

INTERVALOS DE CONFIANÇA BOOTSTRAP PARA O PARÂMETRO d EM MODELOS DE INTEGRAÇÃO FRACIONÁRIA

Glaura Franco

Universidade Federal de Minas Gerais
glaura@est.ufmg.br

Valdério Reisen

Universidade Federal do Espírito Santo
valderio@cce.ufes.br

Neander Almeida

Universidade Federal de Minas Gerais
nfa@est.ufmg.br

Resumo

Este artigo apresenta estudos de construção de intervalos de confiança *bootstrap* para o parâmetro de memória d do modelo de Integração Fracionária FI(d) (Hosking, 1981), calculados através dos estimadores semiparamétricos propostos por Geweke and Porter-Hudak (1983) e Reisen (1994) denotados, respectivamente, por GPH e SPR. Os intervalos *bootstrap* utilizados foram os percentílico e *bootstrap-t* (Efron and Tibshirani, 1993). Simulações Monte Carlo foram realizadas para comparação dos estimadores do modelo e dos intervalos de confiança. Os resultados obtidos mostram que o método *bootstrap* apresenta estimativas muito similares com as obtidas pelos métodos semiparamétricos. As simulações mostram também que os intervalos *bootstrap* apresentam alto percentual de cobertura, sendo que o intervalo percentílico é levemente superior ao *bootstrap-t*, e os intervalos para o GPH apresentam percentual de cobertura mais próxima do nível nominal que os do SPR.

Palavras-chaves: longa dependência; *bootstrap*; intervalos de confiança.

Abstract

This work presents a study related to bootstrap confidence intervals for the memory parameter d in the Fractionally Integrated process FI(d) (Hosking, 1981) based on the semiparametric estimators proposed by Geweke & Porter-Hudak (1983) and Reisen (1994), hereafter denoted by GPH and SPR, respectively. The bootstrap interval techniques used were the percentile and the *bootstrap-t* (Efron & Tibshirani, 1993). Monte Carlo simulations were performed to estimate d and to construct interval comparisons. The results show that the bootstrap method gives similar estimated values compared with the semiparametric estimates. The simulation studies also show that the bootstrap intervals present a reasonably coverage frequency rate, where the *bootstrap-t* interval is slightly superior to the percentile, and the intervals from the GPH method are more precise than the intervals from the SPR estimator.

Keywords: long memory; bootstrap; confidence intervals.

1. INTRODUÇÃO

O modelo ARFIMA(p, d, q) (Fracionário Autoregressivo Integrado Média Móvel) é derivado do modelo ARIMA, proposto por Box & Jenkins (1976), e pertence a uma classe de modelos em que

a característica 'longa dependência' é encontrada. Uma série temporal com longa dependência apresenta correlações significativas mesmo para observações distantes no tempo. O procedimento de modelagem de séries pertencentes à classe de modelos ARFIMA (p,d,q) inclui a estimação do parâmetro d , que neste caso pode assumir valores fracionários.

Existem, na literatura, vários procedimentos para a estimação de d . Dentre eles podemos citar os métodos semi-paramétricos de Geweke & Porter-Hudak (1983), Reisen (1994) e Robinson (1995) e paramétricos de Fox & Taquq (1986) e Sowell (1992). Outras referências podem ser encontradas nos recentes trabalhos que exploram modelos ARFIMA em diferentes situações, tais como Lopes et al. (2004) e Doukhan et al. (2003). Mesmo com uma extensa bibliografia referente à estimação do parâmetro d , a identificação de séries que possuem a característica de memória longa não é uma tarefa muito simples, pois as inferências sobre o parâmetro d são feitas, em geral, utilizando-se a distribuição assintótica dos estimadores, o que algumas vezes não leva a resultados muito precisos, principalmente se as séries utilizadas têm tamanho pequeno.

Neste trabalho, a construção de intervalos de confiança para d será feita utilizando a distribuição assintótica dos estimadores, assim como intervalos baseados no método *bootstrap* (Efron, 1979; Efron & Tibshirani, 1993). Apresentamos também uma breve descrição destes procedimentos, juntamente com um exercício de simulação Monte Carlo para verificarmos a eficácia dos métodos de *bootstrap* neste caso. As séries serão reamostradas pelo procedimento *bootstrap nos resíduos* (Franco & Reisen, 2004) e os intervalos comparados através do percentual de cobertura dos mesmos. A série da BOVESPA é estudada como exercício de aplicação da metodologia explorada empiricamente.

O trabalho foi dividido como segue: Na seção 2 apresentamos o modelo ARFIMA(p,d,q) juntamente com os estimadores semiparamétricos de d . Na seção 3 é descrito o método *bootstrap* utilizado, assim como os resultados da simulação para comparação dos intervalos de confiança. Na seção 4 aplicamos as metodologias exploradas neste trabalho na série BOVESPA, e a seção 5 apresenta a conclusão do trabalho.

2. MODELO E ESTIMADORES

2.1 O Modelo ARFIMA(p,d,q)

Seja $\{X_t\}$ o processo ARFIMA(p,d,q) que é definido por

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)a_t, \quad d \in (-0,5; 0,5) \quad (1)$$

onde $\{a_t\}$ é um processo ruído branco com $E(a_t) = 0$ e variância constante σ^2 ; $\phi_p(B)$ e $\theta_q(B)$ são os polinômios autoregressivo e média móvel, respectivamente; B é o operador de retardo e d é o parâmetro de diferenciação. Para $d \in (0,0 ; 0,5)$, $\{X_t\}$ é um processo estacionário e invertível (ver, por exemplo, Reisen (1994)).

O fato de d assumir valores não inteiros confere aos modelos ARFIMA(p,d,q) a propriedade de *longa dependência* quando $d \in (0,0 ; 0,5)$. O termo *longa dependência* é devido à dependência, não desprezível, mesmo entre observações distantes da série. Isto pode ser visto através da expansão binomial do operador de diferenciação fracionária

$$(1-B)^d = 1 - dB + \frac{d(d-1)B^2}{2!} - \frac{d(d-1)(d-2)B^3}{3!} + \dots \quad (2)$$

Os coeficientes em (2) evidenciam as contribuições das observações no modelo, mesmo quando estas são de *lags* distantes. Isto pode ser justificado através da forma assintótica da função de autocorrelação, ρ_j , $j > 0$, de $\{X_t\}$, que é dada por $\rho_j \sim |j|^{2d-1}$. Este decaimento hiperbólico da função de autocorrelação é típico de *longa dependência*. Maiores detalhes sobre o processo ARFIMA(p,d,q) podem ser encontrados em Reisen (1995) ou Doukhan et al. (2003).

2.2 Função Espectral do Modelo ARFIMA(p,d,q)

Seja $\{X_t\}$ um processo estacionário ARFIMA(p,d,q). A função espectral de $\{X_t\}$ é dada por:

$$f(\omega) = f_U(\omega) [2 \sin(\omega/2)]^{-2d} \quad (3)$$

onde $\omega \in [-\pi, \pi]$ são as frequências de Fourier e $f_U(\omega)$ é a função densidade espectral de um processo ARMA(p,q).

Para uma amostra de tamanho n de $\{X_t\}$, um estimador da função densidade espectral é a função periodograma, $I(\omega)$, definida por

$$I(\omega) = \left[R(0) + 2 \sum_{j=1}^{n-1} R(j) \cos(\omega j) \right] / 2\pi, \quad \omega \in [-\pi, \pi] \quad (4)$$

onde $R(j)$, a autocovariância amostral, é dada por

$$R(j) = \frac{1}{n} \sum_{i=1}^{n-j} (X_i - \bar{X})(X_{i+j} - \bar{X}), \quad j = 0, \pm 1, \dots, \pm(n-1). \quad (5)$$

Um estimador alternativo da densidade espectral $f(\omega)$ é a função periodograma suavizado, $f_s(\omega)$, definido, para $\omega \in [-\pi, \pi]$, por

$$f_s(\omega) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{n-1} \lambda(j) R(j) \cos(\omega j) \quad (6)$$

onde $\lambda(j)$ é uma função ponderadora, conhecida como “lag window”. Diferentes formas de $\lambda(j)$ são sugeridas na literatura (Baillie, 1996).

2.3 Estimação do Parâmetro de Diferenciação d

Na literatura existem vários métodos para se estimar o parâmetro d dos modelos ARFIMA(p,d,q) (Baillie, 1996). Neste trabalho serão utilizados os seguintes métodos semi-paramétricos para obter \hat{d} :

- Método da Regressão Utilizando a Função Periodograma (GPH).
- Método da Regressão Utilizando a Função Periodograma Suavizado (SPR).

Regressão Utilizando a Função Periodograma

O método da regressão, proposto por Geweke e Porter-Hudak (1983), consiste em tomar o logaritmo da função densidade espectral $f(\omega)$, dada em (3), e estimar o parâmetro d por mínimos quadrados através da equação de regressão obtida. Sendo $f(\omega)$ desconhecida, é necessário substituí-la por algum de seus estimadores. Geweke e Porter-Hudak (1983) utilizaram a função periodograma $I(\omega)$.

Assim, para obter o estimador GPH toma-se o logaritmo da equação (3), substituindo-se ω pelas frequências $\omega_j = 2\pi j / n$, com $j = 1, 2, \dots, g(n)$, e $f(\omega)$ por $I(\omega_j)$.

Então,

$$\ln I(\omega_j) = \ln f_u(0) - d \ln \left\{ 2 \sin \left(\frac{\omega_j}{2} \right) \right\}^2 + \ln \left\{ \frac{f_u(\omega_j)}{f_u(0)} \right\} + \ln \left\{ \frac{I(\omega_j)}{f(\omega_j)} \right\}. \quad (7)$$

O número de ordenadas periodograma utilizadas, $g(n)$, é uma função que satisfaz às seguintes condições:

$$\lim_{n \rightarrow \infty} g(n) = \infty, \quad \lim_{n \rightarrow g} \frac{g(n)}{n} = 0, \quad \lim_{n \rightarrow g} \frac{\ln n^2}{g(n)} = 0. \quad (8)$$

Quanto à escolha de $g(n)$, o mais indicado é escolher $g(n) = n^\alpha$, $\alpha \in (0,1)$. Usualmente toma-se $\alpha \in (0,5; 0,8)$. Um estudo sobre a escolha de α pode ser encontrado em Reisen (1994).

Nas proximidades de zero $\ln\{f_u(\omega_j)/f_u(0)\}$ é desprezível se comparado aos demais termos, logo (7) pode ser vista como uma equação do tipo $y_j = a + bx_j + \varepsilon_j$, onde

$$y_j = \ln I(\omega_j), \quad a = \ln f_u(0) - c, \quad b = -d, \quad x_j = \ln \left\{ 2 \sin \left(\frac{\omega_j}{2} \right) \right\}^2, \\ \varepsilon_j = \ln \left\{ \frac{I(\omega_j)}{f(\omega_j)} \right\} + c \quad e \quad c = E \left(- \ln \left\{ \frac{I(\omega_j)}{f(\omega_j)} \right\} \right).$$

Pode-se provar que ε_j 's são aproximadamente *i.i.d* Gumbel $(0, \pi^2/6)$.

Portanto, o estimador GPH de d é dado por

$$\hat{d}_p = -\hat{b} = \frac{-\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2}, \quad (9)$$

obtido pelo método de mínimos quadrados da regressão de $y_1, y_2, \dots, y_{g(n)}$ em $x_1, x_2, \dots, x_{g(n)}$.

Assintoticamente, o valor esperado e a variância de \hat{d}_p são, respectivamente, $E(\hat{d}_p) = d$ e

$$\sigma_{\hat{d}_p}^2 = \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (x_j - \bar{x})^2}, \text{ e sua distribuição assintótica é } \frac{\hat{d}_p - d}{\sqrt{\sigma_{\hat{d}_p}^2}} \rightarrow N(0,1).$$

Regressão Utilizando a Função Periodograma Suavizado

Como o estimador $I(\omega)$ de $f(\omega)$ é inconsistente (Baillie, 1996), uma solução alternativa para estimar d foi proposta por Reisen (1994), que utilizou a função periodograma suavizado $f_s(\omega)$, (6), no lugar de $I(\omega)$ em (7). A função $f_s(\omega)$ é obtida através da janela de Parzen para um coeficiente de suavização $m = n^\beta$ (consideramos $\beta = 0.9$, ver Reisen (1994)). Doukhan et al. (2003, página 263) chamam a atenção para o fato de que, embora este estimador não seja amplamente utilizado, ele foi o primeiro para o qual uma teoria assintótica foi rigorosamente estabelecida.

O estimador de d obtido pelo método da regressão usando a função periodograma suavizado, com a janela de Parzen, é dado por

$$\hat{d}_{sp} = -\hat{b} = \frac{-\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2} \quad (10)$$

onde $y_j = \ln f_s(\omega_j)$ e $\sigma_{\hat{d}_{sp}}^2 = 0,53928 \frac{m}{n \sum_{j=1}^{g(n)} (x_j - \bar{x})^2}$, $\omega \neq 0, \pi$ (ver Reisen, 1995).

Para este estimador temos, também, uma distribuição assintótica normal dada por

$$\frac{\hat{d}_{sp} - d}{\sqrt{\sigma_{\hat{d}_{sp}}^2}} \rightarrow N(0,1).$$

3. O MÉTODO *BOOTSTRAP*

Devido à não independência das observações, o uso de *bootstrap* em séries temporais deve ser cercado de cuidados (Efron, 1979). A série deve ser reamostrada “indiretamente”, pois caso contrário sua estrutura original pode ser perdida nas séries reamostradas. Existem várias abordagens *bootstrap* em modelos ARFIMA(p, d, q) e neste trabalho utilizaremos o *Bootstrap* nos resíduos do modelo ajustado (Franco & Reisen, 2004), que será denominado aqui de *Bootstrap* pseudo-paramétrico.

3.1 *Bootstrap* pseudo-paramétrico e intervalos de confiança

Seja X_t uma série de n observações modelada por um ARFIMA(p, d, q). Após estimar os parâmetros dos vetores ϕ_p , θ_q e d , os resíduos amostrais são obtidos através de

$$\hat{a}_t = \hat{\theta}_q^{-1}(B) \hat{\phi}_p(B) (1-B)^d X_t. \quad (11)$$

Este resíduo deve ser centralizado, diminuindo-se sua média e dividindo pelo desvio-padrão. O resíduo centralizado será denotado por \tilde{a}_t . Em seguida, reamostra-se \tilde{a}_t com reposição, atribuindo a cada \tilde{a}_t , $t=1, 2, \dots, n$, uma massa de probabilidade igual a $1/n$, obtendo assim os resíduos *bootstrap* a_t^* . Então é possível construir recursivamente a série *bootstrap* X_t^* a partir de a_t^* , da seguinte maneira

$$X_t^* = \hat{\theta}_q(B) \hat{\phi}_p^{-1}(B) (1-B)^{-d} a_t^*. \quad (12)$$

Esta é uma abordagem pseudo-paramétrica do método de reamostragem *bootstrap* por não utilizar nenhuma suposição sobre a distribuição dos resíduos para reamostrá-los.

Existem alguns métodos para encontrar intervalos de confiança para um parâmetro desconhecido, tais como os métodos da quantidade pivotal, assintótico e da inversão da função de distribuição, porém contendo algumas restrições como, por exemplo, a dificuldade em se encontrar a quantidade pivotal ou até mesmo, a imprecisão causada por aproximações através da distribuição assintótica. Assim sendo, Efron e Tibshirani (1993) propuseram métodos em que se utiliza o *Bootstrap* para a construção dos intervalos de confiança mais precisos e de maneira mais simplificada, utilizando a distribuição *bootstrap* da quantidade pivotal.

Intervalo *Bootstrap* Percentílico

Os intervalos de confiança *bootstrap* percentílico são intervalos simples de serem obtidos através das replicações *bootstrap* de uma série original.

A princípio são geradas B amostras *bootstrap* da série original, sendo que são estimados, para cada replicação, os valores do parâmetro d de interesse, ordenando-se em seguida essas estimativas. Então, podemos definir os limites inferior e superior de um intervalo de confiança *bootstrap* percentílico ao nível de $(1-2\alpha)$ como sendo

$$\left[\hat{d}^{*(\alpha)} \quad ; \quad \hat{d}^{*(1-\alpha)} \right] \quad (13)$$

onde $\hat{d}^{*(\alpha)}$ é o $100.\alpha$ -ésimo valor das estimativas ordenadas obtidas através das B replicações *bootstrap*.

Intervalo *Bootstrap-t*

Através do uso das técnicas de reamostragem *bootstrap* podemos obter intervalos de confiança precisos sem ter que assumir a normalidade dos dados, estimando a distribuição diretamente a partir dos mesmos, construindo então uma tabela com os quantis. Em posse desta tabela de quantis podemos construir intervalos de confiança assim como fazemos usando as tabelas normal padronizada e t-Student.

Para este procedimento, inicialmente devemos gerar **B** amostras *bootstrap* da série original. Para cada série *bootstrap*, calculamos a estatística

$$T_{(b)}^* = \frac{(\hat{d}_{(b)}^* - \hat{d})}{\hat{se}_{\hat{d}}^*(b)} \quad b = 1, \dots, B \quad (14)$$

onde $\hat{se}^*(b)$ é o erro padrão estimado de d^* para a b -ésima série *bootstrap* X_t^* . Note que para obter o erro padrão $\hat{se}^*(b)$ devemos reamostrar cada série *bootstrap* B_2 vezes. Ou seja, para este procedimento é necessário realizar $B \times B_2$ replicações *bootstrap*. Este método constrói uma tabela com os percentis da distribuição empírica $T_{(b)}^*$, obtida da geração das **B** amostras *bootstrap*.

O intervalo de confiança *bootstrap-t* é dado por

$$\left[\hat{d} - \hat{t}^{(1-\alpha)} \cdot \hat{se}, \quad \hat{d} - \hat{t}^{(\alpha)} \cdot \hat{se} \right] \quad (15)$$

onde $\hat{t}^{(\alpha)}$ e $\hat{t}^{(1-\alpha)}$ são, respectivamente, o $100\alpha\%$ e o $100(1-\alpha)\%$ percentil da distribuição empírica de T^* e \hat{se} é o erro padrão de \hat{d} , que pode ser obtido calculando-se o erro padrão das **B** amostras *bootstrap*.

3.2 Resultados das simulações

Através de procedimentos Monte Carlo foram simuladas inicialmente 500 séries do modelo ARFIMA(0,d,0), com tamanho $n = 300$ e parâmetro $d = 0,3, 0,5$ e $1,0$. A linguagem de programação utilizada foi o FORTRAN. As estatísticas calculadas para avaliar a performance dos estimadores e dos métodos *bootstrap* foram o vício = (valor estimado – valor real) e o Erro Quadrático Médio (EQM). Os intervalos de confiança (IC) foram comparados através do Percentual de Cobertura (PC), para um nível nominal de 95% e do tamanho $L = \text{lsup-linf}$ dos mesmos.

O processo de obtenção das séries *bootstrap* e respectivas estimativas segue o esquema abaixo:

1. Para cada uma das séries geradas, foram encontrados os estimadores GPH e SPR para d e o vetor de resíduos \hat{a}_t ;
2. Utilizando os estimadores GPH e SPR da série original e os resíduos centralizados reamostrados com reposição, obtemos a série *bootstrap*;
3. Para cada uma das séries *bootstrap* foram calculados o estimador GPH, denotado por GPHBoot, e o estimador SPR, denotado por SPRBoot.
4. Foram construídos intervalos de confiança percentílico e *bootstrap-t* para d . A performance dos intervalos foi avaliada através do percentual de cobertura dos mesmos.

Para cada série foram realizados $B_1=500$ *bootstraps* e, para cada um destes, um segundo *bootstrap*, com $B_2=100$, para construção dos intervalos de confiança *bootstrap-t*. O vício e EQM dos estimadores são apresentados na Tabela 1 e os Intervalos de Confiança para os três tamanhos de séries nas Tabelas 2 a 4. Os menores vícios e EQM's estão em negrito, assim como os Percentuais de Cobertura (PC) mais próximos de 95% e os menores tamanhos dos IC.

Através da Tabela 1 podemos ver que (como também observado por Reisen et al., 2001) o estimador GPH apresenta menor vício, mas maior EQM que o SPR, exceto para $d=1,0$, onde o SPR apresentou menor vício e menor EQM. Podemos observar que o *bootstrap* aproxima bem os resultados de cada estimador com competitivos valores dos EQM's. Foi apresentada somente a tabela para o tamanho da série $n=100$, pois os outros tamanhos de amostra apresentam comportamento similar.

Tabela 1: Vício e EQM para os estimadores GPH e SPR de d e seus respectivos *bootstraps* (n=100)

d real	Estimador	d	Vício	EQM
d=0,3	GPH	0,29895	-0,00105	0,01455
	GPHBoot	0,30529	0,00529	0,01546
	SPR	0,27045	-0,02955	0,01134
	SPRBoot	0,24906	-0,05094	0,01334
d=0,5	GPH	0,50564	0,00564	0,01775
	GPHBoot	0,53208	0,03208	0,01946
	SPR	0,47873	-0,02127	0,01189
	SPRBoot	0,47694	-0,02306	0,01265
d=1,0	GPH	1,00843	0,00843	0,01335
	GPHBoot	0,99802	-0,00198	0,00329
	SPR	0,99962	-0,00038	0,00942
	SPRBoot	1,00156	0,00156	0,00432

Observando a Tabela 2 (n =50) percebemos que os intervalos que apresentam menor tamanho são os que possuem Percentuais de Cobertura (PC) mais distantes do nível nominal de 95%. Assim, no geral, o intervalo normal é o que apresenta menor tamanho, porém o intervalo percentílico apresenta PC mais próximo de 95%. Comparando os intervalos construídos para o GPH e o SPR, observamos que o primeiro apresenta PC mais próximo do valor nominal de 95%, mas o segundo apresenta menor tamanho.

Tabela 2: Intervalos de confiança de 95% normal, percentílico e *bootstrap-t* para d . (n=50)

d real	Estimador	IC Normal	PC	IC Bootstrap-t	PC	IC Percentílico	PC
d=0,3	GPH	[-0,061 ; 0,683] (L = 0,745)	97	[-0,052 ; 0,677] (L = 0,729)	98,8	[-0,089 ; 0,666] (L = 0,755)	97
	SPR	[0,092 ; 0,438] (L = 0,346)	77,4	[0,074 ; 0,652] (L = 0,578)	91	[-0,098 ; 0,506] (L = 0,604)	92,2
d=0,5	GPH	[0,138 ; 0,883] (L = 0,745)	97	[0,104 ; 0,859] (L = 0,756)	97,6	[0,127 ; 0,887] (L = 0,760)	96
	SPR	[0,290 ; 0,637] (L = 0,346)	74,8	[0,242 ; 0,839] (L = 0,598)	92,4	[0,113 ; 0,730] (L = 0,617)	92,2
d=1,0	GPH	[0,614 ; 1,358] (L = 0,745)	95	[0,689 ; 1,367] (L = 0,679)	81,4	[0,613 ; 1,276] (L = 0,666)	98,6
	SPR	[0,803 ; 1,149] (L = 0,346)	79,4	[0,735 ; 1,326] (L = 0,593)	79,8	[0,652 ; 1,195] (L = 0,543)	97

Obs.: PC é o Percentual de Cobertura do intervalo e L é o tamanho do intervalo (lsup-linf)

O mesmo comportamento observado para n=50 é observado nas Tabelas 3 e 4, onde n=100 e 300, ou seja, intervalos que possuem melhor cobertura, possuem também maior tamanho. Comparando os Percentuais de Cobertura percebemos que, para n=100, o IC Normal apresenta melhores resultados para o GPH e o Percentílico é melhor para o SPR e, para n=300, o intervalo percentílico tem melhor performance para os dois estimadores. Comparando os intervalos construídos para o GPH e o SPR, observamos novamente que o primeiro apresenta PC mais próximo do valor nominal de 95%, mas o segundo apresenta menor tamanho. Podemos observar também que o tamanho dos intervalos decresce à medida que o tamanho da amostra aumenta e que, para o GPH, os Percentuais de Cobertura se tornam mais próximos de 95%, principalmente para amostras de tamanho n=300.

Tabela 3: Intervalos de confiança de 95% normal, percentílico e *bootstrap-t* para d. (n=100)

d real	Estimador	IC Normal	PC	IC Bootstrap-t	PC	IC Percentílico	PC
d=0,3	GPH	[0,057 ; 0,561] (L = 0,505)	95,2	[0,067 ; 0,558] (L = 0,492)	97	[0,040 ; 0,551] (L = 0,511)	94,2
	SPR	[0,169 ; 0,398] (L = 0,229)	74,4	[0,138 ; 0,538] (L = 0,400)	88,8	[0,039 ; 0,455] (L = 0,417)	93
d=0,5	GPH	[0,263 ; 0,767] (L = 0,505)	93,8	[0,240 ; 0,748] (L = 0,508)	97	[0,260 ; 0,779] (L = 0,520)	93,8
	SPR	[0,370 ; 0,600] (L = 0,229)	73,6	[0,320 ; 0,733] (L = 0,413)	90,4	[0,250 ; 0,680] (L = 0,429)	93,2
d=1,0	GPH	[0,741 ; 1,246] (L = 0,505)	95,2	[0,791 ; 1,252] (L = 0,460)	82,8	[0,740 ; 1,196] (L = 0,456)	99,2
	SPR	[0,878 ; 1,108] (L = 0,229)	76,4	[0,812 ; 1,206] (L = 0,394)	81,6	[0,780 ; 1,151] (L = 0,372)	98,4

Obs.: PC é o Percentual de Cobertura do intervalo e L é o tamanho do intervalo (lsup-linf)

Tabela 4: Intervalos de confiança de 95% normal, percentílico e *bootstrap-t* para d. (n=300)

d real	Estimador	IC Normal	PC	IC Bootstrap-t	PC	IC Percentílico	PC
d=0,3	GPH	[0,152 ; 0,444] (L = 0,293)	95,7	[0,156 ; 0,447] (L = 0,291)	95,3	[0,142 ; 0,438] (L = 0,295)	95,3
	SPR	[0,223 ; 0,349] (L = 0,126)	66,3	[0,192 ; 0,431] (L = 0,239)	90,3	[0,147 ; 0,391] (L = 0,244)	92
d=0,5	GPH	[0,362 ; 0,654] (L = 0,293)	93	[0,351 ; 0,649] (L = 0,298)	96,3	[0,362 ; 0,666] (L = 0,305)	94
	SPR	[0,433 ; 0,559] (L = 0,126)	68	[0,391 ; 0,633] (L = 0,243)	88	[0,365 ; 0,614] (L = 0,249)	92,7
d=1,0	GPH	[0,855 ; 1,148] (L = 0,293)	94,3	[0,884 ; 1,200] (L = 0,317)	84,0	[0,857 ; 1,121] (L = 0,264)	99,3
	SPR	[0,945 ; 1,071] (L = 0,126)	72	[0,890 ; 1,116] (L = 0,226)	81,3	[0,893 ; 1,102] (L = 0,209)	99

Obs.: PC é o Percentual de Cobertura do intervalo e L é o tamanho do intervalo (lsup-linf)

Para fazer uma comparação final, devemos notar que o fato de um intervalo apresentar menor tamanho, mas possuir pior Percentual de Cobertura torna este intervalo menos aceitável, pois apesar de ele ser menor, ele não cobre o verdadeiro valor do parâmetro. Diante desta perspectiva, podemos considerar o Intervalo Percentílico mais apropriado neste caso, pois no geral ele apresentou Percentuais de Cobertura mais próximos do nível nominal de 95%.

4. APLICAÇÃO

Vamos analisar a série com índices mensais de pregões ocorridos na Bolsa de Valores de São Paulo (BOVESPA), entre setembro de 1994 e julho de 2003. A série é composta por 107 observações mensais. A Figura 1 indica não haver o efeito de sazonalidade nem a presença de tendência. O objetivo nesta seção não é encontrar o melhor modelo para descrever a série, mas simplesmente verificar se esta apresenta a característica de longa dependência e, também, aplicar os métodos de construção de intervalos *bootstrap* propostos e investigados, empiricamente, neste artigo.

Os estimadores de d obtidos através dos métodos GPH e SPR, usando $\alpha = 0,5$ e $0,8$, estão apresentados na Tabela 5. Todos os intervalos incluem o valor zero, portanto os testes não indicam evidência significativa para rejeitar a hipótese de que esta série não possui a característica de longa dependência. Notamos uma diferença entre as estimativas de d obtidas pelos diferentes *bandwidths* ($\alpha=0,5$ e $\alpha=0,8$), especialmente os obtidos pelo método GPH. Estas diferenças podem indicar que o modelo possui componentes AR e/ou MA. Como isto não é o objetivo neste trabalho, deixamos esta questão para ser explorada nos estudos futuros. Um estudo empírico com o objetivo de verificar a existência ou não de componentes AR e/ou MA em dados reais é considerado em Reisen, Cribari e Jensen (2003). Neste artigo, os autores analisam a série índice inflacionário brasileiro através de várias abordagens relacionadas ao modelo ARFIMA. Neste contexto, outras referências também são citadas pelos autores.

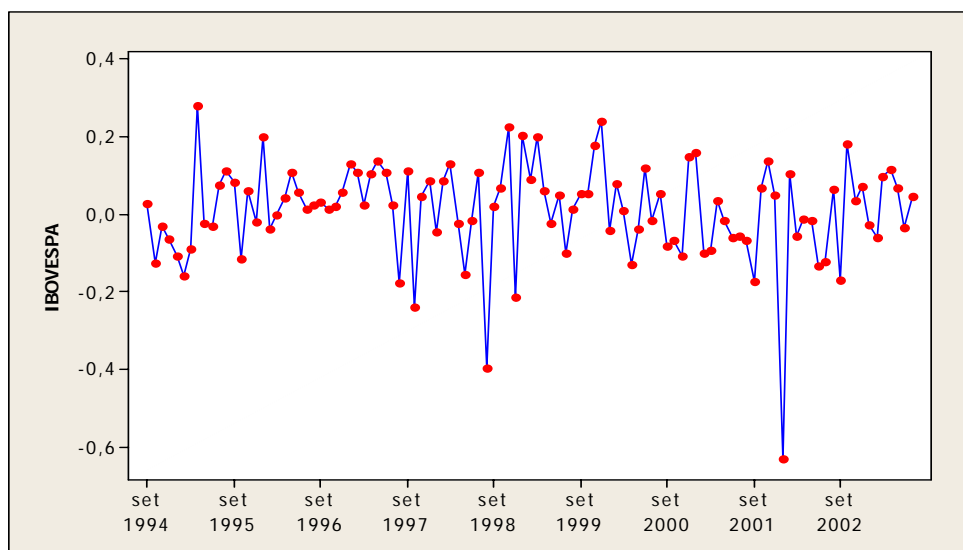


Tabela 5: Intervalos de confiança de 95% normal, percentílico e *bootstrap-t* para d . (BOVESPA)

α	Método	\hat{d}	IC Normal	IC Bootstrap-t	IC Percentílico
0,5	GPH	0,211	[-0,403 ; 0,746] (L = 1,149)	[-0,380 ; 0,792] (L = 1,172)	[-0,544 ; 0,730] (L = 1,274)
	SPR	0,028	[-0,141 ; 0,380] (L = 0,521)	[-0,133 ; 0,857] (L = 0,990)	[-0,535 ; 0,401] (L = 0,936)
0,8	GPH	0,085	[-0,227 ; 0,392] (L = 0,619)	[-0,285 ; 0,451] (L = 0,736)	[-0,345 ; 0,394] (L = 0,739)
	SPR	-0,061	[-0,147 ; 0,134] (L = 0,281)	[-0,057 ; 0,408] (L = 0,465)	[-0,382 ; 0,129] (L = 0,511)

5. CONCLUSÕES

Neste projeto utilizamos o *bootstrap* não-paramétrico nos resíduos para fazer inferência sobre o parâmetro d dos modelos ARFIMA(0, d ,0). Os estimadores de d utilizados foram os semiparamétricos propostos por Geweke and Porter-Hudak (1983) e Reisen (1994). Foram construídos também intervalos de confiança *bootstrap* percentílico e *bootstrap-t* e sua performance comparada através do Percentual de Cobertura (PC) e do tamanho dos intervalos. Os resultados obtidos através de simulações Monte Carlo mostram que o método *bootstrap* empregado aproxima de forma satisfatória o comportamento dos estimadores utilizados, tanto analisando o vício quanto o

Erro Quadrático Médio dos mesmos. As simulações mostram também que os intervalos *bootstrap* apresentam bom PC, sendo que o intervalo percentílico tem uma performance superior aos intervalos normal e *bootstrap-t* e os intervalos para o GPH são mais próximos do nível nominal que o SPR.

Agradecimentos

Os autores agradecem o apoio do CNPq-Brasil. Parte dos resultados apresentados neste artigo foram obtidos no período de bolsa PIBIC-CNPq de Neander F. Almeida.

Referências Bibliográficas

- Baillie, R. T. (1996). Long Memory Time Series Models and Fractional Integration in Econometrics. *Journal of Econometrics*, **73**, 5-59.
- Box, G. E. P., Jenkins G. M. and Reinsel G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Ed. Prentice – Hall, Inc.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, **7**, 1-26.
- Doukham, P., Oppenheim, G. e Taqqu, M.S. (2003). *Theory and applications of long-range dependence*. Birkhäuser, Boston.
- Efron, B and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameters estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517-532.
- Franco, G. C. and Reisen, V. A. (2004). Bootstrap Techniques in Semiparametric Estimation Methods for ARFIMA Models: A Comparison Study. *Computational Statistics*, **19**, 243-259.
- Geweke, J. and Porter-Hudak, S. (1983). The Estimation and Application of Long Memory Time Series Model. *Journal of Time Series Analysis*, **4**(4), 221-238.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, **68**(1), 165-175.
- Lopes, S. R C, Olbermann, B. P. A. and Reisen, V. A. (2004). Comparison of estimation methods in non-stationary ARFIMA process. *Jr. of Statistical Computation and Simulation*, **74**(5) 339-347.
- Doukhan, P., Oppenheim, G. and Taqqu, M. S. (2003). *Theory and Application of Long-Range Dependency*, London, Birkhauser.
- Reisen, V. A. (1994). Estimation of the Fractional Difference Parameter in the ARIMA(p,d,q) Model using the Smoothed Periodogram. *Journal of Time Series Analysis*, **15**(3), 335-350.
- Reisen, V. A. (1995). *ARFIMA – O Modelo ARIMA para o d Fracionário* In: 6^a. Escola de Séries Temporais e Econometria, UFES, Vitória.
- Reisen, V. A., Neto-Cribari, F. and Jensen, M.J.S. (2003). Long memory inflationary dynamics. The case of Brasil. *Studies in nonlinear dynamics and econometrics*, **7**(3), 1-18.
- Reisen, V. A., Abraham, B. and Lopes, S. (2001). Estimation of parameters in ARFIMA process. Simulation study. *Communications in Statistics. Simulation and Computation*, **30**(4), 787-803.
- Robinson, P. M. (1995). Log-Periodogram Regression of Time Series with Long Range Dependence. *Annals of Statistics*, **23**(3), 1048-1072.
- Sowell, F. (1992). Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics*, **53**, 165-188.