

MODELAGEM DO TESTE ESTATÍSTICO DO COEFICIENTE DE CORRELAÇÃO CRUZADA SEM TENDÊNCIA ρ DCCA: TEORIA E APLICAÇÃO

Aloisio Machado da Silva Filho¹, Gilney Figueira Zebende^{1,2}, Arleys Pereira Nunes de Castro¹ e Marcus Fernandes da Silva^{3,4}

SENAI CIMATEC¹
Universidade Estadual de Feira de Santana - UEFS²
Instituto Federal de Educação, Ciência e Tecnologia da Bahia – IFBA³
Universidade Federal da Bahia – UFBA⁴

E-mails: aloisioestatistico@yahoo.com.br, gfzebende@hotmail.com,
arleys.castro@gmail.com, marcusfisico@yahoo.com.br

RESUMO

O objetivo deste artigo é reproduzir o teste estatístico idealizado por Podobnik *et al*, 2011 [1] para o coeficiente de correlação cruzada sem tendência idealizado por Zebende, 2011 [2] e, além disso, propor uma nova metodologia para o referido teste estatístico. A fim de validar a nova proposta metodológica testamos os resultados em séries reais. Os resultados obtidos neste artigo corroboram com os resultados obtidos por [1]. Após os resultados desta pesquisa é possível testar estatisticamente os valores obtidos pelo coeficiente de correlação cruzada sem tendência ρ DCCA, considerando todas as escalas de tempo de tamanho s e também para diferentes tamanhos de séries e magnitudes de confiança.

1. INTRODUÇÃO

Nos dias atuais, com o grande volume de dados e informações que são produzidas nas diversas áreas do conhecimento, é necessária a aplicação e desenvolvimento de métodos que possam auxiliar pesquisadores e profissionais para uma melhor compreensão dos fenômenos estudados. Nesse contexto, o conhecimento e aplicação de métodos estatísticos assumem, nos dias atuais, uma função de extrema relevância, em função da amplitude e complexidade dos fenômenos estudados na área social, econômica, engenharias, entre outras. E também por proporcionar à investigação científica meios para consecução de seus objetivos. Dentre as áreas de estudo da estatística, neste artigo, trataremos sobre métodos relativos à Análise de Séries Temporais que segundo [3], tem como objetivo descrever e analisar os comportamentos passados da série, visando à compreensão do comportamento e conseqüente previsão de movimentos futuros.

Nas últimas duas décadas, têm crescido o estudo e o desenvolvimento de métodos destinados à análise de séries temporais em regime não estacionário como o método DFA – *Detrended Fluctuation Analysis* idealizado por [4], o DCCA – *Detrended Cross-Correlation Analysis* proposto por [5] e o coeficiente de correlação cruzada sem

tendência ρ DCCA implementado por [2]. Alicerçado no que foi discutido até aqui, este artigo objetiva discutir teoricamente o teste estatístico do ρ DCCA modelado por [1], reproduzi-lo e propor uma nova abordagem metodológica para o supracitado teste com aplicações em séries temporais reais.

Para atender aos objetivos deste artigo, a próxima seção aborda os procedimentos metodológicos dos métodos estatísticos contemplados neste artigo. A terceira seção apresenta os resultados e discussões, e por fim, na quarta faremos nossas conclusões.

2. METODOLOGIA

O ρ DCCA modelado por [2], é um expoente capaz de quantificar o nível de correlação cruzada tendo como base o DFA [4] e o DCCA [5]. O expoente λ obtido por meio do DCCA quantifica a correlação cruzada como lei de potência, porém não quantifica o nível da correlação cruzada. O ρ DCCA destina-se à estimação do coeficiente de correlação cruzada sem tendência em diferentes escalas de tamanho s , cujo coeficiente possui sua variação limitada entre -1 a +1 inclusive. Em que -1 representa anti-correlação perfeita (o crescimento de uma série implica o decrescimento da outra, ou vice-versa) e +1 correlação perfeita (o crescimento de uma série implica o crescimento da outra, ou vice-versa). E o valor do ρ DCCA igual a zero significa que não existe correlação cruzada [1] e [9]. É importante relatar que o campo de variação do ρ DCCA foi provado teoricamente por [1], tendo como método a *Desigualdade de Cauchy*.

Segundo [2], para modelar o ρ DCCA é necessário seguir os seguintes passos:

Primeiro passo: Considere duas séries temporais $\{x_t, t=1,2,\dots,N\}$ e $\{y_t, t=1,2,\dots,N\}$, onde N é o número total de valores iguais para as duas séries. Então, pode-se descrever as séries integradas de x_t e y_t obtendo duas novas séries

$$xx_k = \sum_{t=1}^k (x_t - \bar{x}) \quad \text{e} \quad yy_k = \sum_{t=1}^k (y_t - \bar{y}) \quad \text{com} \quad k = 1, 2, \dots, N \quad (1)$$

Em que \bar{x} e \bar{y} são as respectivas médias de todos os dados das séries x_t e y_t

Segundo passo: Dividem-se as duas séries integradas $\{xx_k\}$ e $\{yy_k\}$ em $N_s \equiv \text{int}(N/s)$ segmentos sobrepostos de igual comprimento s . Assim, $2N_s$ segmentos são obtidos conjuntamente. Consideramos $4 \leq s \leq N/4$.

Terceiro passo: Calcula-se a tendência local de cada um dos segmentos por um ajuste dos mínimos quadrados de cada série. A partir desse passo, determina-se a covariância sem tendência

$$F_{xy}(s, \lambda) = \frac{1}{s} \sum_{j=1}^s (xx_{(\lambda-1)s+j} - xP_{\lambda}(j))(yy_{(\lambda-1)s+j} - yP_{\lambda}(j)) \quad (2)$$

Para $\lambda = 1, 2, \dots, N_s$ e

$$F_{xy}(s, \lambda) = \frac{1}{s} \sum_{j=1}^s (xx_{N_s - (\lambda - N_s)s + j} - xP_{\lambda}(j))(yy_{N_s - (\lambda - N_s)s + j} - yP_{\lambda}(j)) \quad (3)$$

Para $\lambda = N_s + 1, N_s + 2, \dots, 2N_s$. Aqui, $xP_{\lambda}(j)$ e $yP_{\lambda}(j)$ são os ajustes polinomiais com ordem $m = 1$ no segmento λ de $\{xx_k\}$ e $\{yy_k\}$, respectivamente.

Quarto passo: Calcula-se a média sobre todos os segmentos para obter a função covariância sem tendência

$$F_{xy}(s) = \left[\frac{1}{N_s} \sum_{\lambda=1}^{N_s} F_{xy}(s, \lambda) \right]^{\frac{1}{2}} \quad (4)$$

Quinto passo: Calcula-se o coeficiente de correlação cruzada sem tendência ρ_{DCCA} usando a razão entre função covariância sem tendência e as funções de variância sem tendência pela seguinte expressão:

$$\rho_{DCCA} = \frac{F_{xy}^2(s)}{F_{xx}(s)F_{yy}(s)} \quad (5)$$

Apesar de ter sido proposto recentemente, o ρ_{DCCA} já foi modelado em diversos campos da ciência. É possível citar:

- No estudo do mercado financeiro [6];
- Em fenômenos climatológicos [7];
- Na análise de dados geofísicos [8];
- Em séries históricas de indicadores de criminalidade e violência [9].

Uma das vantagens do coeficiente de correlação cruzada sem tendência ρ_{DCCA} em relação a outros coeficientes de correlação cruzada como o coeficiente de correlação de Pearson, é a possibilidade de mensurar correlação entre dois sinais em diferentes escalas de tempo s [10]. E, além disso, segundo [9], a análise da correlação entre os sinais é realizada sem a componente tendência que tende a mascarar as verdadeiras correlações. Outra vantagem em relação ao coeficiente de Pearson é a capacidade de analisar séries não lineares. E recentemente os resultados estatísticos obtidos pelo ρ_{DCCA} foram comparados com o coeficiente de Pearson nas séries temporais do mercado de ações dos Estados Unidos por [11]. E os resultados obtidos pelo ρ_{DCCA} apresentaram diversas propriedades interessantes em diferentes escalas de tempo s no mercado de ações dos Estados Unidos que segundo [11], são úteis para o gerenciamento de riscos e seleção ótima de portfólio, especialmente para diversidade do portfólio de ativos.

O ρ_{DCCA} em sua construção leva em consideração a ordem dos pares de valores das séries temporais o que não contraria o princípio básico da análise das séries temporais que é a dependência temporal entre os valores da série fato que não está presente em alguns coeficientes de correlação cruzada como o coeficiente de correlação cruzada de Pearson.

Com objetivo de testar a robustez estatística do ρ_{DCCA} , foi proposto por [1] um teste estatístico para validar a significância dos valores obtidos pelo ρ_{DCCA} (Equação 5). O referido teste estatístico tem como base duas séries temporais independentes e identicamente distribuídas (*iid*), em um processo Gaussiano com média zero e variância igual a 1, com expoentes de autocorrelação do método DFA igual a 0,50 (séries aleatórias) com 95% de confiança. E o teste tem as seguintes hipóteses (ver Figura 1):

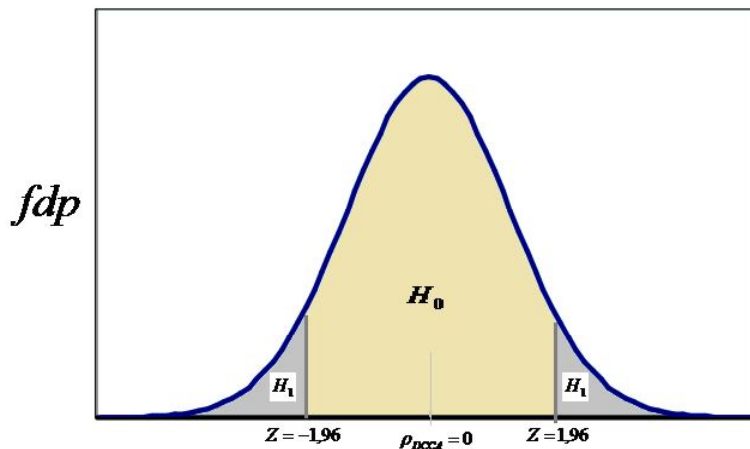


Figura 1: Hipóteses estatísticas do teste proposto por [1] para os valores obtidos pelo ρ_{DCCA} . Aqui *fdp* denota função de densidade de probabilidade.

$H_0 : \rho_{DCCA} = 0$ (não existe correlação cruzada significativa);

$H_1 : \rho_{DCCA} \neq 0$ (existe correlação cruzada significativa).

aqui H_0 denota a hipótese nula ou hipótese de igualdade e H_1 hipótese alternativa ou hipótese de pesquisa.

O teste estatístico supracitado determina a variação dentro da qual a correlação pode ser estatisticamente significativa ou não. O teste de significância do ρ_{DCCA} levou em consideração também o tamanho N da série e o tamanho s do *box* (janela). Tendo como alicerce a distribuição amostral das médias e o Teorema do Limite Central, os valores críticos do ρ_{DCCA} foram definidos pela seguinte expressão:

$$\bar{\rho}_{DCCA} \pm Z\sigma_{\rho_{DCCA}} \tag{6}$$

em que:

$\bar{\rho}_{DCCA}$ denota a média dos ρ_{DCCA} dos 10.000 mil pares de séries *iid*;

Z é quantil que define o nível de significância (5 %, 10%...) (Consultar tabela da Distribuição Normal de Probabilidade);

$\sigma_{\rho_{DCCA}}$ o desvio padrão das séries dos ρ_{DCCA} dos 10.000 pares de séries *iid*.

Maiores detalhes sobre o teste podem ser encontrados em [1] e [12]. Em relação aos conceitos e procedimentos de cálculo utilizados no teste estatístico do ρ_{DCCA} como o Teorema do Limite Central, Distribuição amostral das médias e de outros estimadores é possível consultar [13], [14], [15], [16] e [17]. Os resultados obtidos [1] provaram que os valores do ρ_{DCCA} dependem também do tamanho N da série e do tamanho da escala temporal s (ver Tabela 1, 2 e 3).

3. RESULTADOS E DISCUSSÃO

Com o objetivo de contribuir com o estudo teórico do ρ_{DCCA} reproduzimos o teste estatístico proposto por [1] com os valores críticos do ρ_{DCCA} com 90%, 95% e 99% de confiança para pares de séries independente e identicamente distribuídas (*iid*) para diferentes janelas de tamanho s e séries de diferentes N (ver Tabela 1, 2 e 3 e Figura 2).

Tabela 1 – Valores positivos críticos do coeficiente de correlação cruzada sem tendência ρ_{DCCA} (Equação 6) com 90% de confiança tendo como base 10.000 pares de séries temporais (*iid*) com média zero e variância unitária com ajuste polinomial de grau 1 no *box*.

(iid)	$s=4$	$s=8$	$s=16$	$s=32$	$s=64$	$s=128$	$s=256$
$N=250$	0,11895	0,15574	0,16172	0,16566	0,17675	-	-
$N=500$	0,08354	0,10946	0,11295	0,11400	0,11644	0,12461	-
$N=1.000$	0,05910	0,07939	0,08174	0,08073	0,08116	0,08377	0,08978
$N=2.000$	0,04152	0,05466	0,05641	0,05587	0,05550	0,05587	0,05750
$N=4.000$	0,02938	0,03836	0,03951	0,03918	0,03878	0,03880	0,03923
$N=8.000$	0,02051	0,02737	0,02831	0,02790	0,02762	0,02761	0,02771
$N=16.000$	0,01504	0,01971	0,02000	0,01976	0,01958	0,01950	0,01948

Na tabela 1, 2 e 3 foram demonstrados apenas os valores positivos por se tratar de uma distribuição simétrica. Diante de um par de série temporal de tamanho $N=16.000$ e $s=256$ (Tabela 1), por exemplo, o valor do ρ_{DCCA} será estatisticamente significativo com 90% de confiança para o $|\rho_{DCCA}| > 0,01948$ e neste caso rejeitamos a hipótese nula ($H_0 : \rho_{DCCA} = 0$).

Tabela 2 – Valores positivos críticos do coeficiente de correlação cruzada sem tendência ρ_{DCCA} (Equação 6) com 95% de confiança tendo como base 10.000 pares de séries temporais (*iid*) com média zero e variância unitária com ajuste polinomial de grau 1 no *box*.

(iid)	$s=4$	$s=8$	$s=16$	$s=32$	$s=64$	$s=128$	$s=256$
$N=250$	0,14128	0,18506	0,19222	0,19681	0,20998	-	-
$N=500$	0,09921	0,13014	0,13432	0,13554	0,13845	0,14814	-

$N=1.000$	0,07017	0,09420	0,09667	0,09582	0,09630	0,09936	0,10655
$N=2.000$	0,04937	0,06498	0,06703	0,06640	0,06599	0,06645	0,06839
$N=4.000$	0,03488	0,04562	0,04700	0,04658	0,04609	0,04611	0,04664
$N=8.000$	0,02439	0,03251	0,03363	0,03317	0,03282	0,03280	0,03293
$N=16.000$	0,01782	0,02338	0,02373	0,02346	0,02324	0,02314	0,02312

Tabela 3 – Valores positivos críticos do coeficiente de correlação cruzada sem tendência ρ DCCA (Equação 6) com 99% de confiança tendo como base 10.000 pares de séries temporais (*iid*) com média zero e variância unitária com ajuste polinomial de grau 1 no *box*.

(<i>iid</i>)	$s=4$	$s=8$	$s=16$	$s=32$	$s=64$	$s=128$	$s=256$
$N=250$	0,18593	0,24370	0,25322	0,25911	0,27644	-	-
$N=500$	0,13056	0,17149	0,17705	0,17861	0,08248	0,19520	-
$N=1.000$	0,09231	0,12383	0,12708	0,12599	0,12657	0,13054	0,14009
$N=2.000$	0,06507	0,08562	0,08827	0,08745	0,08697	0,08761	0,09018
$N=4.000$	0,04588	0,06015	0,06196	0,06138	0,06070	0,06073	0,06144
$N=8.000$	0,03216	0,04281	0,04429	0,04371	0,04324	0,04320	0,04336
$N=16.000$	0,02339	0,03072	0,03120	0,03084	0,03054	0,03042	0,03039

Os resultados das simulações do teste estatístico do ρ DCCA realizados nesta pesquisa corroboraram com os de Podobnik *et al*, 2011 [1]. Segundo [1], para obter os valores críticos para tamanhos de N e s não contidos na tabela 1, 2 e 3 os pesquisadores deverão obter por interpolação. Devido a essa limitação modelamos um programa capaz de executar o teste estatístico do ρ DCCA idealizado por [1] para diferentes pares de séries de tamanho N e diferentes escalas de tempo de tamanho s e magnitudes de intervalos de confiança (90%, 95%, ...). O referido programa de computador foi registrado no INPI (Instituto Nacional da Propriedade Intelectual) com co-autoria do SENAI CIMATEC [18].

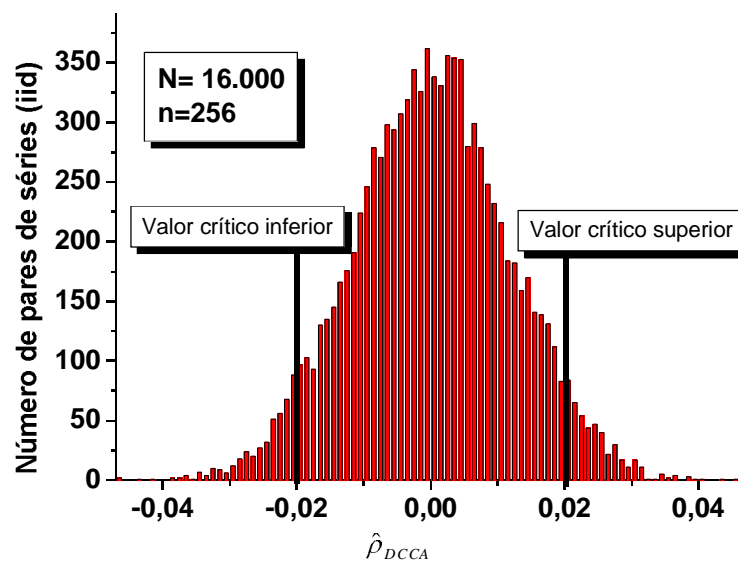


Figura 2: Simulação para o teste estatístico da equação 6 com 10.000 pares de séries *iid* com $N=16.000$ pontos e o tamanho do *box* $s=256$.

Como exemplo de aplicação do teste estatístico do ρ DCCA utilizamos as séries diárias dos homicídios dolosos e homicídios tentados registrados na capital baiana no período 2004 a 2011 da pesquisa realizada por [9] para todas as escalas de tamanho s (ver Figura 3). De acordo com [9] os homicídios dolosos e homicídios tentados são crimes contra pessoa e são amplamente usados nos estudos empíricos sobre os determinantes da criminalidade. Os homicídios dolosos são aqueles praticados voluntariamente ou intencionalmente, por qualquer instrumento ou meio. Já os homicídios tentados são aqueles homicídios cuja execução se iniciou, mas não se consumou por circunstâncias alheias à vontade do/a autor(a) [9].

Em [9] os resultados demonstraram que existe correlação cruzada positiva entre as séries de homicídios dolosos e homicídios tentados para todas as escalas de tempo s , entretanto esta correlação não foi testada estatisticamente pelos os autores. O conhecimento que o coeficiente de correlação cruzada sem tendência ρ DCCA é positivo para qualquer escala de tempo s , permite declarar que se existir um aumento/descréscimo nos homicídios dolosos, espera-se um aumento/descréscimo nos homicídios tentados [9].

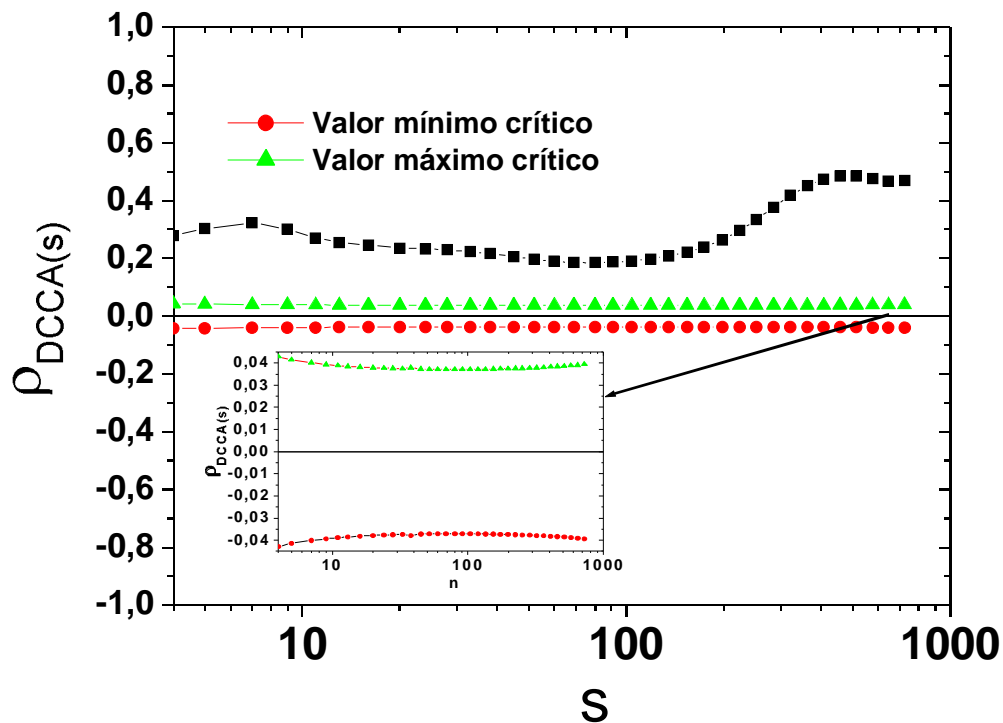


Figura 3: Teste estatístico idealizado por [1] sobre o coeficiente de correlação cruzada sem tendência ρ DCCA com 95% de confiança, nos homicídios dolosos e homicídios tentados registrados diariamente em Salvador-BA de Janeiro de 2004 a dezembro de 2011.

Os resultados obtidos do ρ DCCA (ver Figura 3) foram estatisticamente significativos com 95% de confiança. Sendo assim, rejeitamos a hipótese nula ($H_0 : \rho_{DCCA} = 0$). É possível afirmar que há correlação cruzada entre as séries de homicídios dolosos e homicídios tentados e que esta correlação é significativa para todas as escalas de tempo consideradas com 95% de confiança. E o presente artigo não tem a pretensão de discutir e modelar as variáveis socioeconômicas associadas à criminalidade como educação, saúde e renda.

4. CONCLUSÃO

Neste artigo, realizamos uma breve revisão bibliográfica sobre o coeficiente de correlação cruzada sem tendência ρ DCCA e sobre o seu teste estatístico. Reproduzimos o teste estatístico do ρ DCCA, modelado por [1]. Registramos no INPI o programa que calcula em diferentes escalas de tamanho s o teste estatístico do ρ DCCA.

Após a execução da presente pesquisa é possível testar estatisticamente os valores obtidos pelo coeficiente de correlação cruzada sem tendência ρ DCCA para todas as escalas de tempo e para diferentes tamanhos de séries e magnitudes de confiança.

REFERÊNCIAS

- ¹Podobnik, B; Jiang, Z.-Q.; Zhou, W.-X; Stanley, H. Statistical tests for Power-law cross-correlated processes. *Physical Review E*, v. 84, p. 066118, 2011.
- ²Zebende, G. F. DCCA cross-correlation coefficient: quantifying level of cross-correlation. *Physica A*, Amsterdam, v. 390, p. 614-618, 2011.
- ³Morettin, P. A.; Toloi, C. M. C. *Análise de séries temporais*. São Paulo: ABE – Projeto Fisher. Editora Edgard Blucher LTDA, 2004.
- ⁴Peng, C. K. *et al.* Mosaic organization of DNA nucleotides. *Physical Review E*, College Park, MD, v. 49, p. 1685-1689, 1994.
- ⁵Podobnik, B.; Stanley, H. E. Detrended cross-correlation analysis: a new method for analyzing two non stationary time series. *Physical Review Letters*, v. 100, p. 084102, 2008.
- ⁶Reboredo, J. C.; Rivera-Castro, M. A.; Zebende, G. F. Oil and us dollar exchange rate dependence: a detrended cross-correlation approach. *Energy Economics*, Cambridge, UK, v. 42, p. 132-139, 2014.

⁷Vassoler, R.; Zebende, G. DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A*, Amsterdam, v. 391, p. 2438-2443, 2011.

⁸Marinho, E.; Sousa, A.; Andrade, R. Using detrended cross-correlation analysis in geophysical data. *Physica A*, Amsterdam, v. 392, p. 2195-2201, 2013.

⁹Machado, A.; Silva, M. F.; Zebende, G. F. Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A*, Amsterdam, v. 400, p. 12-19, 2014.

¹⁰Wang, G.-J.; Xie, C. Cross-correlations between renminbi and four major currencies in the renminbi currency basket. *Physica A*, Amsterdam, v. 392, p. 1418-1428, 2013.

¹¹Wang, G.-J. *et al.* Statistical properties of the foreign exchange network at different time scales: Evidence from detrended cross-correlation coefficient and minimum spanning tree. *Entropy*, v. 15, p. 1643-1662, 2013.

¹²Fernandes, M.; Castro, A.; Machado, A. Quantificando a influência do mercado de câmbio nos preços do milho e da soja no município de Barreiras. *Conjuntura & Planejamento*, v. 182, p. 42- 49, 2014.

¹³Bolfarine, H.; Sandoval, M. Introdução à inferência Estatística. Coleção de matemática Aplicada – Sociedade Brasileira de Matemática, 2000.

¹⁴Bussab, W.; Morettin, P. Estatística Básica. São Paulo: Saraiva, 2003.

¹⁵Bolfarine, H.; Bussab, W. O. Elementos de amostragem. São Paulo: ABE-Projeto Fischer Edgard Blucher, 2005.

¹⁶Triola, M. Introdução à Estatística. 10ª ed. Rio de Janeiro: LTC – Livros Técnicos e Científicos, 2008.

¹⁷Morettin, L. G. Estatística Básica Volume único Probabilidade e Estatística. São Paulo: Pearson, 2010.

¹⁸Zebende, G. F.; Castro, A.P.N.; Machado, A.; Fernandes, M. Rhodccasort teste Estatístico. n° do registro 011140000 84 BR 51 2014 000339 6, 2014.