{\frac{2}{5}}n^{\frac{4}{5}}$$

desde que $f_u''(0)\neq 0$.

Com esta escolha de m, o EQM($\hat{d}_{GPH}$) é $O\left(n^{-\frac{4}{5}}\right)$.

No entanto, o m^{opt} não pode ser calculado na prática porque é preciso conhecer a verdadeira função de densidade espectral.

Teorema 3.1.2: (Hurvich et al. (1998)) *Se* $m=o\left(n^{\frac{4}{5}}\right)$ *e* $log^2(n)=o(m)$ *então,*

$$\sqrt{m}\left(\hat{d}_{GPH}-d\right)\xrightarrow{d}N\left(0,\frac{\pi^2}{24}\right)$$

Estimador de Reisen

Reisen (1994) propôs um novo estimador para d, $\hat{d}_{sp}$,o qual é obtido usando o mesmo processo com que foi obtido o estimador $\hat{d}_{GPH}$, mas em vez de se usar o periodograma $I(\omega)$, o qual não é um estimador consistente da função de densidade espectral, vai-se usar o periodograma amaciado

$$f_s(\omega)=\frac{1}{2\pi}\sum_{s=-M}^{M}k\left(\frac{s}{M}\right)R(s)\cos(s\omega)$$

onde $k(u)$ é o gerador da janela de Parzen definida pela expressão

$$k(u)=\begin{cases}1-6u^2+6|u|^3 & |u|\le\frac{1}{2}\\ 2(1-|u|)^3 & \frac{1}{2}<|u|\le 1\\ 0 & |u|>1\end{cases}$$

Portanto,

$$\hat{d}_{sp}=-\frac{\sum_{j=1}^{m}(X_j-\overline{X})Y_j}{\sum_{j=1}^{m}(X_j-\overline{X})^2}$$

onde

$$X_j=\log\left[2\sin\left(\frac{\omega_j}{2}\right)\right]^2$$

e

$$Y_j=\log f_s(\omega_j)\ \text{e}\ \overline{X}=\frac{1}{m}\sum_{j=1}^{m}X_j\ .$$

Reisen (1994) mostrou que para $d<0$ e para $\omega\neq 0,\ \pi$,

$$\text{var}(\hat{d}_{sp})\approx 0.53928\frac{M}{n\sum_{i=1}^{m}(X_i-\overline{X})^2}. \qquad \textbf{(3. 4)}$$

M, geralmente designado por "ponto de truncamento", é função de n e deve ser escolhido de forma a que à medida que $n\to\infty$, $M\to\infty$ e $\frac{M}{n}\to 0$. Reisen (1994) usou $m=n^{\alpha}$ $(0<\alpha<1)$ e $M=n^{\beta}$ $(0<\beta<1)$.

Estimador "pooled"

Observando (3.1), constata-se que o periodograma nas frequências altas ω_s ($s = m+1, ..., \left[\frac{n}{2}\right]$) continua a conter informação sobre d, embora $f_u(\omega_s)$, nestes casos, deixe de poder ser considerada uma constante. Portanto, a regressão usual que se baseia no logaritmo do periodograma pode estar a pôr de parte alguma informação sobre as observações. Pode-se corrigir esta situação usando mais intervalos de frequências, além daquele que corresponde a uma vizinhança de zero e em simultâneo permitir a variação de $f_u(\omega)$.

O método de obtenção do estimador de d, $\hat{d}_{pooled}$ baseia-se na regressão do logaritmo do periodograma usando um maior número de frequências de Fourier, $\omega_s = \frac{2\pi s}{n}$, $s = 1, ..., TL$ com $L \to \infty$ e $\frac{TL}{n} \to 0$, sendo T o número de ordenadas do periodograma usadas na regressão para a obtenção de $\hat{d}_{GPH}$.

Seja $\{X_t\}$ um processo estacionário Gausseano de memória longa com função de densidade espectral $f(\omega)$ onde $-½ < d < ½$ e $f_u(\omega)$ é uma função simétrica, periódica (de período 2π), positiva, contínua, limitada em todo o seu domínio, excepto em zero, e com derivada de terceira ordem finita. No domínio do tempo $\{X_t\}$ tem a forma $(1 - B)^d X_t = u_t$ onde $\{u_t\}$ é uma série temporal estacionária com função de densidade espectral $f_u(\omega)$.

Usando uma representação do logaritmo do periodograma diferente da que foi usada no caso do estimador $\hat{d}_{GPH}$ obtém-se a expressão:

$$\ln(I(\omega_s)) = \ln(f_u(\omega_s)) - 2d \ln\left|1 - e^{i\omega_s}\right| + \ln\left(\frac{I(\omega_s)}{f(\omega_s)}\right)$$

$$= \ln(f_u(\lambda_j)) - 2d \ln\left|2\sin\left(\frac{\omega_s}{2}\right)\right| + \ln\left(\frac{I(\omega_s)}{f(\omega_s)}\right) + \ln\left(\frac{f_u(\omega_s)}{f_u(\lambda_j)}\right), \quad \omega_s \in B_j, \quad \textbf{(3. 5)}$$

onde

$$I(\omega_s) = |w(\omega_s)|^2$$

sendo

$$w(\omega_s) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\omega_s}$$

e

$$B_j = \begin{cases} \left\{ \omega_s : \lambda_j - \dfrac{\pi}{2M} < \omega_s \le \lambda_j + \dfrac{\pi}{2M} \right\} & \lambda_j = \dfrac{(2j+1)\pi}{2M},\ j = 1, \ldots, T^* - 1 \\ \left\{ \omega_s : 0 < \omega_s \le \dfrac{\pi}{M} \right\} & \lambda_0 = 0,\ j = 0 \end{cases}$$

são os conjuntos de frequências de Fourier que vão ser usados.

Esta representação permite que as frequências de ω_s onde o periodograma é calculado estejam na vizinhança de um conjunto de frequências λ_j para j = 0, 1, ..., T^*-1, onde T^* é o parâmetro que determina o número total de conjuntos distintos de frequências de Fourier que vão ser usados.

Shimotsu & Phillips (2002) propuseram estimar o parâmetro d em (3. 5) através de uma regressão linear usando L+1 conjuntos B_0, ..., B_L, onde L é um número tal que à medida que $n \to \infty$, $L \to \infty$ e $L/T^* \to 0$.

Escrevendo (3. 5) da forma

$$Y_{sj} = \mu_j + dX_{sj} + \eta_{sj} + \varepsilon_{sj}, \quad s = 1, \ldots, T;\ j = 0, 1, \ldots, L$$

com

$$Y_{sj} = \ln I(\omega_s), \quad X_{sj} = -\ln\left(4\sin^2\left(\frac{\omega_s}{2}\right)\right), \quad \eta_{sj} = \ln\left(\frac{f_u(\omega_s)}{f_u(\lambda_j)}\right), \quad \varepsilon_{sj} = \ln\left(\frac{I(\omega_s)}{f(\omega_s)}\right) - \psi(1)$$

para $\omega_s \in B_j$ e

$$\mu_j = \ln f_u(\lambda_j) + \psi(1)$$

onde $\psi(1) = \Gamma'(1) = -\gamma$ = -0.57721566... sendo γ a constante de Euler.

Obtém-se o estimador $\hat{d}_{pooled}$ através da seguinte expressão

$$\hat{d}_{pooled} = \frac{\sum_{j=0}^{L} \sum_{\{s:\omega_s \in B_j\}} Y_{sj}\left(X_{sj} - \overline{X}_{.j}\right)}{\sum_{j=0}^{L} \sum_{\{s:\omega_s \in B_j\}} \left(X_{sj} - \overline{X}_{.j}\right)^2} = \sum_{j=0}^{L} \left(\frac{\sum_{\{s:\omega_s \in B_j\}} \left(X_{sj} - \overline{X}_{.j}\right)^2}{\sum_{j=0}^{L} \sum_{\{s:\omega_s \in B_j\}} \left(X_{sj} - \overline{X}_{.j}\right)^2} \right) \hat{d}_j$$

onde ,

$$\overline{X}_{.j} = \frac{1}{T} \sum_{\{s:\omega_s \in B_j\}} X_{sj} = -\frac{1}{T} \sum_{\{s:\omega_s \in B_j\}} \ln\left(4\sin^2\left(\frac{\omega_s}{2}\right)\right)$$

onde $\hat{d}_j$ é o estimador de d obtido usando apenas o conjunto B_j.

Notar que à medida que a dimensão da amostra n aumenta, $\hat{d}_{pooled}$ usa um número cada vez maior de conjuntos B_j, j = 0, ..., L. No entanto, este estimador continua a usar apenas as frequências que estão numa vizinhança da origem, porque $\frac{TL}{n} \to 0$. Portanto este estimador

pertence à classe dos estimadores semi-paramétricos e em simultâneo usa um número crescente de conjuntos B_j.

Pressuposto 3.1.3: $T \to \infty, n \to \infty, \frac{T}{n} = \frac{1}{2T^*} \to 0$

Pressuposto 3.1.4: $\frac{T^*}{T} + \frac{T \ln T}{T} + \frac{\ln^2 n}{T} \to 0$.

Pressuposto 3.1.5:

$$f_u'(0) = 0,\ f_u(\omega) > C_0 > 0,\ |f_u'(\omega)| < C_1 < \infty, |f_u''(\omega)| < C_2 < \infty, |f_u'''(\omega)| < C_3 < \infty, \forall \omega \in [0, \pi]$$

Nos dois teoremas seguintes vai-se assumir que $n = 2TT^*$ com T e T^* inteiros, sendo portanto, T^* definido pela igualdade $T^* = n/(2T)$ (quando n é par). Logo, qualquer escolha de T afecta a amplitude dos intervalos considerados na definição dos conjuntos B_j, a sua velocidade de contracção e o número de conjuntos B_j usados na regressão.

Teorema 3.1.3: (Shimotsu & Phillips (2002)) *Se os pressupostos 3.1.3 – 3.1.5 se verificam, então*

$$E\left(\hat{d}_{pooled} - d\right) = \frac{-\pi^2}{6(1+\Xi)} \frac{f_u''(0)}{f_u(0)} \frac{T^2 L}{n^2} + o\left(\frac{T^2 L}{n^2}\right) + O\left(\frac{\ln^3 T}{T}\right)$$

$$\text{var}\left(\hat{d}_{pooled}\right) = \frac{\pi^2}{24(1+\Xi)T} + o\left(\frac{1}{T}\right)$$

$$EQM\left(\hat{d}_{pooled}\right) =$$

$$\frac{\pi^4}{36(1+\Xi)^2} \left\{\frac{f_u''(0)}{f_u(0)}\right\}^2 \frac{T^4 L^2}{n^4} + \frac{\pi^2}{24(1+\Xi)T} + o\left(\frac{T^4 L^2}{n^4}\right) + O\left(\frac{\ln^6 T}{T^2}\right) + O\left(\frac{TL \ln^3 T}{n^2}\right) + o\left(\frac{1}{T}\right)$$

onde

$$\Xi = \sum_{j=1}^{\infty} \left[- j(j+1)(\ln(j+1) - \ln j)^2 + 1 \right] \approx 0{,}0803.$$

Notar que o EQM tende para zero à medida $n \to \infty$ e portanto $\hat{d}_{pooled}$ é um estimador consistente.

A variância assimptótica deste estimador é inferior à variância assimptótica do estimador GPH, reflectindo o grande número de ordenadas do periodograma que são usadas na

regressão. No entanto, o estimador $\hat{d}_{pooled}$ apresenta um viés assimptótico mais elevado que o estimador GPH devido ao facto de $f_u(\omega)$ não ser constante nos conjuntos B_j para $j > 0$.

Através de um estudo simulado, usando amostras de dimensão finita, realizado por Shimotsu & Phillips (2002) verificou-se que o estimador $\hat{d}_{pooled}$ é melhor que o estimador GPH quando $f_u(\omega)$ tem "picos" perto da origem e tem um desempenho pior quando $f_u(\omega)$ muda monoticamente de $\omega = 0$ para $\omega = \pi$, embora a diferença seja pequena.

O teorema seguinte estabelece a normalidade assimptótica do estimador $\hat{d}_{pooled}$.

Teorema 3.1.4: (Shimotsu & Philips (2002)) *Suponha-se que os pressupostos 3.1.3-3.1.5 se verificam. Além disso, se* $T = O(n^{0.8-\varepsilon})$ *para algum* $\varepsilon > 0$ *e* $L = O(\ln T^*)$, *tem-se*

$$\sqrt{T}\left(\hat{d}_{pooled} - d\right) \xrightarrow{d} N\left(0, \frac{\pi^2}{24(1+\Xi)}\right).$$

Estimador $\hat{d}_r$

Seja $\{X_t\}$ um processo Gausseano estacionário de memória longa com função de densidade espectral

$$g(\omega) = |\omega|^{-2d} g_u(\omega),$$

onde $-\frac{1}{2} < d < \frac{1}{2}$ e $g_u(.)$ é uma função par em $[-\pi, \pi]$ que é contínua em zero com $0 < g_u(0) < \infty$ e $g(\omega)$ é integrável em $]-\pi, \pi[$.

O estimador descrito nesta subsecção é um estimador do parâmetro de memória longa d, $\hat{d}_r$ e é definido como sendo o coeficiente de $-2\log(\omega_j)$ sendo obtido através do método dos mínimos quadrados aplicado à regressão do logaritmo do periodograma, $\log I_j$ calculado em ω_j, numa constante, em X_j, e ω_j^2, ω_j^4, ..., ω_j^{2r} para $j = 1, ..., m$ (sendo m um inteiro positivo tal que $m \leq \left[\frac{n}{2}\right]$ e n é a dimensão da amostra) e para algum inteiro não negativo r, onde $\omega_j = \frac{2\pi j}{n}$, sendo n a dimensão da amostra e $X_j = -2\log(\omega_j)$.

O termo dominante da expressão do viés do estimador $\hat{d}_{GPH}$ (teorema 3.1.1) resulta do termo $\log\left(\frac{f_u(\omega_j)}{f_u(0)}\right)$ em vez do termo $E(\varepsilon_j)$.

Supondo que o pressuposto 3.1.2 se verifica, obtém-se através da expansão em série de Taylor a igualdade

$$\log\left(\frac{f_u(\omega_j)}{f_u(0)}\right) = \frac{\omega_j^2}{2}\frac{f_u''(0)}{f_u(0)} + O(\omega_j^3).$$

É o primeiro termo que se encontra do lado direito da igualdade anterior que é responsável pelo termo dominante do viés do estimador $\hat{d}_{GPH}$. Este termo pode ser eliminado através da adição de ω_j^2 ao modelo de regressão linear

$$\log I_j = \log I(\omega_j) = (\log f_u(0) - C) - 2d\tilde{X}_j + \log\left(\frac{f_u(\omega_j)}{f_u(0)}\right) + \varepsilon_j \qquad \textbf{(3. 6)}$$

onde

$$\varepsilon_j = \log\left(\frac{I(\omega_j)}{f(\omega_j)}\right) + C,$$

$$\tilde{X}_j = \log\left(2\sin\left(\frac{\omega_j}{2}\right)\right)$$

e $C = 0.577216...$ é a constante de Euler.

Além disso, termos de viés adicionais podem ser eliminados através da adição dos termos $\omega_j^4, ..., \omega_j^{2r}$ para $r \geq 2$ ao modelo de regressão linear (3. 6).

Quando $r = 0$, $\hat{d}_r$ é assimptóticamente equivalente ao estimador $\hat{d}_{GPH}$.

Vai-se assumir que a função g_u é "suave" de ordem s na frequência zero. Seja $[s]$ a parte inteira de s. Uma função real h definida na vizinhança de zero é "suave" de ordem $s > 0$ em zero se h é $[s]$ vezes continuamente diferenciável em alguma vizinhança de zero e as suas derivadas de ordem $[s]$ denotadas por $h^{[s]}$ satisfazem a condição

$$\left|h^{([s])}(\omega) - h^{([s])}(0)\right| \leq c|\omega|^{s-[s]}$$

para alguma constante $c < \infty$ e para todo o ω na vizinhança de zero.

Pressuposto 3.1.6: $m = m(n) \to \infty$ *e* $\frac{m}{n} \xrightarrow[n\to\infty]{} 0$.

Pressuposto 3.1.7: *g_u é uma função par em $[-\pi, \pi]$ que é "suave" de ordem s em zero para algum $s \geq 1$, $0 < g_u(0) < \infty$, $-\frac{1}{2} < d < \frac{1}{2}$ e* $\int_{-\pi}^{\pi} |\omega|^{-2d} g_u(\omega) d\omega < \infty$.

O pressuposto anterior permite a seguinte representação em série de Taylor em torno de zero de $\log(g_u(\omega_j))$ até ao termo $[s]$:

$$\log\left(\frac{g_u(\omega_j)}{g_u(0)}\right) = \sum_{k=1}^{[s]} \frac{b_k}{k!}\omega_j^k + O(\omega_j^s) \qquad \textbf{(3. 7)}$$

$$= \sum_{k=1}^{[s/2]} \frac{b_{2k}}{(2k)!}\omega_j^{2k} + O(\omega_j^s) \qquad \textbf{(3. 8)}$$

$$= \sum_{k=1}^{\min\{[s/2],r\}} \frac{b_{2k}}{(2k)!}\omega_j^{2k} + R_j \qquad \textbf{(3. 9)}$$

onde

$$R_j = \sum_{k=\min\{[s/2],r\}+1}^{[s/2]} \frac{b_{2k}}{(2k)!}\omega_j^{2k} + O(\omega_j^s) = 1(s \geq 2+2r)\frac{b_{2+2r}}{(2+2r)!}\omega_j^{2r} + O(\omega_j^a)$$

sendo $a = \min\{s, 4 + 2r\}$, $b_k = \dfrac{d^k}{d\omega^k}\log g_u(\omega)|_{\omega=0}$ e 1 representa a função indicatriz[6].

$O(.)$ verifica-se uniformemente para $j = 1, ..., m$ e para todo o n.

Se s é um inteiro, então (3. 7) – (3. 9) verificam-se com $O(.)$ substituído por $o(.)$.

Seja $Q_{k,j} = \omega_j^k$ onde $\omega_j = \dfrac{2\pi j}{n}$ para $j = 1, ..., m$ e $k = 1, 2, ...$

Sejam $\log I$, X, Q_k, R e ε vectores coluna de dimensão m cujos j-ésimos elementos são $\log I_j$, X_j, $Q_{k,j}$, R_j e e_j, respectivamente, onde $e_j = \log\dfrac{I_j}{f_j} + C$ (C = 0.577216... é a constante de Euler) e $g_j = g(\omega_j)$. Seja Q a matriz $m \times r$ cuja k-ésima coluna é Q_{2k} para k=1, ..., r. Seja 1_m um vector coluna de uns de dimensão m e seja b o vector coluna de dimensão r cujo k-ésimo elemento é $b_{2k}/(2k)!$ para $k = 1, ..., \min\left\{\left[\frac{s}{2}\right], r\right\}$ e 0 para $k = \min\left\{\left[\frac{s}{2}\right], r\right\}+1, ..., r$.

Combinando (3. 9) com (3. 6) (com f_u substituído por g_u e $-2\tilde{X}_j$ por X_j) obtém-se:

$$\log I = (\log g_u(0) - C)1_m + Xd + Qb + R + e\,. \qquad \textbf{(3. 10)}$$

Sejam

$$X^* = X - 1_m\overline{X} \text{ e } Q^* = Q - 1_m\overline{Q}^t$$

onde

---

[6] 1(A) é uma função indicatriz se e só se para x ∈ A, 1(x) = 1 e para x ∉ A, 1(x) = 0.

$$\overline{X} = \frac{1}{m} X^t 1_m \text{ e } \overline{Q} = \frac{1}{m} Q^t 1_m .$$

A igualdade (3. 10) é equivalente ao modelo de regressão linear

$$\log I = K1_m + X^* d + Q^* b + R + e \qquad \textbf{(3. 11)}$$

onde

$$K = \log g(0) - C + \overline{X}d + \overline{Q}^t b .$$

O estimador de d, $\hat{d}_r$, é o coeficiente de X^* obtido pelo método dos mínimos quadrados aplicado à regressão do log(I) em 1_m, X^* e Q^*:

$$\hat{d}_r = \left(X^{*t} M_{Q^*} X^*\right)^{-1} X^{*t} M_{Q^*} \log I$$

onde

$$M_{Q^*} = I_m - Q^* \left(Q^{*t} Q^*\right)^{-1} Q^{*t},$$

sendo I_m a matriz identidade $m \times m$.

(Para $r = 0$, defina-se $M_{Q^*} = I_m$).

O valor esperado de $\hat{d}_r$ é dado pela expressão

$$E\left(\hat{d}_r\right) = d + \left(X^{*t} M_{Q^*} X^*\right)^{-1} X^{*t} M_{Q^*} \left(R + E(\varepsilon)\right).$$

O termo $Q^* b$ de (3. 11), o qual inclui termos ω_j^{2k} para $k = 1, ..., \min\{[s/2], r\}$ na expansão em série de Taylor de $\log\left(\frac{g_u(\omega_j)}{g_u(0)}\right)$ não aparece na igualdade anterior porque é eliminado pela inclusão de Q^* na regressão (3. 11). Consequentemente, o viés de $\hat{d}_r$ é de ordem inferior relativamente ao viés de $\hat{d}_{GPH}$.

Seja μ_r um vector coluna de dimensão r cujo k-ésimo elemento é

$$\mu_{r,k} = \frac{2k}{(2k+1)^2} \text{ para } k = 1, ..., r$$

e seja Γ_r uma matriz $r \times r$ cujo elemento (i, k) é

$$\left[\Gamma_r\right]_{i,k} = \frac{4ik}{(2i+2k+1)(2i+1)(2k+1)} \text{ para } i, k = 1, ..., r.$$

Para $r = 0$ defina-se $\mu_r = 0$ e $\Gamma_r = 1$.

Defina-se a quantidade c_r que vai ser usada nas expressões da variância e do viés do estimador $\hat{d}_r$, no próximo teorema, através da expressão

$$c_r = \left(1 - \mu_r^t \Gamma_r^{-1} \mu_r\right)^{-1}.$$

Seja ξ_r um vector coluna de dimensão r cujo k-ésimo elemento é

$$\xi_{r,k} = \frac{2k(3+2r)}{(2r+2k+3)(2k+1)} \quad \text{para } k = 1, \ldots, r.$$

Defina-se

$$\tau_r = \frac{(2\pi)^{2+2r}(2+2r)c_r}{2(3+2r)!(3+2r)}\left(1-\mu_r^t\Gamma_r^{-1}\xi_r\right).$$

Considere-se o teorema seguinte que estabelece as expressões assimptóticas para o viés e para a variância do estimador $\hat{d}_r$.

Teorema 3.1.5: (Andrews & Guggenberger (2003)) *Suponha-se que os pressupostos 3.1.6 e 3.1.7 se verificam. Então,*

a) $E\left(\hat{d}_r - d\right) = 1(s \geq 2+2r)\tau_r b_{2+2r}\frac{m^{2+2r}}{n^{2+2r}}(1+o(1)) + O\left(\frac{m^a}{n^a}\right) + O\left(\frac{\log^3 m}{m}\right)$ *e*

b) $\text{var}\left(\hat{d}_r\right) = \frac{\pi^2}{24}\frac{c_r}{m} + o\left(\frac{1}{m}\right)$

Se s é um inteiro, a alínea a) verifica-se com $O\left(\frac{m^a}{n^a}\right)$ substituído por $o\left(\frac{m^a}{n^a}\right)$. Em particular, de $s = 2 + 2r$, a alínea a) verifica-se com $O\left(\frac{m^a}{n^a}\right)$ substituído por $o\left(\frac{m^a}{n^a}\right) = o\left(\frac{m^{2+2r}}{n^{2+2r}}\right)$.

Comparando os resultados deste teorema com o teorema 3.1.1. verifica-se que o viés do estimador $\hat{d}_r$ converge mais rapidamente para zero que o viés do estimador $\hat{d}_{GPH}$, enquanto que a variância destes dois estimadores difere apenas por uma constante multiplicativa c_r.

A escolha óptima de m, isto é, a escolha de m que maximiza a convergência para zero do EQM é

$$m \sim n^{2\phi/(2\phi+1)},$$

onde

$$\phi = \min\{s, 2+2r\}$$

e

$$m \sim n^{2\phi/(2\phi+1)}$$

significa que

$$\frac{m}{n^{2\phi/(2\phi+1)}}$$

converge para uma constante positiva finita à medida que $n \to \infty$.

Suponha-se que $m \propto n^{\gamma}$ para algum $0 < \gamma < 1$ e $s \geq 2 + 2r$. Então, o EQM de $\hat{d}_r$ é:

$$EQM(\hat{d}_r) = \tau_r^2 b_{2+2r}^2 \frac{m^{4+4r}}{n^{4+4r}}(1+o(1)) + O\left(\frac{m^{1+2r}\log^3 m}{n^{2+2r}}\right) + \frac{\pi^2}{24}\frac{c_r}{m}(1+o(1))$$

Se $\gamma > (2 + 2r)/ (3 + 2r)$, então o termo $O(.)$ na expressão anterior é de ordem inferior relativamente aos outros dois termos. Ignorando $O(.)$, o valor de m que minimiza o EQM é dado pela expressão

$$m_{opt} = \left[\left(\frac{\pi^2 c_r}{24(4+4r)\tau_r^2 b_{2+2r}^2}\right)^{1/(5+4r)} .n^{(4+4r)/(5+4r)}\right],$$

onde $[A]$ representa a parte inteira de A.

A expressão anterior traduz a escolha óptima de m apenas quando $\tau_r \neq 0$, $b_{2+2r} \neq 0$ e $c_r < \infty$.

Pressuposto 3.1.8: $m = o\left(n^{2\phi/(2\phi+1)}\right)$, *onde* $\phi = min\{s, 2 + 2r\}$, $s \geq 1$, *e* s *é da mesma forma que no pressuposto 3.1.7.*

Teorema 3.1.6: (Andrews & Guggenberger (2003)) *Suponha-se que os pressupostos 3.1.7 e 3.1.8 são válidos. Então,*

$$\sqrt{m}(\hat{d}_r - d) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}c_r\right) \quad quando\ n \to \infty.$$

Estimador de Robinson

Nesta secção vai-se estudar a estimação de modelos de séries temporais múltiplas, os quais permitem que elementos da matriz de densidade espectral tendam para infinito ou para zero na frequência zero.

Seja X_t um vector de dimensão G com g-ésimo elemento X_{gt}, $g = 1, ..., G$. Vai-se assumir que X_t tem matriz de densidade espectral dada por

$$E\left\{(X_1 - E(X_1))(X_{1+j} - E(X_1))'\right\} = \int_{-\pi}^{\pi} e^{ij\omega} f(\omega) d\omega .$$

O (g, h) – ésimo elemento de $f(\omega)$ é a densidade espectral cruzada de X_{gt} e X_{ht} quando $g \neq h$ e é denotada por $f_{gh}(\omega)$. O g-ésimo elemento da diagonal de $f(\omega)$, $f_{gg}(\omega)$ é a função de densidade espectral de X_{gt}.

Para C_g, d_g satisfazendo $0 < C_g < \infty$, $-\frac{1}{2} < d_g < \frac{1}{2}$, assuma-se que à medida que $\omega \to 0^+$

$$f_{gg}(\omega) \sim C_g \omega^{-2d_g}$$

para $g = 1, ..., G$. O símbolo "~" significa que $\dfrac{f_{gg}(\omega)}{C_g \omega^{-2d_g}}$ tende para 1.

O periodograma de X_{gt}, $t = 1, ..., n$ é $I_g(\omega)$, $g = 1, ..., G$.

Seja J um inteiro fixo maior ou igual a 1. Defina-se

$$Y_{gk}^{(J)} = \log\left\{\sum_{j=1}^{J} I_g\left(\omega_{k+j-J}\right)\right\}, \qquad g = 1, ..., G, \quad k = l + J, l + 2J, ..., m.$$

Na teoria assimptótica, à medida que n tende para infinito, l e m também tendem para infinito e $l/m \to 0$.

Vai-se assumir que $\dfrac{m-l}{J}$ é um inteiro.

Definam-se as variáveis não observáveis $U_{gk}^{(J)}$ por

$$Y_{gk}^{(J)} = c_g^{(J)} - d_g\left(2\log\omega_k\right) + U_{gk}^{(J)}, \qquad g = 1, ..., G, \quad k = l + J, l + 2J, ..., m,$$

onde

$$c_g^{(J)} = \log C_g + \psi(J)$$

sendo ψ a função digama definida pela igualdade

$$\psi(z) = \frac{d}{dz}\log\Gamma(z).$$

Os estimadores de $c^{(J)} = \left(c_1^{(J)}, ..., c_G^{(J)}\right)^t$ e $d = (d_1, ..., d_G)^t$ obtidos pelo método dos mínimos quadrados são $\tilde{c}^{(J)}$ e $\tilde{d}^{(J)}$ dados por

$$\begin{bmatrix} \tilde{c}^{(J)} \\ \tilde{d}^{(J)} \end{bmatrix} = vec\left\{Y^{(J)t} Z^{(J)}\left(Z^{(J)t} Z^{(J)}\right)^{-1}\right\}^{7}$$

---

[7] Empilhando as colunas de uma matriz quadrada $L = \left\lfloor L_{ij} \right\rfloor_{i,\ j=1,...,l}$ umas em cima das outras, pode-se definir

$$vec(L) = \begin{bmatrix} L_{\cdot 1} \\ L_{\cdot 2} \\ \vdots \\ L_{\cdot l} \end{bmatrix}$$ onde $L_{\cdot j}$, $j = 1, ..., l$ representa a j-ésima coluna da matriz L.

onde $Z^{(J)} = (Z_{l+J}, Z_{l+2J} \ldots, Z_m)^t$, $Y^{(J)} = (Y_1^{(J)}, \ldots, Y_G^{(J)})$, $Z_k = (1, -2\log(\omega_k))^t$ e $Y_g^{(J)} = \left(Y_{g,l+J}^{(J)}, Y_{g,l+2J}^{(J)}, \ldots, Y_{g,m}^{(J)}\right)^t$.

Seja $\tilde{c}_g^{(J)}$ o g-ésimo elemento de $\tilde{c}^{(J)}$. C_g pode ser estimado através da igualdade

$$\tilde{C}_g^{(J)} = \exp\left(\tilde{c}_g^{(J)} - \psi(J)\right).$$

Pressuposto 3.1.9: *Existem* $C_g \in]0, \infty[$, $d_g \in \left]-\frac{1}{2}, \frac{1}{2}\right[$, $\alpha \in]0, 2]$ *tais que à medida que* $\omega \to 0^+$

$$f_{gg}(\omega) = C_g \omega^{-2d_g} + O\left(\omega^{\alpha - 2d_g}\right),\ g = 1, \ldots, G.$$

Pressuposto 3.1.10: *Numa vizinhança de zero,* $]0, \varepsilon[$, $f_{gh}(\omega)$ *é diferenciável e à medida que* $\omega \to 0^+$

$$\left|\frac{d}{d\omega}\log f_{gh}(\omega)\right| = O\left(\omega^{-1-d_g-d_h}\right),\ g, h = 1, \ldots, G.$$

Seja R_{gh} a coerência entre X_{gt} e X_{ht} sendo definida pela igualdade

$$R_{gh}(\omega) = \frac{f_{gh}(\omega)}{\sqrt{f_{gg}(\omega) f_{hh}(\omega)}}$$

Pressuposto 3.1.11: *Para algum* $\beta \in [0, 2]$ *à medida que* $\omega \to 0^+$,

$$|R_{gh}(\omega) - R_{gh}(0)| = O(\omega^\beta)$$

$g < h = 2, \ldots, G$.

Pressuposto 3.1.12 : $R(0)$ $(= R_{gh}(0))$ *não é singular.*

Pressuposto 3.1.13: $\{X_t: t = 1, 2, \ldots\}$ *é um processo Gausseano.*

Pressuposto 3.1.14: *À medida que* $n \to \infty$

$$\frac{\sqrt{m}\log m}{l} + \frac{l(\log n)^2}{m} + \frac{m^{1+\frac{1}{2}\min(\alpha,\beta)}}{n} \to 0.$$

Teorema 3.1.7: *Suponha-se que os pressupostos 3.1.9 – 3.1.14 se verificam. Então, à medida que* $n \to \infty$

$$\begin{bmatrix} \frac{\sqrt{m}}{\log n}\left(\tilde{c}^{(J)} - c^{(J)}\right) \\ 2\sqrt{m}\left(\tilde{d}^{(J)} - d\right) \end{bmatrix} \xrightarrow{d} N\left[0,\ J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \Omega^{(J)}\right]^{8}$$

e a matriz das covariâncias $\Omega^{(J)}$ *na distribuição limite é consistentemente estimada por*

$$J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \tilde{\Omega}^{(J)}$$

sendo

$$\tilde{\Omega}^{(J)} = \frac{J}{m-l}\sum_{k} \tilde{U}_k^{(J)} \tilde{U}_k^{(J)\,t}$$

com $k = l+J,\ l+2J,\ \ldots,\ m$ *e*

$$\tilde{U}_k^{(J)} = Y_k^{(J)} - \tilde{c}^{(J)} + \tilde{d}^{(J)}\left(2\log \omega_k\right).$$

É de notar que para cada g, $2\sqrt{m}\left(\tilde{d}_g^{(J)} - d_g\right) \xrightarrow{d} N\left(0, J\psi'(J)\right)$, sendo $J\psi'(J)$ uma função decrescente, a qual tende para um à medida que $J \to \infty$.

Estimador de Velasco

Velasco (2000) mostrou a consistência de um estimador que controla a memória longa, d (o qual foi proposto por Robinson(1995a)) de um processo linear (não necessariamente Gausseano), sendo este estimador obtido pelo método dos mínimos quadrados, com base na regressão linear do logaritmo do periodograma, usando apenas as frequências de Fourier que estão na vizinhança de zero, quando se faz associação das ordenadas do periodograma. Velasco (2000) substituiu o periodograma deste estimador pelo periodograma "tapered", o qual usa o "taper" de Hanning e deduz a distribuição assimptótica deste novo estimador.

Seja $\{X_t,\ t = 1, 2, \ldots\}$ um processo estacionário com função de densidade espectral

$$f(\omega) \sim C\omega^{-2d}\ ,\ \omega \to 0^{+}$$

---

[8] O símbolo ⊗ representa o produto de Kronecker. O produto de Kronecker entre duas matrizes quadradas $M = \left[M_{ij}\right]_{i,\,j=1,\ldots,m}$ e $N = \left[N_{ij}\right]_{i,\,j=1,\ldots,n}$ é uma matriz $mn \times mn$ definida pela igualdade

$$M \otimes N = \begin{bmatrix} M_{11}N & M_{12}N & \cdots & M_{1m}N \\ M_{21}N & M_{22}N & \cdots & M_{2m}N \\ \vdots & \vdots & \ddots & \\ M_{m1}N & M_{m2}N & \cdots & M_{mm}N \end{bmatrix}.$$

e sejam X_1, ..., X_n, n observações correspondentes a este processo.

A transformada de Fourier nas frequências $\omega_j = \frac{2\pi j}{n}$, j inteiro, é definida pela expressão

$$W(\omega_j) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} X_t e^{it\omega_j}$$

e o periodograma é definido pela igualdade

$$I(\omega_j) = |W(\omega_j)|^2 .$$

Defina-se para cada J = 1, 2, ... fixo, e para dois inteiros positivos l e m (assumindo que $(m-l)/J$ é um inteiro),

$$Y_k^{(J)} = \log\left(\sum_{j=1}^{J} I(\omega_{k+j-J}) \right), \quad k = l + J, l + 2J, ..., m .$$

O estimador do parâmetro de memória longa considerado em Robinson (1995a) é

$$\hat{d} = \frac{\sum_k (z_k - \bar{z}) Y_k^{(J)}}{\sum_k (z_k - \bar{z})^2}$$

onde

$$z_k = -2\log \omega_k,$$

$$\bar{z} = \frac{J}{m-l} \sum_k z_k .$$

Aqui m é um inteiro menor que n e l é um número de "rectificação". Assimptóticamente m e l tendem para infinito, mas mais lentamente que n.

Velasco (2000) com base nos cinco pressupostos seguintes demonstrou a consistência assimptótica deste estimador.

Pressuposto 3.1.15: *Se* $\{X_t\}$ *é um processo estacionário e se* $d \in \left[0, \frac{1}{2}\right[$, $\alpha \in]0, 2]$ *e* $0 < G < \infty$, $f(\omega) = G\omega^{-2d} + O(\omega^{\alpha-2d})$ *quando* $\omega \to 0^+$.

Pressuposto 3.1.16: *Numa vizinhança de zero,* $]0, \varepsilon[$, $f(\omega)$ *é diferenciável e* $\left| \frac{d}{d\omega} \log f(\omega) \right| = O\left(\frac{1}{\omega}\right)$ *à medida que* $\omega \to 0^+$.

Pressuposto 3.1.17: X_t *satisfaz*

$$X_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \quad \left| \frac{d}{d\omega} \alpha(\omega) \right| = O\left(\frac{|\alpha(\omega)|}{\omega} \right), \quad \alpha(\omega) = \sum_{j=0}^{\infty} \alpha_j e^{ij\omega}$$

à medida que $\omega \to 0^+$, *onde as variáveis aleatórias* ε_t *são i. i. d. com* $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = 1$ *e* $E(\varepsilon_t^4) < \infty$.

Pressuposto 3.1.18: ε_t *tem função característica*

$$\hat{Q}(\theta) = E\left(e^{i\theta\varepsilon_t}\right)$$

satisfazendo

$$\sup_{|\theta| \geq \theta_0} \left| \hat{Q}(\theta) \right| = \delta(\theta_0) < 1, \quad \forall \theta_0 > 0 \quad e \quad \int_{-\infty}^{\infty} \left| \hat{Q}(\theta) \right|^p d\theta < \infty$$

para algum inteiro $p > 1$.

Pressuposto 3.1.19: *À medida que* $n \to \infty$, *para algum* $b > 0$, $J \geq 3$,

$$\frac{\sqrt{m}}{l^{J/(J+2)-b}} + \frac{l(\log n)^2}{m} + \frac{m}{n} \to 0.$$

Teorema 3.1.8: (Velasco (2000)) *Com base nos pressupostos 3.1.15 – 3.1.19, com* $J \geq 3$,

$$\hat{d} \xrightarrow{p} d.$$

Apesar de Velasco (2000) ter conseguido deduzir a consistência assimptótica do estimador proposto por Robinson (1995a), é impossível obter a sua distribuição assimptótica a partir da demonstração do teorema anterior porque o viés do periodograma é grande.

Uma forma de reduzir a dispersão do periodograma é aplicar o "tapering" aos dados e neste caso é um método eficaz de reduzir o viés do periodograma. "Tapering" dá maior relevância às observações que se encontram na parte central da sequência observada, jogando fora as observações extremas através de uma função positiva "suave", h_t, $t = 1, ..., n$, $0 \leq h_t \leq 1$.

Pressuposto 3.1.20: *Seja* $g(\omega) = G|\omega|^{-2d}$ *e assuma-se que para algum* $0 < E_\alpha < \infty$, *à medida que* $\omega \to 0^+$, $\frac{f(\omega)}{g(\omega)} = 1 + E_\alpha \times \omega^\alpha + o(\omega^\alpha)$, $1 < \alpha \leq 2$.

Este pressuposto é satisfeito nas frequências $\omega_j = \frac{2\pi j}{n}$, $j = 1, 2, ..., m$ para n suficientemente grande.

"Tapering" permite reduzir o viés do periodograma para as frequência próximas da origem se se assumir que a função de densidade espectral é suficientemente "suave" naquelas frequências (isto é, $\alpha > 1$ no pressuposto 3.1.20).

A transformada de Fourier discreta "tapered" é definida pela expressão:

$$W^T(\omega_j) = \frac{1}{\sqrt{2\pi \sum_{t=1}^{n} h_t^2}} \sum_{t=1}^{n} h_t X_t e^{i\omega_j t}$$

onde

$$h_t = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi t}{n}\right)\right)$$

e

$$\sum_{t=1}^{n} h_t^2 = \frac{3n}{8}.$$

Então, pode-se escrever a transformada de Fourier "tapered" na frequência ω_j, $2 \le j \le n-2$, como uma combinação linear das transformadas de Fourier usuais, nas frequências ω_j, ω_{j-1} e ω_{j+1},

$$W^T(\omega_j) = \frac{1}{\sqrt{6}}\left(-W(\omega_{j-1}) + 2W(\omega_j) - W(\omega_{j+1})\right).$$

Como consequência, "tapering" destrói as relações de ortogonalidade entre as transformadas de Fourier em diferentes frequências se estas estiverem demasiadamente próximas, porque as transformadas de Fourier "tapered" têm uma componente comum nas frequências de Fourier que estiverem mais próximas que duas frequências básicas. Por esta razão, só vão ser consideradas frequências ω_j e ω_k tais que $k < j - 2$, as quais se espera que não sejam correlacionadas.

Defina-se para $J = 1, 2, \ldots$, fixo (assumindo que $(m-l)/(3J)$ é um inteiro),

$$Y_k^{(T,J)} = \log\left(\bar{I}^T(\omega_k)\right), \quad k = l + 3J, l + 6J, \ldots, m$$

onde

$$\bar{I}^T(\omega_k) = \sum_{j=1}^{J} I^T\left(\omega_{k+3(j-J)}\right), \quad I^T(\omega) = \left|W^T(\omega)\right|^2$$

e

$$\hat{d}^T = \frac{\sum_k (z_k - \bar{z}) Y_k^{(T,J)}}{\sum_k (z_k - \bar{z})^2}$$

onde

$$z_k = -2\log \omega_k \,,\ \bar{z} = \frac{3J}{m-l}\sum_k z_k \,.$$

Aqui considera-se a associação de J ordenadas do periodograma "tapered" e embora cada uma das transformadas de Fourier "tapered" seja uma combinação linear das transformadas de Fourier de três frequências adjacentes, esta definição de $\hat{d}^T$ assegura a não correlação assimptótica de $\bar{I}^T(\omega_k)$ e a independência assimptótica de $Y_k^{(T,J)}$ nas diferentes frequências.

Pressuposto 3.1.21: *À medida que* $n \to \infty$, *para algum* $\alpha > 0$, $\frac{1}{l} + \frac{l(\log n)^2}{m} + \frac{m^{1+1/2\alpha}}{n} \to 0$

Teorema 3.1.9: (Velasco (2000)) *Com base nos pressupostos 3.1.17, 3.1.18, 3.1.20, 3.1.21, se* ε_t *tem momentos de todas as ordens,* $J \geq 3$ *e* $\alpha J/(J+2) > 1$ *então,*

$$\sqrt{m}\left(\hat{d}^T - d\right) \xrightarrow{d} N\left(0, \frac{3J}{4}\psi'(J)\right),$$

onde

$$\psi'(x) = \frac{d}{dx}\psi(x) = \frac{d}{dx}\log\Gamma(x)$$

é a função digama.

3.1.1.2. GSE ou LWE

O estimador semi-paramétrico Gausseano (GSE) ou estimador local de Whittle (LWE) é um estimador semi-paramétrico que só usa a forma paramétrica da função de densidade espectral perto da origem:

$$f(\omega) \sim G\omega^{1-2H} \ , \omega \to 0, \qquad \textbf{(3. 12)}$$

$G \in]0, \infty[$ e $H \in]0, 1[$, sendo H por vezes designado por parâmetro de auto-semelhança.

Este estimador também se baseia no periodograma.

Seja $\{X_t\}$ um processo estacionário e $X_1, \ldots, X_n$, n observações daquele processo.

Defina-se a transformada de Fourier discreta $w(\omega)$ e o periodograma $I(\omega)$ através das expressões

$$w(\omega) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} X_t e^{it\omega}$$

e

$$I(\omega) = |w(\omega)|^2,$$

Sendo calculadas nas frequências de Fourier $\omega_j = \frac{2\pi j}{n}$ para $j = 1, ..., m$, onde m é um inteiro menor ou igual a $0.5n$.

Sejam G_0 e H_0 os verdadeiros parâmetros, e G e H quaisquer valores admissíveis.

Para uma função de densidade espectral satisfazendo (3. 12) considere-se a função análoga à função Q do estimador de Whittle:

$$Q(G,H) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log G\omega_j^{1-2H} + \frac{I(\omega_j)}{G\omega_j^{1-2H}}\right\}$$

onde $\omega_j = \frac{2\pi j}{n}$.

Seja $\Theta = [\Delta_1, \Delta_2]$ o intervalo dos valores possíveis para H_0, onde $0 < \Delta_1 < \Delta_2 < 1$. O estimador pretendido é

$$(\hat{G}, \hat{H}) = \underset{\substack{0<G<\infty \\ H\in\Theta}}{\arg\min}\, Q(G,H).$$

ou

$$\hat{H} = \underset{H\in\Theta}{\arg\min}\, R(H)$$

onde

$$R(H) = Q(\hat{G}, H) - 1 = \log\frac{1}{m}\sum_{j=1}^{m}\frac{I(\omega_j)}{\omega_j^{1-2H}} - (2H-1)\frac{1}{m}\sum_{j=1}^{m}\log\omega_j\,,$$

sendo

$$\hat{G} = \frac{1}{m}\sum_{j=1}^{m}\frac{I(\omega_j)}{\omega_j^{1-2H}}.$$

Robinson (1995b) mostrou que este estimador é consistente e assimptóticamente normal. Na teoria assimptótica a Gausseanidade nunca é assumida.

Pressuposto 3.1.22: *À medida que* $\omega \to 0^+$, $f(\omega) \sim G_0\omega^{1-2H_0}$ *onde* $G_0 \in]0,\infty[$ *e* $H_0 \in [\Delta_1, \Delta_2]$.

Pressuposto 3.1.23: *Numa vizinhança de zero,* $]0, \delta[$, $f(\omega)$ *é diferenciável e*

$$\frac{d}{d\omega}\log f(\omega) = O(\omega^{-1}),\ \omega \to 0^+.$$

Pressuposto 3.1.24: *Tem-se que*

$$X_t - E(X_0) = \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty$$

onde $E(\varepsilon_t \mid F_{t-1}) = 0$, $E(\varepsilon_t^2 \mid F_{t-1}) = 1$, *quase certamente,* $t = 0, \pm 1, \ldots$, *sendo* F_t *o* σ*-campo dos acontecimentos gerados por* ε_s, $s \leq t$, *e existe uma variável aleatória* ε *tal* $E(\varepsilon^2) < \infty$ *e para todo o* $\eta > 0$ *e para algum* $K > 0$,

$$P(|\varepsilon_t| > \eta) \leq KP(|\varepsilon| > \eta).$$

Pressuposto 3.1.25: *À medida que* $n \to \infty$, $\frac{1}{m} + \frac{m}{n} \to 0$.

Teorema 3.1.10: (Robinson (1995b)) *Se os pressupostos de 3.1.22-3.1.25 se verificam, então* $\hat{H} \xrightarrow{p} H$ *quando* $n \to \infty$.

Pressuposto 3.1.26: *Para algum* $\beta \in]0, 2]$,

$$f(\omega) \sim G_0 \omega^{1-2H_0} \left(1 + O(\omega^{\beta})\right), \quad \omega \to 0^+$$

onde $G_0 \in]0, \infty[$ *e* $H_0 \in [\Delta_1, \Delta_2]$.

Pressuposto 3.1.27: *Numa vizinhança de zero,* $]0, \delta[$, $\alpha(\omega)$ *é diferenciável e*

$$\frac{d}{d\omega}\alpha(\omega) = O\left(\frac{|\alpha(\omega)|}{\omega}\right), \quad \omega \to 0^+.$$

Pressuposto 3.1.28: *Assuma-se que o pressuposto 3.1.24 se verifica e que* $E(\varepsilon_t^3 \mid F_{t-1}) = \mu_3$ *quase certamente,* $E(\varepsilon_t^4) = \mu_4$, $t = 0, \pm 1, \ldots$, *para constantes finitas* μ_3 *e* μ_4.

Pressuposto 3.1.29: *À medida que* $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} \to 0.$$

Teorema 3.1.11: *Se os pressupostos 3. 1. 26 –3. 1. 29 se verificam,*

$$\sqrt{m}\left(\hat{H} - H_0\right) \xrightarrow{d} N\left(0, \frac{1}{4}\right) \text{ quando } n \to \infty.$$

3.1.2. Métodos globais

Os métodos de estimação globais constróem estimadores de $f(.)$ usando todas as frequências de Fourier.

3. 1. 2. 1. Estimador FEXP

Tal como o estimador GPH, o estimador FEXP baseia-se na regressão do logaritmo do periodograma e foi originalmente proposto por Janacek (1982).

Este estimador semi-paramétrico de d é obtido através do ajustamento de um modelo FEXP ao logaritmo do periodograma. Portanto, baseia-se no ajustamento de modelos paramétricos potencialmente mal especificados.

Seja $\{X_t\}$ um processo estacionário Gausseano de memória longa com função de densidade espectral $f(\omega)$ onde $d \in]-½, ½[$ .Sejam $X_0, X_1, ..., X_{n-1}$, n observações daquele processo.

Assuma-se nesta subsecção que n é par. Vai-se assumir também que $\log(f_u(.))$ pode ser representado na forma de uma série de Fourier convergente,

$$\log f_u(\omega) = \sum_{k=0}^{\infty} \gamma_k \cos(k\omega) \text{ para } \omega \in [-\pi, \pi]$$

com

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty .$$

Também se vai assumir que $f_u(.)$ é diferenciável em $[-\pi, \pi] \setminus \{0\}$ com

$$|f_u'(\omega)| \leq \frac{C}{|\omega|} \text{ para } \omega \in [-\pi, \pi] \setminus \{0\} \qquad \textbf{(3. 13)}$$

sendo C uma constante finita.

A expressão (3. 13) constitui uma imposição global de regularidade sobre $f_u(\omega)$, a qual é verificada pelos processos FGN, ARFIMA e FEXP.

A expressão (3. 13) juntamente com as imposições feitas a f_u implicam que f é diferenciável em $[-\pi, \pi] \setminus \{0\}$ com

$$|f'(\omega)| \leq \frac{Cf(\omega)}{|\omega|} \text{ para } \omega \in [-\pi, \pi] \setminus \{0\}.$$

Seja $\omega = (\omega_1, ..., \omega_{n^*})'$ sendo $n^* = (n/2) - 1$.

$$\log I(\omega) = \log f(\omega) + \log\left(\frac{I(\omega)}{f(\omega)}\right) = -2d \log|2\sin(\omega/2)| + \log f_u(\omega) - \gamma + e \qquad \textbf{(3. 14)}$$

onde $e = (e_1, e_2, ..., e_{n^*})' = \left(\log\left(\frac{I(\omega_1)}{f(\omega_1)}\right) + \gamma, ..., \log\left(\frac{I(\omega_{n^*})}{f(\omega_{n^*})}\right) + \gamma\right)'$ e $\gamma = 0.577216...$

representa a constante de Euler.

$\{e_t\}$ é um ruído branco Gausseano com $E(e_j) = 0$ e $\text{var}(e_j) = \pi^2 / 6$.

Vai-se ajustar um modelo FEXP ao $\log(I(\omega))$. Se tal modelo se verificar exactamente então $\log f_u(\omega)$ em (3. 14) reduzir-se-á a uma combinação linear de V_0, V_1, ..., V_h onde $V_j = \cos(j\omega)$ para $j = 0, 1, ..., h$. O estimador $\hat{d}_{FEXP}$ é definido como sendo $-\frac{1}{2}$ vezes o coeficiente de $\log|2\sin(\omega/2)|$ numa regressão de mínimos quadrados do $\log(I(\omega))$ em $[V_0, V_1, ..., V_h, \log|2\sin(\omega/2)|]$. De igual modo $-2\hat{d}_{FEXP}$ é o declive numa regressão linear simples (sem intersecção) do $\log(I(\omega))$ em $\boldsymbol{r}$ onde

$$\boldsymbol{r} = \log|2\sin(\omega/2)| - \sum_{k=0}^{h} b_k V_k$$

e

$$\sum_{k=0}^{h} b_k V_k$$

é a projecção ortogonal de $\log|2\sin(\omega/2)|$ no espaço linear definido pela base $[V_0, V_1, ..., V_h]$. Assim,

$$\hat{d}_{FEXP} = (-\boldsymbol{r}' \log I(\omega)) / (2\|\boldsymbol{r}\|^2)$$

onde $\|.\|^2$ é a soma dos quadrados de um vector de dimensão n^*.

Hurvich & Brodsky (2001) demonstraram a consistência do estimador $\hat{d}_{FEXP}$ e obtiveram expressões para o viés e para a variância deste estimador.

Teorema 3.1.12: (Hurvich e Brodsky (2001)) *Se* $\frac{1}{h} + \frac{h}{n}\log^2 n \xrightarrow[n\to\infty]{} 0$,

então

$$viés(\hat{d}_{FEXP}) = \frac{h}{2}\sum_{j=h+1}^{\infty} \frac{\gamma_j}{j \bmod n} \chi\{h+1 \le j \bmod n \le n^*\} +$$

$$\frac{h}{2}\sum_{j=\frac{n}{2}+1}^{\infty} \frac{\gamma_j}{n-j \bmod n} \chi\left\{\frac{n}{2}+1 \le j \bmod n \le n-h-1\right\} + O\left(\frac{h\log^3 n}{n}\right)$$

sendo χ *a função indicatriz.*

Teorema 3.1.13: (Hurvich e Brodsky (2001)) *Defina-se as condições 1-4:*

1. $\frac{h}{n}\log^6 n \xrightarrow[n\to\infty]{} 0$
2. $hn^{\delta-1}\log^4 n \xrightarrow[n\to\infty]{} 0$
3. $hn^{-\delta}\log^3 n \xrightarrow[n\to\infty]{} 0$
4. $hn^{2d(1-\delta)-\delta}\log^4 n \xrightarrow[n\to\infty]{} 0$

Se a condição 1 se verifica e se existe um $\delta \in]0, 1[$ *fixo tal que as condições 2 e 3 se verificam simultaneamente, então*

$$\mathrm{var}\left(\hat{d}_{FEXP}\right) = O\left(\frac{h}{n}\right).$$

Se a condição 1 se verifica e se existe um $\delta \in]0, 1[$ *fixo tal que as condições 2, 3 e 4 se verificam simultaneamente, então*

$$\mathrm{var}\left(\hat{d}_{FEXP}\right) = \frac{\pi^2}{6}\frac{h}{n} + o\left(\frac{h}{n}\right) \qquad \textbf{(3.15)}$$

A equação (3. 15) verifica-se se $h = O(\log n)$. Suponha-se que $h = O(n^{\alpha})$ para algum $\alpha \in \left]0, \frac{1}{2}\right[$ fixo, então a condição 1 verifica-se e as condições 2, 3 e 4 são equivalentes a $\delta < 1-\alpha$, $\delta > \alpha$ e $\delta > (\alpha + 2d)/(1 + 2d)$, respectivamente. O teorema 3.1.12 juntamente com as soluções simultâneas das desigualdades acima descritas permitem concluir que

$$\mathrm{var}\left(\hat{d}_{FEXP}\right) = O\left(\frac{h}{n}\right)$$

para todo o $\alpha \in \left]0, \frac{1}{2}\right[$.

O resultado mais preciso de (3. 15) verifica-se para todos os $\alpha \in \left]0, \frac{1}{2}\right[$ se $d \leq 0$. A equação (3. 15) também se verifica para $0 < \alpha < \frac{1}{2d+2}$ se $d > 0$.

Suponha-se que γ_j decai exponencialmente para zero de tal forma que $\log f_u(.)$ tem um número infinito de derivadas contínuas.

Se h é proporcional a $\log n$, com uma escolha adequada da constante de proporcionalidade, então

$$EQM\left(\hat{d}_{FEXP}\right)=O\left(\frac{\log n}{n}\right).$$

Se h é proporcional a $(\log\ n)^{1+\varepsilon}$ para qualquer $\varepsilon > 0$ fixo e usando uma constante de proporcionalidade positiva qualquer,

$$EQM\left(\hat{d}_{FEXP}\right)=O\left(\frac{(\log n)^{1+\varepsilon}}{n}\right).$$

Suponha-se que

$$\gamma_j \sim j^{2d^*-1}$$

para algum $d^* \in]-\infty, 0[$. Se $d^* < -\frac{1}{4}$,

$$EQM\left(\hat{d}_{FEXP}\right)=O\left(n^{4d^*/(1-4d^*)}\right),$$

o qual é alcançado quando

$$h = O\left(n^{1/(1-4d^*)}\right).$$

A velocidade de convergência do $\mathrm{EQM}\left(\hat{d}_{FEXP}\right)$ pode tornar-se arbitrariamente próxima de $O(n^{-1})$ se d^* for suficientemente negativo, o que está de acordo com o facto de que quanto mais pequeno for d^*, mais "suave" é $f_u(.)$.

Hurvich & Brodsky (2001) deduziram uma expressão da $\mathrm{var}\left(\hat{d}_{FEXP}\right)$ que constitui uma aproximação mais exacta do que aquela que é fornecida pelo teorema 3.1.13:

$$\mathrm{var}\left(\hat{d}_{FEXP}\right) \approx \pi^2/(24\ \|\boldsymbol{r}\|^2).$$

Esta expressão resulta da igualdade

$$\hat{d}_{FEXP} = d - [\boldsymbol{r}^t \log(f^*(\omega))]\ /\ (2\ \|\boldsymbol{r}\|^2) - (\boldsymbol{r}^t\boldsymbol{e})\ /\ (2\ \|\boldsymbol{r}\|^2)$$

juntamente com a aproximação de que $e_1, \ldots, e_n$. são i. i. d. com média zero e variância $\pi^2 / 6$.

Suponhamos que γ_k decai com rapidez exponencial (pelo menos), isto é,

$$|\gamma_k| \le \frac{CA^k}{k^v}$$

onde C, v e A são constantes reais tais que $C > 0$, $v \ge 0$ e $0 < A < 1$.

Pressuposto 3.1.30: $\frac{h}{2}\sum_{k=h+1}^{\infty}\frac{\gamma_k}{k} \sim \frac{A^h}{h^v}G_h$ *quando* $h \to \infty$ *onde* $\sup_h |G_h| < \infty$ *e* $\inf_h |G_h| > 0$.

A soma anterior determina as propriedades do viés($\hat{d}_{FEXP}$).

Hurvich (2001) deduziu outra expressão para o viés de $\hat{d}_{FEXP}$.

A partir dos teoremas 3.1.12 e 3.1.13., se

$$\frac{\log\log n}{h}+\frac{h}{n}\log^6 n \xrightarrow[n\to\infty]{} 0,$$

então

$$vi\acute{e}s\left(\hat{d}_{FEXP}\right)=\frac{h}{2}\sum_{k=h+1}^{\infty}\frac{\gamma_k}{k}+O\left(A^{n^*}\right)+O\left(\frac{h\left(\log^3 n\right)}{n}\right)=\frac{A^h}{h^v}G_h\left(1+o(1)\right)+O\left(\frac{h\left(\log^3 n\right)}{n}\right)$$

e

$$EQM\left(\hat{d}_{FEXP}\right)\sim\frac{A^{2h}}{h^{2v}}G^2{}_h+\sigma^2\frac{h}{n}\quad .$$

Se h for proporcional a log(n) com constante de proporcionalidade superior a $\frac{-1}{2\log A}$,

$EQM\left(\hat{d}_{FEXP}\right)\le C^*\frac{\log n}{n}$ sendo $C^*>0$.

Por outro lado, se h for proporcional a $(\log(n))^{1+\delta}$ para qualquer $\delta>0$ fixo e se se usar uma constante de proporcionalidade positiva qualquer, então

$$EQM\left(\hat{d}_{FEXP}\right)\le C^*\frac{(\log n)^{1+\delta}}{n}.$$

É de notar que $\hat{d}_{FEXP}$ usa todas as frequências de Fourier. Contudo, agora é preciso decidir o número de co-senos, h, a usar. Para valores pequenos de h o viés é grande (porque o modelo paramétrico não tem grau de liberdade suficiente para modelar a componente de memória curta da função de densidade espectral) e a variância é pequena (porque o número de parâmetros a estimar é pequeno). Pelo contrário, valores grandes de h levam a estimadores pouco enviesados, mas com variância elevada.

O número de cosenos deve aumentar com n para que a aproximação $\log f_u(\omega)\approx\sum_{k=0}^{h}\gamma_k\cos(k\omega)$ seja suficientemente boa de forma a se obter um estimador pouco enviesado.

Hurvich (2001) propôs um critério local do critério de Mallows C_L para a selecção do número de parâmetros h do modelo exponencial ajustado.

Para qualquer inteiro positivo h defina-se

$$X_h = [V_0.\ V_1, \ldots, V_h]$$

e

$$X = [X_h, \tilde{V}]$$

onde

$$\tilde{V} = \log |2\sin(\omega/2)|.$$

Seja

$$\hat{\beta} = (X^t X)^{-1} X^t y,$$

o estimador do parâmetro

$$\beta = (\gamma_0, \gamma_1, \ldots, \gamma_h, -2d)^t$$

obtido pelo método dos mínimos quadrados.

Defina-se

$$H = X(X^t X)^{-1} X^t$$

e seja

$$\hat{\boldsymbol{y}} = H\boldsymbol{y} = X\hat{\beta}$$

sendo

$$\boldsymbol{y} = \log I(\omega) + \gamma.$$

Seja m um inteiro positivo tal que $\frac{1}{m} + \frac{m}{n} \xrightarrow[n\to\infty]{} 0$ e seja W uma matriz diagonal com as primeiras m entradas diagonais iguais à unidade e todas as outras entradas iguais a zero. Para qualquer vector $\boldsymbol{v}$ de dimensão n^*, defina-se

$$\|\boldsymbol{v}\|_W^2 = \boldsymbol{v}^t W \boldsymbol{v} = \sum_{j=1}^{m} v_j^2 .$$

O critério local C_L é definido por

$$C_L(h) = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_W^2 + 2\sigma^2 \mathrm{Tr}(WH) = \|\hat{\boldsymbol{y}} - \boldsymbol{\mu}\|_W^2 + \|\boldsymbol{e}\|_W^2 + 2\boldsymbol{\mu}^t (\mathrm{I} - H)W\boldsymbol{e} + 2\,[\sigma^2 \mathrm{Tr}(WH) - \boldsymbol{e}^t HW\,\boldsymbol{e}]$$

onde

$$\boldsymbol{\mu} = \log f(\omega).$$

Moulines & Soulier (1999) construíram uma versão mais generalizada do estimador FEXP com base em condições menos restritivas sob γ_j que aquelas que foram consideradas por Hurvich & Brodsky (2001). Este estimador FEXP generalizado permite o "pooling" do periodograma antes de se aplicar os logaritmos, sendo um estimador assimptóticamente normal.

A expressão da variância (3. 15) tal como as proporções de convergência do estimador $\hat{d}_{FEXP}$ são casos especiais dos resultados obtidos por Moulines & Soulier (1999). O "pooling" por

eles usado reduz a variância assimptótica por uma constante multiplicativa que pode ser arbitrariamente próxima de $6/\pi^2$.

3. 1. 2. 2. Estimador FEXP generalizado

Seja $\{X_t\}$ um processo estacionário com função de densidade espectral $f(.)$.

onde $f_u(.)$ é uma função positiva, contínua e periódica de período 2π.

Moulines & Soulier (1999) usaram uma abordagem um pouco diferente daquela que foi usada por Janacek (1982) para a estimação de d. Eles construíram um estimador de f usando todas as frequências. Isto é conseguido através de um estimador não paramétrico da componente de memória curta. Sob condições de regularidade apropriadas o logaritmo da componente de memória curta $l^* = \log f_u$ pode ser expandido numa série de Fourier

$$l^* = \sum_{j=0}^{\infty} \theta_j h_j \qquad \textbf{(3. 16)}$$

onde

$$h_0(\omega) = \frac{1}{\sqrt{2\pi}}$$

e

$$h_j(\omega) = \frac{\cos(j\omega)}{\sqrt{\pi}} \quad (j > 0).$$

A classe de modelos para a qual a soma (3. 16) tem um número finito de parcelas é a classe dos modelos FEXP. A expansão (3. 16) tem, em geral, um número infinito de coeficientes não nulos. Um estimador não paramétrico de f_u é obtido através da estimação de apenas um número finito de coeficientes de Fourier.

Seja m^* um inteiro fixo e defina-se para todo o n, $n_{m^*} = 2m^*[n/2m^*]$ e $K_n = [n/2m^*]$, onde $[n/2m^*]$ representa a parte inteira de $n/2m^*$. Defina-se a transformada de Fourier discreta, $W_n(\omega)$ e o periodograma $I_n(\omega)$ de $X_1, \ldots, X_n$ através das expressões:

$$W_n(\omega) = \frac{1}{\sqrt{2\pi n_{m^*}}} \sum_{t=1}^{n_{m^*}} X_t e^{it\omega}$$

e

$$I_n(\omega) = |W_n(\omega)|^2.$$

Não foram usados os últimos dados, mas isto é assimptóticamente irrelevante dado que

$$\frac{n_{m^*}}{n} \xrightarrow[n\to\infty]{} 1.$$

A transformada de Fourier e o periodograma são calculadas nas "frequências de Fourier"[9]

$\omega_s = \frac{2\pi s}{n_{m^*}}$, $1 \le s \le n_{m^*}/2$.

As frequências vão ser associadas em blocos disjuntos de m* frequências e vai ser calculada a soma das ordenadas dos periodogramas das m frequências de cada bloco. Para $k = 1, ..., K_n$ seja

$$J_k = \{m^*(k\text{-}1)+1, ...,m^*k\}$$

e

$$Y_{n,k} = \log\left(e^{-\psi(m^*)} \sum_{i\in J_k} I_n(\omega_s) \right) = \log\left(2e^{-\tau_{m^*}} \sum_{i\in J_k} I_n(\omega_i) \right)$$

onde ψ representa a função digama, isto é,

$$\psi(m^*) = \frac{\Gamma'(m^*)}{\Gamma(m^*)}$$

e

$$\tau_{m^*} = E\left(\log \sum_{i=1}^{m^*} \chi_i^2 \right)$$

sendo $(\chi_i^2)_{i=1, ..., m^*}$ uma sequência de qui-quadrados (centrados) independentes com dois graus de liberdade.

Seja

$$l(\omega) = \log f(\omega) = dg^*(\omega) + l^*(\omega)$$

onde

$$g^*(\omega) = -2\log\left|1 - e^{i\omega}\right|$$

e

$$l^*(\omega) = \log f_u(\omega) = \sum_{j=0}^{\infty} \theta_j h_j(\omega).$$

A técnica de regressão do logaritmo do periodograma baseia-se na igualdade

$$Y_{n,k} = \log f(y_k) + \varepsilon_{n,k},\ 1 \le k \le 2K_n$$

---

[9] É de notar que as frequências ω_s não são exactamente as frequências de Fourier pois $n_{m^*} \neq n$. No entanto, esta escolha de frequências é assimptóticamente indiferente porque $\frac{n_{m^*}}{n} \xrightarrow[n\to\infty]{} 1$.

sendo

$$y_k = (2k\text{-}1)\,\pi/\,2K_n$$

e

$$\varepsilon_{n,k} = Y_{n,k} - \log(f(y_k)).$$

Portanto,

$$Y_{n,k} = dg^*(y_k) + \sum_{j=0}^{h} \theta_j h_j(y_k) + l_h^*(y_k) + \varepsilon_{n,k}$$

onde

$$l_h^*(\omega) = \log(f_u(\omega)) - \sum_{j=0}^{h} \theta_j h_j(\omega)\ .$$

Seja $\hat{\Theta}_h = \left(\hat{d}_{h,n}, \hat{\theta}_0, ..., \hat{\theta}_h\right)'$ o estimador obtido pelo método dos mínimos quadrados de $\Theta_h = \left(d_{h,n}, \theta_0, ..., \theta_h\right)^t$.

É de notar que $\hat{d}_{h,n}$ e $\left(\hat{\theta}_i\right)_{i=0,\,...,h}$ dependem de h e de n.

Um estimador semi-paramétrico $\hat{l}_{h,n}$ do logaritmo do periodograma é definido pela igualdade

$$\hat{l}_{h,n} = \hat{d}_{h,n} g^* + \sum_{j=0}^{h} \hat{\theta}_j h_j\ .$$

Seja $\hat{d}_{h,n}$ o estimador obtido pelo método dos mínimos quadrados quando são estimados $h + 1$ coeficientes de Fourier, sendo definido pela igualdade

$$\hat{d}_{h,n} = \underset{\bar{d}, \bar{\theta}_0, ..., \bar{\theta}_h}{\arg\min} \sum_{k=1}^{K_n} \left(Y_{n,k} - \bar{d} g^*(y_k) - \sum_{J=0}^{h} \bar{\theta}_j h_j(y_k) \right)^2 .$$

Com base em alguns pressupostos Moulines & Soulier (1999) provaram a consistência do estimador $\hat{d}_{h,n}$ e deduziram a sua distribuição assimptótica.

Pressuposto 3.1.31: *O processo $\{X_t\}$ é Gausseano.*

Pressuposto 3.1.32: *Na função de densidade espectral f(.) de $\{X_t\}$ $d \in \left]-\frac{1}{2}, \frac{1}{2}\right[$, $f_u(.)$ é positiva e diferenciável em $[-\pi, \pi] \setminus \{0\}$ e existe uma constante finita C^* tal que*

$$\forall \omega \in [-\pi, \pi] \setminus \{0\},\ |f_u'(\omega)| \le \frac{C^*}{|\omega|}\ .$$

Pressuposto 3.1.33: *Existe um número real $\beta > 0$ e um número real finito L tal que*

$$|\theta_0| + \sum_{j=1}^{\infty} j^{\beta} |\theta_j| \leq L .$$

Defina-se

$$\alpha_j = \frac{2\sqrt{\pi}}{j}, \quad j \geq 1$$

e

$$\gamma_h = \sum_{j=h+1}^{\infty} \alpha_j^2 .$$

Então,

$$g^*(\omega) = \sum_{j=1}^{\infty} \alpha_j h_j(\omega).$$

Preposição 3.1.1: *Suponha-se que os pressupostos 3.1.31 – 3.1.33 se verificam. Então,*

$$E\left(\hat{d}_{h,n} - d\right)^2 = \frac{4\pi m^* \psi'(m^*)}{n\gamma_h} + \frac{\left(\sum_{j=h+1}^{\infty} \alpha_j \theta_j\right)^2}{\gamma_h^2} + \varepsilon_{h,n}$$

onde existe uma constante C tal que para todo o $1 \leq h \leq K_n$,

$$|\varepsilon_{h,n}| \leq Chn^{-1}\left(h^{-\beta} \log^3(n) + h \log^6(n).n^{-1}\right)$$

Com base no pressuposto 3.1.33 o viés é limitado por

$$\gamma_h^{-1} \left| \sum_{j=h+1}^{\infty} \alpha_j \theta_j \right| \leq \frac{L}{2\sqrt{\pi}} h^{-\beta} .$$

Portanto, o viés tende para zero à medida que h tende para infinito.

É de notar que o viés não é necessariamente uma função monótona não crescente de h.

A variância de $\hat{d}_{h,n}$ é proporcional a γ_h^{-1} e é assimptóticamente equivalente a $h/(4\pi)$

Portanto, se $h \equiv h_n$ é escolhido de forma que $h_n \xrightarrow[n\to\infty]{} \infty$, $\frac{h_n}{n} \xrightarrow[n\to\infty]{} 0$ e

$$\left(h_n^{-\beta} \log^3(n) + h_n \log^6(n).n^{-1}\right) \xrightarrow[n\to\infty]{} 0,$$

então $\hat{d}_{h,n}$ é um estimador consistente de h.

Sob o pressuposto 3.1.32, o menor EQM que é possível obter é da ordem de $n^{-2\beta/(2\beta+1)}$, o qual é obtido escolhendo h_n proporcional a $n^{1/(2\beta+1)}$. Na prática isto significa que a velocidade de convergência do estimador aumenta com a velocidade de decaimento para zero dos coeficientes de Fourier da componente de memória curta (ou equivalentemente, com a regularidade desta função).

Teorema 3.1.14. (Moulines & Soulier (1999)) *Seja $\{h_n\}$ uma sequência crescente de inteiros tais que*

$$\lim_{n\to\infty} h_n^3 \log^2(n) n^{-2} = 0, \quad \lim_{n\to\infty} n \log^2(n) h_n^{-1-2\beta} = 0.$$

Se o pressuposto 3.1.32 se verifica e se o pressuposto 3.1.33 também se verifica com $\beta > \frac{1}{4}$, *então*

$$\sqrt{\frac{n}{h_n}}\left(\hat{d}_{h_n,n} - d\right) \xrightarrow[d]{} N\left(0, m^* \psi'(m^*)\right),$$

$$\lim_{n\to\infty} n p_n^{-1} E\left(\hat{d}_n - d\right)^2 = m^* \psi'(m^*).$$

Como $m^* \psi'(m^*)$ decresce tendendo para um à medida que $m^* \to \infty$, m^* deve ser escolhido tão grande quanto possível. Por outro lado, para uma amostra de dimensão finita, m^* demasiadamente grande leva a um aumento do viés.

Na prática, a escolha de h_n ocorre com base na amostra. O critério geralmente usado é o critério de Mallows, C_h.

O estimador de d é quase tão eficiente, em termos da velocidade de convergência, que um estimador baseado num modelo paramétrico, mas evita as inconsistências resultantes da má especificação do modelo.

Para escolher a ordem de truncamento vai-se usar a estatística de Mallows C_h, a qual consiste em escolher empiricamente o número de parcelas da série de Fourier (3. 16) a usar da forma

$$\hat{h}_n = \arg\min_{1\le h\le K_n} \left(S_{h,n} + 4\pi \frac{p}{K_n} \sigma^2_{m^*} \right) \qquad \textbf{(3. 17)}$$

onde

$$S_{h,n} = \frac{\pi}{K_n} \sum_{k=1}^{2K_n} \left(Y_{n,k} - \hat{l}_{h,n}(y_k)\right)^2$$

é a média dos quadrados dos resíduos e

$$\sigma^2_{m^*} = \operatorname{var}\log\left(\sum_{i=1}^{m^*} \chi_i^2 \right)$$

sendo $(\chi_1^2,...,\chi_{m^*}^2)$ m^* qui-quadrados (centrados) com dois graus de liberdade.

O teorema seguinte mostra que a escolha de $\hat{h}_n$ é assimptóticamente óptima com base no critério do erro

$$R_{n,k} = \frac{\pi}{K_n} \sum_{k=1}^{2K_n} E\left(l(y_k) - \hat{l}_{h,n}(y_k)\right)^2$$

Teorema 3.1.15: (Moulines & Soulier (2000)) *Suponha-se que a função de densidade espectral f do processo* $\{X_t\}_{t\in\mathbb{Z}}$ *é da forma*

$$l(\omega) \equiv \log f(\omega) = dg^*(\omega) + l^*(\omega)$$

onde

$$g^*(\omega) = -2\log|1-e^{i\omega}|$$

e l é limitada em todo o seu domínio, excepto em zero, e além disso, é diferenciável com derivada limitada.*

Suponha-se também que o processo $\{X_t\}_{t\in\mathbb{Z}}$ *é Gausseano. Assuma-se que existe um número infinito de coeficientes não nulos* θ_j*. Então, para* $\hat{h}_n$ definido em (3. 17) $\mathbb{P}\text{-}\lim_{n\to\infty}\left(R_{\hat{h}_n,n} / \lim_{1\le h\le K_n} R_{h,n}\right) = 1$ onde $\mathbb{P}$ - lim *representa convergência em probabilidade. Além disso,* $\lim_{n\to\infty} \hat{h}_n = \infty$ *com probabilidade 1.*

Corolário 3.1.2. (Moulines & Soulier (2000)) $\hat{d}_{\hat{h}_n}$ *é um estimador consistente de d.*

3.2. Estimação paramétrica

Na abordagem paramétrica (Dahlaus (1989), Sowell (1992), Doornik & Ooms (2003)) é assumido um modelo de dimensão finita para $f_u(.)$.

3. 2. 1. Estimador de máxima verosimilhança exacto

Seja $\{X_t\}$ um processo estacionário Gausseano de memória longa com $d \in]0, ½[$, com média μ e variância σ^2.

Assuma-se também que $\{X_t\}$ é um processo linear, invertível e causal, isto é, $\{X_t\}$ pode ser representado das duas formas seguintes:

$$X_t = \sum_{s=1}^{\infty} b(s) X_{t-s} + \varepsilon_t^*$$

e

$$X_t = \sum_{s=0}^{\infty} a(s) \varepsilon_{t-s}^*$$

onde as inovações ε_t^* são variáveis aleatórias não correlacionadas com média zero e variância $\sigma_{\varepsilon^*}^2$.

Vai-se estimar simultaneamente todos os parâmetros da função de densidade espectral pelo método de máxima verosimilhança.

Considerem-se n observações $X = (X_1, ..., X_n)^t$ do processo $\{X_t\}$. Então a distribuição conjunta de X é

$$h(X, \theta^0) = (2\pi\sigma_{\varepsilon^*}^2)^{-\frac{n}{2}} |\Sigma_n(\theta^0)|^{-\frac{1}{2}} \exp\left[-\frac{(X - \mu I)' \Sigma_n^{-1}(\theta^0)(X - \mu I)}{2\sigma_{\varepsilon^*}^2} \right]$$

onde θ^0 é o vector dos parâmetros do modelo, I é um vector de uns de dimensão n e $\sigma_{\varepsilon^*}^2 \Sigma_n(\theta)$ é a matriz das covariâncias de X constituída pelas auto-covariâncias γ_{i-j}, $i, j = 1, ..., n$, de "lag" i-j.

O logaritmo da função de máxima verosimilhança é dado por

$$L_n(X, \theta^0) = \log h(X, \theta^0) = -\frac{n}{2} \log(2\pi\sigma_{\varepsilon^*}^2) - \frac{1}{2} \log|\Sigma_n(\theta^0)| - \frac{1}{2\sigma_{\varepsilon^*}^2} (X - \mu I)' \Sigma_n^{-1}(\theta^0)(X - \mu I)$$

O estimador de máxima verosimilhança de θ^0, $\hat{\theta}$, é obtido maximizando $L_n(X, \theta^0)$, o qual é solução do sistema de equações $L_n'(X, \hat{\theta}) = 0$.

Pressuposto 3.2.1: *Seja $\{X_t\}$ um processo estacionário Gausseano com média μ e função de densidade espectral $f_\theta(\omega)$, $\theta \in \Theta \subset IR^s$, onde μ e θ são parâmetros desconhecidos. Sejam μ^0 e θ^0 os verdadeiros parâmetros do processo, onde θ^0 pertence ao interior de Θ, o qual se assume que é compacto. Se $\theta \neq \theta'$, supõe-se que o conjunto $\{\omega \mid f_\theta(\omega) = f_{\theta'}(\omega)\}$ tem medida de Lebesgue positiva.*

Existe $\zeta: \Theta \rightarrow]0, 1[$ tal que para cada $\delta > 0$:

Pressuposto 3.2.2: *$f_\theta(\omega)$ é contínua para todo o (ω, θ), $\omega \neq 0$, $f_\theta^{-1}(\omega)$ é contínua para todo o (ω, θ) e $f_\theta(\omega) = O(|\omega|^{-\zeta(\theta)-\delta})$.*

Pressuposto 3.2.3: $\frac{\partial}{\partial\theta_j} f_\theta^{-1}(\omega)$, $\frac{\partial^2}{\partial\theta_j\partial\theta_k} f_\theta^{-1}(\omega)$ *e* $\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f_\theta^{-1}(\omega)$ *são contínuas para todo o (ω, θ) e*

$$\frac{\partial}{\partial\theta_j} f_\theta^{-1}(\omega) = O\left(|\omega|^{\zeta(\theta)-\delta}\right), \quad 1 \le j \le s,$$

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} f_\theta^{-1}(\omega) = O\left(|\omega|^{\zeta(\theta)-\delta}\right), \quad 1 \le j,\ k \le s,$$

$$\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f_\theta^{-1}(\omega) = O\left(|\omega|^{\zeta(\theta)-\delta}\right), \quad 1 \le j,\ k,\ l \le s.$$

Pressuposto 3.2.4: $\frac{\partial}{\partial\omega} f_\theta^{-1}(\omega)$ *e* $\frac{\partial^2}{(\partial\omega)^2} f_\theta^{-1}(\omega)$ *são contínuas para todo o (ω, θ), $\omega \neq 0$ e*

$$\left(\frac{\partial}{\partial\omega}\right)^k f_\theta^{-1}(\omega) = O\left(|\omega|^{\zeta(\theta)-k-\delta}\right)$$

para $k = 0, 1, 2$.

Pressuposto 3.2.5: *As constantes acima podem ser escolhidas independentemente de θ, mas não independentemente de δ.*

Pressuposto 3.2.6: *Assuma-se que α é contínua. Além disso existe uma constante C com*

$$|f_\theta(\omega) - f_{\theta'}(\omega)| \le C|\theta - \theta'| f_{\theta'}(\omega)$$

uniformemente para todo o ω e para todos os θ, θ' tais que $\zeta(\theta) \le \zeta(\theta')$, onde $|.|$ representa a norma Euclideana.

Teorema 3.2.1: (Dahlhaus(1989)) *Suponha-se que os pressupostos 3.2.1 - 3.2.6 se verificam e que $\hat{\mu}$ é um estimador consistente de μ^0, então*

$$\hat{\theta} \xrightarrow{p} \theta^0.$$

Teorema 3.2.2: (Dahlhaus(1989)) *Suponha-se que os pressupostos 3.2.1 - 3.2.6 se verificam e que $\hat{\mu}$ é um estimador $(1 - \zeta(\theta^0))/2$ – consistente de μ^0. Então,*

$$\sqrt{n}\left(\hat{\theta}-\theta^{0}\right)\xrightarrow{d}N\left(0,\ \Delta^{-1}\left(\theta^{0}\right)\right)$$

onde

$$\Delta(\theta)=\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\nabla\log f_{\theta}(\omega)\right)\left(\nabla\log f_{\theta}(\omega)\right)^{t}d\omega$$

sendo

$$\nabla g_{\theta}=\left(\frac{\partial}{\partial\theta_{j}}g_{\theta}\right)_{j=1,\ \ldots,\ p}.$$

Teorema 3.2.3: (Dahlhaus(1989)) *Suponha-se que os pressupostos 3.2.1 – 3.2.6 se verificam. Então $\hat{\theta}$ é um estimador eficiente de θ^{0}.*

O cálculo da função L_n ou das suas derivadas envolve a inversão da matriz Σ_n. Por razões computacionais é preferível evitar esta inversão, dado que esta operação pode ser numericamente instável. A inversão de Σ_n pode ser evitada através da decomposição da função de máxima verosimilhança de dimensão n num produto de distribuições normais unidimensionais (Pai & Ravinshanker (1996)):

$$h(X) = h_1(X_1)h_2(X_2|X_1)h_3(X_3|X_1X_2) \ldots h_n(X_n|X_1X_2 \ldots X_{n-1})$$

onde $h_j(X_j|X_1X_2 \ldots X_{j-1})$ representa a distribuição condicional de X_j dado $X_1, \ldots, X_{j-1}$. Contudo, este processo precisa de calcular as auto-correlações parciais, o que não é computacionalmente pesado para certos processos como é o caso dos ARFIMA(0, d, 0) , mas para os processos ARFIMA(p, d, q) já é.

As implementações iniciais do estimador de máxima verosimilhança exacto para os processos ARFIMA não tomaram em conta a estrutura da matriz das covariâncias, apresentando instabilidade numérica e só podiam ser usadas para pequenas dimensões amostrais.

Actualmente existem vários métodos que permitem calcular as auto-covariâncias de um processo ARFIMA estacionário evitando alguns problemas computacionais.

Seja $\{X_t^*\}$ um processo ARFIMA(p, d, q) Gausseano com média μ.

A função de auto-covariância de um processo ARFIMA estacionário com média μ é

$$\gamma_k = E\left[\left(X_t^* - \mu\right)\left(X_{t-k}^* - \mu\right)\right]$$

A matriz das covariâncias da distribuição conjunta de $X^* = \left(X_1^*, \ldots, X_n^*\right)'$ é da forma

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \cdots & \cdots & \gamma_0 \end{pmatrix} = \Sigma$$

Esta matriz é designada por matriz Toeplitz simétrica e é denotada por $\tau[\gamma_0, ..., \gamma_{n-1}]$.

Supondo que $X^* \sim N(\mu, \Sigma)$, o logaritmo de máxima verosimilhança é

$$\log L(d, \phi, \theta, \beta, \sigma_{\varepsilon^*}^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}X^{**t}\Sigma^{-1}X^{**} \qquad \textbf{(3. 18)}$$

onde $X^{**} = X^* - \mu$.

Parâmetros adicionais da regressão em μ são denotados por β, mas podem ser ignorados inicialmente.

Seja

$$R = \tau[\gamma_0 / \sigma_{\varepsilon^*}^2, ..., \gamma_{n-1} / \sigma_{\varepsilon^*}^2].$$

Na literatura existem vários métodos para calcular as auto-covariâncias de um processo ARFIMA estacionário (Hosking (1981), Ravishanker & Ray (1997), Chan & Palma (1998), Sowell (1992)).

No entanto, o método proposto por Sowell (1992) mostrou-se rápido e preciso.

Doornik & Ooms (2003) propuseram alguns refinamentos àquele método tornando-o computacionalmente mais eficiente e evitando alguns problemas de instabilidade numérica que poderiam surgir no método de Sowell.

O método proposto por Sowell (1992) com os refinamentos anteriormente referidos é descrito a seguir.

Considerem-se as auto-covariâncias de um processo ARFIMA calculadas através da expressão

$$\gamma_i = \sigma_{\varepsilon^*}^2 \sum_{j=1}^{p} \xi_j \left[\sum_{k=-q}^{q} \psi_k C(d, p+k-i, \rho_j) \right] \quad , \quad i = 0, ..., n-1, \qquad \textbf{(3. 19)}$$

proposta por Doornik & Ooms (2003), onde $\rho_1, ..., \rho_p$ são as raízes (possivelmente complexas) do polinómio auto-regressivo e

$$\psi_k = \sum_{s=|k|}^{q} \theta_s \theta_{s-|k|}$$

$$\xi_j^{-1} = \rho_j \left[\prod_{i=1}^{p} (1 - \rho_i \rho_j) \prod_{\substack{m=1 \\ m \neq j}}^{p} (\rho_j - \rho_m) \right] \qquad \textbf{(3. 20)}$$

onde $\theta_0 = 1$.

C é definida pela expressão

$$C(d, h, \rho)=\frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}\frac{(d)_h}{(1-d)_h}\left[\rho^{2p}F(d+h, 1; 1-d+h; \rho)+F(d-h, 1; 1-d-h; \rho)-1\right] \quad \textbf{(3. 21)}$$

sendo $F(a, 1; c; \rho)$ a função Hipergeométrica,

$$F(a, b; c; \rho)=\sum_{i=0}^{\infty}\frac{(a)_i(b)_i}{(c)_i}\frac{\rho^j}{i!}$$

onde

$$(a)_i = a(a + 1)(a + 2) \ldots (a + i - 1), (a)_0 = 1.$$

O cálculo de $F(a,1; c; \rho)$ pode ser feito recursivamente.

Doornik & Ooms (2003) propuseram que essa recursão fosse da frente para trás, ao contrário daquilo que foi proposto por Sowell (1992) (porque quando a recursão ocorre de trás para a frente, a função Hipergeométrica, a partir de certa altura, começa a crescer exponencialmente),

$$F(a-1, 1; c-1; \rho)=\frac{a-1}{c-1}\rho F(a, 1; c; \rho)+1$$

começando no termo final $F(d\text{-}p+q+n\text{-}1, 1; \text{-}d\text{-}p+q+n; \rho_j)$.

Na ausência do parâmetro auto-regressivo, (3. 19) reduz-se a

$$\gamma_j = \sigma_{\varepsilon^*}^2\sum_{k=-q}^{q}\psi_k\frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}\frac{(d)_{k-i}}{(1-d)_{k-i}}.$$

A proporção $\dfrac{(d)_h}{(1-d)_h}$ para $h = p\text{-}q\text{-}n+1, \ldots, 0, \ldots, p+q$ pode ser calculada usando a recursão de trás para a frente para $h > 0$:

$$(d)_h = (d + h - 1)(d)_{h-1};$$

e a recursão da frente para trás para $h < 0$:

$$(d)_h = \frac{(d)_{h-1}}{d-h}.$$

No entanto a expressão ξ_j envolve a divisão por ρ_j, o que pode criar problemas quando ρ_j está próximo de zero. Além disso, a expressão é numericamente instável para uma única raiz perto de zero.

Doornik & Ooms (2003) sugerem que se rescrevam as expressões (3. 19) e (3. 20), movendo o factor ρ_j^{-1} de ξ_j para C. Assim os dois problemas ficam resolvidos escrevendo

$$G(a; c; \rho)=\rho^{-1}\{F(a, 1; c; \rho)-1\}=\sum_{i=0}^{\infty}\frac{(a)_{i+1}}{(c)_{i+1}}\rho^i$$

A recursão da frente para trás da fórmula G é

$$G(a-1;\, c-1;\, \rho)=\frac{a-1}{c-1}\left[1+\rho G(a;\, c;\, \rho)\right]$$

e

$$\rho^{-1}\left[\rho^{2p}F(.)+F(.)-1\right]$$

de (3. 21) é substituído por

$$\rho^{2p}G(d+h;\, 1-d+h;\, \rho)+\rho^{2p-1}+G(d-h;\, 1-d-h;\, \rho).$$

Substitua-se Σ por $R\sigma_\varepsilon^2$ em (3. 18). Obtendo-se a expressão

$$\log L(d,\, \phi,\, \theta,\, \beta,\, \sigma_{\varepsilon_*}^2) \propto -\frac{1}{2}\log|R|-\frac{n}{2}\log\sigma_{\varepsilon_*}^2-\frac{1}{2\sigma_{\varepsilon_*}^2}X^{**t}R^{-1}X^{**}$$

Derivando relativamente a $\sigma_{\varepsilon_*}^2$, igualando a zero a derivada obtida e resolvendo a equação obtém-se

$$\hat{\sigma}_{\varepsilon_*}^2=n^{-1}X^{**t}R^{-1}X^{**}.$$

A função de verosimilhança concentrada é dada pela igualdade

$$l_c(d,\, \phi,\, \theta,\, \beta)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}-\frac{1}{2}\log|R|-\frac{n}{2}\log\left(n^{-1}X^{**t}R^{-1}X^{**}\right).$$

É de notar que esta função coloca $\sigma_{\varepsilon_*}^2$ fora do logaritmo da função de verosimilhança poupando um parâmetro na estimação.

Seja $\mu = X\beta$. Esta igualdade permite que β não faça parte da função de verosimilhança obtendo-se a função NPL ("normal profile log likelihood"),

$$l_P(d,\, \phi,\, \theta)=-\frac{n}{2}(1+\log(2\pi))-\frac{1}{2}\log|R|-\frac{n}{2}\log\left[n^{-1}\hat{X}^{**t}R^{-1}\hat{X}^{**}\right]$$

onde

$$\hat{X}^{**}=X^{*}-X\hat{\beta},\ \hat{\beta}=\left(X^tR^{-1}X\right)^{-1}X^tR^{-1}X^{*}.$$

Então, a função que vai ser usada no processo de maximização é

$$-\frac{1}{2}\left(n^{-1}\log|R|+\log\sigma_{\varepsilon_*}^2\right)$$

a qual pode ser maximizada usando derivadas.

Existem vários métodos para o cálculo do determinante e da inversa de R entre os quais a decomposição de Chelosky, o algoritmo de Levinson e o algoritmo de Durbin, entre outros. Estes dois últimos métodos estão descritos em Golub & Van Loan (1989).

Existem dois métodos alternativos que permitem obter estimadores pelo método de máxima verosimilhança, os os quais procuram evitar o cálculo de n auto-covariâncias para $(d,\, \phi,\, \theta)$.

Chan & Palma (1998) usam a decomposição do erro de previsão da função de verosimilhança, a qual é facilmente calculada usando as recursões do filtro de Kalman.
Ravishaker & Ray (1997) utilizaram a factorização da função de verosimilhança sugerida por Pai & Ravishanker (1996).

3. 2. 2. Estimador de Whittle

O estimador de Whittle foi proposto por Fox & Taqqu (1986).
Suponha-se que o processo $\{X_t\}$ é Gausseano. Então, a distribuição conjunta de $X = (X_1, ..., X_n)^t$ é dada pela função de máxima verosimilhança,

$$h(X;\theta) = (2\pi)^{-\frac{n}{2}} |\Sigma(\theta)|^{-\frac{1}{2}} e^{-\frac{1}{2}X^t\Sigma^{-1}(\theta)X}$$

onde $\theta=(\theta_1, ..., \theta_M)$.
O logaritmo desta função é dado por

$$L_n(X;\theta) = \log h(X;\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_n(\theta)| - \frac{1}{2}X^t\Sigma_n^{-1}(\theta)X \qquad \textbf{(3. 22)}$$

A matriz Σ_n^{-1} é difícil de calcular, mas pode ser substituída pela matriz $A(\theta)$ cujos elementos são mais fáceis de calcular, sendo

$$A(\theta) = [\alpha(j-l)]_{j,l=1,...,n}$$

onde

$$\alpha(j-l) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f(\omega;\theta)}e^{i(j-l)\omega}d\omega .$$

Usando a igualdade que foi mostrada por Grenander & Szegö (1958),

$$\lim_{n\to\infty}\frac{1}{n}\log|\Sigma_n(\theta)| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\omega,\theta)d\omega ,$$

substitui-se

$$\log|\Sigma_n(\theta)| \text{ por } \frac{n}{2\pi}\int_{-\pi}^{\pi}\log f(\omega,\theta)d\omega$$

na expressão (3. 22) e em vez de se usar a matriz Σ^{-1}, usa-se a matriz $A(\theta)$ obtendo-se a aproximação de L_n, L_n^*,

$$L_n^*(X;\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\omega;\theta)d\omega - \frac{1}{2}X^t A(\theta)X .$$

Um estimador de máxima verosimilhança de θ^0 é obtido através da minimização da função de Whittle (1953)

$$L_W(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\omega;\theta)d\omega + \frac{X' A(\theta) X}{n}$$

relativamente a θ.

Minimizar $L_w(\theta)$ é equivalente a resolver o sistema de equações

$$\frac{\partial}{\partial \theta_j} L_w(\theta)|_{\theta=\hat{\theta}} = 0$$

com $j = 1, 2, ..., M$, ou seja, é equivalente a resolver a igualdade

$$\frac{\partial}{\partial \theta_j}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\omega, \theta)\, d\omega + \frac{1}{n}\frac{\partial}{\partial \theta_j} X' A(\theta) X = 0\,, \quad j = 1, ..., M \qquad \textbf{(3. 23)}$$

Esta igualdade pode ser simplificada. Defina-se $\eta = (\theta_2, \theta_3, ..., \theta_M)$ e $\theta^* = (1, \eta)$.

Escolha-se um parâmetro escalar θ_1 tal que:

$$f(\omega, \theta) = \theta_1 f(\omega, \theta^*)$$

e

$$\int_{-\pi}^{\pi} \log f(\omega, \theta^*) d\omega = 0\,. \qquad \textbf{(3. 24)}$$

Sejam

$$\dot{A}_j(\theta) = \left(\dot{\alpha}_j(r-s, \theta)\right)_{r,\, s=1,\, ...,\, n} = \frac{\partial}{\partial \theta_j} A(\theta) \quad (j = 1, ..., M),$$

$$Q(\theta) = X' A(\theta) X = \sum_{r,\, s=1}^{n} \alpha(r-s, \theta) X_r X_s\,,$$

$$\dot{Q}_j(\theta) = X' \dot{A}_j(\theta) X \quad (j = 1, 2, ...)$$

e

$$\dot{Q}(\theta) = \left(\dot{Q}_1(\theta), ..., \dot{Q}_M(\theta)\right)'.$$

Devido a (3. 24) a equação (3. 23) simplifica-se tornando os dois passos seguintes mais simples.

1. Obtenção de $\hat{\eta}$ através da resolução do sistema de equações

$$\dot{Q}(\hat{\eta}) = 0 \qquad \textbf{(3. 25)}$$

2. Estabelecer que a igualdade $\hat{\sigma}_\varepsilon^2 = 2\pi\theta_1$ é equivalente a

$$\hat{\sigma}_\varepsilon^2 = 2\pi Q^*(\hat{\eta}) \qquad \textbf{(3. 26)}$$

sendo σ_ε^2 o erro de previsão óptimo a um passo.

O teorema seguinte é um resumo dos resultados de Fox & Taqqu (1986), Beran (1986) e Dahlhaus (1989).

Teorema 3.2.4: *Seja $\{X_t\}$ um processo linear, estacionário e Gausseano de memória longa, invertível e causal e seja $\hat{\theta}$ o estimador definido por (3.25) e (3. 26). Então*

$$\hat{\theta} \to \theta^0$$

quase certamente, e

$$\sqrt{n}\left(\hat{\theta} - \theta^0\right) \xrightarrow{d} \varphi$$

onde φ é um vector aleatório normal de dimensão K com média zero e matriz de covariâncias C, onde

$$C(\theta^0) = 2D^{-1}(\theta^0),$$

sendo $[D_{ij}]_{i,j=1,\ldots,K}$ definido pela expressão

$$D_{ij}\left(\theta^0\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i}\log f(\omega)\frac{\partial}{\partial\theta_j}\log f(\omega)d\omega\,|_{\theta=\theta^0}.$$

3.2.3. Estimador dos parâmetros do processo ARFIMA com inovações SαS

Seja $\{Z_t\}$ um processo ARFIMA $S\alpha S$ onde a_t são as inovações $S\alpha S$ e $d > 0$.

Kokoszka & Taqqu (1996) estimaram os parâmetros daquele processo usando uma versão do método de Whittle.

Assuma-se que as inovações são i.i.d. com média zero e estão no domínio de atracção de uma lei α-estável com $1 < \alpha < 2$, isto é,

$$P\left(|a_t| > x\right) = x^{-\alpha}L(x)$$

à medida que $x \to \infty$, onde L é uma função de variação lenta e

$$\frac{P(a_t > x)}{P\left(|a_t| > x\right)} \to a, \quad \frac{P(a_t < -x)}{P\left(|a_t| > x\right)} \to b,$$

onde $a \geq 0$, $b \geq 0$ e $a + b = 1$.

Para as inovações a_t descritas existe uma única representação de $\{Z_t\}$ em médias móveis

$$Z_t = \sum_{j=0}^{\infty} c_j a_{t-j}$$

desde que Φ_p e Θ_q não tenham zeros no círculo unitário, não tenham zeros em comum e $d < 1 - \frac{1}{\alpha}$. Os coeficientes c_j são definidos pela expressão

$$\sum_{j=0}^{\infty} c_j z^j = \frac{\Theta_q(z)}{\Phi_p(z)(1-z)^d}, \quad |z| < 1$$

e são assimptóticamente proporcionais a j^{d-1} quando $j \to \infty$.

Pretende-se estimar o vector $\beta_0 = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, d)$ de dimensão $p + q + 1$ onde $\phi_1, ..., \phi_p$ e $\theta_1, ..., \theta_q$ são os coeficientes dos polinómios $\Phi_p(z)$ e $\Theta_q(z)$.

Defina-se o conjunto E como sendo o compacto contido no conjunto

$\{(\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, d) : \phi_p \neq 0, \theta_q \neq 0, \Phi_p(z)$ e $\Theta_q(z)$ não têm zero comuns, $\Phi_p(z)\,\Theta_q(z) \neq 0$ para $|z| \leq 1, d \in \left]0, -\frac{1}{\alpha}\right[\}$.

Os elementos de E são designados por β e o verdadeiro valor dos parâmetros é denotado por β_0.

Sejam,

$$\gamma_n(h) = \frac{1}{n}\sum_{t=1}^{n-|h|} Z_t Z_{t+|h|}$$

a função de auto-covariância amostral e

$$\rho_n(h) = \frac{\sum_{t=1}^{n-|h|} Z_t Z_{t+|h|}}{\sum_{t=1}^{n} Z_t^2} = \frac{\gamma_n(h)}{\gamma_n(0)}$$

a função de auto-correlação amostral.

O periodograma normalizado é definido através da expressão

$$\tilde{I}_n(\omega) = \frac{\left|\sum_{t=1}^{n} Z_t e^{-i\omega t}\right|^2}{\sum_{t=1}^{n} Z_t^2} = \sum_{|h|<n} \rho_n(h) e^{-i\omega h}, \quad -\pi \leq \omega \leq \pi$$

Para $\beta \in E$ defina-se a função

$$g(\omega, \beta) = \left|\frac{\Theta_q(e^{-i\omega}, \beta)}{\Phi_p(e^{-i\omega}, \beta)(1-e^{-i\omega})^{d(\beta)}}\right|^2 = \left|\sum_{j=0}^{\infty} c_j(\beta) e^{-i\omega j}\right|^2.$$

Seguindo os trabalhos de Fox & Taqqu (1986) e Mikosch et al. (1995), Kokoszka & Taqqu (1996) definiram o estimador β_n baseado na amostra $X_1, ..., X_n$, como sendo o valor de β que minimiza

$$\sigma_n^2(\beta) = \int_{-\pi}^{\pi} \frac{\tilde{I}_n(\omega)}{g(\omega, \beta)} d\omega, \quad \beta \in E.$$

Os teoremas seguintes mostram a consistência e a distribuição assimptótica deste estimador.

Teorema 3.2.5: (Kokoszka & Taqqu (1996)) *Se β_0 é o verdadeiro parâmetro e β_n é o verdadeiro valor de β que minimiza $\sigma_n^2(\beta)$, então*

$$\beta_n \xrightarrow{p} \beta_0 \quad e \quad \sigma_n^2(\beta_0) \xrightarrow{p} \frac{2\pi}{\gamma(0)}$$

Suponha-se que a_t é simétrico e está no domínio de atracção normal de uma lei $S\alpha S$, isto é,

$$n^{-\frac{1}{\alpha}} \sum_{j=1}^{n} a_j \xrightarrow{d} Y$$

onde

$$E\left(e^{i\theta Y}\right) = e^{-\sigma^\alpha |\theta|^\alpha}.$$

Teorema 3.2.6:. (Kokoszka & Taqqu (1996)) *Se as inovações a_t são simétricas e satisfazem*

$$\lim_{\omega \to \infty} \omega^\alpha P\left(|a_t| > \omega\right) = C_\alpha \sigma^\alpha,$$

Então

$$\left(\frac{n}{\log n}\right)^{\frac{1}{\alpha}} (\beta_n - \beta_0) \xrightarrow{d} 4\pi W^{-1}(\beta_0) \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} b_k$$

onde Y_0 é $(\alpha/2)$-estável positivo com parâmetro de escala $C_{\alpha/2}^{-2/\alpha}$ e Y_k, $k \geq 1$ são i.i.d. $S\alpha S$ com parâmetro de escala $C_\alpha^{-1/\alpha}$. As variáveis aleatórias Y_0, Y_1, ... são independentes,

$$C_\alpha = \left(\alpha \int_0^\infty (1 - \cos x) \frac{dx}{x^{\alpha+1}}\right)^{-1} = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & se \quad \alpha \neq 1 \\ \dfrac{2}{\pi} & se \quad \alpha = 1 \end{cases}$$

$W(\beta_0)$ é a matriz $(p+q+1)\times(p+q+1)$ com entradas

$$w_{ij} = \int_{-\pi}^{\pi} g(\omega, \beta_0) \frac{\partial^2}{\partial\beta_i \partial\beta_j} g^{-1}(\omega, \beta_0) d\omega \qquad i, j = 1, ..., p+q+1$$

E b_k é um vector de dimensão $p + q + 1$ cuja j-ésima coordenada é

$$(b_k)_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} g(\omega, \beta_0) \frac{\partial}{\partial\beta_j} g^{-1}(\omega, \beta_0) d\omega. \qquad i, j = 1, ..., p+q+1$$

Resumindo, o primeiro estimador semi-paramétrico local obtido a partir da regressão do logaritmo do periodograma foi obtido por Geweke & Porter-Hudak (1983). Reisen (1994) obteve o estimador de d através da regressão do logaritmo do periodograma amaciado, o qual foi obtido usando a janela de Parzen, tendo deduzido a sua variância assimptótica para $d < 0$.

O estimador de *d* proposto por Geweke & Porter-Hudak (1983) foi justificado teoricamente para séries Gausseanas estacionárias e invertíveis por Robinson (1995a) e Hurvich et al. (1998). Foi generalizado para séries lineares (possivelmente não Gausseanas) estacionárias e invertíveis por Velasco (2000). Velasco (2000) usou o "pooling" e o "tapering" para a não Gausseanidade.

Robinson (1995a) mostrou que, no caso de séries temporais Gausseanas, estacionárias e invertíveis, aplicando uma quantidade suficientemente grande de "pooling" a um estimador do tipo GPH obtém-se melhorias em termos de eficiência do estimador. O GSE ("Gaussian semi-parametric estimator") foi justificado para processos lineares estacionários e invertíveis por Robinson (1995b).

Shimotsu & Phillips (2002) construíram um estimador que se baseia na regressão do logaritmo do periodograma usando um maior número de frequências de Fourier que no caso do estimador de Geweke & Porter-Hudak (1983), apresentando uma variância assimptótica inferior relativamente à variância que este estimador exibe, tendo sido justificado teoricamente para séries estacionárias Gausseanas . No entanto, o estimador de Shimotsu & Phillips (2002) tem um viés superior.

Andrews & Guggenberger (2003) construíram um estimador que permite reduzir o viés do estimador de Geweke & Porter-Hudak (1983), tendo sido justificado teoricamente para processos estacionários Gausseanos.

O estimador FEXP é um estimador semi-paramétrico global, tendo sido originalmente proposto por Janacek (1982) e justificado teoricamente para séries Gausseanas estacionárias e invertíveis por Moulines & Soullier (1999) que também usaram o "pooling". Hurvich & Brodsky (2001) derivaram o viés, a variância e o EQM assimptóticos do estimador FEXP não "pooled".

Doornik & Ooms (2003) implementaram a estimação dos processos ARFIMA estacionários pelo método de máxima verosimilhança exacto tendo em conta a estrutura da matriz das covariâncias propondo alguns refinamentos ao método proposto por Sowell (1992) para o cálculo das auto-covariâncias tornando-o computacionalmente mais eficiente e evitando alguns problemas de instabilidade numérica que poderiam surgir no método de Sowell.

Kokoszka & Taqqu (1996) propuseram um estimador paramétrico baseado numa versão do método de Whittle, o qual é constituído pelos parâmetros de um processo ARFIMA com inovações α-estáveis, tendo sido mostrada a consistência assimptótica deste estimador.

4. Aplicação de alguns métodos de estimação a uma série temporal hidrológica

Neste capítulo vão ser aplicados diversos métodos de estimação a uma série hidrológica e vão ser discutidos os diversos estimadores mediante os resultados obtidos.

Na secção 4. 1. vai-se analisar a série em estudo, nas secções 4. 2. a 4. 5. vão ser usados os estimadores de Geweke e Porter-Hudak, Reisen, Robinson e Velasco, respectivamente.

Na secção 4. 6. vai-se discutir a modelação da série em estudo pelo método de máxima verosimilhança e por fim, na secção 4. 7. fazem-se os comentários finais.

4.1. Análise da série em estudo

A série em estudo é uma série hidrológica a qual corresponde às medições diárias do fluxo médio (m^3/s) do rio Shiel, Reino Unido, no período 1 de Agosto de 1995 a 31 de Dezembro de 2002 sendo constituída por 2710 observações. Os dados desta série foram recolhidos na estação de medição Shielfoot e podem ser obtidos no endereço **http://www.nerc-wallingford.ac.uk/ih/nrfa/webdata/092001/g.html** .

O fluxo é calculado através dos dados obtidos para o nível do rio, dado que a sua medição directa é difícil e cara.

Os dados relativos ao nível do rio são recolhidos na estação de medição Shielfoot usando uma variedade de sensores que medem o nível da água. Além disso, também são usados sensores de pressão que traduzem a pressão da água em nível. Flutuações rápidas no nível da água devido às ondas que ocorrem na superfície não são tidas em conta devido à colocação do sensor do nível da água num suporte adequado.

A figura 4. 1 apresenta a série em estudo.

A figura 4. 2 apresenta a auto-correlação amostral e a auto-correlação parcial amostral da série, enquanto que a figura 4. 3 apresenta o periodograma da série referida.

Como se pode observar através da figura 4. 2, verifica-se que a série apresenta sazonalidade anual, a qual vai ser retirada por dois processos. No primeiro processo o factor sazonal para cada mês é calculado como a diferença entre a média de todos os valores medidos para esse mês e a média de todas as observações. O segundo processo usa o método STL ("seasonal-trend decomposition procedure based on loess") proposto por Cleveland et al. (1990), o qual é

uma função do S-Plus. Este método decompõe a série temporal em componentes de variação de frequência, através de uma sequência de "suavizadores" loess.

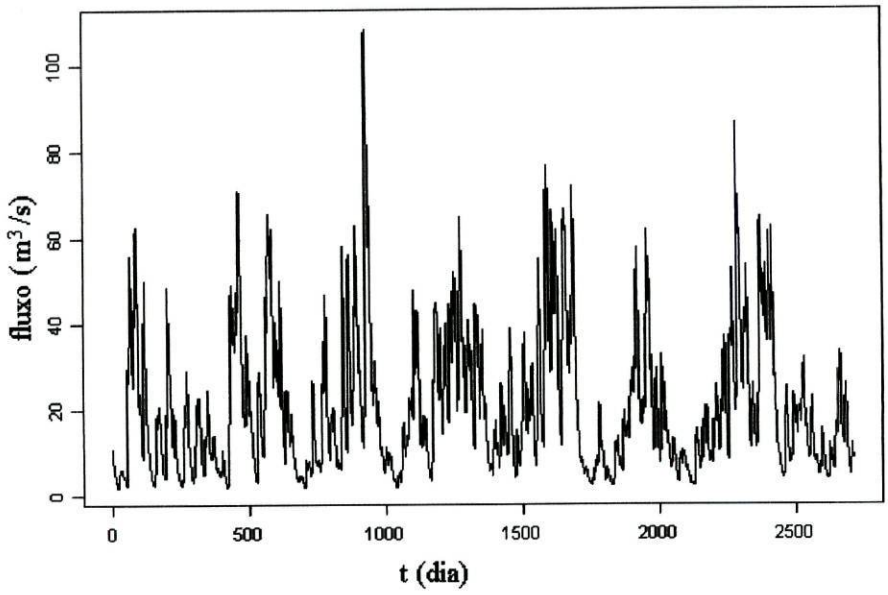


Figura 4.1: Série temporal representativa do fluxo médio diário no rio Shiel entre 1 de Agosto de 1995 e 31 de Dezembro de 2002.

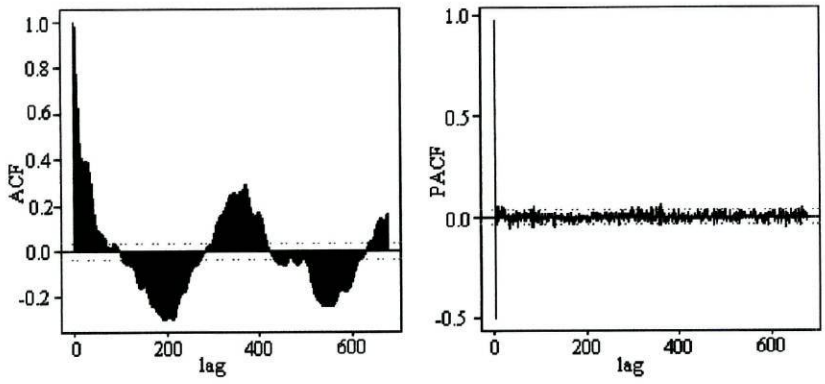


Figura 4. 2: Função de auto-correlação amostral e de auto-correlação parcial amostral da série representada na figura anterior.

A figura 4.4 apresenta do lado esquerdo a série representativa da sazonalidade retirada à série original, pelo primeiro método e do lado direito pelo método STL.

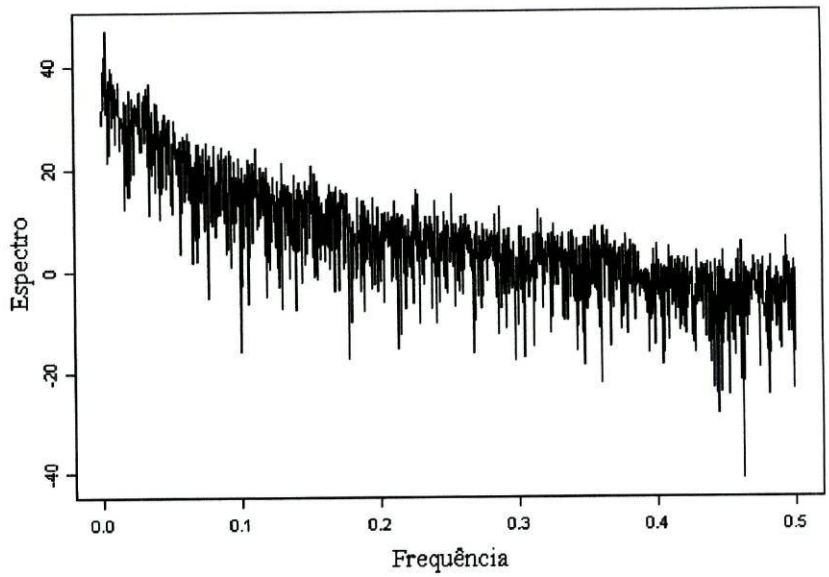


Figura 4. 3: Função de densidade espectral estimada da série representada na figura 4. 1.

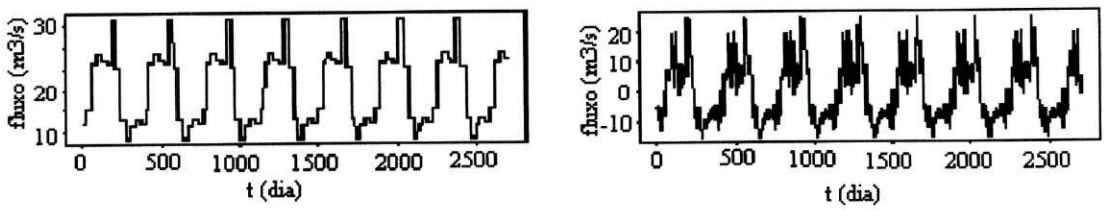


Figura 4. 4: Séries representativas da sazonalidade retirada pelos dois processos referidos.

Na figura 4. 5 está representada a série em estudo após lhe ter sido retirada a sazonalidade pelos dois métodos atrás referidos, estando do lado esquerdo a série cuja sazonalidade foi retirada pelo primeiro processo e do lado direito a série cuja sazonalidade foi retirada pelo método STL. Como se pode observar, as duas séries obtidas têm um comportamento muito semelhante, embora a série que se encontra representada do lado esquerdo parece estar centrada em zero, enquanto que a série do lado direito aparenta estar centrada em 20. Na figura 4. 6, do lado direito, em cima está representada a função de auto-correlação amostral da série obtida pelo método STL e em baixo está representada a função de auto-correlação parcial amostral da mesma série. Do lado esquerdo, estão representadas as funções de auto-correlação amostral e de auto-correlação parcial amostral da série sem sazonalidade obtida pelo primeiro método referido, estando em cima a primeira função e em baixo a segunda.

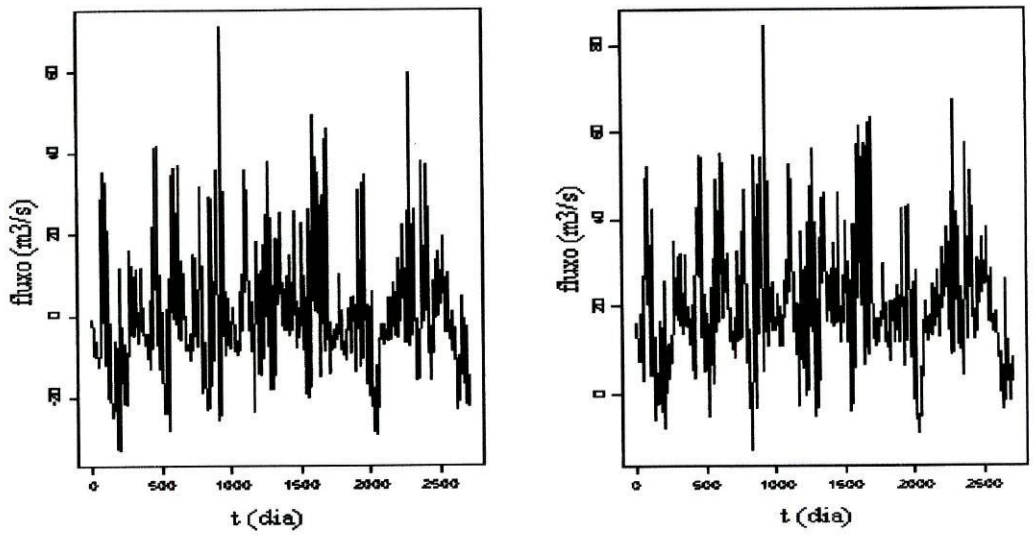


Figura 4.5: Do lado direito está representada a série relativa às medições diárias no rio Shiel após lhe ter sido retirada a sazonalidade pelo método STL e do lado esquerdo está representada a mesma série após lhe ter sido retirada a sazonalidade pelo primeiro processo referido

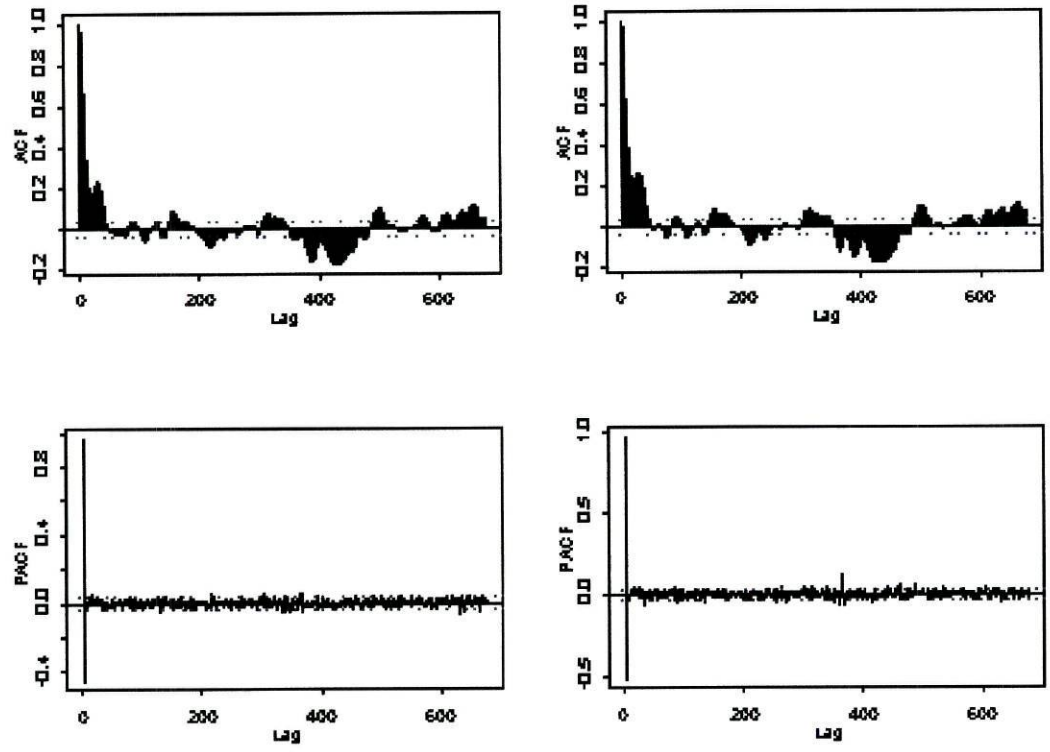


Figura 4. 6: Funções amostrais de auto-correlação e de auto-correlação parcial relativas às séries sem sazonalidade.

Como se pode observar, na figura 4. 6 as funções amostrais de auto-correlação e de auto-correlação parcial são muito semelhantes para aquelas duas séries. Apesar das duas funções de auto-

-correlação apresentarem um comportamento oscilatório, já não apresentam sazonalidade.

No entanto, a média da série obtida pelo primeiro método referido é 5.35055×10^{-6}, enquanto que a média da série obtida pelo método STL é 20.2218. Portanto, no estudo que se segue vai-se usar a série que apresenta a média mais próxima de zero.

4.2. Estimador de Geweke e Porter-Hudak

Seja m o número de observações usadas na regressão para a obtenção do estimador proposto por Geweke e Porter-Hudak, $\hat{d}_{GPH}$.

Como a variância, o viés e o EQM do estimador $\hat{d}_{GPH}$ são fortemente influenciados pelo número de observações usadas na regressão (3. 2), m, foram obtidos diferentes estimativas para o d em função de m (ver tabela 4. 1).

Recorde-se que m deve ser função da dimensão da amostra e que à medida que $m \to \infty$ e $n \to \infty$, $\frac{m}{n} \to 0$. Neste estudo usa-se, $m = n^{\alpha}$, $0 < \alpha < 1$, tendo-se substituído α pelos valores 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 e 0.9.

Os resultados apresentados na tabela 4. 1 foram conseguidos através da aplicação de um programa escrito em Matlab (programa B1 do apêndice B) à série sem sazonalidade.

Como se pode observar na tabela 4. 1 as estimativas de d são muito influenciadas por m, verificando-se $\hat{d}_{GPH}$ a variar entre 0.1141 e 0.9767, ou seja, para certos valores de m obtém-se um $\hat{d}_{GPH}$ que aparentemente caracteriza a série em estudo como sendo uma série de memória longa estacionária, enquanto que para outros valores de m obtém-se um indicativo de que a série em estudo poderá ser não estacionária. No entanto, é preciso ser-se prudente na interpretação dos resultados obtidos.

Como se pode observar através do teorema 3. 1. 1., à medida que m aumenta, a variância diminui, enquanto que o viés aumenta. A diminuição da variância teórica também pode ser observada na tabela 4.1. Aqueles valores foram calculados usando a expressão da variância do teorema 3. 1. 1., negligenciando o último termo. Intuitivamente é fácil compreender porque é que ocorre uma diminuição da variância e um aumento do viés à medida que m aumenta.

Quanto maior for o número de observações usadas na regressão menos influência têm as observações extremas e consequentemente ocorre uma diminuição da variância.

Quando se usa um número de observações muito grande como é o caso de $m = 558$ ou de $m = 1229$, a parcela $\log\frac{f_u(\omega_j)}{f_u(0)}$ da regressão (3. 1) já poderá não ser desprezável relativamente às outras parcelas devido ao facto de certas frequências usadas já estarem consideravelmente distantes de zero. Portanto, nesta situação, a eliminação daquela parcela da expressão (3. 1) poderá contribuir fortemente para o aumento do viés.

m	$\hat{d}_{GPH}$	Var($\hat{d}_{GPH}$)
$m = 2710^{0.3} = 11$	0.4072	0.03738
$m = 2710^{0.4} = 24$	0.2180	0.01713
$m = 2710^{0.5} = 52$	0.1141	0.00791
$m = 2710^{0.6} = 115$	0.2905	0.00358
$m = 2710^{0.7} = 253$	0.3908	0.00163
$m = 2710^{0.8} = 558$	0.6971	0.00074
$m = 2710^{0.9} = 1229$	0.9767	0.00033

Tabela 4. 1: Valores estimados de *d* obtidos a partir das observações sem sazonalidade usando o estimador de Geweke e Porter-Hudak (1983).

Nestas circunstâncias, os valores de $\hat{d}_{GPH}$ de 0.6971 e 0.9767 podem não ser valores de confiança pois o estimador usado nestes casos apresenta um viés elevado. Contudo, para valores de m grandes a variância diminui. Portanto, é preciso ter em atenção o compromisso variância-viés e analisar se ocorre uma diminuição considerável da variância para $m = 558$ e $m = 1229$ relativamente a outros valores de m.

Por outro lado, quando se usa um número de observações demasiadamente pequeno, a variância é consideravelmente superior, apesar do viés ser pequeno devido ao facto do termo que é eliminado da igualdade (3. 1) ser desprezável.

Para $m = 11$ obteve-se um valor de $\hat{d}_{GPH} = 0.4072$. É de notar que naquela situação o estimador usado apresenta (teoricamente) uma grande variabilidade.

Calculando a variância teórica para cada valor de m e testando-se a nulidade dos valores obtidos para d, constata-se que para $m = 24$ e para $m = 52$ as estimativas para d não são estatisticamente significativamente diferentes de zero.

Como não se conhece o modelo que se ajusta à série em estudo, não é possível calcular o m^{opt}, ou seja o m que permitiria obter o $\hat{d}_{GPH}$ com o menor EQM possível.

4.3. Estimador de Reisen

O estimador de Reisen é obtido de forma semelhante ao estimador anterior. A diferença consiste no uso da janela de Parzen e no consequente cálculo do periodograma amaciado.

Os resultados da tabela seguinte foram obtidos através da aplicação de um programa escrito em Matlab (programa B2 do apêndice B) aos dados da série sem sazonalidade em estudo.

Na tabela 4.2. estão representados os diferentes valores de $\hat{d}_{sp}$ em função do número de observações usadas na regressão, m, e da dimensão da janela de Parzen, M e a variância teórica para $\hat{d}_{sp}$.

Tal como Reisen (1994) vai-se usar (3. 4) para $d \in]-½, ½[$, apesar daquela expressão ser válida apenas para $d < 0$. Com base em (3. 4) obtém-se a variância teórica de $\hat{d}_{sp}$ para diferentes valores de m e M.

Como se pode observar através da tabela 4. 2, à medida que M aumenta, a variância teórica de $\hat{d}_{sp}$ também aumenta e à medida que m aumenta, a variância teórica diminui.

Tal como no caso do estimador anterior, para valores elevados de m verificam-se valores de $\hat{d}_{sp}$ elevados indicando a possibilidade da série em estudo não ser estacionária. No entanto, como já foi referido antes, os resultados obtidos podem estar relacionados com o facto do estimador usado poder apresentar um aumento no viés, para valores de m grandes, por se ter desprezado a parcela $\log \frac{f_u(\omega_j)}{f_u(0)}$ da regressão do logaritmo do periodograma.

Para $m = 11$ verifica-se que para quase todos os valores de M usados se obtém $\hat{d}_{sp}$ negativo, excepto na janela de dimensão 1229, que se obtém $\hat{d}_{sp} = 0.5213$. Em todos os restantes casos obtém-se $\hat{d}_{sp}$ positivo.

Todos os valores de d negativos são estatisticamente significativamente diferentes de zero, excepto para $m = 11$ e $M = 253$.

Para $m = 24$ e $M = 253$ e $M = 558$ as estimativas de d obtidas através do estimador de Reisen não são estatisticamente significativamente diferentes de zero.

À medida que m aumenta, para cada M usado verifica-se um aumento de $\hat{d}_{sp}$. Isto não aconteceu quando se usou o estimador de GPH.

M / m	115	253	558	1229
11	$\hat{d}_{sp}$ = -0.0769 var($\hat{d}_{sp}$) = 0.0010	$\hat{d}_{sp}$ = -0.0799 var($\hat{d}_{sp}$) = 0.0023	$\hat{d}_{sp}$ = -0.2362 var($\hat{d}_{sp}$) = 0.0050	$\hat{d}_{sp}$ = 0.5213 var($\hat{d}_{sp}$) = 0.0110
24	$\hat{d}_{sp}$ = 0.0620 var($\hat{d}_{sp}$) = 0.000361	$\hat{d}_{sp}$ = 0.0529 var($\hat{d}_{sp}$) = 0.000794	$\hat{d}_{sp}$ = 0.0371 var($\hat{d}_{sp}$) = 0.0018	$\hat{d}_{sp}$ = 0.199 var($\hat{d}_{sp}$) = 0.0039
52	$\hat{d}_{sp}$ = 0.2447 var($\hat{d}_{sp}$) = 0.000141	$\hat{d}_{sp}$ = 0.2410 var($\hat{d}_{sp}$) = 0.000311	$\hat{d}_{sp}$ = 0.1899 var($\hat{d}_{sp}$) = 0.000685	$\hat{d}_{sp}$ = 0.1004 var($\hat{d}_{sp}$) = 0.0015
115	$\hat{d}_{sp}$ = 0.3680 var($\hat{d}_{sp}$) = 0.0000576	$\hat{d}_{sp}$ = 0.3774 var($\hat{d}_{sp}$) = 0.000127	$\hat{d}_{sp}$ = 0.3999 var($\hat{d}_{sp}$) = 0.000279	$\hat{d}_{sp}$ = 0.2803 var($\hat{d}_{sp}$) = 0.000615
253	$\hat{d}_{sp}$ = 0.5417 var($\hat{d}_{sp}$) = 0.0000247	$\hat{d}_{sp}$ = 0.5501 var($\hat{d}_{sp}$) = 0.0000542	$\hat{d}_{sp}$ = 0.5457 var($\hat{d}_{sp}$) = 0.000120	$\hat{d}_{sp}$ = 0.3060 var($\hat{d}_{sp}$) = 0.000263
558	$\hat{d}_{sp}$ = 0.7664 var($\hat{d}_{sp}$) = 0.0000109	$\hat{d}_{sp}$ = 0.7721 var($\hat{d}_{sp}$) = 0.0000240	$\hat{d}_{sp}$ = 0.7681 var($\hat{d}_{sp}$) = 0.0000530	$\hat{d}_{sp}$ = 0.6735 var($\hat{d}_{sp}$) = 0.000117
1229	$\hat{d}_{sp}$ = 1.0397 var($\hat{d}_{sp}$) = 0.00000512	$\hat{d}_{sp}$ = 1.0453 var($\hat{d}_{sp}$) = 0.0000127	$\hat{d}_{sp}$ = 1.0540 var($\hat{d}_{sp}$) = 0.0000248	$\hat{d}_{sp}$ = 0.9598 var($\hat{d}_{sp}$) = 0.0000547

Tabela 4. 2: Valores estimados do parâmetro d usando o estimador de Reisen em função do número de observações usadas na regressão, m, e em função do tamanho da janela de Parzen, M e a variância teórica de $\hat{d}_{sp}$ para cada caso.

As estimativas para d apresentam uma maior amplitude de valores em função de m para as diferentes dimensões da janela de Parzen usadas, relativamente às estimativas de d quando se usou o estimador $\hat{d}_{GPH}$.

Apenas quando se usa a janela de Parzen de dimensão M = 1229 é que a amplitude das estimativas obtidas diminui obtendo-se estimativas análogas às obtidas quando se usou o estimador de GPH, para os mesmos valores de m.

Isto deve-se ao facto de que à medida que a dimensão da janela vai aumentando, o amaciamento vai diminuindo.

À medida que M aumenta (para o mesmo valor de m) verifica-se que a variância teórica também aumenta, sendo no entanto, sempre inferior à variância teórica do estimador $\hat{d}_{GPH}$ para os mesmos valores de m.

Reisen (1994) constatou através de um estudo simulado que à medida que M aumenta verifica-se uma redução do viés.

Estes resultados estão de acordo com o facto de que um aumento de M conduz a uma diminuição do viés e a um aumento da variância do estimador espectral.

Portanto, é preciso ter em atenção a relação viés-variância na interpretação dos resultados obtidos para os diferentes valores de M.

4.4. Estimador de Robinson

Nesta secção vai-se aplicar o estimador de Robinson à série em estudo. Este estimador baseia-se na regressão do logaritmo da soma de ordenadas do periodograma correspondentes a frequências de Fourier adjacentes.

Velasco (2000) demonstrou a consistência assimptótica deste estimador para $J \geq 3$ para séries temporais lineares (não necessariamente Gausseanas).

Este estimador pode ser aplicado de uma forma mais abrangente dado que ele continua a ser assimptóticamente "bem comportado" mesmo que a série em estudo não seja Gausseana.

A tabela 4.3 apresenta as diferentes estimativas obtidas através do estimador $\hat{d}_{Robinson}$ para diferentes valores de m, J e l. Os resultados apresentados naquela tabela foram obtidos através da aplicação do estimador $\hat{d}_{Robinson}$ à série em estudo usando-se o programa B3 escrito em Matlab que se encontra no apêndice B. É de notar que m deve ser escolhido de tal forma que $\frac{m-l}{J}$ seja inteiro. Se tal não acontecer o programa procura um $m1$ que seja o maior valor inteiro menor que m para o qual $\frac{m1-l}{J}$ é inteiro.

Para $l = 0$ e $J = 1$, $\hat{d}_{Robinson} = \hat{d}_{GPH}$ e como se pode constatar os resultados obtidos são exactamente iguais aos obtidos na tabela 4. 1 para $m = 11$, $m = 24$ e $m = 52$. Para os restantes

	$l=0$		$l=1$		$l=2$		$l=3$	
	m	$\hat{d}_{Robinson}$	m	$\hat{d}_{Robinson}$	m	$\hat{d}_{Robinson}$	m	$\hat{d}_{Robinson}$
$J=1$	11	0.4072	11	0.4590	11	0.4316	11	0.6931
	24	0.2180	24	0.2144	24	0.1734	24	0.2361
	52	0.1141	52	0.1044	52	0.0776	52	0.0987
	115	0.2904	115	0.2835	115	0.2686	115	0.2769
	253	0.3904	253	0.3851	253	0.3759	253	0.3784
	558	0.6936	558	0.6898	558	0.6841	558	0.6842
	1229	0.9515	1229	0.9482	1229	0.9443	1229	0.9434
$J=2$	11	0.1763	11	0.4726	11	0.1529	11	0.8536
	24	0.1598	24	0.2247	24	0.1261	24	0.2833
	52	0.1187	52	0.1206	52	0.0947	52	0.1288
	115	0.2852	115	0.2977	115	0.2701	115	0.2974
	253	0.4171	253	0.4107	253	0.4066	253	0.4072
	558	0.7085	558	0.7150	558	0.7010	558	0.7109
	1229	0.9509	1229	0.9651	1229	0.9452	1229	0.9610
$J=3$	11	-0.0539	11	0.3364	11	0.7421	11	0.4122
	24	0.1279	24	0.2556	24	0.2660	24	0.1798
	52	0.1239	52	0.1309	52	0.1447	52	0.1336
	115	0.2947	115	0.2833	115	0.3027	115	0.2933
	253	0.4265	253	0.4084	253	0.4249	253	0.4211
	558	0.7300	558	0.7241	558	0.7219	558	0.7240
	1229	0.9731	1229	0.9635	1229	0.9606	1229	0.9671
$J=4$	11	-0.4517	11	0.1697	11	0.3831	11	1.4427
	24	0.1081	24	0.2746	24	0.2820	24	0.3139
	52	0.1217	52	0.2135	52	0.1718	52	0.1227
	115	0.2989	115	0.3190	115	0.3099	115	0.2816
	253	0.4303	253	0.4468	253	0.4356	253	0.4065
	558	0.7360	558	0.7468	558	0.7012	558	0.7254
	1229	0.9738	1229	0.9848	1229	0.9732	1229	0.9664

Tabela 4. 3: Valores estimados do parâmetro d usando o estimador de Robinson para diferentes valores de m, J e l.

valores não se obtêm valores exactamente iguais, apenas semelhantes, embora a diferença entre as estimativas de d tenha tendência a aumentar à medida que m também aumenta. Isto deve-se ao facto de no caso do estimador de Robinson se estar a usar a aproximação

$2\sin\left(\frac{\omega}{2}\right) \approx \omega$, a qual só é válida na vizinhança de zero. Portanto, à medida que m aumenta estão-se a usar um número crescente de frequências, as quais estão cada vez mais afastadas de zero e consequentemente a aproximação usada deixa de ser tão precisa. Daí a discrepância nos resultados.

Para $J = 2$ está-se a associar as ordenadas adjacentes do periodograma duas a duas, para $J = 3$ está-se a associar as ordenadas adjacentes do periodograma três a três e assim sucessivamente.

A associação das ordenadas adjacentes do periodograma reduz a variância do estimador. Contudo, a associação de um número demasiadamente grande de ordenadas do periodograma conduz a um aumento do viés nas estimativas. Portanto, é necessário agir com prudência.

À medida que J aumenta o número de observações usadas na regressão para o mesmo m ($m = 2710^{\alpha}$) diminui. Logo, na prática, para os valores de m mais baixos, à medida que J aumenta a variância também aumenta porque vai-se usar um número muito reduzido de observações na regressão.

Por exemplo, para $m = 11$, $l = 0$ e $J = 4$ apenas se usam duas observações na regressão. Portanto, a estimativa obtida para d não deve ser considerada pois o erro quadrático médio associado deve ser elevado. Já para o caso de $m = 1229$, $l = 0$ e $J = 4$ usam-se na regressão 307 observações.

Portanto, é preciso cuidado na interpretação dos valores obtidos. Tal como nos estimadores anteriores, o número de observações usadas na regressão influencia fortemente as estimativas obtidas para d. A grande variabilidade nos valores obtidos ilustra claramente esta influência.

Para um J fixo, à medida que l aumenta, vai haver uma redução no número de observações usadas na regressão, pois as primeiras l ordenadas do periodograma não são usadas, podendo levar a um aumento acentuado da variância para os valores de m mais baixos.

Para $J > 1$, o aumento de l para valores pequenos de m provoca maior variabilidade nos resultados obtidos. Este facto provavelmente está relacionado com o número reduzido de observações usadas na regressão.

Para $J = 1$, a variabilidade dos dados, à medida que l aumenta, é pouco acentuada comparativamente com os casos $J > 1$, mesmo para valores mais baixos de m. Esta situação poderá estar relacionada com o maior número de observações que são usadas na regressão no caso $J = 1$.

Para cada l fixo, à medida que J aumenta observam-se estimativas muito diferentes para $m = 11$ e $m = 24$ para os diferentes valores de J, o que traduz uma variabilidade elevada nos dados, a qual é consequência do número reduzido de observações usadas na regressão.

4.5. Estimador de Velasco

O estimador de Velasco é obtido usando o mesmo processo com que se obteve o estimador de Robinson.

No entanto, aquele estimador usa o periodograma "tapered" usando o "taper" de Hanning. Além disso, as frequências de Fourier adjacentes usadas na regressão estão mais afastadas que no estimador anterior para garantir que as relações de ortogonalidade entre as transformadas de Fourier continuam válidas.

O estimador de Velasco pode ser aplicado a qualquer série linear independentemente desta ser Gausseana ou não.

Na tabela 4. 4 encontram-se as diferentes estimativas para d obtidas pelo estimador $\hat{d}_{Velasco}$ para diferentes valores de m, J e l. Os valores obtidos resultam da aplicação do estimador $\hat{d}_{Velasco}$ às observações da série em estudo através da aplicação de um programa (programa B4) escrito em Matlab que se encontra no apêndice B.

Para $j = 1$ e $m = 11$ e para l igual a 0, 1 ou 2, apenas são usadas 3 observações na regressão do logaritmo do periodograma "tapered" enquanto que para $l = 3$ são usadas apenas duas observações.

Portanto, nestes casos, tal como seria de esperar, observa-se uma grande variabilidade nos dados obtidos. Logo as estimativas para estas situações extremas não devem ser consideradas.

Para $J = 1$ e $m = 24$ são usadas 8 observações na regressão do logaritmo do periodograma "tapered", enquanto que para $J = 1$ e $m = 1229$ são usadas 409 observações. Portanto, tal como nos estimadores anteriores deve-se ter cuidado com o valor de m usado pois para m demasiadamente pequeno verifica-se que o estimador apresenta uma variabilidade elevada. Tal como nos casos anteriores, para $m = 558$ e $m = 1229$ obtêm-se estimativas elevadas para d. No entanto, no caso deste estimador verifica-se que à medida que J aumenta as estimativas para d para $m = 558$ e para $m = 1229$ aumentam consideravelmente relativamente aos outros valores atribuídos a m, para todos os valores de l.

Para $m = 11$ não é possível obter estimativas para d quando se considera $J > 1$ porque é um valor demasiadamente pequeno.

Para $m = 24$ e $J = 2$ são usadas na regressão do logaritmo do periodograma "tapered" associado apenas 4 observações enquanto que para $J = 3$ e $J = 4$ apenas são usadas 2 observações.

	$l=0$		$l=1$		$l=2$		$l=3$	
	m	$\hat{d}_{Velasco}$	m	$\hat{d}_{Velasco}$	m	$\hat{d}_{Velasco}$	m	$\hat{d}_{Velasco}$
$J=1$	11	2.2580	11	1.3624	11	0.2101	11	3.3392
	24	0.0027	24	0.3055	24	-0.1440	24	-0.3978
	52	0.0671	52	-0.0074	52	-0.0734	52	-0.0997
	115	0.1837	115	0.2211	115	0.1402	115	0.1057
	253	0.2524	253	0.3045	253	0.3314	253	0.2127
	558	0.6374	558	0.7183	558	0.7205	558	0.6164
	1229	0.9232	1229	0.9559	1229	0.9648	1229	0.9103
$J=2$	11	---	11	---	11	---	11	---
	24	-0.1001	24	-0.1083	24	-0.2339	24	0.1598
	52	0.3293	52	0.1024	52	-0.1090	52	-0.0418
	115	0.3254	115	0.3202	115	0.1643	115	0.2183
	253	0.3472	253	0.3604	253	0.3030	253	0.2965
	558	0.7046	558	0.7606	558	0.7145	558	0.6700
	1229	0.9771	1229	0.9927	1229	0.9795	1229	0.9444
$J=3$	11	---	11	---	11	---	11	---
	24	-0.3698	24	-0.6058	24	-0.2748	24	0.2891
	52	0.4014	52	0.2118	52	-0.0233	52	0.0680
	115	0.3183	115	0.2740	115	0.1387	115	0.2969
	253	0.3655	253	0.3754	253	0.3168	253	0.3607
	558	0.7178	558	0.7530	558	0.7296	558	0.7340
	1229	0.9943	1229	0.9973	1229	0.9923	1229	1.0029
$J=4$	11	---	11	---	11	---	11	---
	24	-0.1603	24	---	24	---	24	---
	52	0.2644	52	0.1183	52	-0.1542	52	0.1079
	115	0.3749	115	0.3318	115	0.1850	115	0.3933
	253	0.4115	253	0.3907	253	0.3015	253	0.3751
	558	0.7587	558	0.7754	558	0.7403	558	0.7472
	1229	1.0183	1229	1.0293	1229	1.0124	1229	1.0101

Tabela 4. 4: Valores estimados do parâmetro d usando o estimador de Velasco para diferentes valores de m, J e l.

Portanto, os resultados obtidos nestes casos resultam da aplicação de um estimador com uma variância alta naquelas circunstâncias. Para m = 52, quando j = 3 o número de observações usadas é 5 e quando J = 4 o número de observações usadas é 4, o que é um número bastante

pequeno, tendo em conta a dimensão da série em estudo. Como se pode observar na tabela 4. 4 os resultados obtidos nestes casos apresentam uma grande variabilidade.

É de notar que à medida que l aumenta também há uma redução no número de observações usadas na regressão, embora o impacto dessa redução não seja tão acentuado como acontece com o aumento de J.

O aumento de l tem maior influência quando o número de observações usadas na regressão é baixo. Quando aquele número é alto, à medida que l aumenta verifica-se uma variabilidade pouco acentuada nos dados.

Conclui-se então que à medida que J aumenta, o valor de m usado também deve ser maior de forma a obtermos uma estimativa de d que seja o mais precisa possível.

Tal como era de esperar os valores obtidos com o estimador de Velasco, apresentam uma maior variabilidade que os que foram obtidos usando o estimador anterior porque para os mesmos valores de m e de J existe uma redução acentuada no número de observações usadas na regressão do logaritmo "tapered" e na regressão do logaritmo "tapered" associado.

4.6. Estimação pelo método de máxima verosimilhança

Para se obter a estimação de d pelo método de máxima verosimilhança é preciso especificar o modelo completo.

Nesta secção vai-se procurar ajustar um modelo ARFIMA à série em estudo após lhe ter sido retirada a sazonalidade, pelo método de máxima verosimilhança, usando a função *arima.fracdiff()* do S-Plus. Esta função estima os parâmetros de um modelo ARIMA(p, d, q), obtendo-se valores de máxima verosimilhança aproximados.

Começa-se por ajustar modelos simples como é o caso dos ARFIMA(1, d, 0) e ARFIMA(0, d, 1) e sequencialmente vai-se aumentando a complexidade do modelo.

O logaritmo da verosimilhança, *loglike*, permite a comparação dos modelos e permite escolher o número de parâmetros AR e MA, com base no maior *loglike*. Para diferentes valores do logaritmo da verosimilhança correspondentes a modelos ARFIMA com diferentes parâmetros escolhe-se o modelo que apresente menor AIC. No entanto, para valores de AIC que não sejam muito diferentes escolhe-se o modelo mais parcimonioso.

Contudo, quando se aplicou alguns modelos ARFIMA aos dados correspondentes à série em estudo constatou-se que a matriz das covariâncias dos parâmetros estimados apresentava valores não admissíveis para a variância para alguns dos parâmetros.

No apêndice C apresentam-se alguns dos resultados obtidos após se ter aplicado certos modelos ARFIMA à série em estudo.

Os valores obtidos poderão significar que os modelos ARFIMA não se ajustam adequadamente aos dados. É de salientar uma dificuldade que surgiu durante a tentativa de modelar os dados através de um ARFIMA. Esta dificuldade consiste na validação do modelo escolhido, a qual deveria ser feita com base na análise dos resíduos. No entanto não é possível obter os resíduos no programa usado, o S-Plus.

4.7. Comentários finais

Verificou-se através da aplicação dos diversos estimadores semi-paramétricos à série em estudo uma grande variabilidade nas estimativas de d indicando a forte influência que o número de observações usadas na regressão tem nos resultados obtidos.

Os diversos estimadores semi-paramétricos usados parecem concordar que $d > 0.5$ indicando que a série em estudo possivelmente é uma série não estacionária de memória longa. Portanto, o fluxo de água no rio Shiel num dado dia pode estar relacionado com o fluxo que este rio apresentou à muito tempo atrás. No entanto, é difícil saber qual o valor exacto de d.

Apêndice A

Definição A1. (Hosking (1981)) *Um processo ARFIMA(1, d, 0), $\{X_t\}$, é um processo que satisfaz a igualdade:*

$$(1-\phi B)(1-B)^d X_t = \varepsilon_t$$

onde $\{\varepsilon_t\}$ é um ruído branco.

Para garantir a estacionaridade e invertibilidade de$\{X_t\}$ assume-se que $|\phi|<1$ e $|d|<\frac{1}{2}$.

As correlações de $\{X_t\}$ são expressas em termos da função hipergeometrica:

$$F(a,b;c;z)=1+\frac{ab}{c.1}z+\frac{a(a+1)b(b+1)}{c(c+1).1.2}z^2+....$$

Lema A1. (Hosking (1981)) *Seja $\{X_t\}$ um processo ARFIMA(1, d, 0), $(1-\phi B)(1-B)^d X_t = \varepsilon_t$ estacionário e invertível e seja $Y_t = (1 - \phi B)X_t$.*

a) *Os pesos ψ e π das representações de médias móveis infinita e auto-regressiva infinita de $\{X_t\}$ são dados por:*

$$\psi_k = \frac{(k+d-1)!}{k!(d-1)!}F(1,-k;1-d-k;\phi) \sim \frac{k^{d-1}}{(1-\phi)(d-1)!}$$

$$\pi_k = \frac{(k-d-2)!}{(k-1)!(-d-1)!}\left\{1-\phi-\frac{1+d}{k}\right\} \sim \frac{1-\phi}{(-d-1)!}k^{-d-1}.$$

b) *A função de densidade espectral de $\{X_t\}$ é* $f_{ARFIMA}(\omega)=\dfrac{\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d}}{1+\phi^2-2\phi\cos\omega} \sim \dfrac{\omega^{-2d}}{(1-\phi)^2}$.

c) *Sejam $\gamma^x(k)$, $\rho^x(k)$, $\gamma^y(k)$, $\rho^y(k)$ as funções de auto-covariância e de auto- correlação de $\{X_t\}$ e de $\{Y_t\}$, respectivamente; então:*

$$\gamma^x(k)=\frac{\gamma^y(k)\{F(1,d+k;1-d+k;\phi)+F(1,d-k;1-d-k;\phi)-1\}}{1-\phi^2}$$

$$\rho^x(k)=\frac{\rho^y(k)\{F(1,d+k;1-d+k;\phi)+F(1,d-k;1-d-k;\phi)-1\}}{(1-\phi)F(1,1+d;1-d;\phi)}$$

$$\rho^x(k) \sim \frac{(-d)!}{(d-1)!}\frac{1+\phi}{(1-\phi)^2}\frac{k^{2d-1}}{F(1,1+d;1-d;\phi)}$$

Em particular,

$$\gamma^x(0) = \frac{(-2d)!}{[(-d)!]^2} \frac{F(1, 1+d; 1-d; \phi)}{1+\phi}$$

$$\rho^x(1) = \frac{(1+\phi^2)F(1, d; 1-d; \phi) - 1}{\phi[2F(1, d; 1-d; \phi) - 1]}$$

Definição A2. (Hosking (1981)) *Um processo ARFIMA(0, d, 1), $\{X_t\}$, é um processo que satisfaz a igualdade:*

$$(1-B)^d X_t = (1-\theta B)\varepsilon_t$$

onde $\{\varepsilon_t\}$ é o ruído branco.

Este processo é estacionário e invertível para $|\theta| < 1$ e $|d| < \frac{1}{2}$.

Lema A2. (Hosking (1981)) *Seja $\{X_t\}$ um processo ARFIMA(0, d, 1) estacionário e invertível; seja $(1-B)^d Y_t = \varepsilon_t$ de tal forma que $X_t = (1-\theta B)Y_t$; sejam $\gamma^x(k)$, $\rho^x(k)$, $\gamma^y(k)$, $\rho^y(k)$ as funções de auto-covariância e de auto-correlação de $\{X_t\}$ e de $\{Y_t\}$ respectivamente.*

a) *Os pesos ψ e π de $\{X_t\}$ são iguais, respectivamente, aos pesos π e ψ de um ARFIMA(1, -d, 0) com parâmetro autoregressivo θ.*

b) *A função de densidade espectral de $\{X_t\}$ é*

$$f_{ARFIMA}(\omega) = (1+\theta^2 - 2\theta\cos\omega)\left[2\sin\left(\frac{\omega}{2}\right)\right]^{-2d} \sim (1-\theta)^2 \omega^{-2d}$$

c) A função de auto-covariância de $\{X_t\}$ é

$$\gamma^x(k) = \gamma^y(k) \frac{(1-\theta)^2 k^2 - (1-d)[(1-d)(1+\theta^2) - 2\theta d]}{k^2 - (1-d)^2}$$

e a função de auto-correlação de $\{X_t\}$ é $\rho^x(k) = \rho^y(k) \frac{ak^2 - (1-d)^2}{k^2 - (1-d)^2}$ onde

$$a = \frac{(1-\theta)^2}{1+\theta^2 - \frac{2\theta d}{1-d}}.$$

Em particular,

$$\gamma^x(0) = \left(1 + \theta^2 - \frac{2\theta d}{1-d}\right)\frac{(-2d)!}{[(-d)!]^2},$$

$$\rho^x(1) = \frac{a - (1-d)^2}{(1-d)(2-d)} = \frac{(1+\theta^2)d(2-d) - 2\theta(1-d+d^2)}{(1-d)(2-d)\left[1 + \theta^2 - \frac{2\theta d}{1-d}\right]}.$$

Além disso,

$$\rho^x(k) \sim \frac{(-d)!}{(-d-1)!} a k^{2d-1}.$$

Apêndice B

B1)

```
function[d]=GPH(s,m)

% Este programa elimina o periodograma calculado na frequência zero porque a função do
% Matlab periodogram calcula o periodograma na frequência zero, atribuindo-lhe o valor
zero.
% s - representa a série
% m - é o n.º de observações usadas na regressão
% d - é o parâmetro de diferenciação


[P, W]=periodogram(s);
I=P(2:(m+1));
w=W(2:(m+1));

for j=1:m
   y(j)=log(I(j));
   x(j)=log((2*sin(0.5*w(j)))^2);
end
mediax=mean(x);
N=0;
D=0;
for j=1:m
   soma1=(x(j)-mediax).*y(j);
   N=N+soma1;
   soma2=(x(j)-mediax)^2;
   D=D+soma2;
end
d=(-N)/D
```

B2)

```
function[d]=Reisen(s,m,M)

% s - representa a serie
% d - é o parâmetro de diferenciação
% m - é o n.º de observações da regressão
% A janela de Parzen tem de ser um vector com o mesmo n.º de elementos da série.
% Para poder centrar a janela vai-se retirar o 1.º elemento da serie, porque
% esta é constituída por um número par de observações.
% M representa a dimensão da janela de Parzen.

S=0;
for r=-1354:1354
   r1=abs(r);
   S=S+1;
```

```
  if r1<=M/2
     P(S)=1-6*(r/M)^2+6*(r1/M)^3;
   elseif M/2<=r1<=M
     P(S)=2*(1-r1/M)^3;
   else r1>M
     P(S)=0;
   end
end
R=[P(1:2709)'];
[A, W]=periodogram(s,R);
I=A(2:(m+1));
w=W(2:(m+1));

for j=1:m
   y(j)=log(I(j));
   x(j)=log((2*sin(0.5*w(j)))^2);
end
mediax=mean(x);
N=0;
D=0;
for j=1:m
   soma1=(x(j)-mediax).*y(j);
   N=N+soma1;
   soma2=(x(j)-mediax)^2;
   D=D+soma2;
end
d=(-N)/D
```

B3)

```
function[d]=Robinson(s,m,J,l)

% s representa a série
% m representa o n.º de observações usadas na regressão se (m-l)/J for inteiro.
% Se aquela divisão não for inteira o programa substitui o m inicial por um m1,o qual é o
% maior inteiro inferior a m tal que (m1-l)/J é inteiro.
% J representa o número de ordenadas do periodograma que vão ser associadas.
% l representa o número de ordenadas do periodograma correspondentes às frequências de
% Fourier mais próximas de zero que não vão ser usadas na regressão.
% d é o parâmetro de diferenciação.

m1=m-rem(m-l,J);
[P,W]=periodogram(s);
I=P(2:(m1+1));
w=W(2:(m1+1));
k=l+J;
i=1;
while k<=m1
   soma=0;
   for j=1:J
```

```
      soma=soma+I(k+j-J);
   end
   y(i)=log(soma);
   i=i+1;
   k=k+J;
end
for i=1:((m1-l)/J)
   x(i)=-2*log(w(i));
end
mediax=mean(x);
N=0;
D=0;
for j=1:((m1-l)/J)
   soma1=(x(j)-mediax).*y(j);
   N=N+soma1;
   soma2=(x(j)-mediax)^2;
   D=D+soma2;
end
d=N/D
```

B4)

```
function[d]=Velasco(s,m,J,l)

% s representa a série
% m representa o n.º de observações usadas na regressão se (m-l)/J for inteiro.
% Se aquela divisão não for inteira o programa substitui o m inicial por um m1,o qual é o
% maior inteiro inferior a m tal que (m1-l)/J é inteiro.
% J representa o número de ordenadas do periodograma que vão ser associadas.
% l representa o número de ordenadas do periodograma correspondentes às frequências de
% Fourier mais próximas de zero que não vão ser usadas na regressão.
% d é o parâmetro de diferenciação.

m1=m-rem(m-l,3*J);
h=hann(2710);
[P,W]=periodogram(s,h);
I=P(2:(m1+1));
w=W(2:(m1+1));
k=l+3*J;
i=1;
while k<=m1
   soma=0;
   for j=1:J
      soma=soma+I(k+3*(j-J));
   end
   y(i)=log(soma);
   i=i+1;
   k=k+3*J;
end
for i=1:((m1-l)/(3*J))
```

```
   x(i)=-2*log(w(i));
end
mediax=mean(x);
N=0;
D=0;
for j=1:((m1-l)/(3*J))
   soma1=(x(j)-mediax).*y(j);
   N=N+soma1;
   soma2=(x(j)-mediax)^2;
   D=D+soma2;
end
d=N/D
```

Apêndice C

```
> arima.fracdiff(ss,model=list(ar=0.5,ma=0))
$model:
$model$ar:
[1] 0.8402459

$model$ma:
[1] -0.3254914

$model$d:
[1] 0.2635778
```

$var.coef:

	d	ar1	ma1
d	valor não admissível	0.00041994566	0.00004028872
ar1	0.00041994566	valor não admissível	0.00001514292
ma1	0.00004028872	0.00001514292	0.00012204676

```
$loglik:
[1] -6986.604

$h:
[1] 0.00007362637

$d.tol:
[1] 0.0001220703

$M:
[1] 100
```

$hess:

	d	ar1	ma1
d	valor não admissível	-5329.954	2356.463
ar1	-5329.954	valor não admissível	2135.510
ma1	2356.463	2135.510	valor não admissível

```
$call:
arima.fracdiff(x = ss, model = list(ar = 0.5, ma = 0))

> arima.fracdiff(ss,model=list(ar=0,ma=0.5))
$model:
$model$ar:
[1] 0.840246

$model$ma:
[1] -0.3254915
```

$model$d:
[1] 0.2635776

$var.coef:

	d	ar1	ma1
d	valor não admissível	0.0004199461	0.00004028873
ar1	0.00041994612	valor não admissível	0.00001514280
ma1	0.00004028873	0.0000151428	0.00012204665

$loglik:
[1] -6986.604

$h:
[1] 0.00007362637

$d.tol:
[1] 0.0001220703

$M:
[1] 100

$hess:

	d	ar1	ma1
d	valor não admissível	-5329.955	2356.463
ar1	-5329.955	valor não admissível	2135.510
ma1	2356.463	2135.510	valor não admissível

$call:
arima.fracdiff(x = ss, model = list(ar = 0, ma = 0.5))

> arima.fracdiff(ss,model=list(ar=0.4,ma=-0.2))
$model:
$model$ar:
[1] 0.8402459

$model$ma:
[1] -0.3254914

$model$d:
[1] 0.2635778

$var.coef:

	d	ar1	ma1
d	valor não admissível	0.0004199458	0.00004028873
ar1	0.00041994577	valor não admissível	0.00001514290
ma1	0.00004028873	0.0000151429	0.00012204675

$loglik:
[1] -6986.604

$h:
[1] 0.00007362637

$d.tol:
[1] 0.0001220703

$M:
[1] 100

$hess:

	d	ar1	ma1
d	valor não admissível	-5329.954	2356.463
ar1	-5329.954	valor não admissível	2135.510
ma1	2356.463	2135.510	valor não admissível

$call:
arima.fracdiff(x = ss, model = list(ar = 0.4, ma = -0.2))

> arima.fracdiff(ss,model=list(ar=c(0.1,0.5),ma=0)))
Problem: Syntax error: No opening parenthesis before
unbalanced (")") on input line 1
> arima.fracdiff(ss,model=list(ar=c(0.1,0.5),ma=0))
$model:
$model$ar:
[1] 1.0343115 -0.1481854

$model$ma:
[1] -0.2257384

$model$d:
[1] 0.1735519

$var.coef:

	d	ar1	ar2	ma1
d	valor não admissível	0.00019653247	-0.0001730917	-0.00020475268
ar1	0.00019653247	0.00001602444	-0.0003088306	-0.00002774629
ar2	-0.00017309173	-0.00030883057	0.0009571336	0.00039157966
ma1	-0.00020475268	-0.00002774629	0.0003915797	0.00011728608

$loglik:
[1] -6986.001

$h:
[1] 0.00007362002

$d.tol:
[1] 0.0001220703

$M:
[1] 100

$hess:

	d	ar1	ar2	ma1
d	valor não admissível	-6566.631	-4135.8032	2446.717
ar1	-6566.6313	valor não admissível	-514.7535	-13216.125
ar2	-4135.8032	2446.717	valor não admissível	2191.924
ma1	-514.7535	-13216.125	2191.9244	valor não admissível

$call:
arima.fracdiff(x = ss, model = list(ar = c(0.1, 0.5), ma = 0))

> arima.fracdiff(ss,model=list(ar=c(0.3,-0.4,0.5,-0.2,0.6),ma=0))
$model:
$model$ar:
[1] 0.955737849 0.005562629 -0.035617856 -0.016717523
[5] -0.017893239

$model$ma:
[1] -0.3665815

$model$d:
[1] 0.1088884

$var.coef:

	d	ar1	ar2
d	valor não admissível	-9.373199e-006	2.716201e-004
ar1	-9.373199e-006	1.164193e-007	-3.359366e-006
ar2	2.716201e-004	-3.359366e-006	9.693579e-005
ar3	1.193395e-004	9.366050e-006	-2.710609e-004
ar4	-9.516804e-005	5.980051e-006	-1.731557e-004
ar5	3.739160e-007	-1.730534e-008	6.723857e-007
ma1	-1.563538e-004	-1.784065e-006	5.181918e-005

	ar3	ar4	ar5
d	1.193395e-004	-9.516804e-005	3.739160e-007
ar1	9.366050e-006	5.980051e-006	-1.730534e-008
ar2	-2.710609e-004	-1.731557e-004	6.723857e-007
ar3	3.154429e-004	3.483591e-005	-1.430036e-007
ar4	3.483591e-005	2.935545e-004	-1.148961e-006
ar5	-1.430036e-007	-1.148961e-006	4.496862e-009
ma1	2.481828e-004	-8.174885e-005	3.198557e-007

	ma1
d	-1.563538e-004
ar1	-1.784065e-006
ar2	5.181918e-005
ar3	2.481828e-004
ar4	-8.174885e-005

ar5 3.198557e-007
ma1 9.005843e-005

$loglik:
[1] -6984.37

$h:
[1] 0.00007360283

$d.tol:
[1] 0.0001220703

$M:
[1] 100

$hess:

	d	ar1	ar2	ar3
d	valor não admissível	-7059.32952	-4723.026	-3666.5379
ar1	-7059.3295	valor não admissível	-12709.911	-2356.7410
ar2	-4723.0255	-3003.41712	valor não admissível	186.8704
ar3	-12709.9107	-673515.79678	-169038210.105	valor não admissível
ar4	-3666.5379	-36962.36031	-5849867.593	2308.7534
ar5	-2356.7410	-74.12462	5741.023	-9764.9166
ma1	186.8704	-2495.77477	-12596.481	-12745.5320

	ar4	ar5	ma1
d	-3003.41712	-2495.775	2308.753
ar1	-673515.79678	-169038210.105	-9764.917
ar2	-36962.36031	-5849867.593	-12745.532
ar3	-74.12462	5741.023	2019.341
ar4	valor não admissível	-12596.481	-11227.173
ar5	2019.34074	valor não admissível	-14208.882
ma1	-11227.17292	-14208.882	valor não admissível

$call:
arima.fracdiff(x = ss, model = list(ar = c(0.3, -0.4, 0.5,
-0.2, 0.6), ma = 0))

REFERÊNCIAS

[1] AGIAKLOGLOU, C., NEWBOLD, P., WOHAR, M. (1993) Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, **14**, 235-246.

[2] ANDĚL, J. (1986) Long memory time series models. *Kibernetika,* **22(2)**, 105-223.

[3] ANDREWS, D. W. K., GUGGENBERGER, P. (2003) A bias-reduced log-periodogram regression estimator for the long memory parameter. *Econometrica*, **71(2)**, 675-712.

[4] ARTECHE, J., ROBINSON, P. M. (2000) Semi parametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, **21(1)**, 1-25.

[5] BERAN, J. (1986) Estimation, testing and prediction for self-similar and related processes. *PhD Thesis*, ETH *Zürich.*

[6] BERAN, J (1992) Statistical methods for data with long range dependence, with discussion. *Statistical Science,* **7(4)**, 404-427.

[7] BERAN, J (1993) Fitting Long-memory models by generalized linear regression. *Biometrika* **80**, 817-822.

[8] BERAN, J. (1994) Statistics for long-memory processes. Chapman & Hall.

[9] BLOOMFIELD, P. (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217-226.

[10] BOLLERSLEV, T., JUBINSKI, D. (1999) Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics,* **17**, 9-21.

[11] BOX, G. E. P., JENKINS, G. M. (1976) Time series analysis: forecasting and control. 2nd edition. *San Francisco: Holden-Day.*

[12] CHEUNG, Y., DIEBOLD, F. (1994) On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics*, **62**, 301-316.

[13] CHAN, N. H., PALMA, W. (1998) State space modelling of long-memory processes. *Annals of Statistics*, **26**, 719-740.

[14] CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E., TERPENING, I. (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, **6**, 3-73.

[15] DAHLHAUS, R. (1989) Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, **17**, 1749-1766.

[16] DAVIES, R. B., HARTE, D. S. (1987) Tests for Hurst effect. *Biometrika,* **74**, 95-101.

[17] DOORNIK, J. A., OOMS, M. (2003) Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Comp. Stat. Data An.*, **42(3)**, 333-348.

[18] FEDER, J. (1988) *Fractals*. Plenum Press, New York.

[19] FOX, R., TAQQU, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *The Annals of Statistics*, **14(2)**, 517-532.

[20] GARRETT, M., WILLINGER, W. (1994) Analysis, modelling and generation of self-similar VBR video traffic. *Proceedings of SSIGCOMM'94.*

[21] GEWEKE, J., PORTER-HUDAK, S. (1983) The estimation and application of long memory time series model. *Journal of Time Series Analysis,* **1(1)**, 15-29.

[22] GIRAITIS, L., LEIPUS, R. (1995) A generalized fractionally differencing approach in long-memory modelling. *Lietuvos Matematikos Rinkinys*, **35(1)**, 65-81.

[23] GIRAITIS, L., ROBINSON, P., SAMAROV, A. (1997) Rate optimal semi parametric estimation of the memory parameter of the Gaussian time series with long range dependence. *Journal of Time Series Analysis*, **18**, 49-61.

[24] GOLUB, G. H., VAN LOAN, C. F. (1989) Matrix computations. *Baltimore: The Johns Hopkins University Press.*

[25] GRANGER, C. W. (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14,** 227-238.

[26] GRANGER, C. W., JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis,* **1**, 15-29.

[27] GRENANDER, U., SZEGÖ, G. (1958) Toeplitz forms and their application. *University of California Press, Berkeley.*

[28] HASLETT, J., RAFTERY, A. (1989) Space time modelling with long-memory dependence: assessing Ireland's wind power resource. *Appl. Statist.*, **38(1)**, 1-50

[29] HASSLER, U. (1994) (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis,* **15(1)**, 19-30.

[30] HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika,* **68**, 165-176.

[31] HOSKING, J. R. M. (1982) Some models of persistence in time series. *Time Series Analysis: Theory and Practice 1*, 641-653.

[32] HOUSER, M. A. (1999) Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference*, **80(1-2)**, 229-255.

[33] HURST, H. E. (1951) Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, **116**, 770-799.

[34] HURVICH, C. M., DEO, R. S., BRODSKY, J. (1998) The mean squared error of Geweke and Porter-Hudak's estimation of the memory parameter of a long memory time series. *Journal of Time Series Analysis* **19(1)**, 19-46.

[35] HURVICH, C. M., DEO, R. S. (1999) Plug-in selection of the number of the frequencies in regression estimates of the memory parameter of a long-memory time series. *Journal of Time Series Analysis,* **20(3)**, 331-341.

[36] HURVICH, C. M (2001) Model selection for broadband semiparametric estimation of long memory in time series. *Journal of Time Series Analysis,* **22(6)**, 679-709.

[37] HURVICH, C. M., BRODSKY, J. (2001) Broadband semi-parametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis,* **22(2)**, 221-249.

[38] JANACEK, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis,* **3**, 177-183.

[39] KOKOSZKA, P. S., TAQQU, M. S. (1995) Fractional ARIMA with stable innovations. *Stochastic Processes and Their Applications,* **60**, 19-47.

[40] KOKOSZKA, P. S., TAQQU, M. S. (1996) Parameter estimation for infinite variance fractional ARIMA, *The Annals of Statistics*, **24(5)**, 1880-1913.

[41] KOKOSZKA, P. S., TAQQU, M. S. (2001) Can one use the Durbin-Levinson algorithm to generate infinite variance fractional ARIMA time series ? *Journal of Time series Analysis,* **22(3),** 317-337.

[42] LEDESMA, S., LIU, D. (2000) A fast method for generating self-similar network traffic. In *Proceedings of the 2000 International Conference on Communication Technologies,*54-61, Beijing, China.

[43] MANDELBROT, B. B. (1965) Une classe de processus stochastiques homothétiques à soi : application à la loi climatológique de H. E. Hurst. *Comptes Rendus Acad. Sci. Paris*, **260**, 3274-3277.

[44] MANDELBROT, B. B., VAN NESS, J. W. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM rev*, **10(4)**, 422-437.

[45] MANDELBROT, B. B., WALLIS, J. R. (1969) Computer experiments with fractional Gaussian noises. *Water Resour. Res.*, **5**, 228-267.

[46] MIKOSCH, T., GADRICH, T., KLÜPPELBERG, C. & ADLER, R. J. (1995) Parameter estimation for ARMA models with infinite variance innovations, *The Annals of Statistics*, **23(1)**, 305-326.

[47] MOULINES, E., SOULIER, P. (1999) Broadband log-periodogram regression of time series with long-range dependence. *The Annals of Statistics,* **27(4),** 1415-1439.

[48] MOULINES, E., SOULIER, P. (2000) Data driven order selection for projection estimator of the spectral density of time series with long range dependence. *Journal of Time Series Analysis,* **21(2)**, 193-18.

[49] PAI, J. S., RAVISHANKER, N. (1996) Exact likelihood function forms for an ARFIMA process. In *Data Analysis and Information Systems, Statistical and Conceptual Approaches. Proceedings of the 19th Annual Conference of the Gesellschaft für Klassifikation e V., University of Basel, March 8-10, 1995,*323-331. Springer Verlag, Berlin.

[50] PAXON, V. (1997) Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *Computer Communication Review*, **27(5)**, 5-18.

[51] PEITGEN, H. -O., SOUPE, D. EDITORES (1998) *The Science of Fractal Images.* Springer-Verlag, New York.

[52] RAVISHANKER, N., RAY, B. K. (1997) Bayesian analysis of vector ARFIMA processes. *Australian Journal of Statistics*, **39**, 295-312.

[53] 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**, 335-350.

[54] ROBINSON, P. M. (1990) Time series with long range dependence. *Advances in Econometrics. Proceedings of the Sixth World Congres,*.**1,** 47-95. Cambridge University Press (1994).

[55] ROBINSON, P. M. (1994a) Time series with strong dependence, vol. 1, Cambridge University Press, ch. 2 of Advances in Econometrics, Sixth World Congress, 47- 95.

[56] ROBINSON, P. M. (1994b) Semi parametric analysis of long-memory time series analysis. *The Annals of Statistics*, **22**, 515-539.

[57] ROBINSON, P. M. (1995a) Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23(3)**, 1048-1072.

[58] ROBINSON, P. M. (1995b) Gaussian semi-parametric estimation of the long range dependence. *The Annals of Statistics*, **23(5)**, 1630-1661.

[59] SAMARODNITSKY, G., TAQQU, M. (1994) Stable non-Gaussian random processes, *Chapman & Hall.*

[60] SHIMOTSU, K., PHILLIPS, P. C. B. (2002) Pooled log periodogram regression. *Journal of Time Series Analysis*, **23(1)**, 57-93.

[61] SMITH, J., TAYLOR, N., YADA, V. S. (1997) Comparing the bias and misspecification in ARFIMA models. *Journal of Time Series Analysis*, **18**, 507-527.

[62] SOWELL, F. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165-188.

[63] TAQQU, M., TEVEROVSKI, V. (1996) Semi-parametric graphical estimation techniques for long-memory data, in Proceedings of the Athens Conference on Applied Probability and Time-Series Analysis, vol. 115 of Lecture Notes in Statistics, Springer-Verlag, 420-432.

[64] TAQQU, M., TEVEROVSKI, V. (1997) Robustness of Whittle-type estimators of time-series of long-range dependence. *Communications in Statistics: Stochastic Models.*

[65] VELASCO, C. (2000) Non-Gaussian log-periodogram regression. *Econometric Theory,* **16**, 44-79.

[66] VOSS, R. F. (1988) Fractals in nature: from characterization to simulation. In H. –O. Peitgen e D. Soupe, editores. *The Science of Fractal Ima*ges, capítulo 1, 21-70. Springer-Verlag, New York.

[67] WHITTLE, P. (1953) Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-434.

[68] WOODWARD, W. A., CHENG, Q. C., GRAY, H. L. (1998) A k-factor GARMA long-memory model. *Journal of Time Series Analysis*, **19(4)**, 485-504.