Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Número de regressores do Método DFA

Raquel Romes Linhares [1]
Sílvia Regina Costa Lopes [2]

1 Introdução

O método da *análise de flutuações destendenciadas* (*Detrended Fluctuation Analysis* - DFA), proposto por Peng et al. (1994), é um exemplo de metodologia recente, sendo utilizada em um crescente número de aplicações, para identificar longa dependência em séries temporais (ver Peng et al., 1994; Ben-Avraham e Havlin, 2000; Bardet e Kammoun, 2008; Crato et al., 2010). O método DFA tem como objetivo o cálculo de uma flutuação estatística $F(l)$, onde l representa o tamanho de uma janela, para mapear um conjunto de medidas. Variando o tamanho de l, as flutuações podem ser caracterizadas através de um expoente de escala obtido a partir da curva ajustada ao gráfico $\ln(F(l))$ versus $\ln(l)$, com $l \in \{4, 5, \cdots, g(n)\}$. Para aplicar o método DFA, primeiro divide-se a série temporal em blocos com l observações. Em cada bloco, calculam-se as somas parciais $\{Y_t\}_{t=1}^{l}$, ajusta-se uma reta $Y_t^l = a + bt$ para estas somas parciais e então calcula-se a soma $\sum_{t=1}^{l}(Y_t - \widehat{Y}_t^l)^2$, para cada l observações. Crato et al. (2010) provam a consistência do estimador DFA, baseados nas condições usuais $i.i.d.$, considerando $l \in \{4, 5, \cdots, g(n)\}$. Na literatura a escolha mais utilizada para $g(n)$ é $\lfloor \frac{n}{10} \rfloor$, mas não existe uma regra específica para esta escolha (ver Hu et al., 2001). Aqui realizamos algumas simulações de Monte Carlo em modelo ARFIMA $(0, d, 0)$, para investigar o efeito do número de regressores do Método DFA. O trabalho está organizado da seguinte forma. Na Seção 2, apresentamos os processos ARFIMA e descrevemos o método DFA. Os resultados e discussões são apresentados na Seção 3. Seção 4 dá as conclusões.

2 Métodos

2.1 ARFIMA(p, d, q)

Definição 2.1 (Processos ARFIMA(p, d, q)) Sejam $\{\varepsilon_t\}_{t \in Z}$ um processo ruído branco e $\mathcal{B}$ o *operador defasagem*, i.e., $\mathcal{B}^k(X_t) = X_{t-k}$. Se $\{X_t\}_{t \in Z}$ é um processo estacionário tal que

$$\Phi(\mathcal{B})(1 - \mathcal{B})^d (X_t - \mu) = \Theta(\mathcal{B})\varepsilon_t, \; t \in Z, \tag{2.1}$$

[1]FAMAT - UFU. e-mail: *raquelromeslinhares@gmail.com*
[2]PPGMAT - UFRGS.

para $d \in (-0.5, 0.5)$, onde μ é a média do processo e $\Phi(\cdot)$ e $\Theta(\cdot)$ são os polinômios autoregressivos e média móvel, de ordens p e q, respectivamente, então $\{X_t\}_{t \in Z}$ é denominado um *processo geral com diferenciação fracionária*, denotado por ARFIMA(p,d,q).

Se $0 < d < 0.5$, o processo ARFIMA(p,d,q) apresenta a propriedade de *longa dependência*.

2.2 DFA

O método das análises de flutuações destendenciadas ("*Detrended Fluctuation Analysis*" - DFA) foi proposto por Peng et al. (1994). Para utilizar este método em uma série temporal qualquer é preciso seguir os seguintes passos. Dada uma série temporal $\{X_t\}_{t=1}^n$, o primeiro passo do método DFA consiste em calcular a soma parcial $Y_t = \sum_{j=1}^{t} X_j$ para cada $t \in \{1, 2, \cdots, n\}$. No segundo passo divide-se a série temporal $\{Y_t\}_{t=1}^n$ em $\lfloor \frac{n}{l} \rfloor$ blocos não sobrepostos ($\lfloor \cdot \rfloor$ é a função maior inteiro), onde cada bloco contém exatamente l observações. No terceiro passo, para cada bloco ajusta-se, pelo método dos mínimos quadrados, uma reta aos dados. Denotamos por $\widehat{Y}_t^l$, para $t = 1, \cdots, n$, a ordenada y dos segmentos de reta ajustados nos blocos de tamanho l. Em seguida, destendencia-se a série temporal $\{Y_t\}_{t=1}^n$, ou seja, em cada bloco calcula-se $Z_t^l = Y_t - \widehat{Y}_t^l$, para todo $t \in \{1, \cdots, n\}$. O desvio padrão dos resíduos obtidos da diferença entre Y_t e $\widehat{Y}_t^l$ é denominado de *raíz da flutuação média quadrática* e é definido por

$$\mathrm{F}(l) = \sqrt{\frac{1}{\tilde{n}} \sum_{t=1}^{\tilde{n}} (Z_t^l)^2}, \tag{2.2}$$

onde $\tilde{n} = l \cdot \lfloor \frac{n}{l} \rfloor$. Por fim, para cada $l \in \{4, 5, \cdots, g(n)\}$, calcula-se *a raíz da flutuação média quadrática*, dada na expressão (2.2). Observe que $F(l)$ aumenta quando l cresce. Uma relação linear em um gráfico ln versus ln indica presença de escala

$$\mathrm{F}(l) \sim \varphi l^{\beta}. \tag{2.3}$$

Aplicando logaritmo em ambos os lados da expressão (2.3) obtemos

$$\ln(\mathrm{F}(l)) \sim \ln(\varphi) + \beta \ln(l). \tag{2.4}$$

A expressão (2.4) é da forma de uma equação de regressão linear simples, dada por

$$y_j = a + b x_j + \varepsilon_j, \qquad j \in \{1, \cdots, m\}, \tag{2.5}$$

onde $y_j = \ln(\mathrm{F}(l))$, $a = \ln(\varphi)$, $b = \mathrm{H}$, $x_j = \ln(l)$, $l = j + 3$, com $l \in \{4, 5, \cdots, g(n)\}$ e $m = g(n) - 3$. Não consideramos $l = 3$ pelo aumento significativo no tempo computacional. A relação da escala β com o parâmetro H do efeito de Hurst (ver Hurst, 1951) da série temporal

$\{X_t\}_{t=1}^n$ é dada por $\beta = \mathrm{H}$. Segue-se então que o estimador de H, pelo método dos mínimos quadrados baseado na regressão linear (2.5), é dado por

$$\widehat{\mathrm{H}}_{\mathrm{DFA}} = \frac{\sum_{j=1}^{m}(x_j - \overline{x})y_j}{\sum_{j=1}^{m}(x_j - \overline{x})^2}, \tag{2.6}$$

onde $y_j = \ln(\mathrm{F}(j+3))$, $x_j = \ln(j+3)$, $\overline{x} = \frac{1}{m}\sum_{j=1}^{m} x_j$ e $m = g(n) - 3$. O estimador para parâmetro fracionário d, pelo método DFA, é dado por $\hat{d}_{\mathrm{DFA}} = \widehat{\mathrm{H}}_{\mathrm{DFA}} - \frac{1}{2}$, onde $\widehat{\mathrm{H}}_{\mathrm{DFA}}$ é dado pela expressão 2.6.

3 Resultados e discussões

Tabela 3.1: Estimador $\hat{d}_{DFA}$ para ARFIMA$(0, 0.1, 0)$ com Diferentes $g(n)$.

$n = 1000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.2391	0.1554	0.1222	0.1068	0.0995	0.1015
vício	0.1391	0.0554	0.0222	0.0068	0.0005	0.0015
eqm	0.0229	0.0042	0.0014	0.0012	0.0018	0.0017
var	0.0035	0.0012	0.0009	0.0011	0.0018	0.0017
$n = 5000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.1743	0.1243	0.1076	0.1004	0.0948	0.0918
vício	0.0743	0.0243	0.0076	0.0004	-0.0052	-0.0082
eqm	0.0058	0.0008	0.0003	0.0004	0.0009	0.0012
var	0.0003	0.0002	0.0003	0.0004	0.0008	0.0011
$n = 10000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.1560	0.1177	0.1034	0.0986	0.0970	0.0970
vício	0.0560	0.0177	0.0034	-0.0014	-0.0030	-0.0030
eqm	0.0032	0.0004	0.0001	0.0003	0.0007	0.0009
var	0.0001	0.0001	0.0001	0.0003	0.0007	0.0009

Para investigar o efeito de $g(n) = \lfloor n^{\beta} \rfloor$, para algum $0 < \beta < 1$, no método da análise de flutuações destendenciadas, geramos séries temporais obtidas de processos ARFIMA$(0, d, 0)$, com 200 replicações e tamanho $n \in \{1000, 5000, 10000\}$.

Consideramos diferentes valores para $g(n)$, incluindo o valor mais utilizado na literatura, dado por $\lfloor \frac{n}{10} \rfloor$ (ver Hu et al., 2001). Para cada $g(n)$, calculamos o valor médio ($\hat{d}_{DFA}$), o vício, o erro quadrático médio (eqm) e a variância (var). Os resultados dos experimentos, são dados nas Tabelas 3.1 a 3.3. Podemos observar nas Tabelas 3.1 a 3.3, que para cada $g(n)$ fixo, quando o tamanho da amostra cresce, o erro quadrático médio e a variância tendem a zero. Em nosso experimento, $g(n) = \lfloor n^{0.3} \rfloor$, parece gerar maior vício para o parâmetro d. Observe que a escolha

Tabela 3.2: Estimador $\hat{d}_{DFA}$ para ARFIMA$(0, 0.2, 0)$ com Diferentes $g(n)$.

$n = 1000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.3031	0.2332	0.2099	0.2018	0.1939	0.1973
vício	0.1031	0.0332	0.0099	0.0018	-0.0061	-0.0027
eqm	0.0144	0.0025	0.0014	0.0015	0.0018	0.0018
var	0.0038	0.0014	0.0013	0.0015	0.0018	0.0018
$n = 5000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.2454	0.2057	0.1976	0.1955	0.1907	0.1883
vício	0.0454	0.0057	-0.0024	-0.0045	-0.0093	-0.0117
eqm	0.0023	0.0002	0.0003	0.0005	0.0010	0.0014
var	0.0003	0.0002	0.0002	0.0004	0.0010	0.0012
$n = 10000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.2300	0.2026	0.1966	0.1959	0.1965	0.1953
vício	0.0300	0.0026	-0.0034	-0.0041	-0.0035	-0.0047
eqm	0.0010	0.0001	0.0002	0.0003	0.0007	0.0012
var	0.0001	0.0001	0.0002	0.0003	0.0007	0.0012

Tabela 3.3: Estimador $\hat{d}_{DFA}$ para ARFIMA$(0, 0.3, 0)$ com Diferentes $g(n)$.

$n = 1000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.3625	0.3088	0.2937	0.2890	0.2839	0.2868
vício	0.0625	0.0088	-0.0063	-0.0110	-0.0161	-0.0132
eqm	0.0086	0.0014	0.0014	0.0020	0.0026	0.0024
var	0.0047	0.0013	0.0014	0.0019	0.0024	0.0022
$n = 5000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.3211	0.2930	0.2893	0.2934	0.2921	0.2900
vício	0.0211	-0.0070	-0.0107	-0.0066	-0.0079	-0.0100
eqm	0.0008	0.0003	0.0004	0.0006	0.0011	0.0015
var	0.0003	0.0003	0.0003	0.0006	0.0011	0.0014
$n = 10000$						
$g(n)$	$\lfloor n^{0.3} \rfloor$	$\lfloor n^{0.4} \rfloor$	$\lfloor n^{0.5} \rfloor$	$\lfloor n^{0.6} \rfloor$	$\lfloor n^{0.7} \rfloor$	$\lfloor \frac{n}{10} \rfloor$
$\hat{d}_{DFA}$	0.3102	0.2935	0.2925	0.2938	0.2959	0.2969
vício	0.0102	-0.0065	-0.0075	-0.0062	-0.0041	-0.0031
eqm	0.0003	0.0002	0.0002	0.0004	0.0008	0.0013
var	0.0002	0.0001	0.0002	0.0003	0.0008	0.0013

mais utilizada na literatura, $g(n) = \lfloor \frac{n}{10} \rfloor$, gera boas estimativas para o parâmetro d. No entanto, $g(n) = \lfloor n^{\beta} \rfloor$ apresenta melhores resultados do que $g(n) = \lfloor \frac{n}{10} \rfloor$, para $d \in \{0.1, 0.2, 0.3\}$ se $\beta \in \{0.5, 0.6, 0.7\}$. Para n suficientemente grande, a escolha $g(n) = \lfloor n^{\beta} \rfloor$, sob o ponto de vista computacional, tem a vantagem de exigir um número menor de regressores no método DFA.

4 Conclusões

Para investigar o efeito de $g(n) = \lfloor n^{\beta} \rfloor$, para algum $0 < \beta < 1$, no método da análise de flutuações destendenciadas, geramos séries temporais obtidas de processos ARFIMA$(0,d,0)$. Consideramos diferentes valores para $g(n)$, incluindo aquele mais utilizado na literatura, dado por $\lfloor \frac{n}{10} \rfloor$. Em nosso experimento, $g(n) = \lfloor n^{0.3} \rfloor$, parece gerar maior vício. Observamos que a escolha mais utilizada na literatura, $g(n) = \lfloor \frac{n}{10} \rfloor$, gera boas estimativas para o parâmetro d. No entanto, $g(n) = \lfloor n^{\beta} \rfloor$ apresenta melhores resultados do que $g(n) = \lfloor \frac{n}{10} \rfloor$, para $d \in \{0.1, 0.2, 0.3\}$ se $\beta \in \{0.5, 0.6, 0.7\}$. Para n suficientemente grande, a escolha $g(n) = \lfloor n^{\beta} \rfloor$, para finalidades computacionais, tem a vantagem de exigir um número menor de regressores no método DFA.

Referências

[1] BEN-AVRAHAN; D. e HAVLIN, S. **Diffusion and Reactions in Fractals and Disordered Systems**. Cambridge: Cambridge University Press, 2000.

[2] BARDET, J.M. e KAMMOUN, I. Asymptotic Properties of the Detrended Fluctuation Analysis of Long-Range Dependence Processes. **IEEE Transactions on Information Theory**, v. 54(5), p. 2041-2052, 2008.

[3] CRATO, N., LINHARES; R.R. e LOPES, S.R.C. Statistical Properties of Detrended Fluctuation Analysis. **Journal of Statistical Computation and Simulation**, v. 80(6), p. 625-641, 2010.

[4] HU, K.; IVANOV, P.C; CHEN, Z.; CARPENA, P. e STANLEY, H.E. Effect of Trends on Detrended Fluctuation Analysis. **Phyical Review E**, v. 64, 011114, 2001.

[5] HURST, H.R. Long-Term Storage in Reservoirs. **Trans. Am. Soc. Civil Eng.**, v. 116, p. 770-799, 1951.

[6] PENG, C.; BULDYREV, S. H.; SIMONS, M.; STANLEY, H.E., e GOLDBERGER, A.L. Mosaic Organization of DNA Nucleotides. **Physical Review E**, v.49(5), p. 1685-1689, 1994.