



Instituto de
MATEMÁTICA
E ESTATÍSTICA

UFRGS



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

DEPARTAMENTO DE ESTATÍSTICA

O método da correlação cruzada destendenciada em séries temporais

Lucas de Freitas Gonçalves

Porto Alegre
2018

Lucas de Freitas Gonçalves

O método da correlação cruzada destendenciada em séries temporais

Trabalho de Conclusão de Curso submetido
como requisito parcial para a obtenção do
grau de Bacharel em Estatística.

Orientador Metodológico
Prof. Dr. Guilherme Pumi

Coorientador Metodológico
Prof.^a Dr.^a Taiane S. Prass

Porto Alegre
2018

Instituto de Matemática e Estatística
Departamento de Estatística

O método da correlação cruzada destendenciada em séries temporais

Lucas de Freitas Gonçalves

Banca examinadora:

Prof. Dr. Guilherme Pumi
UFRGS

Prof.^a Dr.^a Taiane S. Prass (Coorientadora)
UFSM

Prof. Dr. Marcio Valk
UFRGS

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador Prof. Guilherme Pumi, juntamente com minha coorientadora Prof. Taiane S. Prass, que tornou este trabalho possível, me passando conhecimentos e lições que levarei em minha carreira e vida pessoal. Ao final deste ciclo, vejo neste trabalho uma consolidação desta formação que escolhi e que é parte fundamental para o que almejo para o meu futuro. Também agradeço a UFRGS como um todo, através do Instituto de Matemática e Estatística, que te

m profissionais de classe mundial e que foram fundamentais ao que me tornei hoje. O sucesso profissional que já tenho e virei a alcançar é em grande parte mérito de vocês.

Mas nada disto seria sequer possível se uma pessoa não tivesse lutado pela minha educação, desde que me conheço por gente. Mãe, obrigado por sempre me incentivar e não medir esforços para que eu obtivesse a educação que tenho hoje.

Também não posso deixar de mencionar aqui aqueles amigos que fazem o meu dia-a-dia ser mais alegre, não só pensando neste ciclo que se encerra, mas em toda a minha vida. Vagner e Luiz, aqueles irmãos que a vida nos traz. Colegas que a UFRGS me apresentou e se tornaram grandes amigos: Allan, Douglas, Felipe, Jorge e Tiago. E aos conhecidos do SESC: bora!

Além de tudo isso, o que seria do ser humano sem aquela pessoa. Aquela com a qual nós construímos planos, comemoramos conquistas e superamos dificuldades. Ou seja, vivemos. Da melhor e mais intensa forma possível. Posso dizer que encontrei a minha e que ela foi fundamental na conclusão deste trabalho. Apoio, compreensão, carinho e aqueles lanchinhos durante as horas que se estendiam. Obrigado, meu amor, Carina.

Por fim, gostaria de agradecer a mim mesmo. Eu consegui!

O preço de qualquer coisa é a quantidade de vida que você troca por isso.
Henri David Thoreau

RESUMO

Um dos grandes objetivos do estudo de séries temporais multivariadas é entender a estrutura de dependência entre séries. Para este fim, uma das maneiras mais utilizadas é através do estudo da estrutura de autocorrelação cruzada entre duas séries temporais. Para séries reais, este método só é possível sob a hipótese de estacionariedade das séries em estudo, uma hipótese comumente inválida. Uma maneira alternativa de estudar a estrutura de dependência em séries não-estacionárias é o método conhecido como Detrended Cross-Correlation Analysis (DCCA), cuja finalidade é verificar a existência de dependência cruzada em séries temporais não-estacionárias. Neste trabalho nos propomos a estudar o comportamento amostral da DCCA no contexto de séries temporais bivariadas através de simulações de Monte Carlo. Foram considerados três cenários básicos para as simulações: séries temporais independentes e identicamente distribuídas (i.i.d) e independentes entre si; séries temporais i.i.d e dependentes entre si e séries temporais com correlação cruzada não-nula. Além do estudo da DCCA, apresentamos ainda a comparação desta com o coeficiente de correlação cruzada tradicional. Através de simulações no software R, foram evidenciados os resultados teóricos esperados e identificados comportamentos importantes de serem mais investigados, que se juntarão a um estudo maior que pretende utilizar este método para a análise de correlação cruzada entre séries temporais não-estacionárias e também na presença de dados faltantes.

Palavras-chave: Séries temporais, estrutura de dependência, correlação cruzada, DCCA, DFA.

5

ABSTRACT

One of the main objectives in the study of multivariate time series is to understand and describe the dependence structure between time series. The classical approach to this problem is through the well-known concept of cross correlation between two time series. In applications, however, in order for this method to be useful, it is essential that the time series in study are stationary, which is often a questionable requirement to make in applications. An alternative way to study the dependence structure in non-stationary time series is the Detrended Cross-Correlation Analysis (DCCA). In this work we aim to study the finite sample behavior of the DCCA by means of Monte Carlo simulations. In the simulations we consider three main scenarios: bivariate independent and identically distributed (i.i.d) time series, i.i.d time series correlated with each other and bivariate non-i.i.d time series with non-null cross correlation. We also compare our finds for the DCCA with the traditional cross correlation coefficient. Our results show that the finite sample behavior of the DCCA is very close to the expected theoretical one. This work is part of a bigger project aiming towards the study of the DCCA in the context of non-stationary time series with missing data.

Keywords: Time series, dependence structure, cross-correlation coefficient, DCCA, DFA.

SUMÁRIO

1	INTRODUÇÃO	9
2	METODOLOGIAS	11
2.1	Cross-correlation analysis ou análise de correlação cruzada	11
2.2	Detrended fluctuation analysis (DFA) ou análise de flutuação destendenciada	11
2.3	Detrended cross-correlation analysis (DCCA) ou análise de correlação cruzada destendenciada	12
2.4	Detrended cross-correlation coefficient ou coeficiente de correlação cruzada destendenciada	13
2.5	Teste de normalidade Shapiro-Wilk	14
3	SIMULAÇÕES	14
3.1	Cenário 1 – Séries i.i.d e independentes entre si	15
3.2	Cenário 2 – Séries i.i.d e dependentes entre si	16
3.2.1	Sinal + Ruído	16
3.2.2	Normal bivariada com correlação 0.5	17
3.2.3	Normal bivariada com correlação 0.8	17
3.3	Cenário 3 – Séries com correlação cruzada não-nula	17
4	RESULTADOS	19
4.1	Cenário 1 – Séries i.i.d e independentes entre si	20
4.2	Cenário 2 – Séries i.i.d e dependentes entre si	23
4.2.1	Sinal + Ruído	23
4.2.2	Normal bivariada com correlação 0.5	23
4.2.3	Normal bivariada com correlação 0.8	28
4.3	Cenário 3 – Séries com correlação cruzada não-nula	32
5	CONSIDERAÇÕES FINAIS	35
6	CONCLUSÃO	36
	REFERÊNCIAS	37
	APÊNDICE	38

1. INTRODUÇÃO

Uma série temporal é qualquer conjunto de observações ordenadas no tempo (Morettin e Toloí, 2004).

O estudo destas, de maneira univariada, pode envolver objetivos como previsão de valores futuros da série, seja de curto ou longo prazo, e descrição do comportamento da série, tal qual existência de tendências e sazonalidade (Morettin e Toloí, 2004). Além disto, pensando em um cenário multivariado, pode-se ter o interesse de entender a estrutura de dependência entre as séries.

Para a análise de séries temporais, assim como em outros métodos, é comum que alguns pressupostos devam ser atendidos para que as propriedades sejam válidas para a interpretação dos resultados. Conforme Morettin e Toloí (2004), “uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável”.

Em contrapartida, considerando as mais diversas informações disponíveis nos campos de estudo, seja ele qual for, uma grande quantidade de séries não apresentará esta condição, sendo assim não-estacionária (Morettin e Toloí, 2004).

Como uma maneira alternativa para a análise de correlação cruzada na presença da condição de não-estacionariedade, Podobnik e Stanley (2008) propuseram o método Detrended Cross-Correlation Analysis (DCCA) ou Análise de Correlação Cruzada Destendenciada. O DCCA é uma generalização do método Detrended Fluctuation Analysis (DFA) ou Análise de Flutuação Destendenciada, proposto por Peng et al. (1994). Ambos os métodos têm como objetivo verificar a correlação de longo alcance em séries temporais não-estacionárias.

Neste contexto, o presente trabalho se propõe a investigar o método DCCA e seus resultados baseado em simulações, buscando verificar como ele se comporta em diferentes cenários de estrutura de dependência entre as séries temporais, comparando com a análise de correlação cruzada tradicional.

Foram estabelecidos três cenários para análise neste trabalho. No primeiro, temos duas séries temporais independentes e igualmente distribuídas (i.i.d) e independentes entre si. Nesta situação espera-se que os resultados para ambos os métodos seja de correlação cruzada zero. No segundo, as duas séries temporais também são i.i.d, porém possuem dependência entre si, apenas no lag zero. Neste, espera-se que a correlação cruzada calculada seja zero para os lags diferentes de zero, na correlação cruzada tradicional, e que na correlação cruzada destendenciada o resultado pra todos os lags seja em torno do valor da correlação no lag zero. Por último, as duas séries temporais terão correlação cruzada não-nula até um lag estabelecido e os resultados esperados são mais complexos e serão detalhados na seção específica de simulações.

É importante saber que neste trabalho estão sendo analisadas séries temporais estacionárias, tendo em vista que este é o começo de um projeto que busca compreender melhor o método e suas possíveis aplicações nos diversos cenários possíveis, como, além dos propostos neste trabalho, a presença da não-estacionariedade e de dados faltantes nas séries temporais.

Para alcançar o que se propõe, o trabalho está estruturado em seis capítulos. No capítulo 2, estão conceituadas as metodologias que foram necessárias para sua realização, isto é, a análise de correlação cruzada, os métodos DCCA e DFA e o coeficiente de correlação cruzada destendenciada (ρ_{DCCA}). Apresentamos ainda o teste de normalidade Shapiro-Wilk, tendo em conta que também este foi utilizado durante o processo. No terceiro capítulo apresentaremos a metodologia das simulações, bem como a definição das séries temporais que foram simuladas para este estudo. No quarto são expostos os resultados para cada cenário, fazendo considerações sobre o que se obteve e o que se esperava obter. Por fim, nos capítulos cinco e seis, as considerações finais e a conclusão do trabalho, respectivamente.

2. METODOLOGIAS

Neste capítulo serão apresentadas as metodologias utilizadas no presente trabalho, tais como correlação cruzada, análise de flutuação destendenciada, análise de correlação cruzada destendenciada, coeficiente de correlação cruzada destendenciada e teste de normalidade Shapiro-Wilk.

2.1. Cross-correlation analysis ou análise de correlação cruzada

Sejam duas séries temporais $\{X_{1,t}\}_{t=0}^{\infty}$ e $\{X_{2,t}\}_{t=0}^{\infty}$ estacionárias com médias $\mu_1 = E(X_{1,t})$ e $\mu_2 = E(X_{2,t})$ e variâncias $\sigma_1^2 = Var(X_{1,t})$ e $\sigma_2^2 = Var(X_{2,t})$ para todo $t = 1, 2, \dots, n$. As funções de covariância cruzada, denotada por $\gamma_{1,2}(h)$, e correlação cruzada, denotada por $\rho_{1,2}(h)$, são definidas por

$$\gamma_{1,2}(h) = E[(X_{1,t} - \mu_1)(X_{2,t+h} - \mu_2)]$$

e

$$\rho_{1,2} = \frac{E[(X_{1,t} - \mu_1)(X_{2,t+h} - \mu_2)]}{\sigma_1^2 \sigma_2^2} = \frac{1}{\sigma_1^2 \sigma_2^2} \gamma_{1,2}(h),$$

onde $h \in Z$. Observe que quando $h = 0$, $\rho_{1,2}(0)$ será o coeficiente de correlação de Pearson, geralmente utilizado quando se tem interesse de mensurar a relação linear entre duas variáveis aleatórias $X_{1,t}$ e $X_{2,t}$.

$$\rho_{1,2}(h) = \frac{Cov(X_{1,t}, X_{2,t})}{\sigma_1 \sigma_2}.$$

2.2. Detrended fluctuation analysis (DFA) ou análise de flutuação destendenciada

Proposto por Peng et al. (1994), a análise de flutuação destendenciada tem o objetivo de quantificar a correlação cruzada em séries temporais não-estacionárias. Para utilizar o método, devem-se realizar as etapas a seguir.

Seja $\{X_t\}_{t=0}^{\infty}$ uma série temporal e seja x_1, \dots, x_n uma amostra de tamanho n advinda desta série. Primeiramente calcula-se o sinal integrado $R_k = \sum_{i=1}^k x_i$, sendo $k = 1, \dots, n$.

Em seguida, deve-se dividir a série integrada em $n - m$ intervalos de tamanho m . Para cada intervalo, começando em i e terminando em $i + m$, sendo $i = 1, \dots, n - m$, estima-se $R_i(k)$ através de um polinômio de grau ≥ 1 . É importante saber que, no artigo original onde é definido a DFA, eram utilizados intervalos não sobrepostos. Porém, quando trabalhamos com a DCCA, calculamos a DFA a partir de intervalos sobrepostos.

Após, deve-se calcular a variância dos resíduos para cada intervalo de tamanho $m + 1$, denotada por $f_{DFA}^2(m)$,

$$f_{DFA}^2(m, i) = \frac{1}{m+1} \sum_{k=1}^{i+m} (R_k - R_i(k))^2.$$

Por fim, calcula-se a variância destendenciada através da média das variâncias dos resíduos de todas as $n - m$ janelas de tamanho $m + 1$, denotada por $F_{DFA}^2(m)$,

$$F_{DFA}^2(m) = \frac{1}{n-m} \sum_{k=1}^{n-m} f_{DFA}^2(m, i).$$

2.3. Detrended cross-correlation analysis (DCCA) ou análise de correlação cruzada destendenciada

Proposto por Podobnik (2008), a análise de correlação cruzada destendenciada é uma generalização do método análise de flutuação destendenciada e tem o objetivo de quantificar a correlação cruzada entre duas séries temporais não-estacionárias com a mesma quantidade de observações. Embora recente, este método tem sido utilizado para diversos campos de estudo, como Jun E Da-Qing (2012) no estudo de diferença entre resultados de eletroencefalograma de indivíduos jovens e idosos, e Machado et al. (2014) na relação entre homicídios e tentativas de homicídios.

Para utilizar o método, devem ser feitos os passos a seguir. Sejam duas séries temporais $\{X_{1,t}\}_{t=0}^{\infty}$ e $\{X_{2,t}\}_{t=0}^{\infty}$ e sejam $x_{1,1}, \dots, x_{1,n}$ e $x_{2,1}, \dots, x_{2,n}$ amostras de igual tamanho n advindas destas séries, devem-se calcular dois sinais integrados $R_k = \sum_{i=1}^k x_{1,i}$ e $R'_k = \sum_{i=1}^k x_{2,i}$, sendo $k = 1, \dots, n$. Dividem-se as séries R_k e R'_k em $n - m$ janelas, cada uma contendo tamanho $m + 1$. Para estas duas séries, cada janela começa em i e termina em $i + m$. Em cada janela, estima-se uma reta através do processo de mínimos quadrados ordinários (MQO) e obtêm-se as retas ajustadas,

denotadas por $\tilde{R}_{k,i}$ e $\tilde{R}'_{k,i}$. A seguir, deve-se calcular a covariância dos resíduos para cada janela, denotada por $f_{DCCA}^2(m, i)$.

$$f_{DCCA}^2(m, i) = \frac{1}{n-1} \sum_{k=i}^{i+m} (R_k - \tilde{R}_{k,i})(R'_k - \tilde{R}'_{k,i}).$$

Por fim, calcula-se a covariância destendenciada através da média das covariâncias dos resíduos de todas as $n-m$ janelas de tamanho $m+1$, denotada por $F_{DCCA}^2(m)$.

$$F_{DCCA}^2(m) = \frac{1}{(n-m)} \sum_{i=1}^{n-m} f_{DCCA}^2(m, i).$$

Conforme mencionado anteriormente, sendo o DCCA uma generalização do DFA, quando $R_k = R'_k$, isto é, temos apenas uma série temporal, a covariância $F_{DCCA}^2(m)$ será reduzida a variância $F_{DFA}^2(m)$.

2.4. Detrended cross-correlation coefficient ou coeficiente de correlação cruzada destendenciada

Proposto por Zebende (2011), o coeficiente de correlação cruzada destendenciada, denotado por $\rho_{DCCA}(m)$ sendo m o lag, é a relação entre os coeficientes $F_{DCCA}^2(m)$ e $F_{DFA}^2(m)$ e tem como objetivo quantificar a correlação cruzada em séries temporais não-estacionárias.

Os resultados do coeficiente podem variar entre -1 e 1. Neste caso, quando $\rho_{DCCA}(m) = 0$ temos duas séries temporais sem correlação. Para valores acima de zero, significa presença de correlação cruzada, sendo perfeita caso seja igual a 1. Da mesma maneira, valores abaixo de zero indicam anti correlação cruzada.

Para calcular o coeficiente de correlação cruzada destendenciada, basta tomar a razão entre a covariância $F_{DCCA}^2(m)$ com o produto das variâncias $F_{1,DFA}^2(m)$ e $F_{2,DFA}^2(m)$.

$$\rho_{DCCA}(m) = \frac{F_{DCCA}^2(m)}{\sqrt{F_{1,DFA}^2(m)F_{2,DFA}^2(m)}}.$$

Para fins de facilitar as leituras na sessão Resultados, este coeficiente foi denotado como coeficiente de correlação DCCA, enquanto o tradicional como correlação cruzada.

2.5. Teste de normalidade Shapiro-Wilk

Proposto por Shapiro e Wilk (1965), o teste Shapiro-Wilk tem o objetivo de verificar a hipótese nula de que uma amostra provém de uma população que possui distribuição Normal.

A estatística de teste, denotada por W , é dada por

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

onde $x_{(i)}$ é a i -ésima estatística de ordem, \bar{x} é a média amostral e as constantes a_i são valores obtidos pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal.

3. SIMULAÇÕES

As simulações foram realizadas no software livre R (R Core Team, 2017), onde foi implementado o método DCCA e também criadas as séries temporais.

Os resultados teóricos apresentados neste capítulo fazem parte dos estudos preliminares relativos ao projeto “Análise de correlação de séries temporais não-estacionárias e na presença de dados faltantes” da professora Taiane Schaedler Prass da UFSM (Auxílio ARD Edital 01/2017).

Para cada cenário, utilizou-se o método de Monte Carlo com 1000 replicações, isto é, cada série temporal foi criada 1000 vezes. Com o objetivo de replicar os resultados gerados, se necessário, para todas as replicações foram utilizadas sementes, com notação $100 + r$, sendo r o número da replicação. Em programação, as sementes servem para memorizar e reproduzir cada geração aleatória de valores.

Todas as séries temporais foram criadas com tamanho $n = 200$ e por isso foram calculadas as correlações cruzadas com lag m , sendo $m = 3, \dots, 20$. O tamanho máximo 20 foi escolhido devido a critério padrão do software R que é $10 * \log_{10} \left(\frac{n}{s} \right)$, sendo n a quantidade de observações das séries temporais e s a quantidade de séries.

Com relação aos resultados esperados para o coeficiente de correlação DCCA, é importante saber que se pode mostrar que, se duas séries temporais $\{X_{1,t}\}_{t=0}^{\infty}$ e $\{X_{2,t}\}_{t=0}^{\infty}$ são estritamente estacionárias e ergódicas, com momento de quarta ordem finito, então

$$F_{k,DFA}^2(m) \xrightarrow{P} E(f_{k,DFA}^2(m, 1)) = \frac{1}{m+1} \text{tr}(K_{m+1}\Gamma_k(m)), \text{ quando } n \rightarrow \infty,$$

e

$$F_{DCCA}^2(m) \xrightarrow{P} E(f_{DCCA}^2(m, 1)) = \text{tr}(K_{m+1}\Gamma_{1,2}(m)), \text{ quando } n \rightarrow \infty,$$

onde $K_{m+1} = J'_{m+1}Q_{m+1}J_{m+1}$, $\Gamma_k(m) = \text{Var}(X_{1,2}(m))$ e $\Gamma_{1,2}(m) = \text{Cov}(X_{1,1}(m), X_{2,1}(m))$, com $X_{k,1}(m) = (X_{k,1}, \dots, X_{k,m+1})'$. Além disto, J_{m+1} é a $(m+1) \times (m+1)$ matriz cujo (r, s) ésimo elemento é dado por

$$J_{r,s} = \begin{cases} 1, & \text{se } 1 \leq s \leq r \leq m+1, \\ 0, & \text{se } 1 \leq r \leq s \leq m+1, \end{cases}$$

e $Q_{m+1} := I_{m+1} - X(X'X)^{-1}X'$, onde I_{m+1} é a $(m+1) \times (m+1)$ matriz identidade e

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & m+1 \end{pmatrix}.$$

Aplicando o teorema da função contínua

$$\rho_{DCCA}(m) = \frac{F_{DCCA}^2(m)}{\sqrt{F_{1,DFA}^2(m)F_{1,DFA}^2(m)}} \xrightarrow{P} \frac{\text{tr}(K_{m+1}\Gamma_{1,2}(m))}{\sqrt{\text{tr}(K_{m+1}\Gamma_1(m))\text{tr}(K_{m+1}\Gamma_2(m))}}, \text{ quando } n \rightarrow \infty.$$

A partir disto, serão expostos os comportamentos esperados do coeficiente de correlação DCCA para cada um dos cenários.

3.1. Cenário 1 – Séries i.i.d e independentes entre si

Neste cenário, analisamos o coeficiente de correlação DCCA e correlação cruzada entre duas séries temporais i.i.d e independentes entre si, sendo elas $X_{1,t}$ e $X_{2,t}$ com distribuição comum $N(0,1)$.

Neste caso,

$$\text{Cov}(X_{k,t}, X_{k,t+h}) = I(h=0) \text{ e } \text{Cov}(X_{1,t}, X_{2,t+h}) = 0, \forall h \in \mathbb{Z}.$$

Segue que

$$\Gamma_k = I_{m+1}, k=1,2 \text{ e } \Gamma_{1,2} = 0_{m+1,m+1}, \forall m > 0.$$

Portanto

$$F_{k,DFA}^2(m) \xrightarrow{P} \frac{tr(K_{m+1})}{m+1}, F_{DCCA}^2(m) \xrightarrow{P} 0, \text{ quando } n \rightarrow \infty,$$

e

$$\rho_{DCCA}(m) \xrightarrow{P} 0, \text{ quando } n \rightarrow \infty, \forall m > 0.$$

3.2. Cenário 2 – Séries i.i.d e dependentes entre si

Para este cenário, analisamos o coeficiente de correlação DCCA e a correlação cruzada entre duas séries temporais i.i.d, mas dependentes entre si, isto é, com correlação cruzada não-nula no lag zero. Consideramos três estruturas de dependência para explorar este cenário.

3.2.1. Sinal + ruído

Duas séries temporais $X_{1,t}$ e $X_{2,t}$, sendo $X_{1,t} \sim N(0,1)$ e $X_{2,t}$ em função de $X_{1,t}$.

$X_{2,t} = 2 * X_{1,t} + 3 + \varepsilon_t$, onde $\varepsilon_t \sim N(0,2^2)$, com $X_{1,t}$ e ε_t independentes para todo $t \in Z$.

Neste caso,

$$Cov(X_{k,t}, X_{k,t+h}) = a_k I(h=0) \text{ e } Cov(X_{1,t}, X_{2,t+h}) = 2I(h=0), \forall h \in Z,$$

onde $a_1 = 1$ e $a_2 = 8$. Segue que

$$\Gamma_k = a_k I_{m+1}, k=1,2 \text{ e } \Gamma_{1,2} = 2I_{m+1}, \forall m > 0.$$

Logo

$$F_{k,DFA}^2(m) \xrightarrow{P} \frac{a_k tr(K_{m+1})}{m+1}, F_{DCCA}^2(m) \xrightarrow{P} \frac{2tr(K_{m+1})}{m+1}, \text{ quando } n \rightarrow \infty,$$

e

$$\rho_{DCCA}(m) \xrightarrow{P} \frac{\sqrt{2}}{2} \approx 0,7071, \text{ quando } n \rightarrow \infty, \forall m > 0.$$

3.2.2. Normal bivariada com correlação 0.5

Duas séries temporais $X_{1,t}$ e $X_{2,t}$ geradas através de uma Normal Bivariada, com marginais $N(0,1)$ e correlação 0,5.

Neste caso,

$$Cov(X_{k,t}, X_{k,t+h}) = I(h = 0) \text{ e } Cov(X_{1,t}, X_{2,t+h}) = \rho I(h = 0), \forall h \in \mathbb{Z}.$$

Segue que

$$\Gamma_k = I_{m+1}, k=1,2 \text{ e } \Gamma_{1,2} = \rho I_{m+1}, \forall m > 0.$$

Logo

$$F_{k,DFA}^2(m) \xrightarrow{P} \frac{tr(K_{m+1})}{m+1}, F_{DCCA}^2(m) \xrightarrow{P} \frac{\rho tr(K_{m+1})}{m+1}, \text{ quando } n \rightarrow \infty,$$

e

$$\rho_{DCCA}(m) \xrightarrow{P} \rho = 0,5, \text{ quando } n \rightarrow \infty, \forall m > 0.$$

3.2.3. Normal bivariada com correlação 0.8

Duas séries temporais $X_{1,t}$ e $X_{2,t}$ geradas através de uma Normal Bivariada, com marginais $N(0,1)$ e correlação 0,8.

Neste caso, análogo ao anterior,

$$\rho_{DCCA}(m) \xrightarrow{P} \rho = 0,8, \text{ quando } n \rightarrow \infty, \forall m > 0.$$

3.3. Cenário 3 – Séries com correlação cruzada não-nula

Neste cenário, analisamos o coeficiente de correlação DCCA e a correlação cruzada entre duas séries temporais que possuem correlação cruzada não-nula. Para isto, gerou-se uma série $X_{1,t}$ que possui distribuição $N(0,1)$ com tamanho amostral $n + 20$. Foi necessário ampliar o tamanho de amostra em 20 devido ao fato de se ter escolhido criar uma segunda série temporal $X_{2,t}$ sendo um MA(20) desta série $X_{1,t}$. Isto é, $X_{2,t} = \sum_{j=0}^{20} \theta_j X_{1,t-j}$, onde $\theta_0 = 1$ e $\theta_j = \frac{21-j}{10}$, se $1 \leq j \leq 20$. Desta maneira, teremos correlação cruzada não-nula até o lag 20.

Neste caso,

$$\text{Cov}(X_{1,t}, X_{1,t+h}) = I(h = 0),$$

$$\gamma_2(h) = \text{Cov}(X_{2,t}, X_{2,t+h}) = \begin{cases} \sum_{j=0}^{20-|h|} \theta_j \theta_{j+|h|}, & -20 \leq h \leq 20, \\ 0, & \text{caso contrário} \end{cases}$$

e

$$\text{Cov}(X_{1,t}, X_{2,t+h}) = \begin{cases} \theta_h, & 0 \leq h \leq 20, \\ 0, & \text{caso contrário.} \end{cases}$$

Segue que

$$\Gamma_1 = I_{m+1}, \Gamma_2 = \sum_{j=0}^T \gamma_2(j) M_j \text{ e } \Gamma_{1,2} = \sum_{j=0}^T \theta_j M_j^*, \forall m > 0,$$

onde $T = \min\{m, 20\}$ e M_j e M_j^* são matrizes de tamanho $(m+1) \times (m+1)$ para as quais os (r, s) ésimos coeficientes são dados, respectivamente, por

$$[M_j]_{r,s} = \begin{cases} 1, & \text{se } |r-s| = j \\ 0, & \text{caso contrário} \end{cases} \text{ e } [M_j^*]_{r,s} = \begin{cases} 1, & \text{se } s-r = j \\ 0, & \text{caso contrário.} \end{cases} \quad (1)$$

Logo

$$\text{tr}(K_{m+1} \Gamma_2) = \sum_{j=0}^T \gamma_2(j) \text{tr}(K_{m+1} M_j) \text{ e } \text{tr}(K_{m+1} \Gamma_{1,2}) = \sum_{j=0}^T \theta_j \text{tr}(K_{m+1} M_j^*), \forall m > 0.$$

É fácil mostrar que K_{m+1} é a matriz simétrica para cada $[K_{m+1}]_{r,s} = 0$, sempre que $r = 1$ ou $s = 1$. Então, de (1), $M_0 = I_{m+1}$, de modo que $\text{tr}(K_{m+1} M_0) = \text{tr}(K_{m+1})$ e

$$\text{tr}(K_{m+1} M_j) = \sum_{r=1}^{m+1} \sum_{s=1}^{m+1} [K_{m+1}]_{r,s} [M_j]_{s,r} = 2 \sum_{r=2}^{m+1-j} [K_{m+1}]_{r,r+j}, \quad 1 \leq j \leq T,$$

e, de maneira análoga, $M_0^* = I_{m+1}$, de modo que $\text{tr}(K_{m+1} M_0^*) = \text{tr}(K_{m+1})$ e

$$\alpha_j^{(m)} := \text{tr}(K_{m+1} M_j^*) = \sum_{r=2}^{m+1-j} [K_{m+1}]_{r,r+j} = \frac{1}{2} \text{tr}(K_{m+1} M_j), \quad 1 \leq j \leq T.$$

No caso particular, sendo $m \leq 20$, $\text{tr}(K_{m+1} M_m) = 0$ e $\text{tr}(K_{m+1} M_m^*) = 0$. Segue que

$$\rho_{DCCA}(m) \xrightarrow{P} \begin{cases} \frac{\alpha_0^{(m)} + \sum_{j=1}^{m-1} \alpha_j^{(m)} \theta_j}{\sqrt{\alpha_0^{(m)} [\alpha_0^{(m)} \gamma_2(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m-1)} \gamma_2(j)]}}, & \text{se } m \leq 20 \\ \frac{\alpha_0^{(m)} + \sum_{j=1}^{20} \alpha_j^{(m-1)} \theta_j}{\sqrt{\alpha_0^{(m)} [\alpha_0^{(m)} \gamma_2(0) + 2 \sum_{j=1}^{20} \alpha_j^{(m)} \gamma_2(j)]}}, & \text{se } m > 20 \end{cases}, \text{ quando } n \rightarrow \infty.$$

Com $\alpha_0^{(m)} := tr(K_{m+1})$. De maneira equivalente, pode ser escrito como

$$\rho_{DCCA}(m) \xrightarrow{P} \begin{cases} \frac{1 + \sum_{j=1}^{m-1} \beta_j^{(m)} \theta_j}{\sqrt{\sum_{j=0}^{20} \theta_j^2 + 2 \sum_{j=1}^{20} \sum_{i=0}^{20-j} \beta_j^{(m)} \theta_i \theta_{i+j}}}, & \text{se } m \leq 20 \\ \frac{1 + \sum_{j=1}^{20} \beta_j^{(m)} \theta_j}{\sqrt{\sum_{j=0}^{20} \theta_j^2 + 2 \sum_{j=1}^{20} \sum_{i=0}^{20-j} \beta_j^{(m)} \theta_i \theta_{i+j}}}, & \text{se } m > 20 \end{cases}, \text{ quando } n \rightarrow \infty,$$

com $\beta_0^{(m)} := \frac{\alpha_j^{(m)}}{\alpha_0^{(m)}}$. Além disto, $\sum_{j=0}^{20} \theta_j^2 + 2 \sum_{j=1}^{20} \sum_{i=0}^{20-j} \beta_j^{(m)} \theta_i \theta_{i+j} = (\sum_{j=0}^{20} \theta_j)^2$, de modo que

$$\beta_j^{(m)} \xrightarrow[m \rightarrow \infty]{} 1 \Rightarrow \rho_{DCCA}(m) \xrightarrow{P} 1, \text{ quando } n, m \rightarrow \infty.$$

4. RESULTADOS

Os resultados serão apresentados através de tabelas, contendo as seguintes informações sobre o coeficiente de correlação cruzada DCCA e a correlação cruzada: média e desvio padrão, quantis $Q_{0.025}$ e $Q_{0.0975}$, p-valor do teste de normalidade Shapiro-Wilk para a distribuição dos coeficientes e limites inferior e superior do Intervalo de Confiança, e também boxplots dos coeficientes de correlação cruzada DCCA e correlação cruzada.

O p-valor do teste de normalidade Shapiro-Wilk tem o objetivo de saber se a distribuição dos coeficientes possui distribuição Normal e assim utilizar os respectivos intervalos de confiança.

4.1. Cenário 1 – Séries i.i.d e independentes entre si

De acordo com a estrutura das duas séries temporais analisadas, esperava-se obter coeficientes de correlação DCCA igual a zero. A partir da Tabela 1 e Gráfico 1, identificamos que de fato está em torno de zero, independente do lag. Também, notamos que para todos os lags o zero está contido nos respectivos intervalos de confiança. Analisando o desvio padrão, percebe-se que este aumenta à medida que o lag também aumenta, fazendo assim com que os intervalos de confiança para os coeficientes fiquem com amplitude maior. Acredita-se que isto se deve ao fato de que no método DCCA, à medida que o lag aumenta, temos janelas de tamanho maior, porém uma menor quantidade de janelas, fazendo com que tenhamos menos observações na hora de calcular as variâncias. Analisando o boxplot, nota-se que as medianas do coeficiente não possuem uma alteração visível graficamente entre os lag, se mantendo no mesmo nível. Nota-se também que existe presença de valores atípicos, tanto inferiores quanto superiores. Avaliando os p-valores, em nenhum lag temos a rejeição da hipótese nula no teste de Shapiro-Wilk, sendo assim, não há evidência de que a distribuição dos coeficientes não seja Normal. Isto pode ser evidenciado também através dos quantis que demonstram uma simetria em torno do zero.

Na Tabela 2 e Gráfico 2, observamos os resultados para a correlação cruzada tradicional. Assim como no DCCA, os resultados estão em torno de zero para todos os lags, indicando ausência de correlação cruzada. Contudo, a variabilidade dos dados possui uma maior estabilidade entre os lags, não aumentando a amplitude à medida que o lag aumenta, o contrario do DCCA. Também, a mediana parece possuir uma pequena, mas visível oscilação entre os lags. Assim como no DCCA, podemos evidenciar valores atípicos, tanto inferiores quanto superiores e os p-valores do teste Shapiro-Wilk também indicam normalidade dos dados, considerando 5% de confiança. Como um resultado interessante, percebe-se que apenas para o lag 17 o intervalo de confiança não contém o zero.

Tabela 1 – DCCA do cenário 1

M	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,0021	0,0814	-0,1567	0,1634	0,2322	-0,0029	0,0072
4	0,0018	0,0799	-0,1578	0,1655	0,1332	-0,0031	0,0068
5	0,0016	0,0802	-0,1558	0,1601	0,1554	-0,0034	0,0066
6	0,0015	0,0817	-0,1549	0,1681	0,3260	-0,0036	0,0065
7	0,0013	0,0835	-0,1591	0,1660	0,5322	-0,0039	0,0065
8	0,0008	0,0857	-0,1669	0,1695	0,7022	-0,0045	0,0061
9	0,0003	0,0882	-0,1758	0,1712	0,9320	-0,0052	0,0057
10	-0,0002	0,0909	-0,1830	0,1730	0,9110	-0,0058	0,0054
11	-0,0005	0,0938	-0,1836	0,1795	0,9235	-0,0063	0,0053
12	-0,0007	0,0968	-0,1932	0,1846	0,8342	-0,0067	0,0053
13	-0,0008	0,1001	-0,2000	0,1901	0,6816	-0,0070	0,0054
14	-0,0008	0,1033	-0,2083	0,1994	0,8003	-0,0072	0,0056
15	-0,0007	0,1065	-0,2141	0,2073	0,8460	-0,0073	0,0059
16	-0,0005	0,1097	-0,2189	0,2131	0,8252	-0,0073	0,0063
17	-0,0004	0,1128	-0,2242	0,2157	0,8177	-0,0073	0,0066
18	-0,0002	0,1159	-0,2303	0,2219	0,8455	-0,0074	0,0070
19	0,0000	0,1189	-0,2360	0,2295	0,8664	-0,0074	0,0074
20	0,0002	0,1219	-0,2382	0,2360	0,8887	-0,0074	0,0077

Gráfico 1 – DCCA do cenário 1

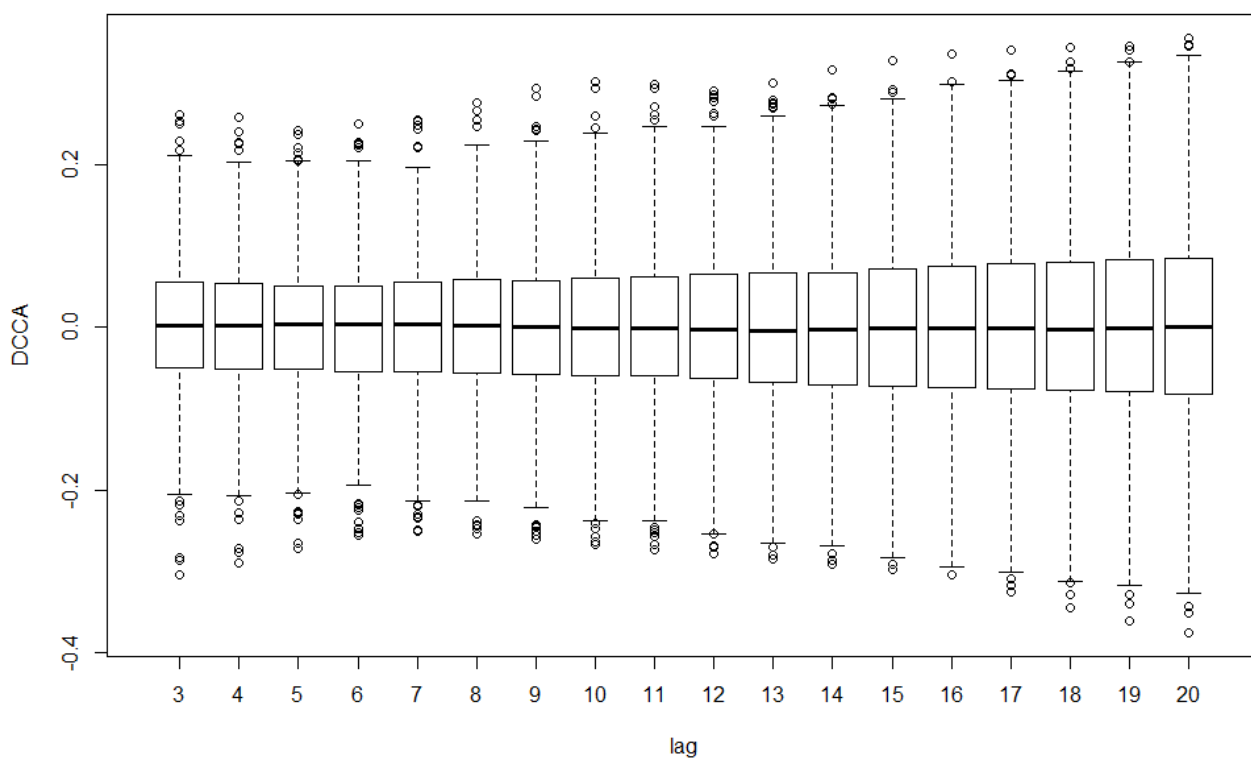
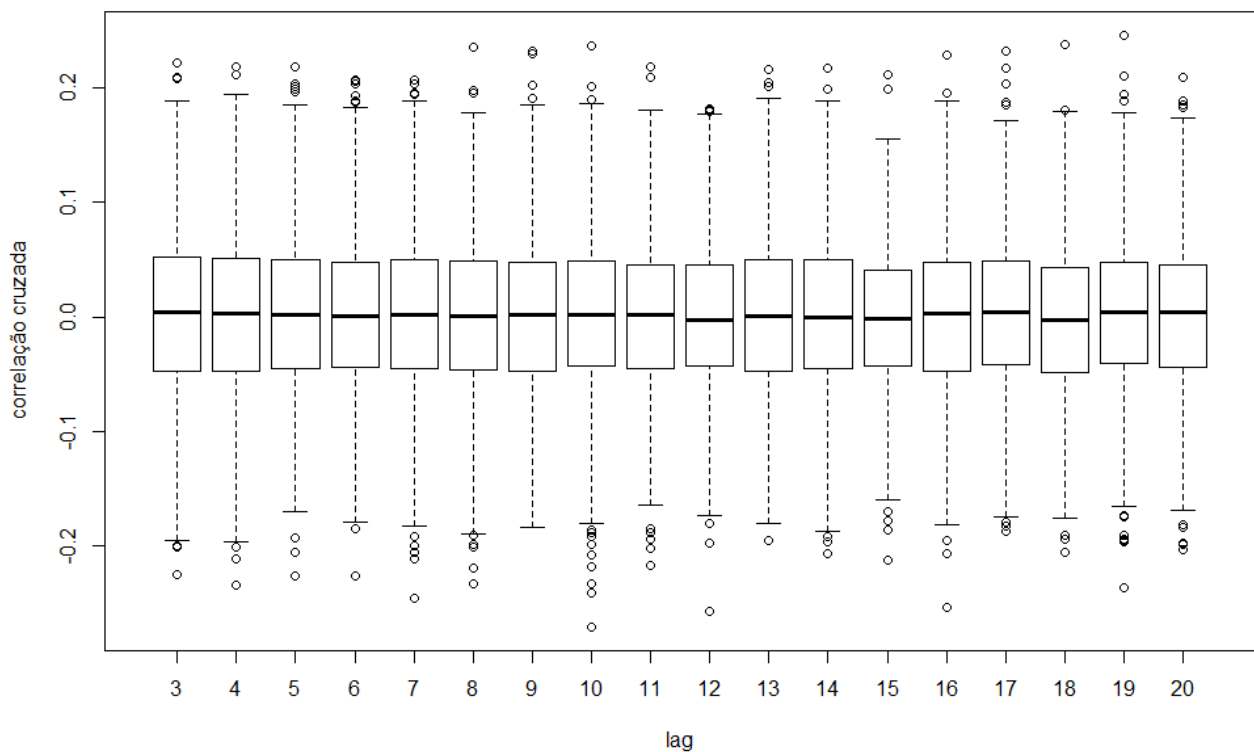


Tabela 2 – Correlação cruzada do cenário 1

m	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,0019	0,0713	-0,1378	0,1280	0,2723	-0,0025	0,0063
4	0,0025	0,0722	-0,1418	0,1369	0,6868	-0,0019	0,0070
5	0,0027	0,0697	-0,1311	0,1369	0,8992	-0,0016	0,0070
6	0,0023	0,0702	-0,1348	0,1398	0,6774	-0,0021	0,0066
7	0,0035	0,0698	-0,1289	0,1448	0,3601	-0,0008	0,0078
8	0,0001	0,0711	-0,1398	0,1398	0,9922	-0,0043	0,0045
9	0,0018	0,0696	-0,1273	0,1326	0,3713	-0,0025	0,0061
10	0,0021	0,0689	-0,1276	0,1361	0,0419	-0,0022	0,0064
11	0,0010	0,0675	-0,1306	0,1294	0,8726	-0,0031	0,0052
12	-0,0002	0,0678	-0,1410	0,1355	0,3537	-0,0044	0,0040
13	0,0022	0,0688	-0,1253	0,1383	0,1093	-0,0020	0,0065
14	-0,0002	0,0679	-0,1300	0,1331	0,7402	-0,0044	0,0040
15	-0,0006	0,0630	-0,1237	0,1270	0,7733	-0,0046	0,0033
16	0,0013	0,0692	-0,1414	0,1399	0,9168	-0,0030	0,0056
17	0,0045	0,0682	-0,1253	0,1362	0,5763	0,0003	0,0087
18	-0,0021	0,0666	-0,1365	0,1254	0,9064	-0,0063	0,0020
19	0,0013	0,0667	-0,1331	0,1306	0,4626	-0,0028	0,0054
20	0,0005	0,0677	-0,1377	0,1312	0,2092	-0,0037	0,0047

Gráfico 2 – Correlação cruzada do cenário 1



4.2. Cenário 2 – Séries i.i.d e dependentes entre si

4.2.1. Sinal + Ruído

Para este cenário, esperava-se que através do método DCCA encontrássemos $\rho_{DCCA}(m) \approx 0,7071$. Podemos observar na Tabela 3 e Gráfico 3 que isto de fato acontece, além de que para todos os lags este valor está dentro dos intervalos de confiança. Também, há um leve decaimento do coeficiente a partir da terceira casa decimal à medida que o lag aumenta. Novamente, assim como no cenário 1, o desvio padrão aumenta, ocasionando intervalos de maior amplitude para lags maiores. Analisando o boxplot, fica evidente a grande presença de valores atípicos inferiores e a inexistência de valores atípicos superiores para lags acima de 11. Podemos perceber também que para lags maiores temos assimetria na distribuição dos coeficientes, o que é um indicativo importante para os resultados do teste de Shapiro-Wilk, que traz indícios de que os dados não sigam uma distribuição Normal.

A Tabela 4 e Gráfico 4 trazem informações a respeito da correlação cruzada tradicional, que para todos os lags oscila em torno de zero. Analisando os intervalos de confiança, observamos que para quatro lags não consta o zero. De acordo com os resultados do teste Shapiro-Wilk e análise do boxplot, possuímos indícios que estes coeficientes seguem distribuição Normal.

4.2.2. Normal bivariada com correlação 0.5

Neste caso, conforme a Tabela 5 e Gráfico 5, temos um comportamento semelhante ao caso Sinal + Ruído, em que o coeficiente está de acordo com o esperado e contido dentro de todos intervalos de confiança, assim como a distribuição destes não possui indícios de seguir distribuição Normal.

Para a correlação cruzada tradicional, Tabela 6 e Gráfico 6, também temos similaridade com o primeiro caso deste cenário, em que a correlação está próximo do zero, mas em 4 lags não está no intervalo de confiança. Há forte indicativos para a distribuição Normal dos dados através do teste Shapiro-Wilk e também do comportamento nos boxplots.

Tabela 3 – DCCA do cenário 2

M	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,7070	0,0415	0,6174	0,7847	0,0007	0,7044	0,7096
4	0,7067	0,0415	0,6147	0,7839	0,0001	0,7042	0,7093
5	0,7066	0,0418	0,6149	0,7834	0,0000	0,7040	0,7092
6	0,7066	0,0423	0,6163	0,7853	0,0001	0,7040	0,7092
7	0,7065	0,0430	0,6156	0,7842	0,0007	0,7038	0,7092
8	0,7063	0,0440	0,6150	0,7876	0,0020	0,7036	0,7090
9	0,7060	0,0451	0,6121	0,7898	0,0044	0,7032	0,7088
10	0,7057	0,0463	0,6065	0,7898	0,0036	0,7029	0,7086
11	0,7054	0,0478	0,6029	0,7918	0,0014	0,7025	0,7084
12	0,7052	0,0495	0,5973	0,7943	0,0003	0,7021	0,7083
13	0,7051	0,0513	0,5958	0,7934	0,0001	0,7019	0,7082
14	0,7049	0,0531	0,5934	0,7959	0,0000	0,7017	0,7082
15	0,7049	0,0550	0,5928	0,7985	0,0000	0,7015	0,7083
16	0,7048	0,0569	0,5839	0,8007	0,0000	0,7013	0,7083
17	0,7048	0,0588	0,5757	0,8034	0,0000	0,7011	0,7084
18	0,7047	0,0606	0,5717	0,8057	0,0000	0,7009	0,7085
19	0,7046	0,0625	0,5672	0,8099	0,0000	0,7008	0,7085
20	0,7046	0,0643	0,5647	0,8116	0,0000	0,7006	0,7086

Gráfico 3 – DCCA do cenário 2

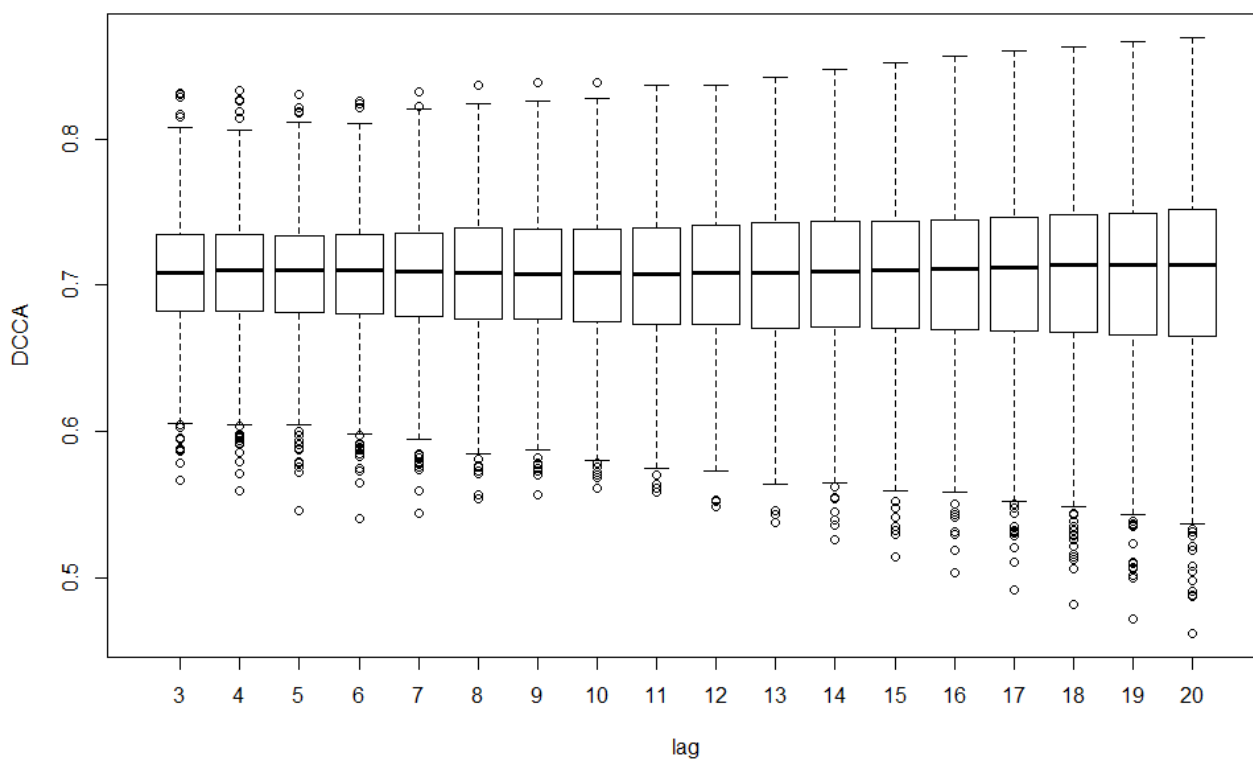


Tabela 4 – Correlação cruzada do cenário 2

m	média	dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	-0,0010	0,0697	-0,1416	0,1317	0,7314	-0,0053	0,0033
4	-0,0035	0,0715	-0,1408	0,1332	0,7888	-0,0079	0,0010
5	-0,0010	0,0708	-0,1405	0,1438	0,2765	-0,0054	0,0034
6	-0,0025	0,0708	-0,1426	0,1381	0,4759	-0,0069	0,0019
7	-0,0014	0,0689	-0,1338	0,1305	0,4255	-0,0057	0,0029
8	-0,0015	0,0701	-0,1344	0,1337	0,6287	-0,0059	0,0028
9	-0,0032	0,0678	-0,1428	0,1269	0,2435	-0,0074	0,0010
10	-0,0045	0,0667	-0,1437	0,1305	0,0451	-0,0087	-0,0004
11	-0,0043	0,0677	-0,1372	0,1224	0,0506	-0,0085	-0,0001
12	-0,0063	0,0680	-0,1366	0,1247	0,5357	-0,0105	-0,0021
13	-0,0029	0,0683	-0,1368	0,1245	0,8718	-0,0071	0,0013
14	-0,0026	0,0634	-0,1315	0,1258	0,2093	-0,0066	0,0013
15	-0,0036	0,0640	-0,1283	0,1224	0,7244	-0,0076	0,0004
16	-0,0037	0,0678	-0,1420	0,1194	0,3389	-0,0079	0,0005
17	-0,0026	0,0687	-0,1338	0,1256	0,5620	-0,0069	0,0016
18	-0,0039	0,0655	-0,1395	0,1160	0,2734	-0,0079	0,0002
19	-0,0054	0,0662	-0,1349	0,1245	0,1679	-0,0095	-0,0013
20	0,0005	0,0662	-0,1381	0,1304	0,1319	-0,0036	0,0046

Gráfico 4 – Correlação cruzada do cenário 2

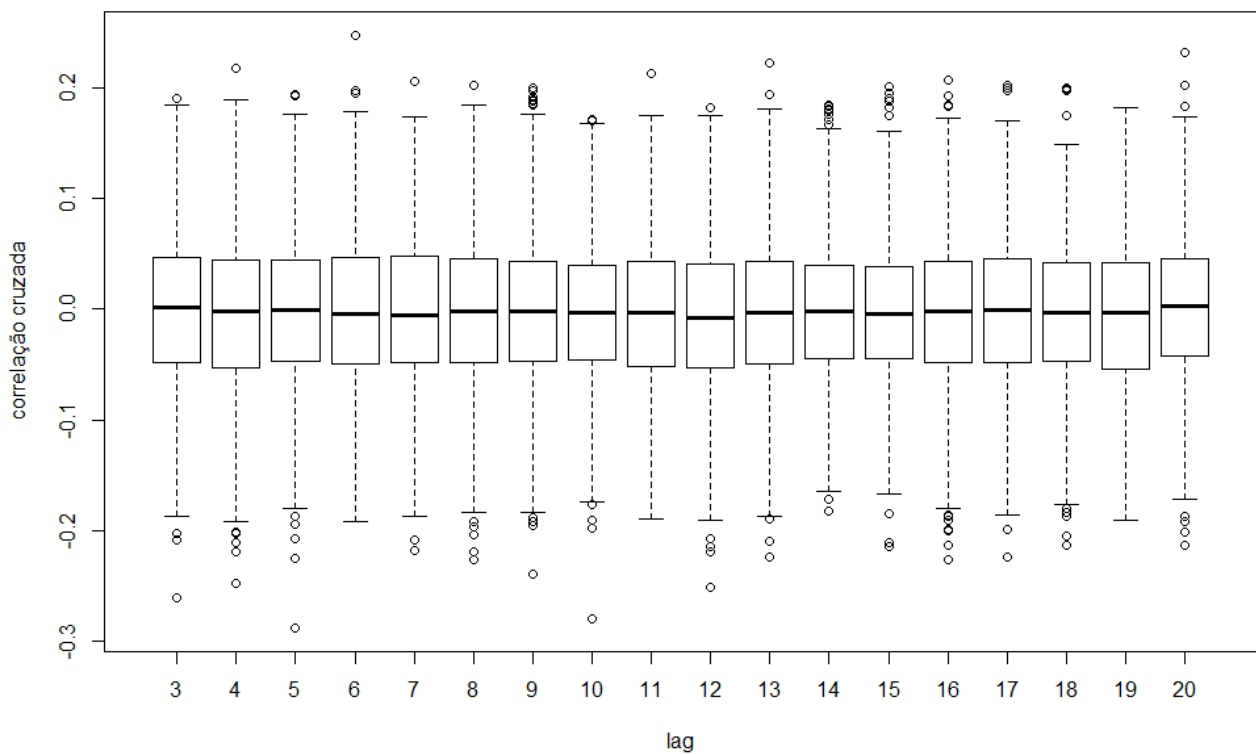


Tabela 5 – DCCA do cenário 2 correlação 0.5

m	Média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,4996	0,0620	0,3686	0,6152	0,0291	0,4958	0,5035
4	0,4993	0,0616	0,3657	0,6132	0,0029	0,4955	0,5031
5	0,4992	0,0621	0,3642	0,6133	0,0032	0,4954	0,5031
6	0,4992	0,0632	0,3659	0,6135	0,0050	0,4953	0,5031
7	0,4992	0,0647	0,3643	0,6177	0,0036	0,4952	0,5032
8	0,4991	0,0666	0,3688	0,6270	0,0034	0,4950	0,5033
9	0,4990	0,0685	0,3614	0,6278	0,0049	0,4947	0,5032
10	0,4987	0,0703	0,3560	0,6288	0,0054	0,4944	0,5031
11	0,4985	0,0723	0,3471	0,6311	0,0076	0,4940	0,5030
12	0,4984	0,0743	0,3422	0,6315	0,0074	0,4938	0,5030
13	0,4983	0,0765	0,3436	0,6364	0,0061	0,4936	0,5030
14	0,4983	0,0787	0,3389	0,6370	0,0047	0,4934	0,5032
15	0,4983	0,0810	0,3344	0,6427	0,0043	0,4933	0,5033
16	0,4983	0,0833	0,3279	0,6487	0,0026	0,4931	0,5035
17	0,4983	0,0857	0,3218	0,6560	0,0016	0,4930	0,5036
18	0,4982	0,0881	0,3105	0,6591	0,0010	0,4928	0,5037
19	0,4982	0,0905	0,3077	0,6590	0,0005	0,4926	0,5038
20	0,4983	0,0928	0,3023	0,6604	0,0002	0,4925	0,5040

Gráfico 5 – DCCA do cenário 2 correlação 0.5

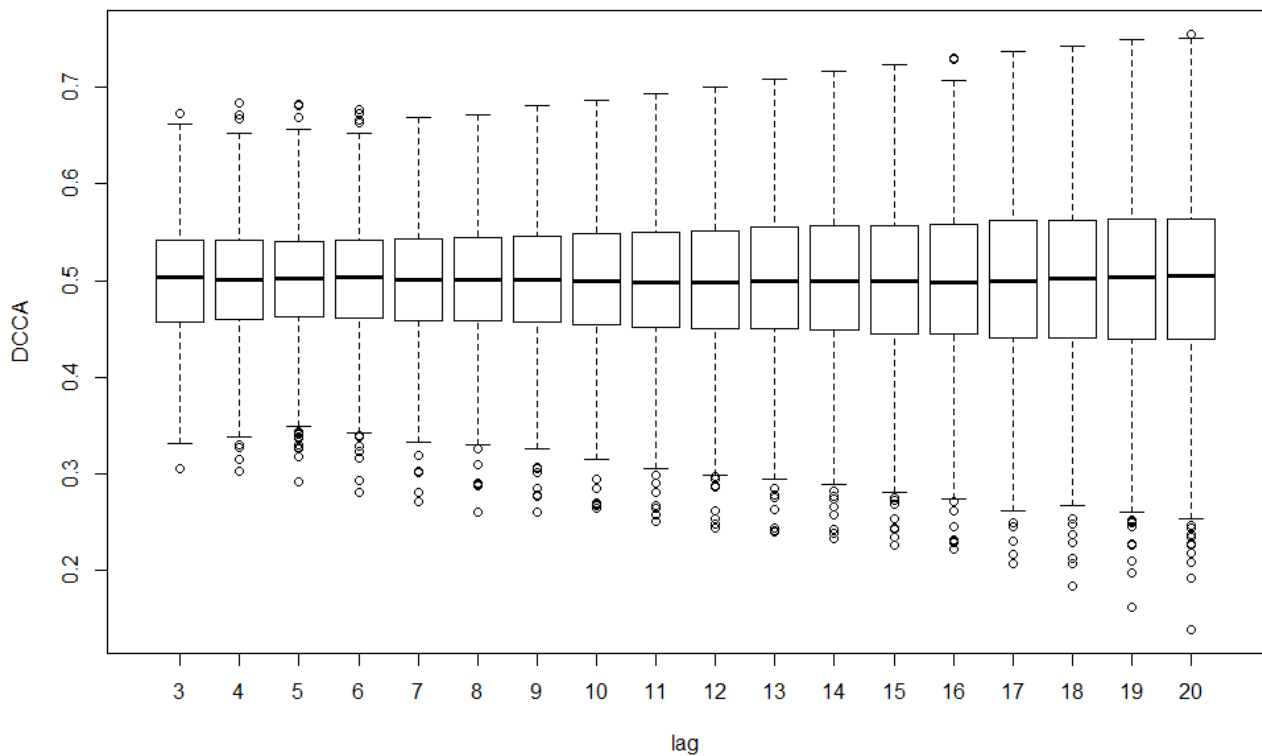
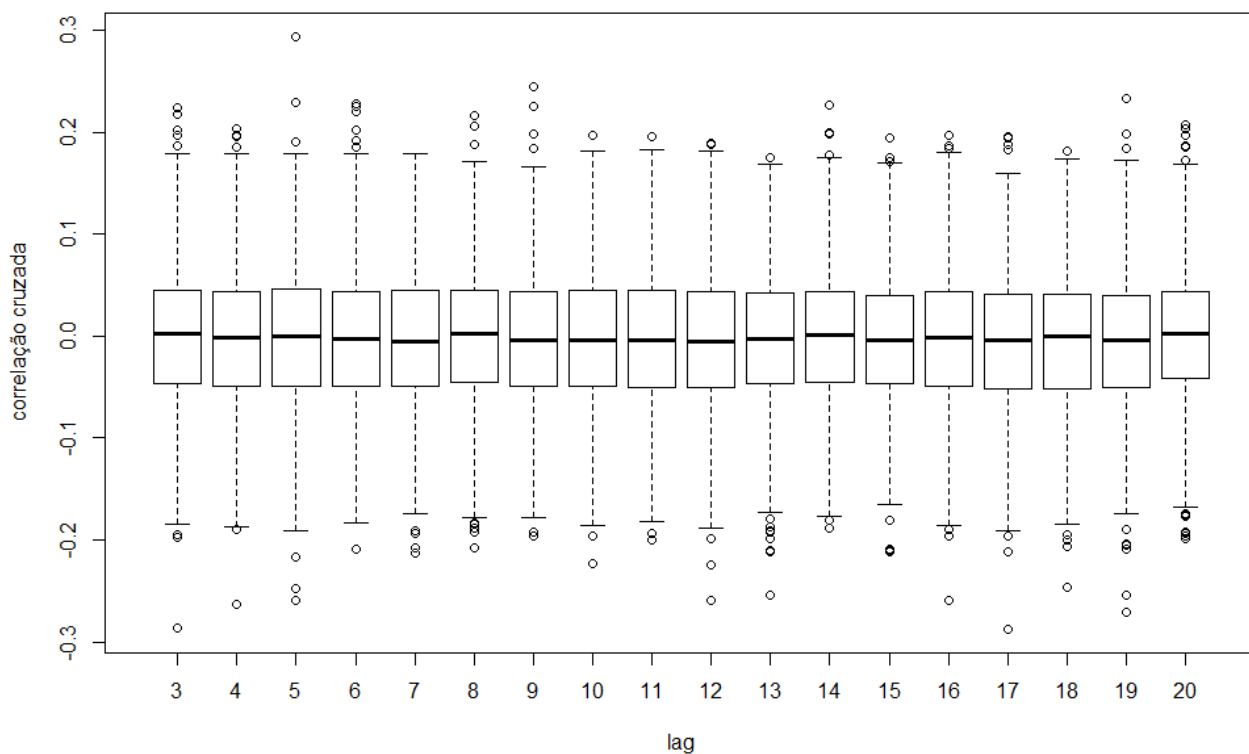


Tabela 6 – Correlação cruzada do cenário 2 correlação 0.5

m	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,0003	0,0709	-0,1436	0,1406	0,7730	-0,0041	0,0047
4	-0,0027	0,0692	-0,1404	0,1353	0,8174	-0,0070	0,0016
5	-0,0007	0,0719	-0,1371	0,1324	0,5389	-0,0052	0,0037
6	-0,0023	0,0699	-0,1416	0,1296	0,6031	-0,0066	0,0021
7	-0,0033	0,0691	-0,1433	0,1254	0,0915	-0,0076	0,0010
8	-0,0005	0,0690	-0,1349	0,1287	0,4426	-0,0048	0,0037
9	-0,0026	0,0681	-0,1380	0,1229	0,5658	-0,0069	0,0016
10	-0,0020	0,0683	-0,1383	0,1370	0,7155	-0,0062	0,0023
11	-0,0041	0,0676	-0,1363	0,1259	0,3898	-0,0083	0,0001
12	-0,0048	0,0685	-0,1331	0,1315	0,7876	-0,0090	-0,0006
13	-0,0031	0,0676	-0,1364	0,1272	0,2742	-0,0073	0,0011
14	-0,0011	0,0650	-0,1266	0,1287	0,2979	-0,0051	0,0029
15	-0,0031	0,0641	-0,1319	0,1262	0,9073	-0,0070	0,0009
16	-0,0027	0,0680	-0,1347	0,1313	0,7252	-0,0069	0,0015
17	-0,0044	0,0686	-0,1442	0,1231	0,5325	-0,0087	-0,0002
18	-0,0044	0,0674	-0,1330	0,1248	0,2434	-0,0086	-0,0002
19	-0,0055	0,0666	-0,1338	0,1267	0,3490	-0,0096	-0,0014
20	0,0010	0,0648	-0,1304	0,1250	0,1280	-0,0030	0,0050

Gráfico 6 – Correlação cruzada do cenário 2 correlação 0.5



4.2.3. Normal bivariada com correlação 0.8

Neste caso, através da Tabela 7 e Gráfico 7, podemos reforçar as análises feitas nos demais casos de coeficiente de correlação DCCA entre duas séries i.i.d e dependentes, em que os resultados dos coeficientes para todos os lags estão em torno dos esperado e constando dentro de todos intervalos de confiança. O comportamento de aumento do desvio padrão e, conseqüentemente, da amplitude dos intervalos de confiança também se repete. Novamente os coeficientes não apresentam indicativos de seguirem distribuição Normal para este cenário de dependência entre as séries. Para investigar isto, suspeitando que a quantidade de replicações possa impactar na convergência a normalidade, foi realizada a simulação de Monte Carlo para 5000 replicações apenas o lag 15, escolhido de maneira aleatória e considerando que para os demais casos será análogo. Também foi feita análise, tanto para 1000 quanto 5000 replicações, retirando-se os valores atípicos. Conforme Gráfico 8 e Quadro 1, percebemos que mantemos a evidência de que os coeficientes não seguem distribuição Normal. Acredita-se que isto se deve ao fato do tamanho de amostra alto exigir um ajuste muito perfeito dos dados à distribuição para que não haja a rejeição da hipótese nula no teste de Shapiro-Wilk.

Tabela 7 – DCCA do cenário 2 correlação 0.8

m	Média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,7994	0,0299	0,7340	0,8536	0,0001	0,7976	0,8013
4	0,7993	0,0298	0,7322	0,8520	0,0000	0,7974	0,8011
5	0,7992	0,0301	0,7315	0,8522	0,0000	0,7974	0,8011
6	0,7992	0,0306	0,7323	0,8521	0,0000	0,7973	0,8011
7	0,7992	0,0313	0,7312	0,8540	0,0000	0,7972	0,8011
8	0,7991	0,0323	0,7338	0,8583	0,0000	0,7971	0,8011
9	0,7990	0,0332	0,7297	0,8587	0,0000	0,7970	0,8011
10	0,7989	0,0341	0,7267	0,8588	0,0000	0,7968	0,8010
11	0,7987	0,0351	0,7223	0,8599	0,0000	0,7966	0,8009
12	0,7986	0,0361	0,7195	0,8610	0,0000	0,7964	0,8009
13	0,7986	0,0371	0,7200	0,8622	0,0000	0,7963	0,8009
14	0,7985	0,0382	0,7179	0,8633	0,0000	0,7962	0,8009
15	0,7985	0,0393	0,7160	0,8659	0,0000	0,7961	0,8010
16	0,7985	0,0405	0,7123	0,8687	0,0000	0,7960	0,8010
17	0,7984	0,0416	0,7087	0,8711	0,0000	0,7959	0,8010
18	0,7984	0,0428	0,7030	0,8721	0,0000	0,7957	0,8010
19	0,7983	0,0440	0,7007	0,8723	0,0000	0,7956	0,8011
20	0,7983	0,0452	0,6989	0,8724	0,0000	0,7955	0,8011

Gráfico 7 – DCCA do cenário 2 correlação 0.8

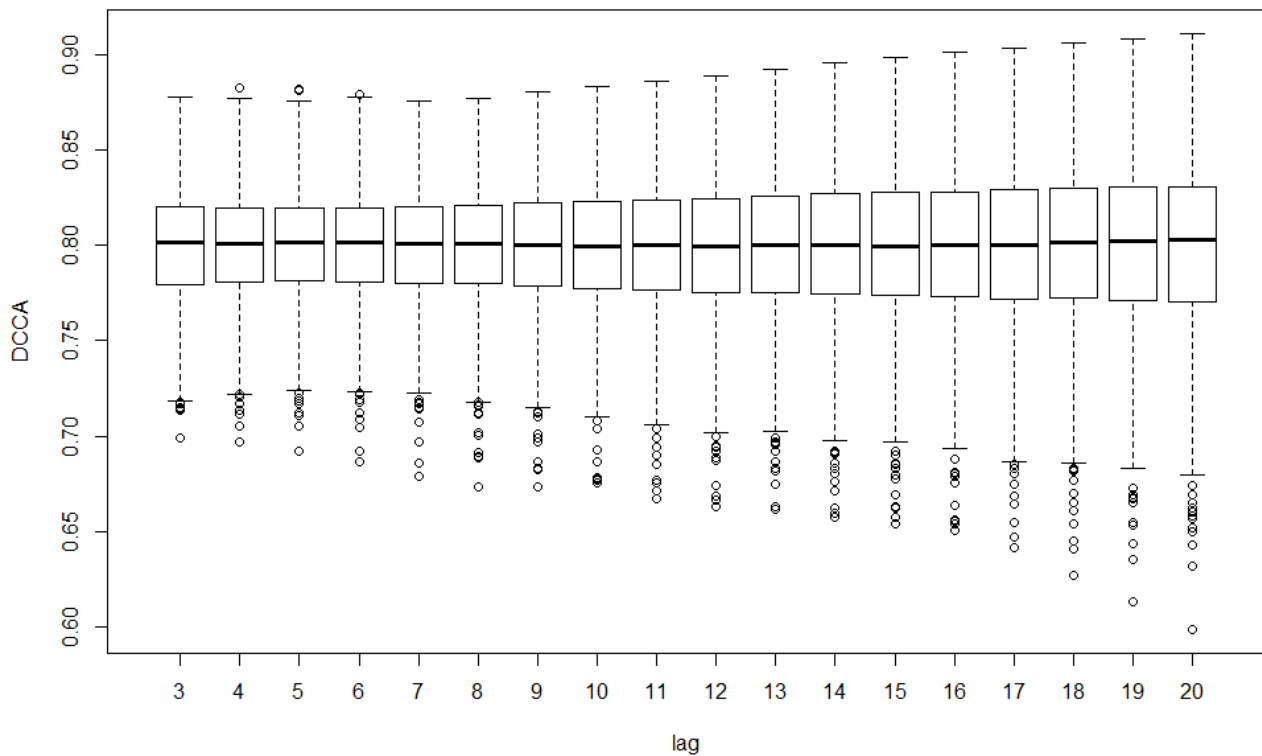
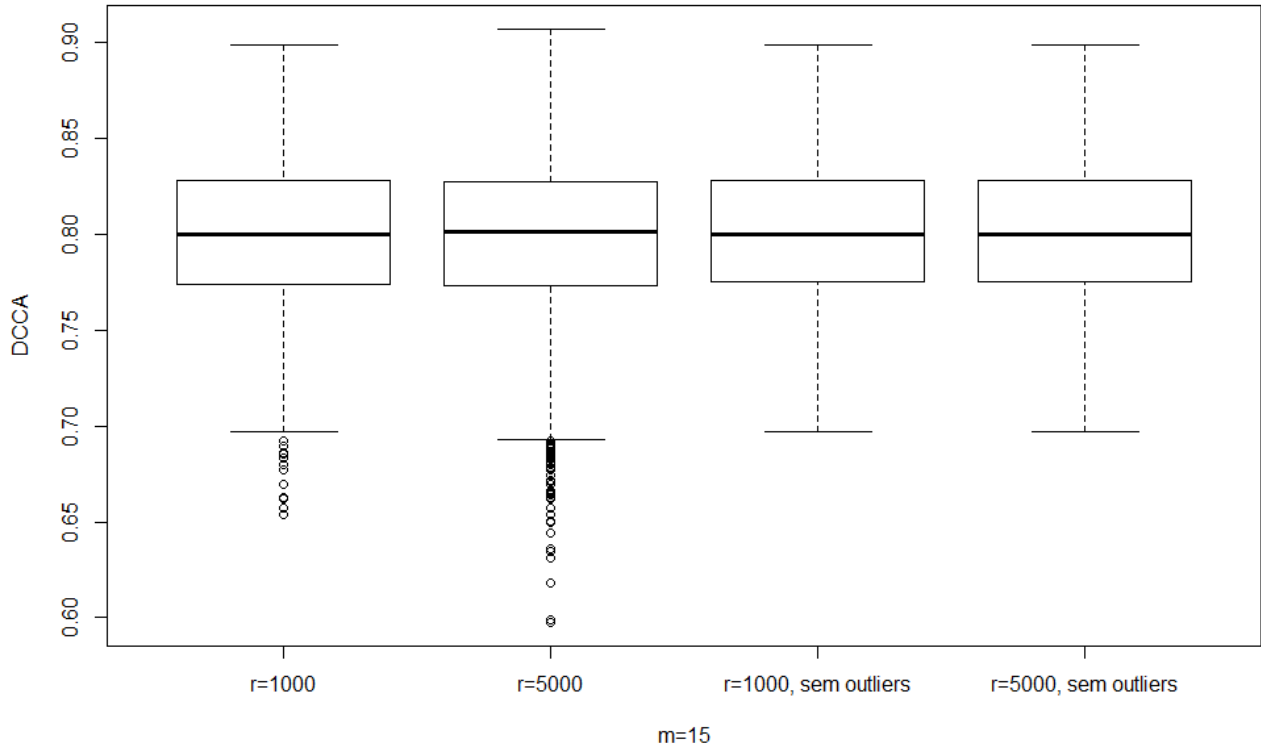


Gráfico 8 – DCCA do cenário 2 correlação 0.8 para m=15 considerando cenários para convergência à normalidade



Quadro 1 – Teste de Shapiro-Wilk considerando cenários para convergência à normalidade

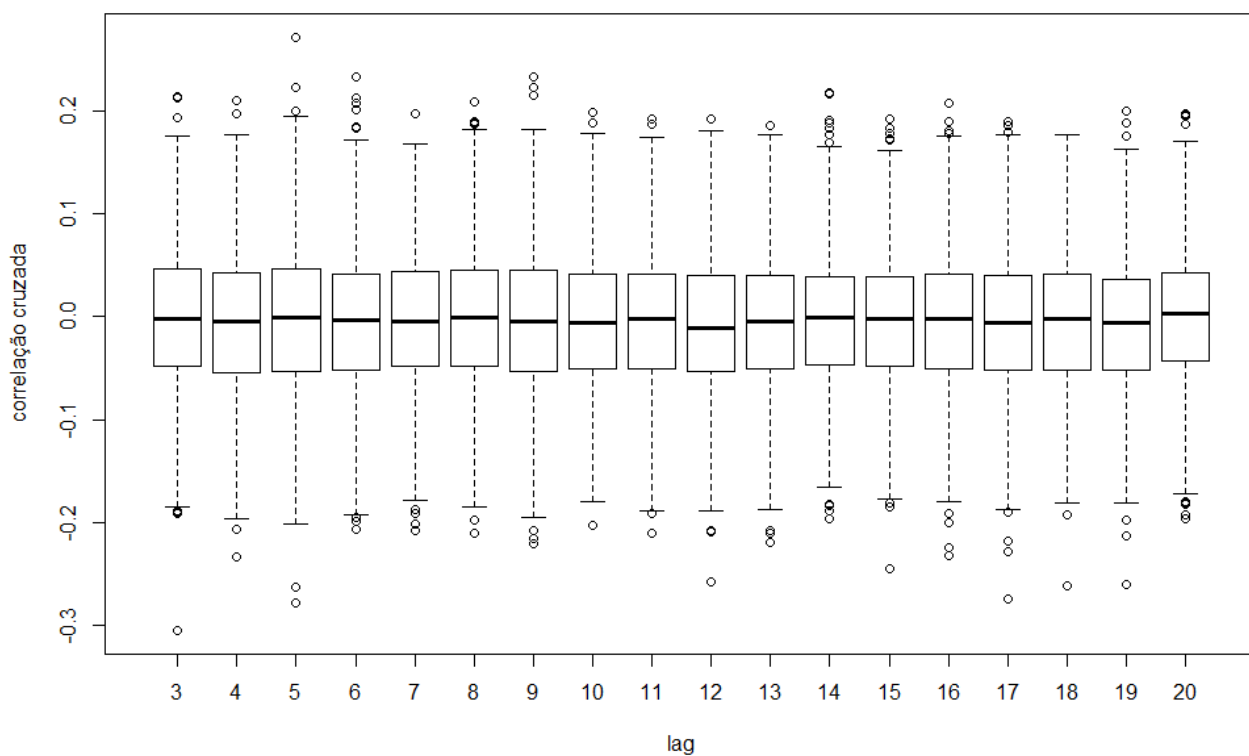
m=15	r=1000	r=5000	r=1000, sem outliers	r=5000, sem outliers
p-valor	$2,5 \times 10^{-7}$	$2,8 \times 10^{-17}$	0,001432553	0,001432553

Com relação aos resultados da correlação cruzada tradicional neste caso, Tabela 8 e Gráfico 9, o único comportamento divergente dos demais casos é quanto aos intervalos de confiança, em que houve um aumento expressivo para 11 lags em que o zero não está contido.

Tabela 8 – Correlação cruzada do cenário 2 correlação 0.8

m	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	-0,0013	0,0703	-0,1369	0,1335	0,4662	-0,0057	0,0030
4	-0,0052	0,0689	-0,1364	0,1301	0,9055	-0,0094	-0,0009
5	-0,0022	0,0721	-0,1420	0,1331	0,6836	-0,0066	0,0023
6	-0,0042	0,0708	-0,1474	0,1320	0,3550	-0,0086	0,0002
7	-0,0046	0,0686	-0,1428	0,1267	0,1403	-0,0089	-0,0004
8	-0,0020	0,0697	-0,1396	0,1317	0,8111	-0,0063	0,0024
9	-0,0045	0,0692	-0,1412	0,1250	0,6650	-0,0088	-0,0002
10	-0,0050	0,0683	-0,1391	0,1280	0,4021	-0,0093	-0,0008
11	-0,0057	0,0673	-0,1371	0,1209	0,2452	-0,0098	-0,0015
12	-0,0070	0,0678	-0,1417	0,1309	0,5481	-0,0112	-0,0028
13	-0,0050	0,0673	-0,1346	0,1264	0,8377	-0,0092	-0,0009
14	-0,0028	0,0640	-0,1319	0,1250	0,5221	-0,0067	0,0012
15	-0,0040	0,0645	-0,1263	0,1212	0,9061	-0,0080	0,0000
16	-0,0046	0,0676	-0,1344	0,1249	0,8623	-0,0088	-0,0004
17	-0,0063	0,0683	-0,1403	0,1264	0,7802	-0,0105	-0,0020
18	-0,0042	0,0664	-0,1337	0,1250	0,2779	-0,0083	-0,0001
19	-0,0076	0,0663	-0,1405	0,1231	0,7232	-0,0117	-0,0035
20	0,0005	0,0644	-0,1372	0,1337	0,0920	-0,0035	0,0045

Gráfico 9 – Correlação cruzada do cenário 2 correlação 0.8



4.3. Cenário 3 – Séries com correlação cruzada não-nula

Através da Tabela 9 e Gráfico 10, observamos que os resultados obtidos nas simulações para o coeficiente de correlação DCCA estão muito próximos aos valores esperados, sendo iguais até a terceira casa decimal e também que todos estão contidos nos respectivos intervalos de confiança. O padrão já identificado de aumento do desvio padrão à medida que aumenta o lag também se repete neste cenário. Cabe notar que a normalidade dos coeficientes apresentou um comportamento diferente, em que apenas 8 lags possuem indícios de seguirem distribuição Normal, de acordo com o teste de normalidade Shapiro-Wilk. No cenário 1 temos evidências de normalidade para todos os lags e, no cenário 2, o oposto. Observa-se também que os valores atípicos neste cenário aparecem na parte inferior e superior e que a distância interquartil e entre os bigodes aumenta expressivamente com o aumento dos lags, refletindo um aumento da variância do coeficiente conforme o lag aumenta.

Para a correlação cruzada tradicional, Tabela 10 e Gráfico 11, os resultados obtidos possuem certa diferença em relação ao esperado, sendo que esta diferença aumenta de maneira expressiva de acordo como aumento dos lags. Devido a isto, nenhum dos valores esperados consta nos respectivos intervalos de confiança. A estabilidade dos desvios padrões mantém-se assim como nos demais cenários. Também conforme os demais cenários temos indicativos para a normalidade dos coeficientes para todos os lags de acordo com o teste Shapiro-Wilk e são percebidos valores atípicos superiores e inferiores.

Tabela 9 – DCCA do cenário 3

m	esperado	média	Dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,0169	0,0166	0,0069	0,0029	0,0296	0,0115	0,0162	0,0171
4	0,0301	0,0300	0,0078	0,0156	0,0452	0,0001	0,0295	0,0305
5	0,0419	0,0419	0,0091	0,0244	0,0592	0,0782	0,0413	0,0425
6	0,0529	0,0529	0,0108	0,0319	0,0733	0,3343	0,0523	0,0536
7	0,0634	0,0634	0,0126	0,0395	0,0883	0,0157	0,0626	0,0642
8	0,0737	0,0737	0,0143	0,0480	0,1027	0,0014	0,0728	0,0746
9	0,0838	0,0838	0,0163	0,0550	0,1186	0,0004	0,0828	0,0848
10	0,0939	0,0938	0,0183	0,0616	0,1318	0,0007	0,0927	0,0950
11	0,1039	0,1037	0,0205	0,0679	0,1470	0,0007	0,1025	0,1050
12	0,1139	0,1136	0,0229	0,0741	0,1617	0,0017	0,1122	0,1150
13	0,1240	0,1234	0,0254	0,0789	0,1778	0,0038	0,1219	0,1250
14	0,1341	0,1333	0,0279	0,0833	0,1926	0,0151	0,1315	0,1350
15	0,1443	0,1432	0,0306	0,0884	0,2077	0,0503	0,1413	0,1451
16	0,1545	0,1532	0,0333	0,0930	0,2228	0,1361	0,1511	0,1553
17	0,1649	0,1633	0,0360	0,0977	0,2362	0,2652	0,1611	0,1656
18	0,1754	0,1736	0,0387	0,1027	0,2492	0,3559	0,1712	0,1759
19	0,1861	0,1838	0,0413	0,1073	0,2662	0,4093	0,1813	0,1864
20	0,1969	0,1943	0,0438	0,1125	0,2818	0,4821	0,1915	0,1970

Gráfico 10 – DCCA do cenário 3 vs valor esperado (linha)

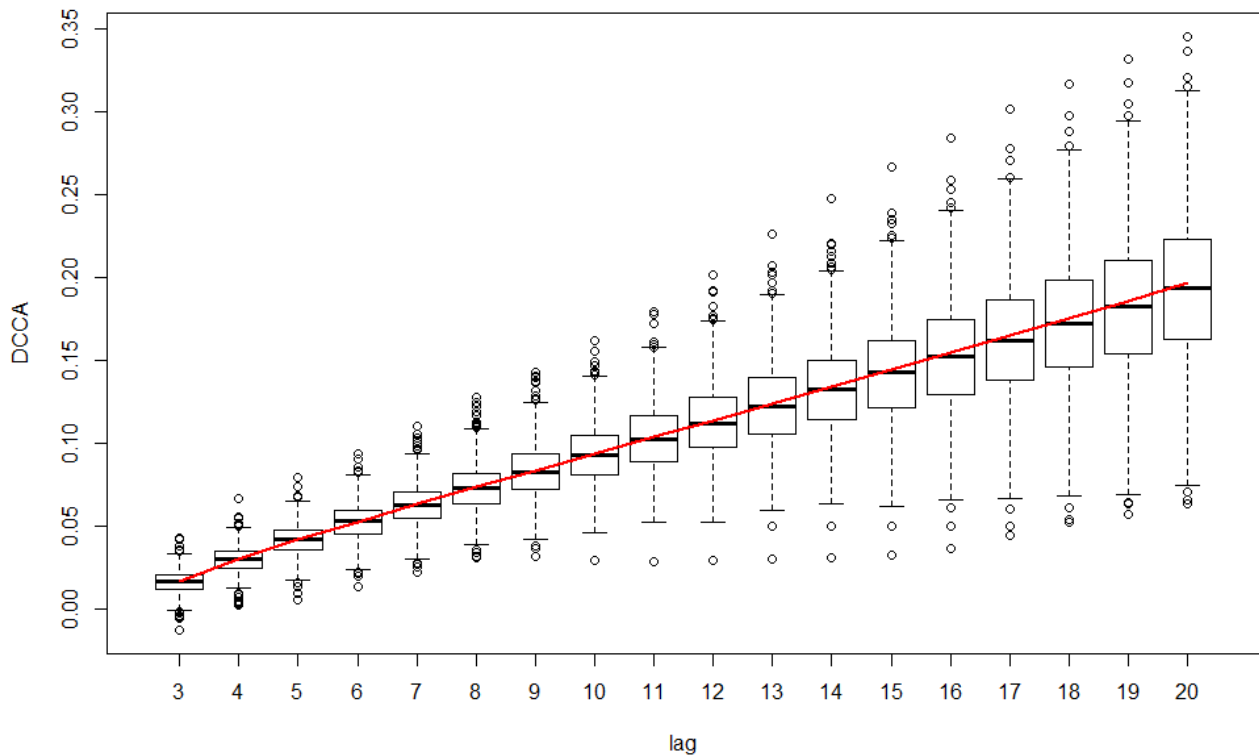
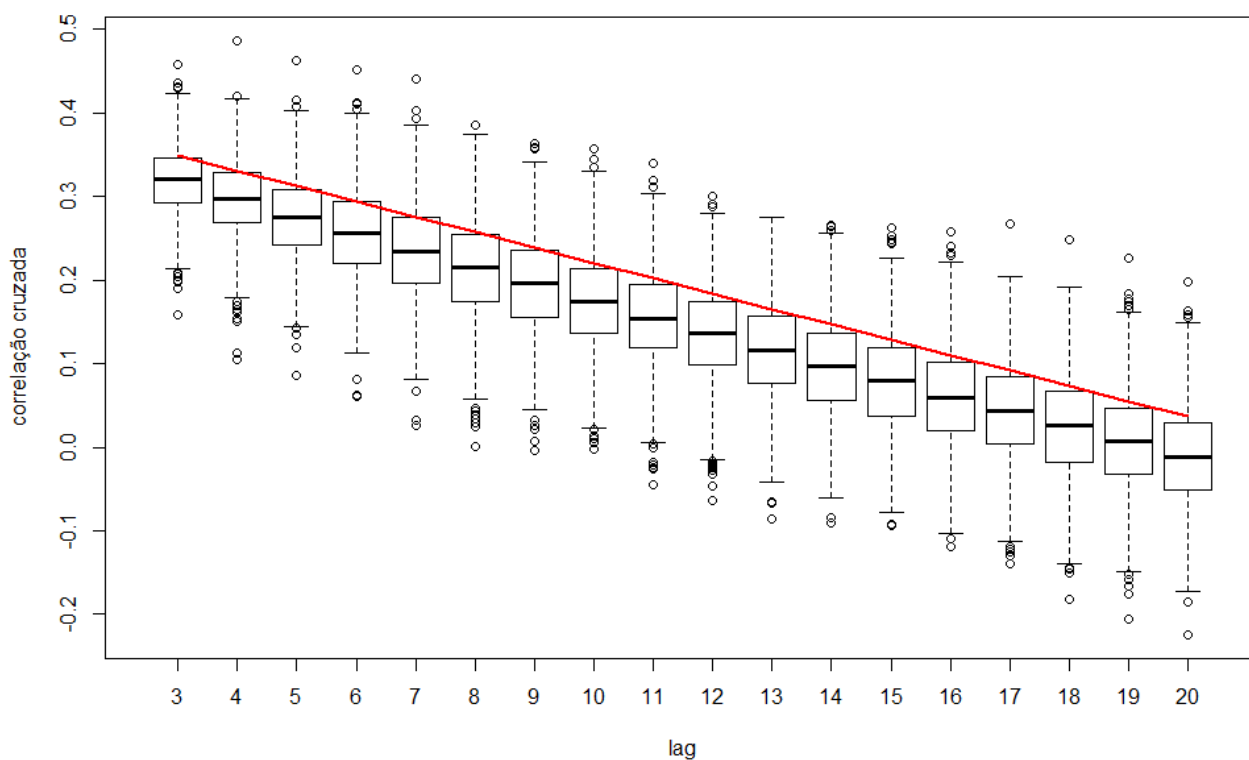


Tabela 10 – Correlação cruzada do cenário 3

m	esperado	média	dp	Q _{0.025}	Q _{0.975}	p-valor	IC _{inf}	IC _{sup}
3	0,3303	0,3189	0,0408	0,2400	0,3978	0,3679	0,3164	0,3214
4	0,3119	0,2969	0,0476	0,2027	0,3889	0,0512	0,2939	0,2998
5	0,2936	0,2754	0,0517	0,1695	0,3740	0,6544	0,2722	0,2786
6	0,2752	0,2550	0,0554	0,1450	0,3598	0,8664	0,2516	0,2585
7	0,2569	0,2341	0,0584	0,1208	0,3466	0,4286	0,2305	0,2378
8	0,2385	0,2140	0,0597	0,0960	0,3293	0,3194	0,2103	0,2177
9	0,2202	0,1944	0,0598	0,0729	0,3124	0,3640	0,1907	0,1981
10	0,2018	0,1741	0,0594	0,0509	0,2863	0,1217	0,1705	0,1778
11	0,1835	0,1541	0,0580	0,0297	0,2617	0,1562	0,1505	0,1577
12	0,1651	0,1341	0,0572	0,0170	0,2422	0,1186	0,1306	0,1377
13	0,1468	0,1145	0,0570	0,0015	0,2250	0,1270	0,1110	0,1180
14	0,1284	0,0960	0,0576	-0,0181	0,2040	0,6722	0,0924	0,0995
15	0,1101	0,0784	0,0581	-0,0321	0,1860	0,4904	0,0748	0,0820
16	0,0917	0,0610	0,0586	-0,0522	0,1719	0,6974	0,0574	0,0646
17	0,0734	0,0433	0,0581	-0,0712	0,1558	0,8331	0,0397	0,0469
18	0,0550	0,0255	0,0594	-0,0877	0,1406	0,8151	0,0219	0,0292
19	0,0367	0,0076	0,0606	-0,1106	0,1288	0,9685	0,0038	0,0113
20	0,0183	-0,0099	0,0618	-0,1289	0,1154	0,8755	-0,0138	-0,0061

Gráfico 11 – Correlação cruzada do cenário 3 vs valor esperado (linha)



5. CONSIDERAÇÕES FINAIS

Analisando de maneira geral o método DCCA, frente à correlação cruzada tradicional e considerando os três cenários investigados, percebemos comportamentos importantes de serem mais investigados.

Considerando séries i.i.d e independentes entre si, ambos os métodos encontram correlação cruzada em torno de zero, de acordo com o esperado. A diferença fica principalmente para o comportamento do desvio padrão dos coeficientes à medida que o lag aumenta, em que na correlação cruzada tradicional se mantém constante enquanto que na DCCA temos um aumento. Acredita-se que isto é devido à maneira que o método é realizado.

Para as séries i.i.d e dependentes entre si, o comportamento esperado é diferente entre os métodos e ambos foram verificados nas simulações. Através das simulações desenvolvidas neste cenário, percebeu-se que a convergência dos coeficientes à normalidade, verificada com uso do teste Shapiro-Wilk, tende a ser mais difícil em relação aos outros dois cenários. Acreditamos que isto seja relacionado com a dependência das séries e também devido ao número de replicações utilizadas e que compunham a amostra. Outro aspecto interessante é quanto à presença de valores atípicos inferiores e ausência dos superiores, a partir de certo lag. Nos intervalos de confiança, no método DCCA para todos os lags, os valores esperados estavam contidos, diferente do método tradicional em que para alguns isto não aconteceu. A estrutura de aumento do desvio padrão foi a mesma do cenário 1.

No cenário 3, que possui estrutura de correlação cruzada entre as séries, observamos que novamente os resultados das simulações se comportaram de acordo com o esperado e também temos o comportamento de aumento do desvio padrão à medida que o lag aumenta, no método DCCA. Destacamos para este cenário o resultado da análise visual entre os métodos, que para o método DCCA teve um ajuste praticamente perfeito e enquanto que a correlação cruzada tradicional teve uma diferença expressiva.

6. CONCLUSÕES

Através deste trabalho foi possível conhecer as metodologias DCCA, DFA e o coeficiente de correlação cruzada destendenciada. Investigamos o método DCCA frente à correlação cruzada tradicional através de resultados de simulações e nos três cenários propostos foram evidenciados os resultados teóricos esperados. Também, foram identificados comportamentos importantes e passíveis de investigação, que se juntam a um estudo maior, conforme já mencionado. Neste projeto, serão trabalhados, além destes cenários já investigados inicialmente, casos em que há de fato a presença da não-estacionariedade nas séries temporais e também de dados faltantes.

Este estudo se torna importante quando traz uma maneira alternativa frente ao método tradicional, que tem como pressuposto a característica da estacionariedade, muitas vezes ausente nos mais diversos casos de estudos. Ainda, contudo, além da compreensão dos resultados, cabe também para trabalhos futuros principalmente a compreensão da análise prática dos resultados obtidos em dados reais, diferente dos simulados aqui trabalhados.

REFERÊNCIAS

- Shapiro, S. e Wilk, M. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*, 52 (3 e 4), 591-611.
- Peng, C., Buldyrev, S., Havlin, S., Simons, M., Stanley, H e Goldberger, A. (1994). "Mosaic organization of DNA nucleotides". *Physical Review E*, 49 (2), 1685-1689.
- Brockwell, P. e Davis, R. (2002). *Introduction to time series and forecasting*. Springer, ISBN 978-0387953519.
- Morettin, P. e Toloí, C. (2004). *Análise de séries temporais*. ABE – Projeto Fisher. Editora Edgard Blucher LTDA, São Paulo.
- Podobnik, B. e Stanley, E. (2008). "Detrended Cross-Correlation Analysis: A New Method for Analysing Two Nonstationary Time Series". *Physical Review Letters*, 100, 084102.
- Podobnik, B., Jiang, Z., Zhou, W. e Stanley, E. (2011). "Statistical tests for power-law cross-correlated process". *Physical Review E*, 84, 066118.
- Zebende, G.(2011). "DCCA cross-correlation coefficient: Quantifying level of cross-correlation". *Physica A*, 390, 614.
- Jun, W. e Da-Qing, Z. (2012). "Detrended cross-correlation of electroencephalogram". *Chinese Physical B*, Vol. 21, No. 2, 028703.
- Zebende, G., Silva, M. e Machado, A. (2013). "DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches". *Physica A*, 392, 1756-1761.
- Machado, A., Silva, M. e Zebende, G. (2014). "Autocorrelation and cross-correlation in time series of homicide and attempted homicide". *Physica A*, 400, 12-19.
- R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

APÊNDICE

Os códigos abaixo foram implementados no R para a análise de correlação cruzada destendenciada (DCCA) e a correlação cruzada tradicional já possui a função *ccf* implementada.

```
dcca <- function(x,y,m){
  rx=cumsum(x) #série integrada da série x
  ry=cumsum(y) #série integrada da série x

  n = length(rx) #tamanho da séries, que devem ser iguais

  acfr = numeric(n-m) #vetor que irá receber a covariância destendenciada
  varerx = numeric(n-m) #vetor que irá receber a variância destendenciada da série rx
  varery = numeric(n-m) #vetor que irá receber a variância destendenciada da série ry

  for(i in 1:(n-m)){
    t = c(1:(m+1)) #janela
    lmrx = lm(rx[i:(i+m)] ~ t) #estimação da reta para a i-ésima janela da série rx
    erx = lmrx$residuals #resíduos da reta estimada para a i-ésima janela da série rx
    lmry = lm(ry[i:(i+m)] ~ t) #estimação da reta para a i-ésima janela da série ry
    ery = lmry$residuals #resíduos da reta estimada para a i-ésima janela da série ry

    acfr[i] = cov(erx,ery) #cálculo da covariância dos resíduos para cada janela
    varerx[i] = var(erx) cov(erx,ery) #cálculo da variância dos resíduos para cada janela para a série x
    varery[i] = var(ery) #cálculo da variância dos resíduos para cada janela para a série y
  }

  dccacor = mean(acfr)/sqrt(mean(varerx)*mean(varery)) #cálculo do coeficiente de correlação cruzada
destendenciada

  return(dccacor) #a função retorna o coeficiente de correlação cruzada destendenciada
}
```