



Trabalho de Conclusão de Curso

**Análise de Correlação Cruzada Destendenciada
em Processos Estacionários com Longa
Dependência**

Lígia Amaral Latorre

19 de dezembro de 2018

Lígia Amaral Latorre

**Análise de Correlação Cruzada Destendenciada em
Processos Estacionários com Longa Dependência**

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientadores:

Prof. Dr. Guilherme Pumi

Prof^a. Dr^a. Taiane Schaedler Prass

Porto Alegre
Dezembro de 2018

Lígia Amaral Latorre

Análise de Correlação Cruzada Destendenciada em Processos Estacionários com Longa Dependência

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pelos Orientadores e pela Banca Examinadora.

Orientador: _____

Prof. Dr. Guilherme Pumi, UFRGS

Doutor pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Orientadora: _____

Prof^ª. Dr^ª. Taiane Schaedler Prass, UFRGS

Doutora pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Banca Examinadora:

Prof^ª. Dr^ª. Taiane Schaedler Prass, UFRGS

Doutora pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Prof. Dr. Marcio Valk, UNICAMP

Doutor pela Universidade Estadual de Campinas - Campinas, SP

Porto Alegre
Dezembro de 2018

Agradecimentos

Aos meus pais, por me proporcionarem uma ótima educação, por todo o amor e por me apoiarem mesmo nos momentos em que não ainda entendiam a minha escolha pelo curso de Estatística. Obrigada.

A minha irmã Luísa, pelo amor, amizade, confiança e apoio sempre. Tu és meu exemplo.

A minha cachorra Pitty (*in memoriam*), por ter sido minha companheira de quase 13 anos e que perdi durante essa jornada. Te amarei eternamente. A cada dia que passa, sinto mais a tua falta.

À Aline, pela amizade dentro e fora da universidade, essencial pra que eu pudesse chegar até aqui. Tu és uma inspiração pra mim e agradeço tua força e teu apoio incondicional. Todas as horas de sono perdidas, fazendo trabalho até tarde, seja por skype ou pessoalmente, com certeza valeram a pena. A gente conseguiu, amiga!

Aos meus orientadores professores Guilherme Pumi e Taiane Prass, por aceitarem meu pedido de orientação e me apresentarem o assunto deste trabalho. Agradeço a paciência e a dedicação que vocês tiveram comigo.

Por fim, a todas as pessoas que, de uma forma ou de outra, contribuíram para tornarem possível a realização do meu sonho. Obrigada a todos.

Resumo

No contexto de séries temporais multivariadas, há interesse em compreender a estrutura de dependência entre duas séries, a qual pode ser estudada através do método da análise de correlação cruzada. Porém, este método exige que os dados atendam ao pressuposto de estacionariedade, o que não acontece na maior parte das séries que encontramos na prática. O método da análise de correlação cruzada destendenciada (Detrended Cross-Correlation Analysis – DCCA) aparece como uma alternativa, visto que tem como objetivo identificar e quantificar a associação cruzada entre duas séries potencialmente não-estacionárias. Neste trabalho, propomos analisar o comportamento amostral da DCCA em quatro cenários envolvendo séries temporais estacionárias com longa dependência, baseando-se em resultados provenientes de simulações. Também foi desenvolvido neste trabalho um aplicativo Shiny para o cálculo do coeficiente DCCA, o qual está disponível online.

Palavras-Chave: Séries temporais, longa dependência, DCCA, correlação cruzada, medidas de associação, Shiny.

Abstract

In the context of multivariate time series, there is interest in understanding the dependence structure between two time series, which can be studied through the cross-correlation analysis method. However, for this method to be useful, it is essential that the time series in study are stationary, which is uncommon in most time series encountered in real life. An alternative is the Detrended Cross-Correlation Analysis (DCCA) method, which aims to identify and quantify the cross-association between two potentially non-stationary time series. In this work, our goal is to study the finite sample behavior of the DCCA in different scenarios involving long range dependent time series, by means of simulations. We also developed a Shiny app that calculates the detrended cross-correlation coefficient, which is available online.

Keywords: Time series, long range dependence, DCCA, cross-correlation, measures of association, Shiny.

Sumário

1	Introdução	8
2	Conceitos e Definições	10
2.1	Análise de Correlação Cruzada (Cross-correlation Analysis) . .	10
2.2	Análise de Correlação Cruzada Destendenciada (Detrended Cross-Correlation Analysis – DCCA)	11
2.3	Processos Com Longa Dependência	13
3	Simulações	15
3.1	Cenário 1	16
3.2	Cenário 2	17
3.3	Cenário 3	19
3.4	Cenário 4	20
4	Resultados	22
4.1	Cenário 1	22
4.2	Cenário 2	26
4.3	Cenário 3	30
4.4	Cenário 4	34
5	Conclusões	38
6	Shiny	40
	Referências Bibliográficas	44

1 Introdução

Uma série temporal é definida como uma sequência de observações ao longo do tempo. Analisando o comportamento de séries temporais é possível identificar padrões, descrever a dependência dos dados em relação ao tempo, bem como fazer previsões de valores futuros. O principal objetivo da análise de séries temporais é desenvolver modelos que forneçam descrições plausíveis para os dados ([Shumway e Stoffer, 2006](#)).

Em se tratando de séries multivariadas, também há interesse em compreender a estrutura de dependência entre duas séries. Por exemplo, podemos estar interessados em como a taxa de emprego afeta ou é afetada pela inflação contemporânea ou inflação passada. Através do método da análise de correlação cruzada (cross-correlation analysis), é possível estudar o comportamento conjunto de duas séries ([Shumway e Stoffer, 2006](#)). No entanto, este método exige que os dados atendam ao pressuposto de estacionariedade, o que não acontece na maior parte das séries que encontramos na prática ([Morettin e Tolo, 2004](#)), tornando-se um problema.

Como uma possível solução para este problema, [Podobnik e Stanley \(2008\)](#) propuseram o método da análise de correlação cruzada destendenciada (Detrended Cross-Correlation Analysis – DCCA). O DCCA é uma generalização do método da análise de flutuação destendenciada (Detrended Fluctuation Analysis – DFA), proposto por [Peng et al. \(1994\)](#), e visa identificar e quantificar a associação cruzada entre duas séries temporais potencialmente não-estacionárias. Este método aparece como uma opção quando a suposição de estacionariedade não for atendida e suas aplicações têm sido estudadas em diferentes áreas, tais como climatologia ([Vassoler e Zebende, 2012](#)), geofísica ([Marinho et al., 2013](#)) e mercado de ações ([Lin et al., 2012](#)).

[Gonçalves \(2018\)](#) identificou que o método da DCCA possui comportamento bastante distinto do método da correlação cruzada tradicional, no contexto de séries estacionárias com curta dependência. No presente trabalho, seguindo a linha do trabalho de [Gonçalves \(2018\)](#), temos como objetivo investigar o comportamento da DCCA no contexto de séries temporais com longa dependência. Para isso, propomos analisar o comportamento amostral do referido método em cenários envolvendo séries estacionárias com longa dependência e com uma estrutura de dependência cruzada entre elas. Neste trabalho, as análises serão baseadas em resultados provenientes de

simulações. Os resultados teóricos estão sendo desenvolvidos paralelamente, mas, devido a sua complexidade, não são objeto de estudo deste trabalho. Também serão comparados os resultados obtidos para os métodos da DCCA e da correlação cruzada tradicional.

Quatro cenários foram determinados para serem analisados neste trabalho. No primeiro, temos duas séries temporais com longa dependência, as quais possuem inovações independentes e identicamente distribuídas (i.i.d.) e independentes entre si, o que implica que as duas séries serão independentes entre si. Neste cenário, espera-se que ambos os métodos resultem em correlação cruzada igual a zero.

No segundo cenário, temos duas séries temporais com longa dependência, as quais possuem inovações i.i.d. e correlacionadas no lag zero, o que implica que as duas séries vão possuir correlação cruzada apenas no lag zero. No terceiro cenário, temos uma série temporal i.i.d. com distribuição $\mathcal{N}(0, 1)$ e uma série temporal com longa dependência que possui inovações i.i.d. dadas pela primeira série, o que implica que as duas séries vão apresentar correlação cruzada não-nula.

No quarto e último cenário, temos duas séries temporais com longa dependência, as quais possuem inovações i.i.d. geradas a partir de uma mesma série temporal i.i.d. com distribuição $\mathcal{N}(0, 1)$, o que implica que as duas séries terão correlação cruzada não-nula. Nestes três últimos cenários, os resultados esperados são mais complexos e serão detalhados na seção específica.

O trabalho está dividido como segue. No Capítulo 2, apresentamos os conceitos utilizados na realização do trabalho, quais sejam, os métodos de análise de correlação cruzada, DFA, DCCA e o coeficiente DCCA, bem como as definições do processo ruído branco e dos processos ARFIMA. No Capítulo 3, será descrito o processo de obtenção dos dados simulados, assim como serão apresentados os cenários e os resultados esperados para cada um deles. O Capítulo 4 traz os resultados obtidos em cada um dos cenários e discussões sobre o que se obteve e o que se esperava obter. As conclusões serão apresentadas no Capítulo 5 e, no Capítulo 6, encerramos o presente trabalho apresentando o aplicativo Shiny desenvolvido para o cálculo do coeficiente DCCA entre duas séries temporais de igual tamanho.

2 Conceitos e Definições

Neste capítulo, serão apresentados os conceitos e definições necessários para realização deste trabalho. Primeiramente, descrevemos o método da análise de correlação cruzada e, na sequência, o método da análise de correlação cruzada destendenciada, que engloba o método da análise de flutuação destendenciada e o coeficiente de correlação cruzada destendenciada. Logo após, introduzimos os processos com longa dependência.

2.1 Análise de Correlação Cruzada (Cross-correlation Analysis)

Sejam $\{X_{1,t}\}_{t \in \mathbb{Z}}$ e $\{X_{2,t}\}_{t \in \mathbb{Z}}$ duas séries temporais conjuntamente estacionárias, com médias $\mu_1 = \mathbb{E}(X_{1,t})$ e $\mu_2 = \mathbb{E}(X_{2,t})$ e variâncias $\sigma_1^2 = \text{Var}(X_{1,t})$ e $\sigma_2^2 = \text{Var}(X_{2,t})$. As funções de covariância cruzada, denotada por $\gamma_{1,2}(h)$, e de correlação cruzada, denotada por $\rho_{1,2}(h)$, são dadas por

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

e

$$\rho_{1,2}(h) = \frac{\gamma_{1,2}(h)}{\sqrt{\gamma_1(0)\gamma_2(0)}} = \frac{\gamma_{1,2}(h)}{\sqrt{\sigma_1^2\sigma_2^2}} = \frac{\gamma_{1,2}(h)}{\sigma_1\sigma_2},$$

onde $h \in \mathbb{Z}$ e $\gamma_k(h)$ é a função de autocovariância da série $\{X_{k,t}\}_{t \in \mathbb{Z}}$, para $k = 1, 2$.

Note que, quando $h = 0$,

$$\rho_{1,2}(0) = \frac{\gamma_{1,2}(0)}{\sigma_1\sigma_2} = \frac{\text{Cov}(X_{1,t}, X_{2,t})}{\sigma_1\sigma_2},$$

que é o coeficiente de correlação de Pearson, geralmente utilizado para medir a relação linear entre duas variáveis aleatórias.

Dadas duas séries temporais $\{x_{1,t}\}_{t=1}^n$ e $\{x_{2,t}\}_{t=1}^n$, os estimadores amostrais da cova-

riância cruzada e da correlação cruzada são, respectivamente,

$$\hat{\gamma}_{1,2}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (x_{1,t} - \bar{x}_1)(x_{2,t+h} - \bar{x}_2), & 0 \leq h \leq n-1, \\ \frac{1}{n} \sum_{t=1-h}^n (x_{1,t} - \bar{x}_1)(x_{2,t+h} - \bar{x}_2), & -n+1 \leq h \leq 0, \end{cases}$$

$$\hat{\rho}_{1,2}(h) = \frac{\hat{\gamma}_{1,2}(h)}{\hat{\sigma}_1 \hat{\sigma}_2}, \quad -n+1 \leq h \leq n-1,$$

onde

$$\bar{x}_k = \frac{1}{n} \sum_{t=1}^n x_{k,t} \quad \text{e} \quad \hat{\sigma}_k^2 = \frac{1}{n} \sum_{t=1}^n (x_{k,t} - \bar{x}_k)^2, \quad k = 1, 2.$$

2.2 Análise de Correlação Cruzada Destendenciada (Detrended Cross-Correlation Analysis – DCCA)

O método da análise de correlação cruzada destendenciada (Detrended Cross-Correlation Analysis – DCCA) foi proposto por [Podobnik e Stanley \(2008\)](#) como uma generalização do método da análise de flutuação destendenciada (Detrended Fluctuation Analysis – DFA) e é baseado na covariância destendenciada. Ambos os métodos têm como objetivo estudar a associação cruzada de longo alcance em séries temporais não-estacionárias. O método da DFA foi definido por [Peng et al. \(1994\)](#) para análise de sequências de DNA, mas vem sendo utilizado em vários campos de estudos. Esse método é realizado em etapas, as quais são descritas na sequência.

Seja $\{X_t\}_{t \in \mathbb{Z}}$ uma série temporal e x_1, \dots, x_n uma amostra de tamanho n proveniente desta série. Primeiro, calcula-se o sinal integrado $R_t = \sum_{i=1}^t x_i$, onde $t = 1, \dots, n$. A série integrada obtida é então dividida 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 $R_t(i)$ através de um polinômio de grau ≥ 1 .

Em seguida, calcula-se a variância dos resíduos da etapa anterior para cada intervalo de tamanho $m + 1$, denotada por $f_{DFA}^2(m, i)$:

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

Por fim, a variância destendenciada, denotada por $F_{DFA}^2(m)$ é calculada como a

média das variâncias dos resíduos de todas as $n - m$ janelas de tamanho $m + 1$:

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

O método da DCCA, apesar de recente, vem sendo aplicado a diversas áreas, tais como climatologia (Vassoler e Zebende, 2012), geofísica (Marinho et al., 2013), mercado de ações (Lin et al., 2012), e também em diferentes campos de estudo, como na relação entre homicídios e tentativas de homicídios (Machado et al., 2014) e no estudo de diferença entre resultados de eletroencefalograma de indivíduos jovens e idosos (Jun e Da-Qing, 2012). O método tem por objetivo mensurar a correlação cruzada entre duas séries temporais não-estacionárias com a mesma quantidade de observações. O cálculo da DCCA envolve várias etapas, as quais são descritas abaixo.

Sejam $\{X_{1,t}\}_{t \in \mathbb{Z}}$ e $\{X_{2,t}\}_{t \in \mathbb{Z}}$ duas séries temporais e $x_{1,1}, \dots, x_{1,n}$ e $x_{2,1}, \dots, x_{2,n}$ amostras de tamanho n provenientes destas séries. Inicialmente, devem ser calculados dois sinais integrados $R_{k,t} = \sum_{j=1}^t x_{k,j}$, $k = 1, 2$ e $t = 1, \dots, n$. Note que os sinais integrados serão somas parciais de $x_{k,j}$. Divide-se cada uma das séries integradas em $n - m$ janelas (ou lags), sendo que cada uma delas possui tamanho $m + 1$.

Para as duas séries integradas, cada janela começa em i e termina em $i + m$. Em cada uma delas, estima-se uma reta através do processo de mínimos quadrados ordinários (MQO) e obtêm-se as retas ajustadas, denotadas por $\tilde{R}_{k,t}(i)$. A seguir, a covariância dos resíduos para cada janela, denotada por $f_{DCCA}(m, i)$, é calculada por:

$$f_{DCCA}(m, i) = \frac{1}{m + 1} \sum_{t=i}^{i+m} (R_{1,t} - \tilde{R}_{1,t}(i))(R_{2,t} - \tilde{R}_{2,t}(i)).$$

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

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

O método da DCCA é considerado uma generalização do método da DFA uma vez que, quando $X_{1,t} = X_{2,t}$, ou seja, temos apenas uma série temporal, a covariância destendenciada $F_{DCCA}(m)$ será reduzida à variância destendenciada $F_{DFA}^2(m)$. É importante ressaltar que a DFA foi definida, em seu artigo original, utilizando-se intervalos não sobrepostos, no entanto, quando trabalhamos com o método da DCCA, calculamos a DFA a partir de intervalos sobrepostos.

Zebende (2011) propôs o coeficiente de correlação cruzada destendenciada (detrended cross-correlation coefficient), denotado por $\rho_{DCCA}(m)$, onde m denota o lag entre as variáveis no qual o coeficiente é calculado. A ideia é combinar os métodos da DCCA e da DFA, assim como fazemos para a correlação de Pearson. Utiliza-se

este coeficiente para calcular a correlação cruzada entre duas séries temporais.

O coeficiente de correlação cruzada destendenciada no lag m é definido como a razão entre a covariância $F_{DCCA}(m)$ com o produto das variâncias $F_{1,DFA}^2(m)$ e $F_{2,DFA}^2(m)$, i.e.,

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

Os valores de $\rho_{DCCA}(m)$ variam entre $-1 \leq \rho_{DCCA}(m) \leq 1$ e sua interpretação, em termos de seus valores, é similar à correlação cruzada tradicional. Se $\rho_{DCCA}(m) = 0$, significa que as duas séries temporais não possuem correlação cruzada. Se $\rho_{DCCA}(m) > 0$, as duas séries possuem correlação cruzada positiva, sendo o caso de correlação cruzada positiva perfeita se $\rho_{DCCA}(m) = 1$. Analogamente, se $\rho_{DCCA}(m) < 0$, as duas séries possuem correlação cruzada negativa, sendo o caso de correlação cruzada negativa perfeita se $\rho_{DCCA}(m) = -1$.

2.3 Processos Com Longa Dependência

A evolução dos estudos dos processos com longa dependência faz dessa área, atualmente, uma parte importante da análise de séries temporais. Os estudos surgiram no início dos anos 50, com o trabalho pioneiro de [Hurst \(1951\)](#), no qual foi analisada a série temporal de níveis do rio Nilo. Esta série caracteriza-se pela presença de correlação entre as observações separadas por um longo período de tempo.

Na década de 80, [Granger e Joyeux \(1980\)](#) e [Hosking \(1981\)](#) introduziram os primeiros modelos para séries temporais com longa dependência, dentre eles a classe de processos ARFIMA. Os modelos ARFIMA, sigla em inglês para modelos autorregressivos fracionalmente integrados de médias móveis, são uma das classes de processos com longa dependência mais difundidas na literatura. Primeiramente, definimos o processo ruído branco e, em seguida, os processos ARFIMA.

Definição 2.1 (Ruído Branco). Seja $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ um processo estocástico. O processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é dito ser um *processo ruído branco* com média 0 e variância σ_ε^2 se, e somente se,

$$\mathbb{E}(\varepsilon_t) = 0 \quad \text{e} \quad \gamma_\varepsilon(h) = \begin{cases} \sigma_\varepsilon^2, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

onde $\gamma_\varepsilon(h) = \text{Cov}(\varepsilon_t, \varepsilon_{t+h})$ é a função de autocovariância do processo $\{\varepsilon_t\}_{t \in \mathbb{Z}}$.

Definição 2.2 (Processos ARFIMA). Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico. Então, $\{X_t\}_{t \in \mathbb{Z}}$ é um processo ARFIMA(p, d, q), com $d \in (-0, 5; 0, 5)$, se for uma solução fracamente estacionária de

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t,$$

onde L denota o operador de defasagem (*backward shift operator*) $L^k(X_t) = X_{t-k}$, para $k \in \{1, 2, \dots\}$, ε_t é um processo ruído branco, $\phi(\cdot)$ e $\theta(\cdot)$ são polinômios dados por

$$\phi(z) = -\sum_{i=0}^p \phi_i z^i, \quad \theta(z) = \sum_{j=0}^q \theta_j z^j, \quad \forall z \in \mathbb{C},$$

onde $\phi_0 = -1$ e $\theta_0 = 1$. O termo fracionário $(1 - L)^d$ é definido pela sua expansão binomial, a qual, em sua forma mais útil é, dada por

$$(1 - L)^{-d} = \sum_{k=0}^{\infty} c_k L^k, \quad \text{onde } c_k = \frac{\Gamma(k + d)}{\Gamma(k + 1)\Gamma(d)} = \prod_{j=1}^k \frac{j - 1 + d}{j}, \quad k \geq 1, \text{ e } c_0 = 1.$$

Observe que, se $\{X_t\}_{t \in \mathbb{Z}}$ é um ARFIMA(0, d , 0), temos que $\{X_t\}_{t \in \mathbb{Z}}$ possui representação causal dada por

$$X_t = (1 - L)^{-d} \varepsilon_t = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k}.$$

3 Simulações

As simulações do presente trabalho foram realizadas no software livre R, versão 3.5.1 (R Core Team, 2018). O código implementado por Gonçalves (2018) foi utilizado para o cálculo do $\rho_{DCCA}(m)$ e foi implementada uma função para criação das séries temporais da classe ARFIMA, as quais foram geradas a partir de sua representação causal, com ponto de truncamento igual a 50.000.

Todas as séries temporais foram criadas com tamanho $n = 1.000$ e foram feitas 1.000 replicações de cada um dos cenários. As janelas consideradas para os cálculos da correlação cruzada foram $m = 3, \dots, 100$. Caso haja interesse em replicar exatamente os resultados aqui gerados, utilizou-se sementes para todas as replicações, com notação $100+r$, sendo r o número da replicação. Em programação, as sementes são úteis para memorizar e reproduzir cada geração aleatória de valores.

Em relação aos resultados teóricos esperados, pode ser mostrado que, se $\{X_{1,t}\}_{t \in \mathbb{Z}}$ e $\{X_{2,t}\}_{t \in \mathbb{Z}}$ são duas séries temporais conjuntamente estritamente estacionárias e ergódicas, com quarto momento finito, então

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

e

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

onde $K_{m+1} = J'_{m+1}Q_{m+1}J_{m+1}$,

$$\Gamma_k(m) = \text{Var}(\mathbf{X}_{k,1}(m)), \quad \Gamma_{1,2}(m) = \text{Cov}(\mathbf{X}_{1,1}(m), \mathbf{X}_{2,1}(m)),$$

com $\mathbf{X}_{k,1}(m) = (X_{k,1}, \dots, X_{k,m+1})'$, J_{m+1} é a $(m+1) \times (m+1)$ matriz cujo (r, s) -ésimo elemento é dado por

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

$Q_{m+1} := I_{m+1} - X(X'X)^{-1}X'$, 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, concluímos que

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

quando $n \rightarrow \infty$.

A seguir, apresentamos cada um dos cenários estudados neste trabalho e os comportamentos esperados do ρ_{DCCA} para cada um deles.

3.1 Cenário 1

Sejam $\{X_{1,t}\}_{t=1}^n$ e $\{X_{2,t}\}_{t=1}^n$ duas séries temporais independentes entre si que seguem o modelo ARFIMA(0, d_k , 0), $k = 1, 2$, com representação causal dada por

$$X_{k,t} = \sum_{j=0}^{\infty} c_{k,j} \varepsilon_{k,t-j}, \quad (3.1)$$

onde as inovações $\{\varepsilon_{k,t}\}_{t \in \mathbb{Z}}$ são independentes entre si e $\varepsilon_{k,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Sob esse cenário, a função de autocovariância de $\{X_{k,t}\}_{t \in \mathbb{Z}}$ é dada por

$$\begin{aligned} \gamma_k(h) &= \text{Cov}(X_{k,t}, X_{k,t+h}) \\ &= \text{Cov}\left(\sum_{i=1}^{\infty} c_{k,i} \varepsilon_{k,t-i}, \sum_{j=0}^{\infty} c_{k,j} \varepsilon_{k,t-j}\right) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} c_{k,i} c_{k,j} \text{Cov}(\varepsilon_{k,t-i}, \varepsilon_{k,t+h-j}) \\ &= \sum_{i=0}^{\infty} c_{k,i} c_{k,i+|h|}, \end{aligned} \quad (3.2)$$

onde a última igualdade é válida pois

$$\text{Cov}(\varepsilon_{k,t-i}, \varepsilon_{k,t+h-j}) \neq 0 \Leftrightarrow t-i = t+h-j \Leftrightarrow j = i+h,$$

e a função de covariância cruzada é $\gamma_{1,2}(h) = 0$, para todo $h \in \mathbb{Z}$. Segue que

$$\Gamma_{1,2}(m) = 0_{(m+1) \times (m+1)}, \quad \forall m > 0,$$

onde $0_{(m+1) \times (m+1)}$ é uma matriz de zeros de tamanho $(m+1) \times (m+1)$, e o (i, j) -ésimo elemento de Γ_k é dado por

$$[\Gamma_k]_{i,j} = \text{Cov}(X_{k,i}, X_{k,j}) = \text{Cov}(X_{k,t}, X_{k,t+(i-j)}) = \gamma_k(i-j). \quad (3.3)$$

Além disso, $[\Gamma_k]_{j,i} = \gamma_k(j-i) = \gamma_k(i-j) = [\Gamma_k]_{i,j}$. Em particular, $[\Gamma_k]_{i,i} = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} c_{k,j}^2$, e, portanto,

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

As simulações foram realizadas para $(d_1, d_2) \in \{(0, 1; 0, 2), (0, 1; 0, 4), (0, 3; 0, 4)\}$.

3.2 Cenário 2

Sejam $\{X_{1,t}\}_{t=1}^n$ e $\{X_{2,t}\}_{t=1}^n$ duas séries temporais correlacionadas que seguem o modelo ARFIMA(0, d_k , 0), $k = 1, 2$, com representação causal dada pela expressão (3.1) e inovações $\{\varepsilon_{k,t}\}_{t \in \mathbb{Z}}$ dadas por $(\varepsilon_{1,t}, \varepsilon_{2,t}) \stackrel{i.i.d}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$, onde

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Neste cenário, o (i, j) -ésimo elemento de Γ_k é dado pela equação (3.3), com $\gamma_k(h) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} c_{k,i} c_{k,i+|h|}$, e o (i, j) -ésimo elemento de $\Gamma_{1,2}$ é dado por

$$\begin{aligned} [\Gamma_{1,2}]_{i,j} &= \text{Cov}(X_{1,i}, X_{2,j}) \\ &= \text{Cov}(X_{1,t}, X_{2,t+(j-i)}) \\ &= \gamma_{1,2}(j-i), \end{aligned} \quad (3.4)$$

onde

$$\begin{aligned} \gamma_{1,2}(h) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{1,i} c_{2,j} \text{Cov}(\varepsilon_{1,t-i}, \varepsilon_{2,t+h-j}) \\ &= \begin{cases} \sum_{i=0}^{\infty} c_{1,i} c_{2,i+h} \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}); & h \geq 0, \\ \sum_{i=0}^{\infty} c_{1,i+|h|} c_{2,i} \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}); & h < 0. \end{cases} \\ &= \begin{cases} \rho \sum_{i=0}^{\infty} c_{1,i} c_{2,h+i}; & h \geq 0, \\ \rho \sum_{j=0}^{\infty} c_{1,j-h} c_{2,j}; & h < 0. \end{cases} \end{aligned} \quad (3.5)$$

Segue que

$$\Gamma_1(m) = \sum_{j=0}^m \gamma_1(j)M_j, \quad \Gamma_2(m) = \sum_{j=0}^m \gamma_2(j)M_j \quad e$$

$$\Gamma_{1,2}(m) = \sum_{j=0}^m \gamma_{1,2}(j)M_j^* + \sum_{j=1}^m \gamma_{1,2}(-j)(M_j^*)',$$

para todo $m > 0$, onde 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

$$[M_j]_{r,s} = \begin{cases} 1, & \text{se } |r-s| = j, \\ 0, & \text{caso contrário,} \end{cases} \quad e \quad [M_j^*]_{r,s} = \begin{cases} 1, & \text{se } s-r = j, \\ 0, & \text{caso contrário,} \end{cases} \quad (3.6)$$

e $(M_j^*)'$ denota a matriz transposta de M_j^* . Então

$$\text{tr}(K_{m+1}\Gamma_1(m)) = \sum_{j=0}^m \gamma_1(j)\text{tr}(K_{m+1}M_j), \quad \text{tr}(K_{m+1}\Gamma_2(m)) = \sum_{j=0}^m \gamma_2(j)\text{tr}(K_{m+1}M_j)$$

$$e \quad \text{tr}(K_{m+1}\Gamma_{1,2}(m)) = \sum_{j=0}^m \gamma_{1,2}(j)\text{tr}(K_{m+1}M_j^*) + \sum_{j=1}^m \gamma_{1,2}(-j)\text{tr}(K_{m+1}(M_j^*)'),$$

para todo $m > 0$. É fácil mostrar que K_{m+1} é uma matriz simétrica tal que $[K_{m+1}]_{r,s} = 0$, sempre que $r = 1$ ou $s = 1$. Então, de (3.6), $M_0 = I_{m+1}$, de modo que $\text{tr}(K_{m+1}M_0) = \text{tr}(K_{m+1})$ e

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

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

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

e

$$\alpha_0^{(m)} := \text{tr}(K_{m+1}), \quad 1 \leq j \leq m.$$

Em particular, $\text{tr}(K_{m+1}M_m) = \text{tr}(K_{m+1}M_m^*) = 0$. Segue que

$$\Gamma_{1,2}(m) = \gamma_{1,2}(0)I_{m+1} + \sum_{j=1}^m \gamma_{1,2}(j)M_j^* + \sum_{j=1}^m \gamma_{1,2}(-j)(M_j^*)', \quad (3.7)$$

$$\text{tr}(K_{m+1}\Gamma_1(m)) = \alpha_0^{(m)}\gamma_1(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)}\gamma_1(j) \quad (3.8)$$

e

$$\text{tr}(K_{m+1}\Gamma_2(m)) = \alpha_0^{(m)}\gamma_2(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)}\gamma_2(j). \quad (3.9)$$

Segue que

$$\begin{aligned} \text{tr}(K_{m+1}\Gamma_{1,2}(m)) &= \gamma_{1,2}(0)\text{tr}(K_{m+1}) + \sum_{j=1}^m [\gamma_{1,2}(j) + \gamma_{1,2}(-j)]\text{tr}(K_{m+1}M_j^*) \\ &= \gamma_{1,2}(0)\text{tr}(K_{m+1}) + \sum_{j=1}^{m-1} [\gamma_{1,2}(j) + \gamma_{1,2}(-j)]\alpha_j^{(m)} \\ &= \alpha_0^{(m)}\gamma_{1,2}(0) + \sum_{j=1}^{m-1} [\gamma_{1,2}(j) + \gamma_{1,2}(-j)]\alpha_j^{(m)} \end{aligned} \quad (3.10)$$

e

$$\begin{aligned} \rho_{DCCA}(m) &\xrightarrow{P} \frac{\alpha_0^{(m)}\gamma_{1,2}(0) + \sum_{j=1}^{m-1} [\gamma_{1,2}(j) + \gamma_{1,2}(-j)]\alpha_j^{(m)}}{\sqrt{\alpha_0^{(m)}\gamma_1(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)}\gamma_1(j)} \sqrt{\alpha_0^{(m)}\gamma_2(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)}\gamma_2(j)}} \\ &= \frac{\alpha_0^{(m)}\gamma_{1,2}(0) + \sum_{j=1}^{m-1} [\gamma_{1,2}(j) + \gamma_{1,2}(-j)]\alpha_j^{(m)}}{\sqrt{\gamma_1(0) + 2 \sum_{j=1}^{m-1} \beta_j^{(m)}\gamma_1(j)} \sqrt{\gamma_2(0) + 2 \sum_{j=1}^{m-1} \beta_j^{(m)}\gamma_2(j)}}, \end{aligned}$$

quando $n \rightarrow \infty$, onde $\beta_j^{(m)} = \frac{\alpha_j^{(m)}}{\alpha_0^{(m)}}$, para todo $m > 0$. As simulações foram realizadas para $(d_1, d_2) \in \{(0, 1; 0, 2), (0, 1; 0, 4), (0, 3; 0, 4)\}$ e foi utilizado $\rho = 0, 5$.

3.3 Cenário 3

Seja $\varepsilon_t \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$ e definimos a série $\{X_{1,t}\}_{t=1}^n = \{\varepsilon_t\}_{t=1}^n$. Seja $\{X_{2,t}\}_{t=1}^n$ uma série temporal que segue o modelo ARFIMA(0, d , 0) e possui representação causal dada pela expressão (3.1), com $c_{k,j} = c_j$ e inovações dadas por $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Neste cenário, o (i, j) -ésimo elemento de Γ_k , para $k = 2$, é dado pela equação (3.3), com $\gamma_k(h) = \sum_{i=0}^{\infty} c_i c_{i+|h|}$, e o (i, j) -ésimo elemento de $\Gamma_{1,2}$ é dado por

$$\begin{aligned} [\Gamma_{1,2}]_{i,j} &= \text{Cov}(\varepsilon_t, X_{2,t+j-i}) \\ &= \gamma_{1,2}(j-i), \end{aligned}$$

onde

$$\gamma_{1,2}(h) = \sum_{j=0}^{\infty} c_j \text{Cov}(\varepsilon_t, \varepsilon_{t+h-j}) = \begin{cases} 0, & h < 0, \\ c_h \sigma_\varepsilon^2, & h \geq 0. \end{cases}$$

Note que não precisamos ter $\gamma_{1,2}(h) = \gamma_{1,2}(-h)$. Segue que

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

onde M_j e M_j^* são dadas em (3.6). Então

$$\text{tr}(K_{m+1} \Gamma_1(m)) = \text{tr}(K_{m+1}), \quad \text{tr}(K_{m+1} \Gamma_2(m)) = \sum_{j=0}^m \gamma_2(j) \text{tr}(K_{m+1} M_j)$$

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

Segue que

$$\text{tr}(K_{m+1} \Gamma_1(m)) = \alpha_0^{(m)}, \quad \text{tr}(K_{m+1} \Gamma_2(m)) = \alpha_0^{(m)} \gamma_2(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)} \gamma_2(j),$$

$$\begin{aligned} \text{tr}(K_{m+1} \Gamma_{1,2}(m)) &= \sum_{j=0}^{m-1} c_j \text{tr}(K_{m+1} M_j^*) \\ &= \sum_{j=0}^{m-1} c_j \alpha_j^{(m)}, \end{aligned}$$

e

$$\rho_{DCCA}(m) \xrightarrow{P} \frac{\sum_{j=0}^{m-1} \alpha_j^{(m)} c_j}{\sqrt{\alpha_0^{(m)} \sqrt{\alpha_0^{(m)} \gamma_2(0) + 2 \sum_{j=1}^{m-1} \alpha_j^{(m)} \gamma_2(j)}}} = \frac{1 + \sum_{j=1}^{m-1} \beta_j^{(m)} c_j}{\sqrt{\gamma_2(0) + 2 \sum_{j=1}^{m-1} \beta_j^{(m)} \gamma_2(j)}},$$

quando $n \rightarrow \infty$, onde $\beta_j^{(m)} = \frac{\alpha_j^{(m)}}{\alpha_0^{(m)}}$, para todo $m > 0$. Neste cenário, as simulações foram realizadas para $d \in \{0, 1; 0, 25; 0, 4\}$.

3.4 Cenário 4

Sejam $\{X_{1,t}\}_{t=1}^n$ e $\{X_{2,t}\}_{t=1}^n$ duas séries temporais correlacionadas que seguem o modelo ARFIMA(0, d_k , 0), $k = 1, 2$, com representação causal dada pela expressão (3.1) e inovações dadas por $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, onde $\varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Neste cenário, (i, j) -ésimo

elemento de Γ_k e o (i, j) -ésimo elemento de $\Gamma_{1,2}$ são dados, respectivamente, pelas equações (3.3) e (3.4), com $\gamma_k(h)$ dado por (3.2) e

$$\begin{aligned} \gamma_{1,2}(h) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{1,i} c_{2,j} \text{Cov}(\varepsilon_{t-i}, \varepsilon_{t+h-j}) \\ &= \begin{cases} \sum_{i=0}^{\infty} c_{1,i} c_{2,h+i}; & h \geq 0, \\ \sum_{j=0}^{\infty} c_{1,j-h} c_{2,j}; & h < 0, \end{cases} \end{aligned}$$

onde a segunda igualdade é válida pois

$$\text{Cov}(\varepsilon_{t-i}, \varepsilon_{t+h-j}) \neq 0 \Leftrightarrow t-i = t+h-j \Leftrightarrow j = h+i.$$

Segue que

$$\begin{aligned} \Gamma_1(m) &= \sum_{j=0}^m \gamma_1(j) M_j, \quad \Gamma_2(m) = \sum_{j=0}^m \gamma_2(j) M_j \quad \text{e} \\ \Gamma_{1,2}(m) &= \sum_{j=0}^m \gamma_{1,2}(j) M_j^* + \sum_{j=1}^m \gamma_{1,2}(-j) (M_j^*)', \end{aligned}$$

para todo $m > 0$, onde M_j e M_j^* são dadas em (3.6). Então

$$\text{tr}(K_{m+1} \Gamma_1(m)) = \sum_{j=0}^m \gamma_1(j) \text{tr}(K_{m+1} M_j), \quad \text{tr}(K_{m+1} \Gamma_2(m)) = \sum_{j=0}^m \gamma_2(j) \text{tr}(K_{m+1} M_j)$$

$$\text{e } \text{tr}(K_{m+1} \Gamma_{1,2}(m)) = \sum_{j=0}^m \gamma_{1,2}(j) \text{tr}(K_{m+1} M_j^*) + \sum_{j=1}^m \gamma_{1,2}(-j) \text{tr}(K_{m+1} (M_j^*)'),$$

para todo $m > 0$. Temos que $\Gamma_{1,2}(m)$ pode ser escrito como em (3.7) e as equações (3.8), (3.9) e (3.10) são válidas aqui neste cenário também. Então

$$\rho_{DCCA}(m) \xrightarrow{P} = \frac{\alpha_0^{(m)} \gamma_{1,2}(0) + \sum_{j=1}^{m-1} [\gamma_{1,2}(j) + \gamma_{1,2}(-j)] \alpha_j^{(m)}}{\sqrt{\gamma_1(0) + 2 \sum_{j=1}^{m-1} \beta_j^{(m)} \gamma_1(j)} \sqrt{\gamma_2(0) + 2 \sum_{j=1}^{m-1} \beta_j^{(m)} \gamma_2(j)}},$$

quando $n \rightarrow \infty$, onde $\beta_j^{(m)} = \frac{\alpha_j^{(m)}}{\alpha_0^{(m)}}$, para todo $m > 0$. As simulações foram realizadas para $(d_1, d_2) \in \{(0, 1; 0, 2), (0, 1; 0, 4), (0, 3; 0, 4)\}$.

4 Resultados

Neste capítulo apresentamos os resultados obtidos nas simulações realizadas no presente trabalho. Em todos os cenários, observa-se que o $\rho_{DCCA}(m)$ apresenta maior variabilidade, à medida que o tamanho da janela m aumenta. Isto deve-se ao fato de que, no método da DCCA, com o aumento do m , temos janelas de tamanho maior, mas em menor quantidade, produzindo menos observações para o cálculo das covariâncias. Em relação às correlações cruzadas, nota-se que há uma estabilidade na variabilidade, independente do lag, em todos os cenários.

4.1 Cenário 1

Neste cenário, para todos os valores de d_1 e d_2 , os valores obtidos para o $\rho_{DCCA}(m)$ e para a correlação cruzada estão em torno de zero em todos os lags, como era esperado, tendo em vista que as duas séries eram independentes entre si. As Figuras 4.1 a 4.6 apresentam os resultados para o $\rho_{DCCA}(m)$ e para a correlação cruzada no cenário 1.

Nota-se que o comportamento do $\rho_{DCCA}(m)$ no cenário 1 (Figuras 4.1, 4.3 e 4.5) é similar em todos os casos analisados. Observa-se que há presença de valores extremos inferiores e superiores ao longo dos lags, com exceção de alguns lags, assim como não parece haver diferença considerável nas medianas, entre os lags.

Nas Figuras 4.2, 4.4 e 4.6, que apresentam os resultados da correlação cruzada, observamos que, para os diferentes valores de d_1 e d_2 , os resultados são semelhantes neste cenário. Percebe-se que, nos casos analisados, há valores extremos, tanto inferiores quanto superiores em todos os lags, e as medianas aparentam oscilar levemente entre os lags.

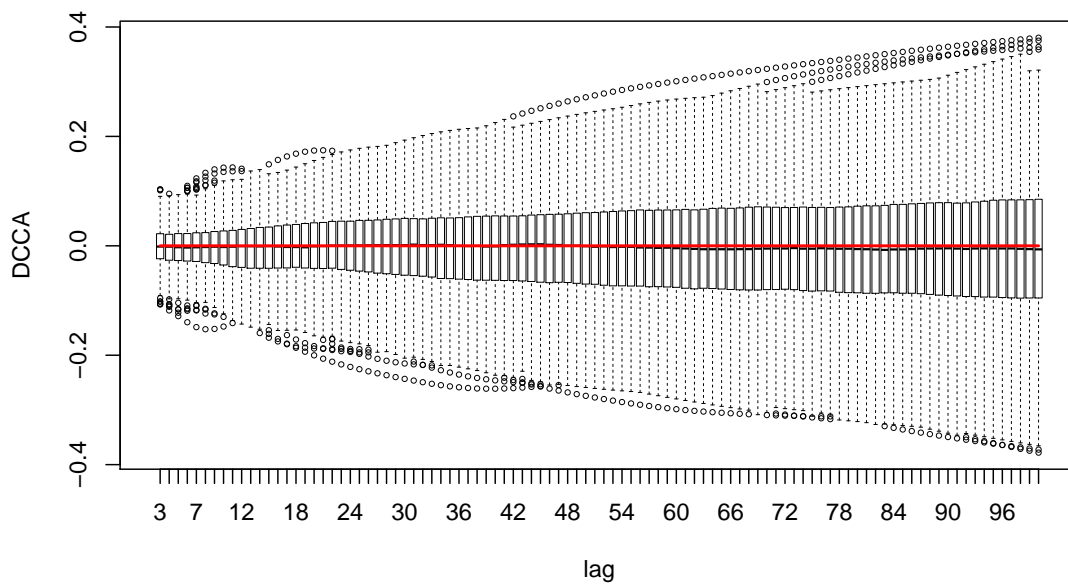


Figura 4.1: Resultados das simulações do cenário 1 do $\rho_{DCCA}(m)$, com $d_1 = 0,1$ e $d_2 = 0,2$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

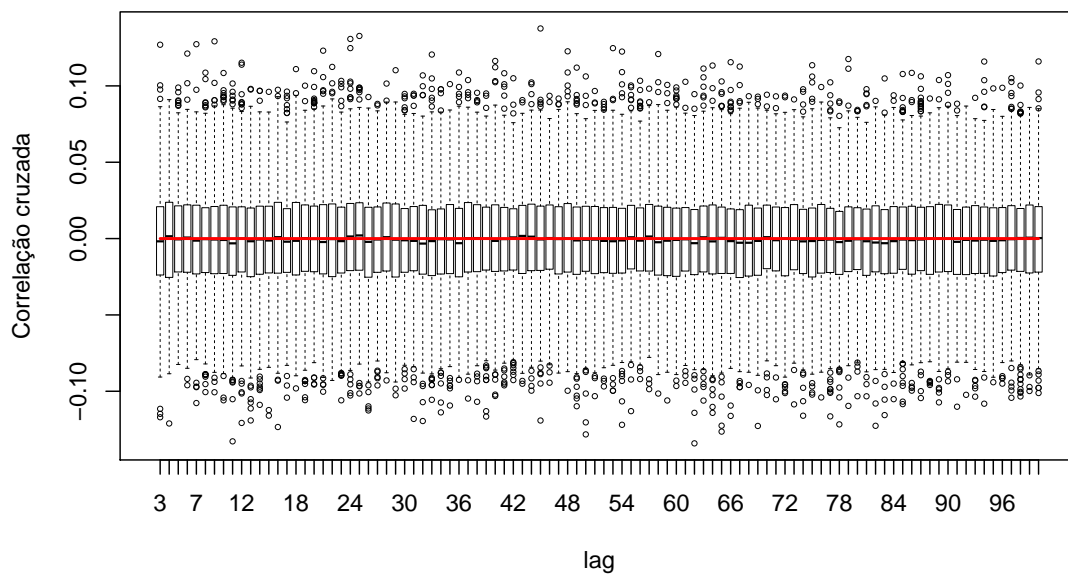


Figura 4.2: Resultados das simulações do cenário 1 da correlação cruzada, com $d_1 = 0,1$ e $d_2 = 0,2$, para os lags $m = \{3, \dots, 100\}$

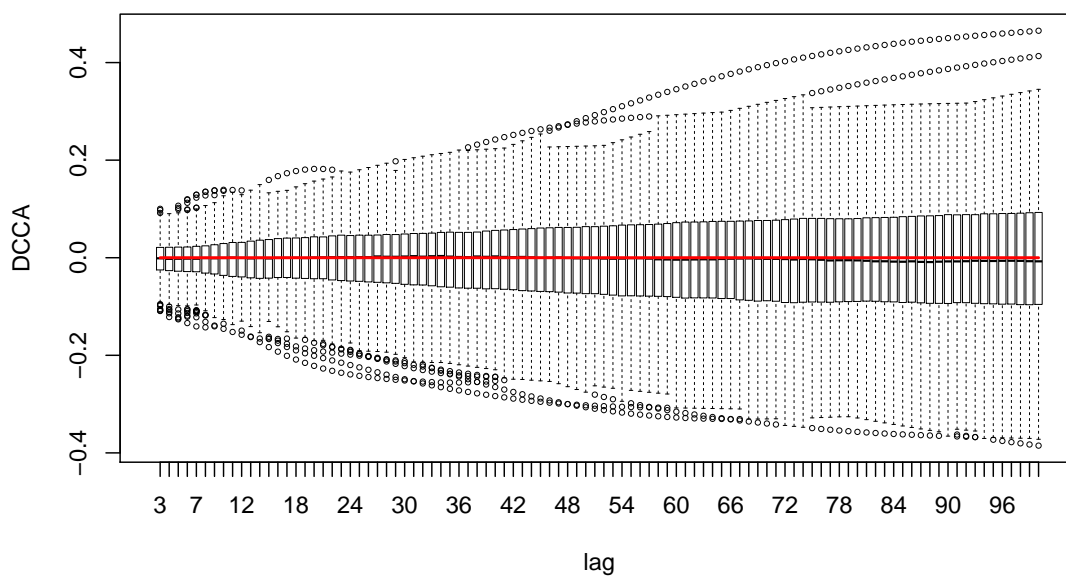


Figura 4.3: Resultados das simulações do cenário 1 do $\rho_{DCCA}(m)$, com $d_1 = 0,1$ e $d_2 = 0,4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

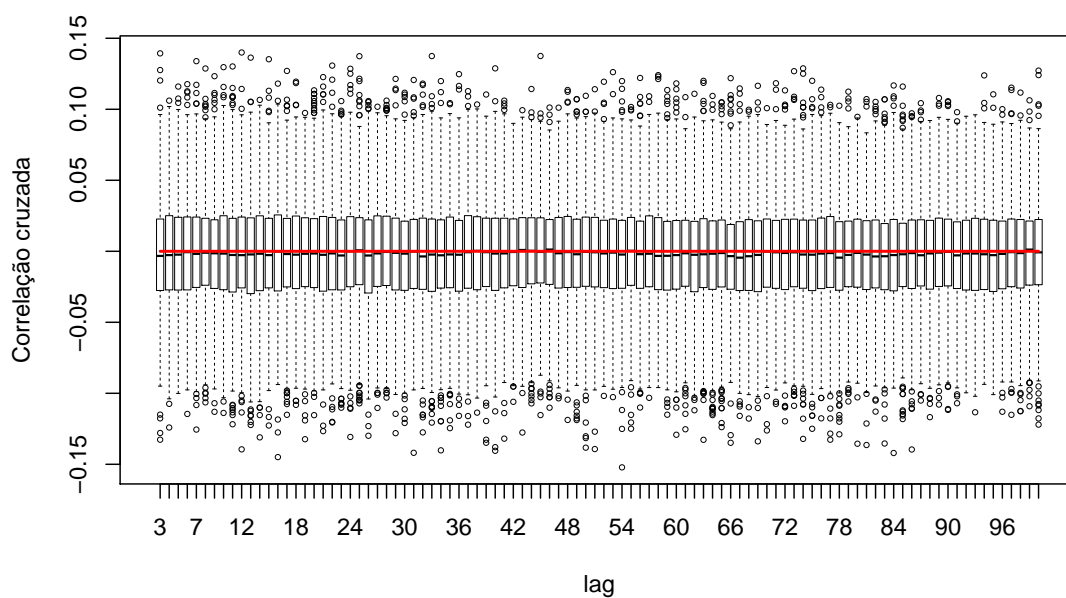


Figura 4.4: Resultados das simulações do cenário 1 da correlação cruzada, com $d_1 = 0,1$ e $d_2 = 0,4$, para os lags $m = \{3, \dots, 100\}$

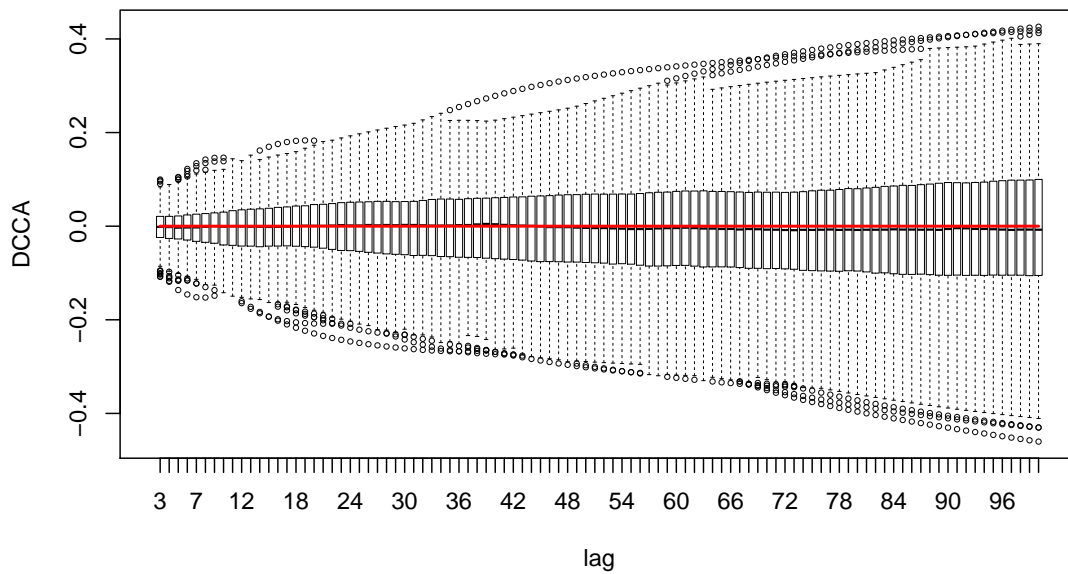


Figura 4.5: Resultados das simulações do cenário 1 do $\rho_{DCCA}(m)$, com $d_1 = 0,3$ e $d_2 = 0,4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

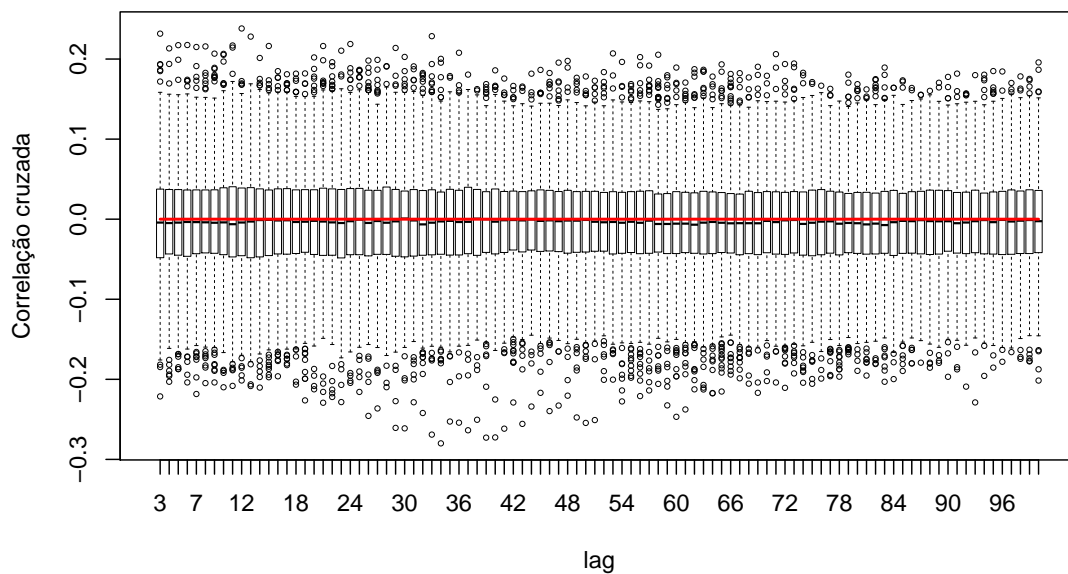


Figura 4.6: Resultados das simulações do cenário 1 da correlação cruzada, com $d_1 = 0,3$ e $d_2 = 0,4$, para os lags $m = \{3, \dots, 100\}$

4.2 Cenário 2

Neste cenário, os valores obtidos para o $\rho_{DCCA}(m)$ foram muito próximos dos valores esperados, para todos os valores de d_1 e d_2 . Em relação à correlação cruzada, os valores obtidos foram mais próximos dos valores esperados quando $d_1 = 0,1$ e $d_2 = 0,2$, e mais longe quando $d_1 = 0,3$ e $d_2 = 0,4$. As Figuras 4.7 a 4.12 apresentam os resultados para o $\rho_{DCCA}(m)$ e para a correlação cruzada no cenário 2.

Quando $d_1 = 0,1$ e $d_2 = 0,2$ neste cenário, observa-se que os valores de $\rho_{DCCA}(m)$ estão em torno de 0,5 em todos os lags e as medianas não parecem apresentar diferença visível graficamente entre os lags (Figura 4.7). Além disso, a partir do lag 47, não há valores extremos superiores. Em relação ao comportamento da correlação cruzada, para estes mesmos valores de d_1 e d_2 no cenário 2, percebe-se, a partir da Figura 4.8, que os valores obtidos estão relativamente próximos dos esperados e parecem estar convergindo para uma constante com valor próximo de zero. Ainda, nota-se que estão presentes valores extremos inferiores e superiores em todos os lags.

Quando $d_1 = 0,1$ e $d_2 = 0,4$, nota-se que os valores obtidos para o $\rho_{DCCA}(m)$ no cenário 2 convergem para uma constante com valor entre 0,4 e 0,5 (Figura 4.9). Ainda, alguns lags não possuem valores extremos inferiores e outros não possuem valores extremos superiores. Para estes mesmos valores de d_1 e d_2 , na Figura 4.10, observamos que há uma pequena diferença entre os valores esperados e os valores obtidos para a correlação cruzada, os quais decrescem para uma constante próxima de zero. Nota-se também a presença de valores extremos inferiores e superiores.

A Figura 4.11 apresenta os resultados para o $\rho_{DCCA}(m)$ no cenário 2 com $d_1 = 0,3$ e $d_2 = 0,4$ e observa-se que os valores aparentam convergir para uma constante com valor em torno de 0,5 e, a partir do lag 41, não há valores extremos superiores. Em relação aos resultados deste cenário para a correlação cruzada com estes mesmos valores de d_1 e d_2 , observamos, pela Figura 4.12, que os valores obtidos estão relativamente longe dos valores esperados e estão decrescendo para uma constante com valor entre 0 e 0,1. Além disso, há presença de valores extremos inferiores e superiores no lags, percebemos que a quantidade de valores extremos superiores é consideravelmente maior do que a de inferiores.

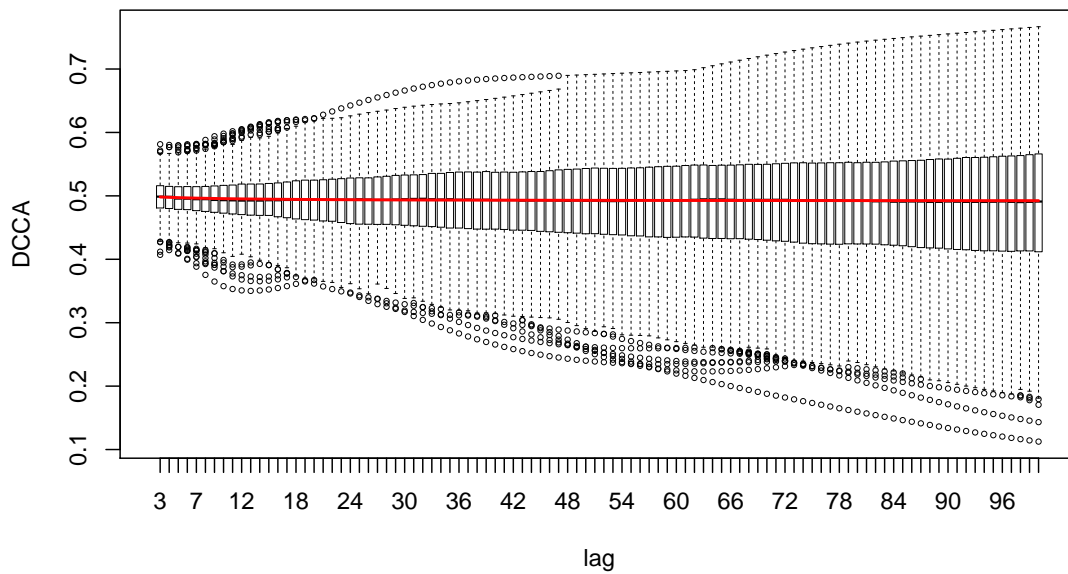


Figura 4.7: Resultados das simulações do cenário 2 do $\rho_{DCCA}(m)$, com $d_1 = 0,1$ e $d_2 = 0,2$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

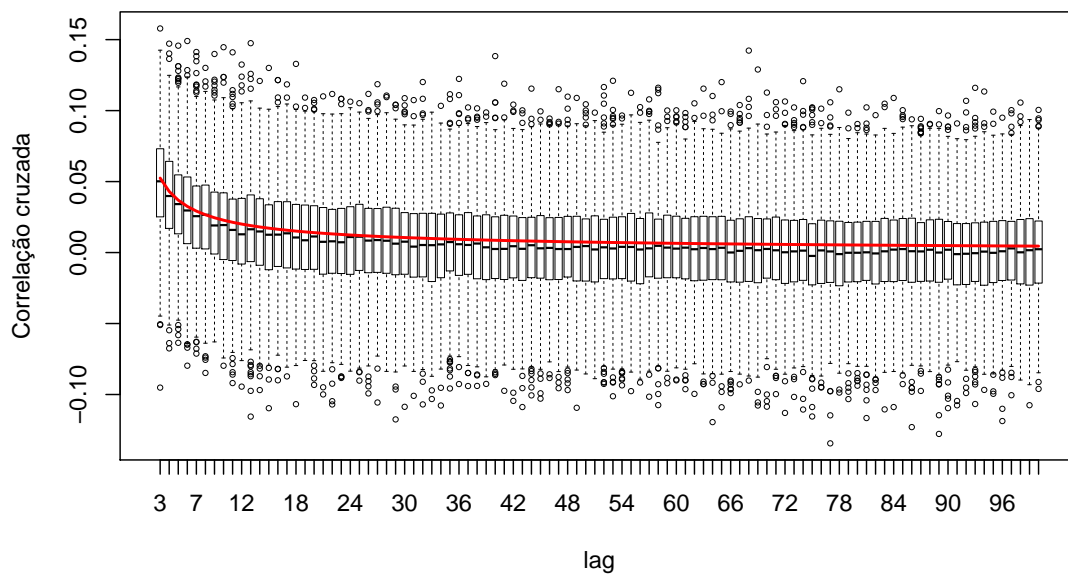


Figura 4.8: Resultados das simulações do cenário 2 da correlação cruzada, com $d_1 = 0,1$ e $d_2 = 0,2$, para os lags $m = \{3, \dots, 100\}$

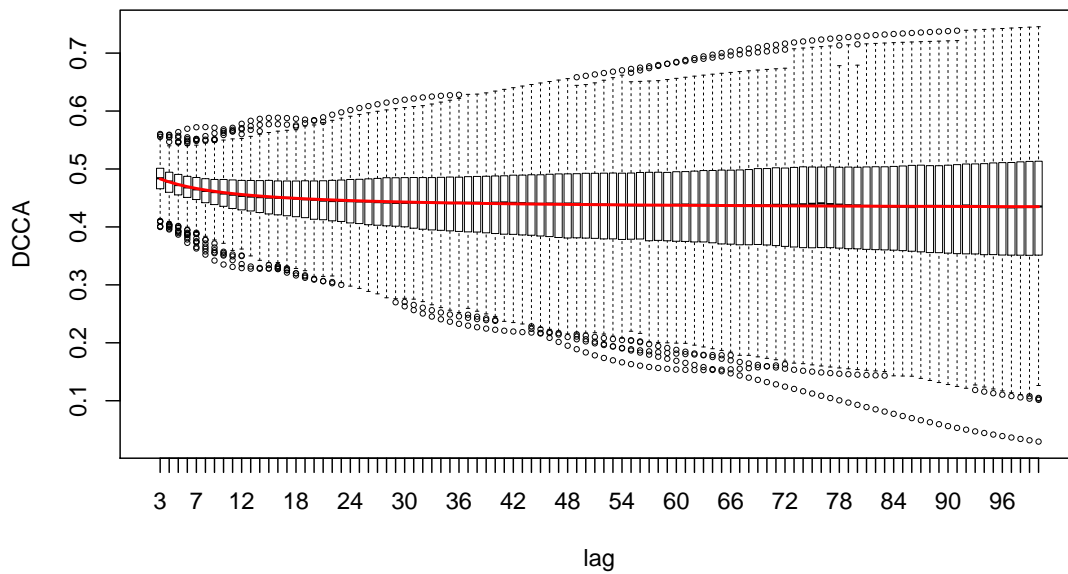


Figura 4.9: Resultados das simulações do cenário 2 do $\rho_{DCCA}(m)$, com $d_1 = 0,1$ e $d_2 = 0,4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

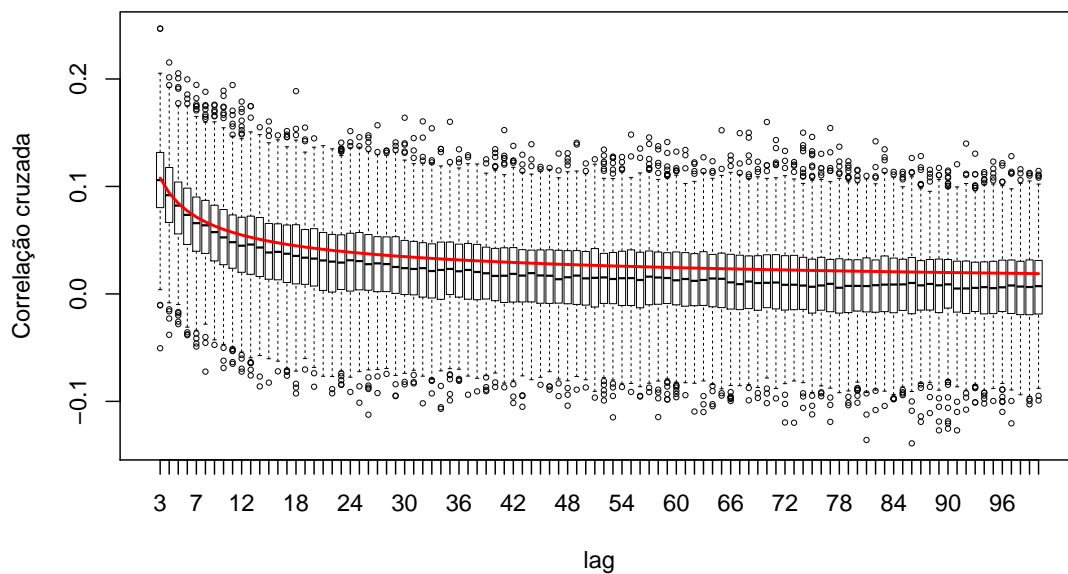


Figura 4.10: Resultados das simulações do cenário 2 da correlação cruzada, com $d_1 = 0,1$ e $d_2 = 0,4$, para os lags $m = \{3, \dots, 100\}$

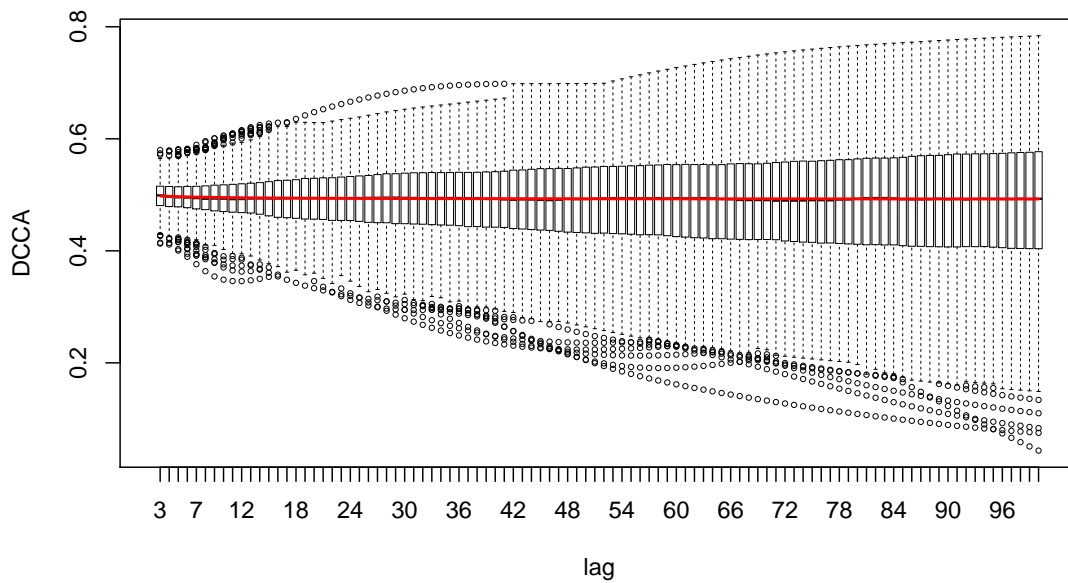


Figura 4.11: Resultados das simulações do cenário 2 do $\rho_{DCCA}(m)$, com $d_1 = 0,3$ e $d_2 = 0,4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

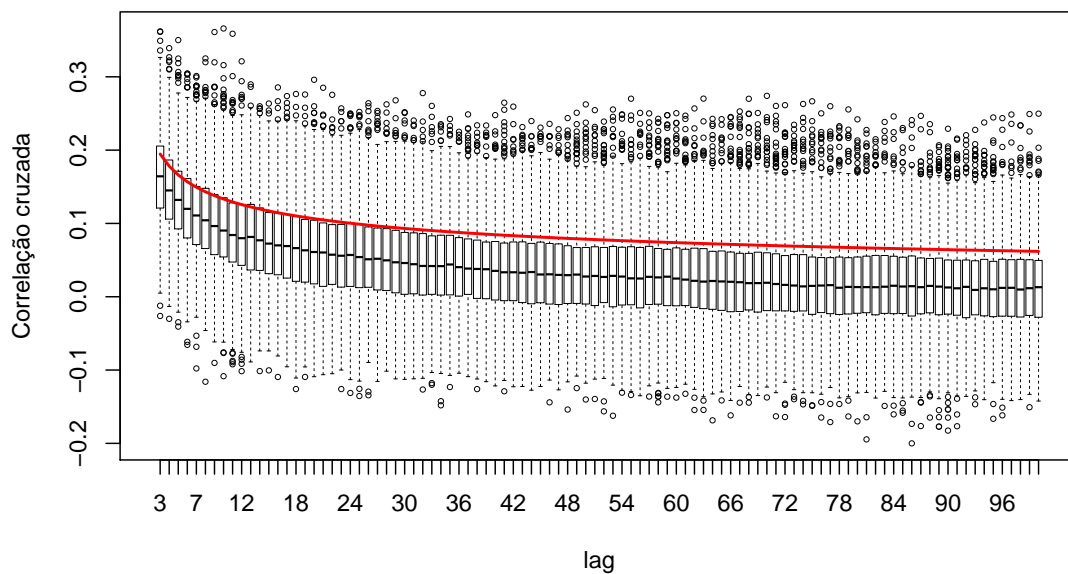


Figura 4.12: Resultados das simulações do cenário 2 da correlação cruzada, com $d_1 = 0,3$ e $d_2 = 0,4$, para os lags $m = \{3, \dots, 100\}$

4.3 Cenário 3

Neste cenário, em relação aos resultados do $\rho_{DCCA}(m)$, observa-se que os valores obtidos foram muito próximos do esperado, para os diferentes valores de d analisados. É importante observar também que, à medida que o lag vai aumentando, há uma pequena diferença entre os valores obtidos e os valores esperados. Este resultado está relacionado com o tamanho amostral da série, visto que os resultados assintóticos disponíveis assumem n muito maior que m . Em relação aos valores obtidos para a correlação cruzada, estes foram bem próximos do esperado quando $d = 0, 1$, mas tiveram uma pequena diferença quando $d = 0, 25$ e uma considerável diferença quando $d = 0, 4$. Percebe-se que, à medida que o d aumenta, o viés da correlação cruzada amostral também aumenta.

As Figuras 4.13, 4.15 e 4.17 apresentam os resultados para o $\rho_{DCCA}(m)$ no cenário 3, com $d = 0, 1$, $d = 0, 25$ e $d = 0, 4$, respectivamente. Observa-se que, em todos os casos, os valores de $\rho_{DCCA}(m)$ aparentam estar convergindo para uma constante. Além disso, nota-se também a presença de valores extremos inferiores e superiores, para todos os valores de d analisados, em todos os tamanhos de janelas, com exceção de alguns que não possuem valores extremos superiores.

As Figuras 4.14, 4.16 e 4.18 dizem respeito aos resultados da correlação cruzada no cenário 3, com $d = 0, 1$, $d = 0, 25$ e $d = 0, 4$, respectivamente. É importante observar que, para todos os casos analisados, os valores obtidos para a correlação cruzada parecem convergir para valores próximos de zero. Percebe-se ainda que, com todos os valores de d utilizados, estão presentes valores extremos inferiores e superiores.

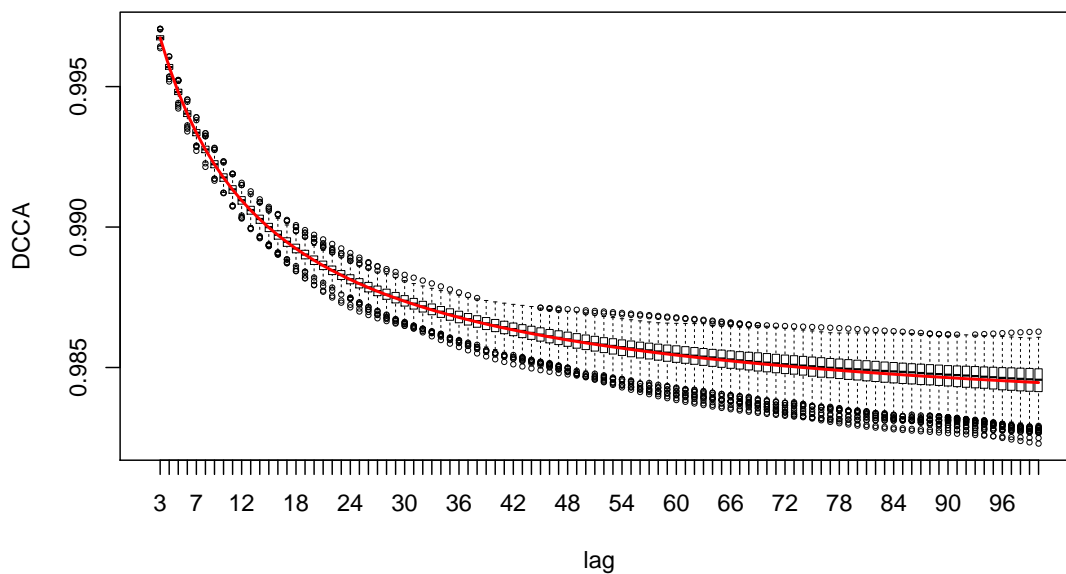


Figura 4.13: Resultados das simulações do cenário 3 do $\rho_{DCCA}(m)$, com $d = 0, 1$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

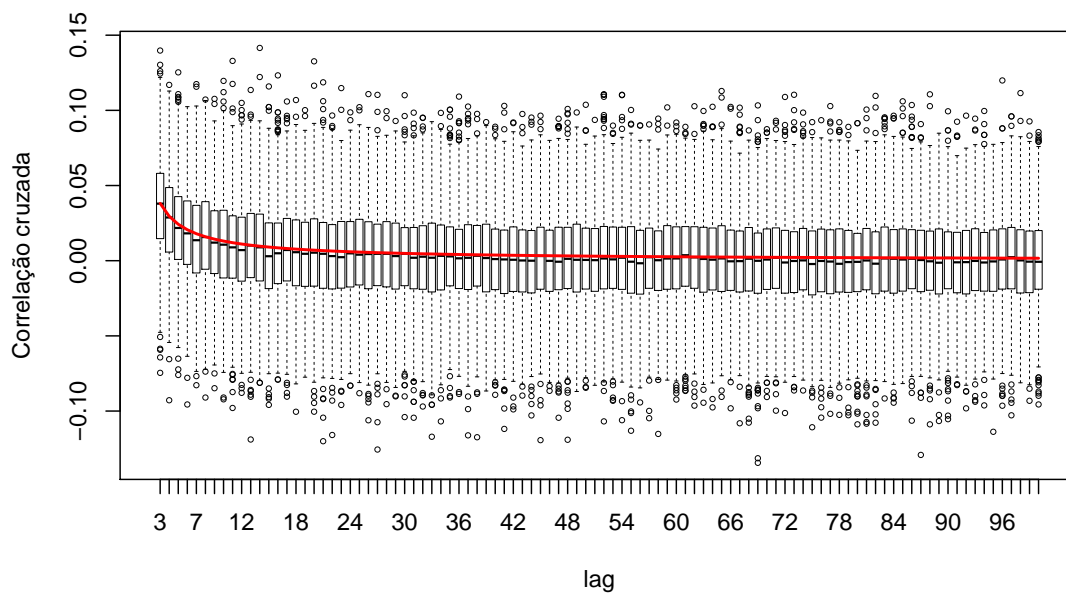


Figura 4.14: Resultados das simulações do cenário 3 da correlação cruzada, com $d = 0, 1$, para os lags $m = \{3, \dots, 100\}$

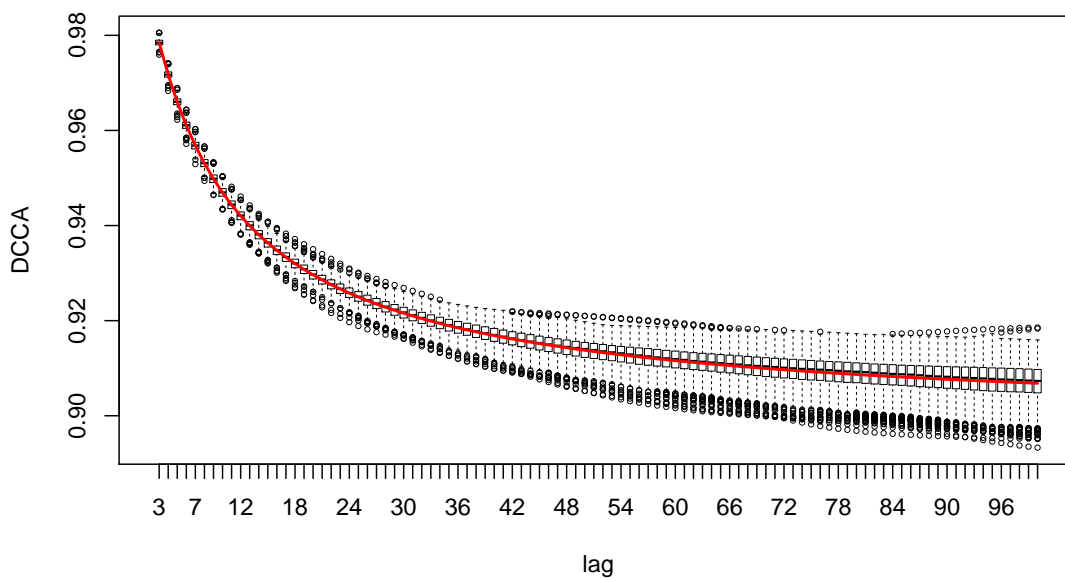


Figura 4.15: Resultados das simulações do cenário 3 do $\rho_{DCCA}(m)$, com $d = 0, 25$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

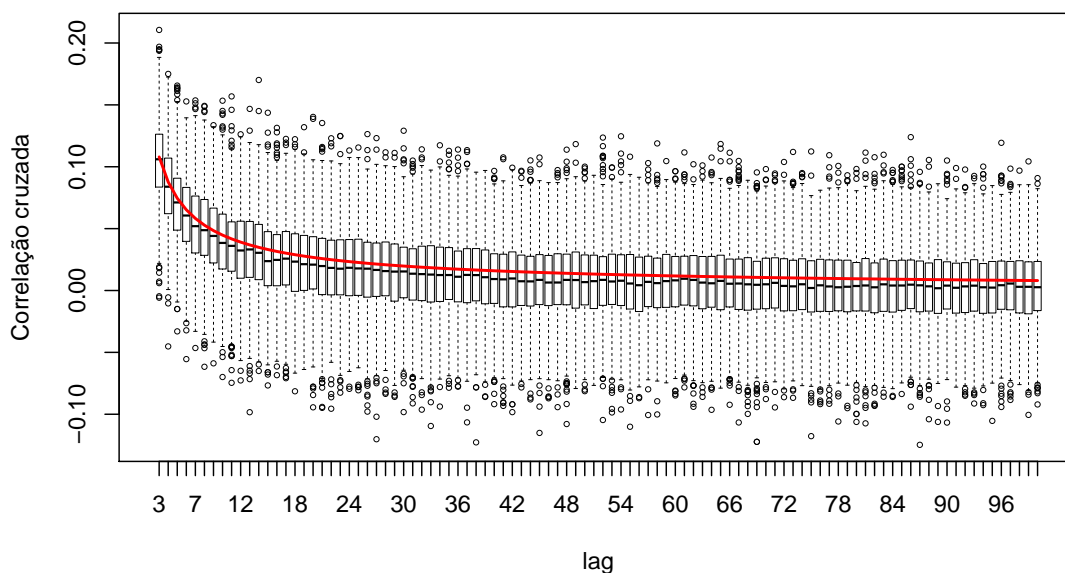


Figura 4.16: Resultados das simulações do cenário 3 da correlação cruzada, com $d = 0, 25$, para os lags $m = \{3, \dots, 100\}$

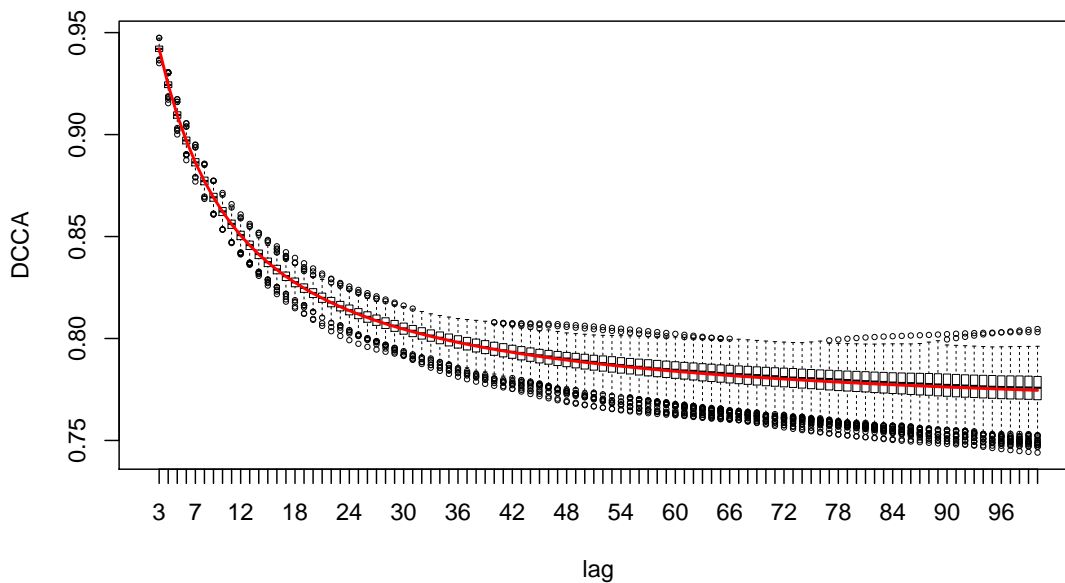


Figura 4.17: Resultados das simulações do cenário 3 do $\rho_{DCCA}(m)$, com $d = 0, 4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

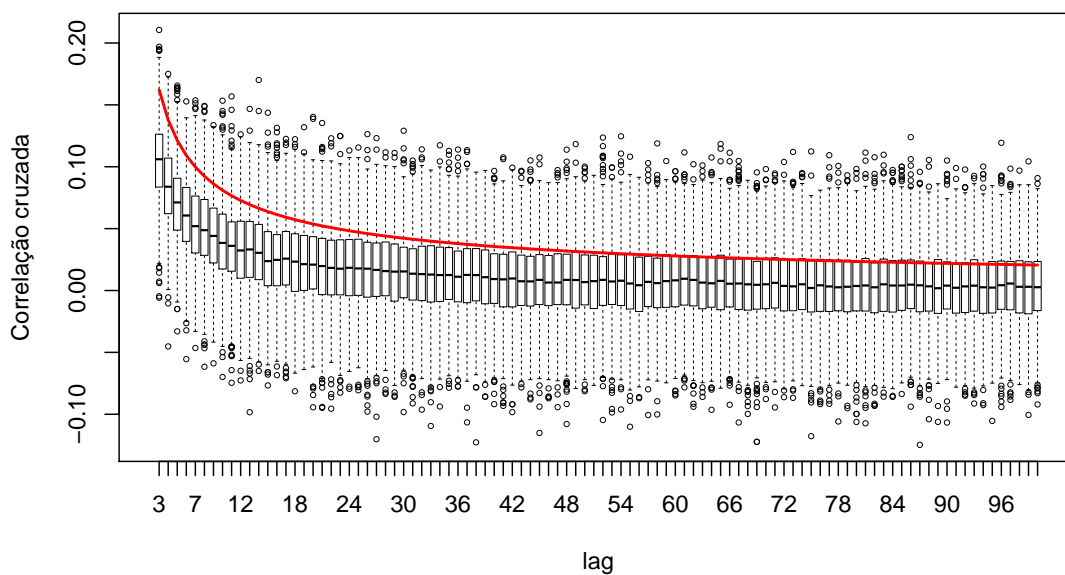


Figura 4.18: Resultados das simulações do cenário 3 da correlação cruzada, com $d = 0, 4$, para os lags $m = \{3, \dots, 100\}$

4.4 Cenário 4

Neste cenário, assim como nos demais, os valores obtidos para o $\rho_{DCCA}(m)$ foram muito próximos do esperado, para todos os valores de d_1 e d_2 . Percebe-se que, com o aumento do lag, há uma pequena diferença entre os valores obtidos e os valores esperados. Novamente, isto deve-se ao tamanho amostral da série e os resultados assintóticos disponíveis. Por outro lado, os valores obtidos para a correlação cruzada não foram muito próximos do esperado, para nenhum dos valores de d_1 e d_2 .

As Figuras 4.19, 4.21 e 4.23 trazem os resultados do $\rho_{DCCA}(m)$ no cenário 4 para $d_1 = 0, 1$ e $d_2 = 0, 2$, $d_1 = 0, 1$ e $d_2 = 0, 4$, e $d_1 = 0, 3$ e $d_2 = 0, 4$, respectivamente. Nota-se que o comportamento do $\rho_{DCCA}(m)$ neste cenário é análogo ao cenário 3, pois, para todos os valores de d_1 e d_2 , os valores obtidos para $\rho_{DCCA}(m)$ parecem convergir para um constante. Também observa-se que há valores extremos tanto inferiores quanto superiores em todos os lags, com exceção de alguns que não possuem valores extremos superiores.

Os resultados da correlação cruzada no cenário 4 para $d_1 = 0, 1$ e $d_2 = 0, 2$, $d_1 = 0, 1$ e $d_2 = 0, 4$, e $d_1 = 0, 3$ e $d_2 = 0, 4$ estão apresentados nas Figuras 4.20, 4.22 e 4.24, respectivamente. Nota-se que, quando $d_1 = 0, 1$ e $d_2 = 0, 2$, os valores obtidos para a correlação cruzada aparentam convergir para um valor próximo de zero, enquanto que, nos demais casos, os valores obtidos para a correlação cruzada aparentam convergir para um valor entre 0 e 0,1. Além disso, observamos a presença de valores extremos inferiores e superiores nos três casos analisados, em todos os lags. Cabe notar que, quando $d_1 = 0, 3$ e $d_2 = 0, 4$, há um número expressivamente maior de valores extremos superiores do que inferiores.

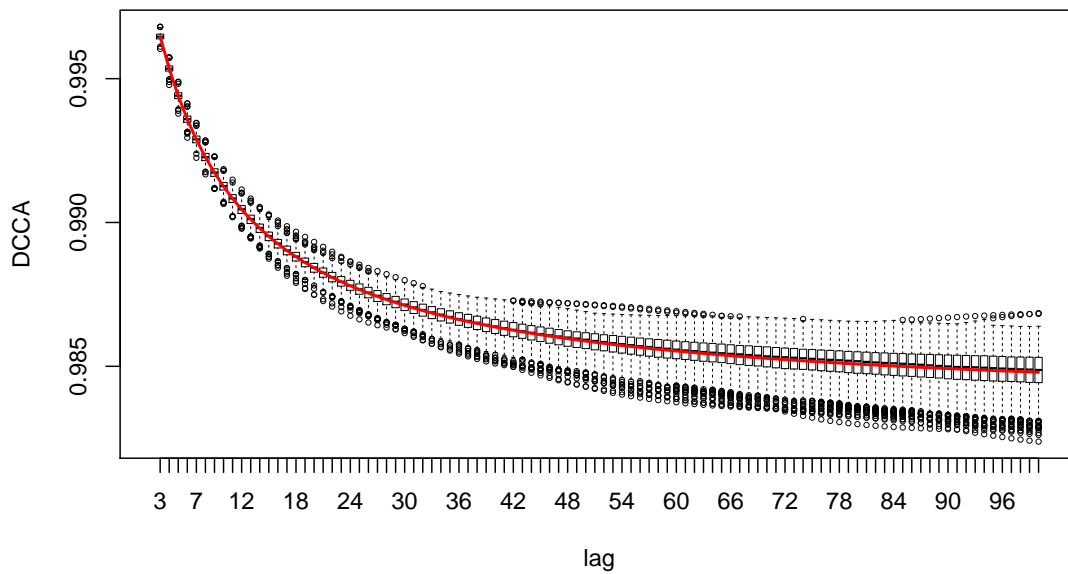


Figura 4.19: Resultados das simulações do cenário 4 do $\rho_{DCCA}(m)$, com $d_1 = 0,1$ e $d_2 = 0,2$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

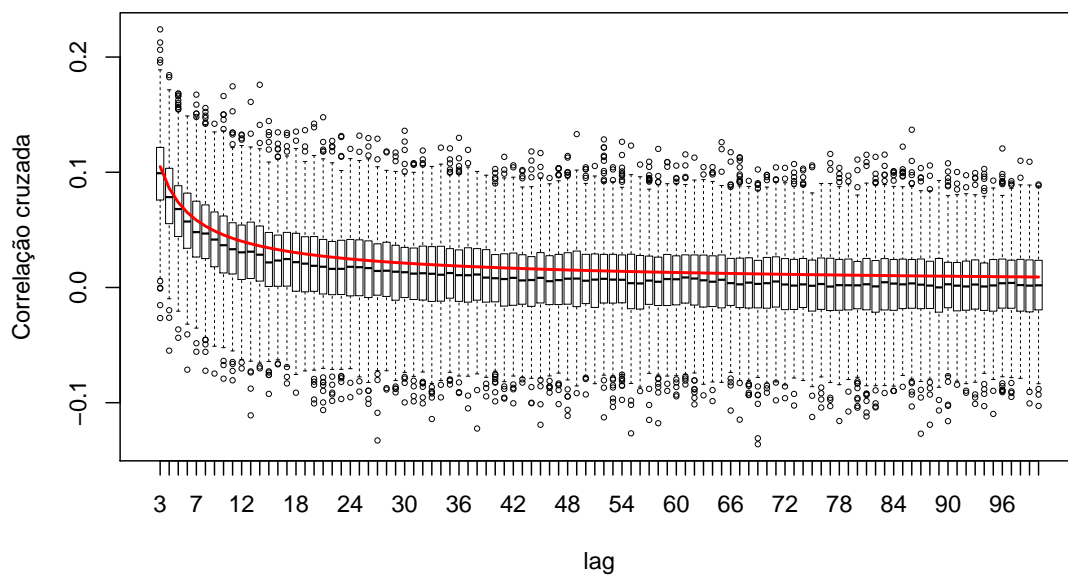


Figura 4.20: Resultados das simulações do cenário 4 da correlação cruzada, com $d_1 = 0,1$ e $d_2 = 0,2$, para os lags $m = \{3, \dots, 100\}$

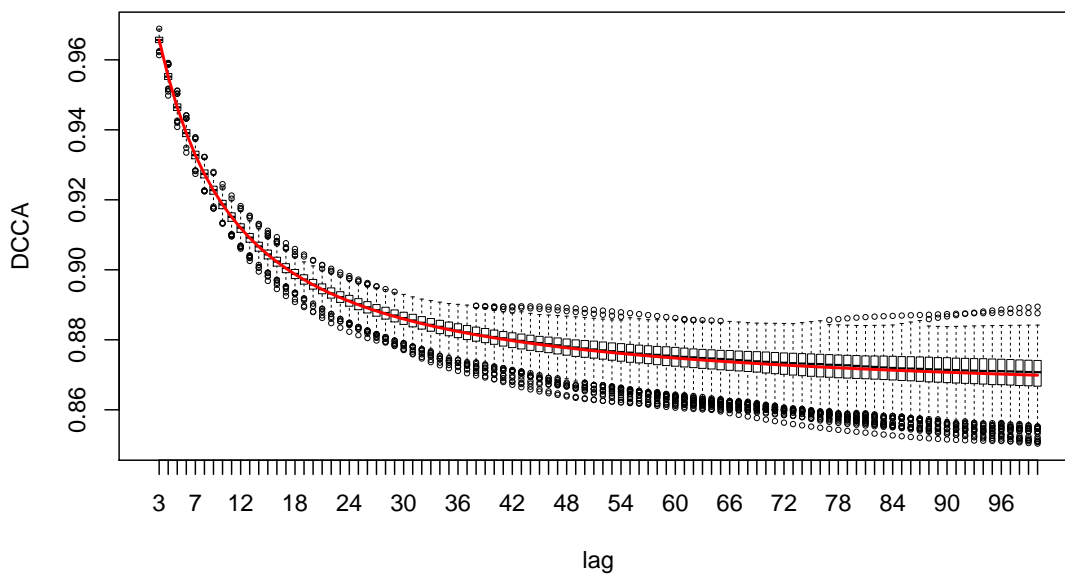


Figura 4.21: Resultados das simulações do cenário 4 do $\rho_{DCCA}(m)$, com $d_1 = 0, 1$ e $d_2 = 0, 4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

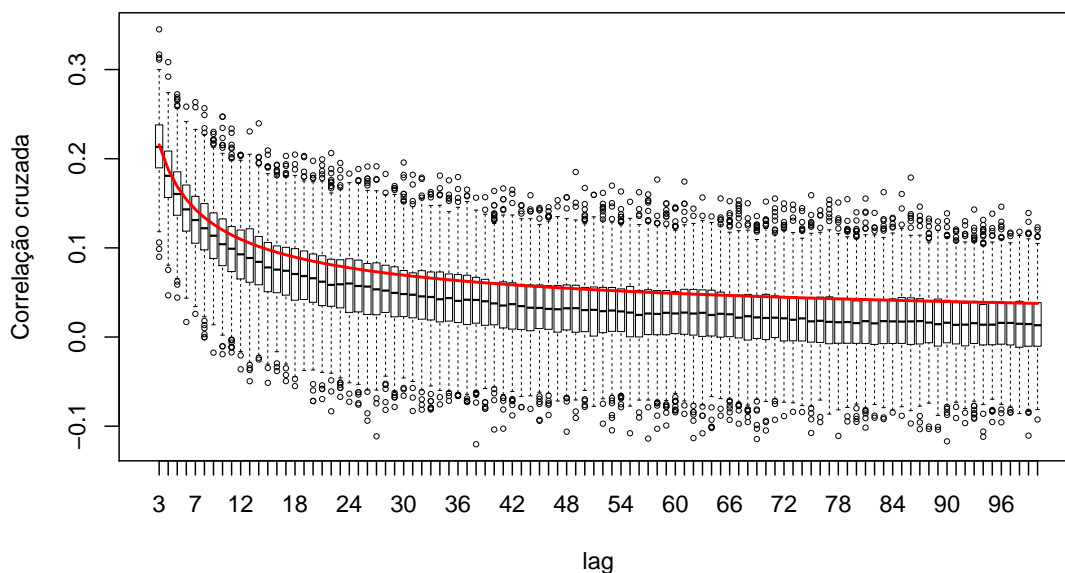


Figura 4.22: Resultados das simulações do cenário 4 da correlação cruzada, com $d_1 = 0, 1$ e $d_2 = 0, 4$, para os lags $m = \{3, \dots, 100\}$

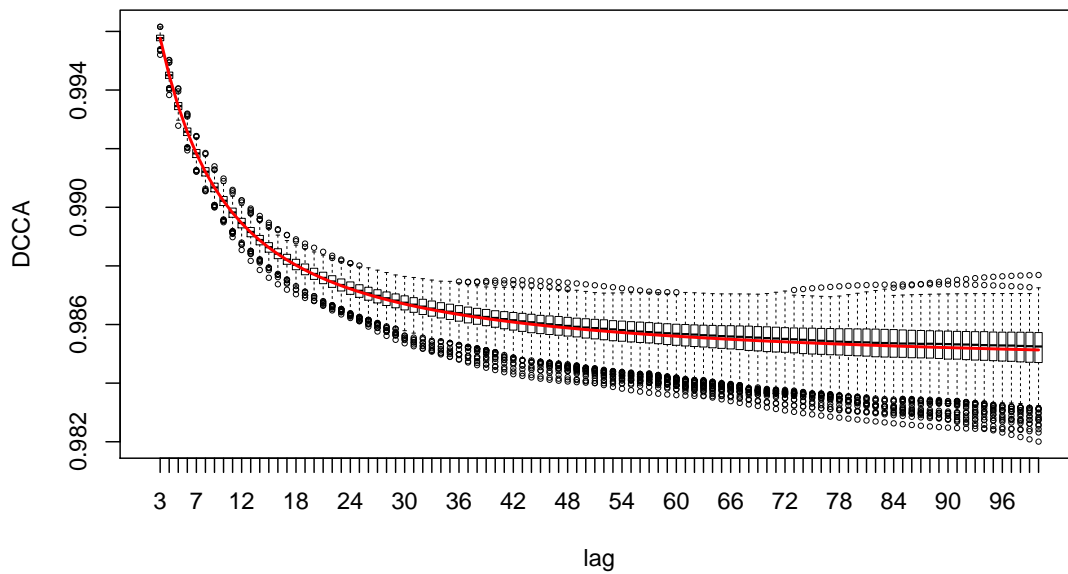


Figura 4.23: Resultados das simulações do cenário 4 do $\rho_{DCCA}(m)$, com $d_1 = 0,3$ e $d_2 = 0,4$, e os valores esperados (linha vermelha) para as janelas de tamanho $m = \{3, \dots, 100\}$

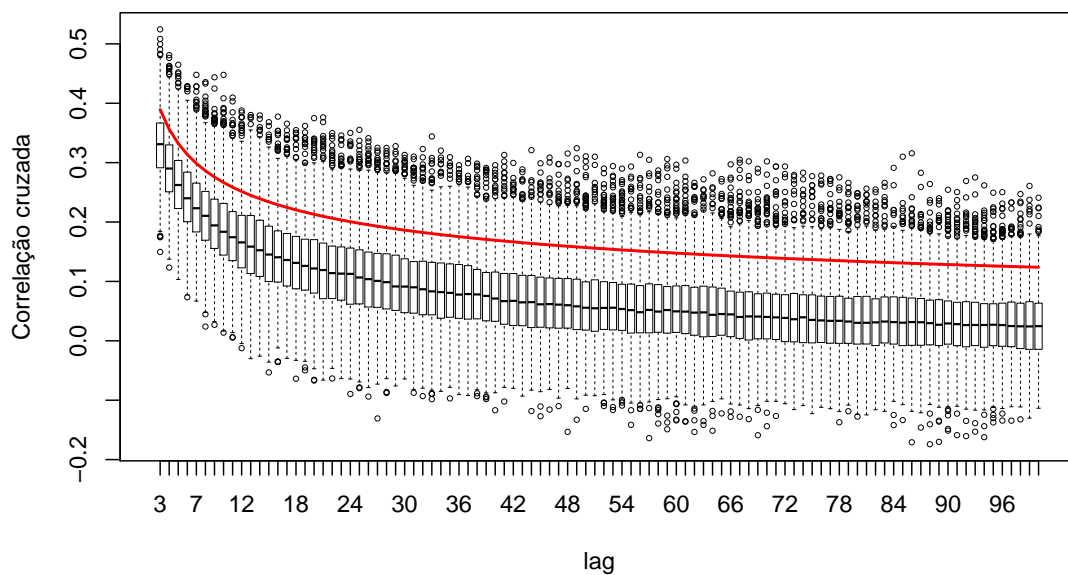


Figura 4.24: Resultados das simulações do cenário 4 da correlação cruzada, com $d_1 = 0,3$ e $d_2 = 0,4$, para os lags $m = \{3, \dots, 100\}$

5 Conclusões

No presente trabalho, estudamos o comportamento do método da DCCA no contexto de séries estacionárias com longa dependência. Analisamos o comportamento amostral do $\rho_{DCCA}(m)$ em quatro cenários, com diferentes valores de d_1 , d_2 e d . No primeiro, tínhamos duas séries temporais com longa dependência, as quais possuíam inovações i.i.d. e independentes entre si, o que implica que as duas séries eram independentes entre si. Neste cenário, esperava-se que ambos os métodos resultassem em correlação cruzada igual a zero.

No segundo cenário, tínhamos duas séries temporais com longa dependência, as quais possuíam inovações i.i.d. e correlacionadas no lag zero, o que implica que as duas séries possuíam correlação cruzada apenas no lag zero. No terceiro cenário, tínhamos uma série temporal i.i.d. com distribuição $\mathcal{N}(0, 1)$ e uma série temporal com longa dependência que possuía inovações i.i.d. dadas pela primeira série, o que implica que as duas séries apresentavam correlação cruzada não-nula. No quarto cenário, tínhamos duas séries temporais com longa dependência, as quais possuíam inovações i.i.d. geradas a partir de uma mesma série temporal i.i.d. com distribuição $\mathcal{N}(0, 1)$, o que implica que as duas séries tinham correlação cruzada não-nula. Nos cenários 2, 3 e 4, os resultados esperados foram detalhados na seção específica.

De maneira geral, ficou evidenciado que o método da DCCA produz valores muito próximos aos valores teóricos esperados, diferentemente da correlação cruzada, a qual aproxima-se dos valores esperados apenas no cenário em que as duas séries temporais são independentes entre si. Conforme [Gonçalves \(2018\)](#) já havia identificado em séries com curta dependência, observou-se a diferença dos dois métodos em relação ao comportamento da variabilidade ao longo dos lags. Enquanto o método da correlação cruzada não varia muito, é visível o aumento expressivo da variabilidade do $\rho_{DCCA}(m)$, à medida que o lag aumenta. Como já mencionado antes, isto é devido à maneira como o método é realizado.

A partir dos resultados obtidos neste trabalho, concluímos que o $\rho_{DCCA}(m)$, nos casos analisados, decresce e converge para uma constante. Essa constante ainda não foi determinada teoricamente para séries com longa dependência, no entanto, para séries com curta dependência, $\rho_{DCCA}(m) \xrightarrow{P} 1$, quando $n, m \rightarrow \infty$ ([Gonçalves, 2018](#)). Além disso, os resultados para os cenários 3 e 4 do $\rho_{DCCA}(m)$ apontam que

há influência do tamanho amostral das séries nos lags maiores.

É importante ressaltar que, apesar do método da DCCA ter sido proposto para séries não-estacionárias, o presente estudo foi realizado a partir de análises de séries temporais estacionárias, pois seguimos a linha do estudo de [Gonçalves \(2018\)](#). Como projeto futuro, temos como objetivo investigar o método em séries temporais não-estacionárias, bem como analisar os resultados obtidos com dados reais.

6 Shiny

O Shiny é um pacote do R que permite a construção de aplicativos da web interativos diretamente do R. Com ele, cria-se aplicativos de maneira fácil e intuitiva sem precisar ter conhecimento de qualquer outra linguagem de programação, além da linguagem R. O Shiny combina o poder computacional do R com a interatividade da web, pois o usuário pode modificar as entradas dos dados e as saídas são atualizadas automaticamente.

A estrutura básica de um aplicativo Shiny é gerada por um script que possui três componentes. São eles: o UI, sigla em inglês para *user interface*, que é o objeto que controla o layout e a aparência do aplicativo; o server, que é a função que contém as instruções necessárias para construção do aplicativo; e o shinyApp, que é a função que cria o aplicativo Shiny.

Neste trabalho, desenvolvemos um aplicativo para calcular o coeficiente de correlação cruzada destendenciada, denotado por coeficiente DCCA, entre duas séries temporais de igual tamanho contidas em um arquivo carregado pelo usuário. A extensão do arquivo deve ser .txt ou .csv e o arquivo deve conter pelo menos duas séries. Os exemplos apresentados aqui foram realizados com um arquivo denominado “exemplo.csv”, o qual contém duas das séries simuladas para o cenário 3 com $d = 0, 4$.

A tela inicial do aplicativo está apresentada na Figura 6.1. O painel na lateral esquerda é para entradas dos dados, onde o usuário, primeiramente, seleciona o arquivo que contém as séries e os detalhes do arquivo (header e separador). Após, seleciona quais serão consideradas a “Série 1” e a “Série 2” para o cálculo do coeficiente e escolhe os tamanhos mínimo e máximo dos intervalos. Há um aviso de que o tamanho máximo não deve ser maior que $n - 1$, onde $n =$ tamanho das séries.

O painel principal está dividido em quatro abas: “DADOS”, “DCCA”, “DESCRITIVAS” e “SOBRE O COEFICIENTE”. A aba “DADOS” é útil apenas para visualização dos dados, pois, quando o arquivo é carregado, nela é gerada uma tabela com os dados, conforme vemos na Figura 6.2, com o banco “exemplo.csv”.

Cálculo do coeficiente DCCA

Figura 6.1: Tela inicial do aplicativo “Cálculo do coeficiente DCCA”

Cálculo do coeficiente DCCA

x1	x2
0.39449197	0.55752157
-0.60807980	-0.47232259
-0.76477950	-0.99663816
1.04737027	0.52088724
0.45467195	0.43603119
0.30251492	0.37922345
-1.65421159	-1.54741947
-2.03001642	-2.66655998
-0.23414186	-1.52854449
-0.62422198	-1.70553800
1.56847236	0.41447017
0.34866471	0.00028718
0.82371523	0.53196694
-0.03981407	-0.06077219
-0.83222596	-0.98573569
-0.51280774	-1.04624024
-0.90333322	-1.56789347
0.28326843	-0.64958368
0.69404232	0.02951510
-1.00427739	-1.34445181

Figura 6.2: Aba “DADOS” do aplicativo “Cálculo do coeficiente DCCA”, com o banco “exemplo.csv”

De acordo com as séries selecionadas como “Série 1” e “Série 2” e com os valores dos tamanhos mínimo e máximo dos intervalos, na aba “DCCA” é gerada uma tabela dos coeficientes DCCA para cada tamanho das janelas. O usuário pode selecionar o número de casas decimais do coeficiente, de 0 a 15, e o número de linhas da tabela, 5, 10, 15 ou 20, que deseja visualizar, e selecionar ou não a opção de mostrar o gráfico dos coeficientes. A Figura 6.3 mostra como fica a aba “DCCA” com o banco

“exemplo.csv”, com o tamanho mínimo do intervalo igual a 3, tamanho máximo do intervalo igual a 100, coeficientes com 5 casas decimais, tabela com 10 linhas e opção de mostrar o gráfico dos coeficientes selecionada.

Cálculo do coeficiente DCCA

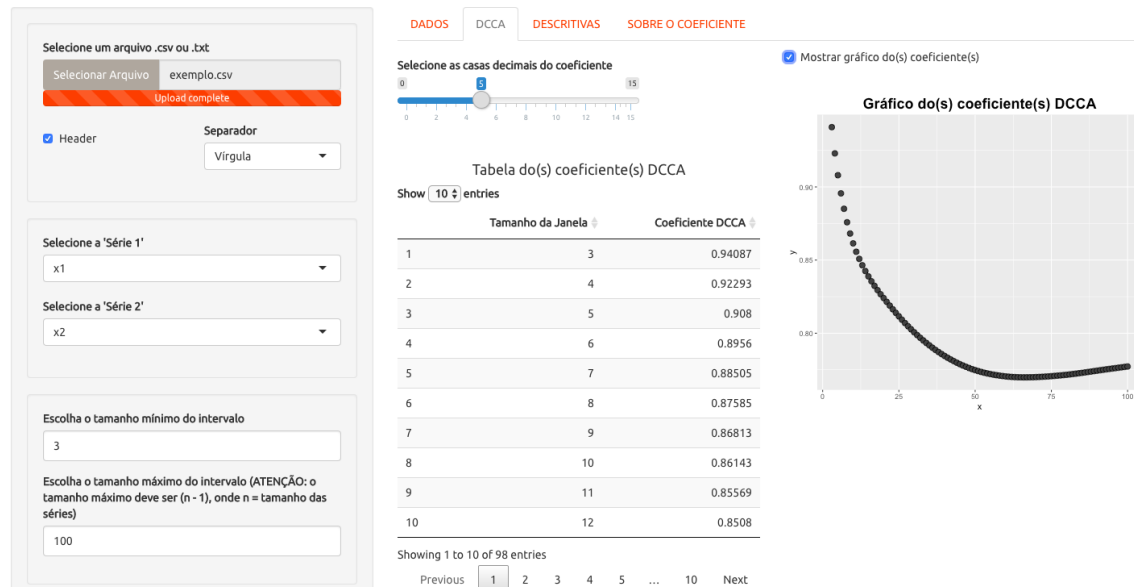


Figura 6.3: Aba “DCCA” do aplicativo “Cálculo do coeficiente DCCA”, com o banco “exemplo.csv”

As estatísticas descritivas dos dados, além de gráficos das duas séries temporais analisadas, são exibidos na aba “DESCRITIVAS” (Figura 6.4). A aba “SOBRE O COEFICIENTE” traz uma breve explicação sobre o método da DCCA e do coeficiente DCCA, assim como algumas referências (Figura 6.5).

Aplicativos do Shiny podem ser executados localmente, através do RStudio (RStudio Team, 2016), que é um ambiente que facilita o desenvolvimento dos códigos R, ou em um servidor online, acessando o aplicativo através de um navegador web por qualquer dispositivo com acesso à internet, sem precisar ter instalado o R, RStudio ou Shiny em seu computador. O aplicativo desenvolvido neste trabalho está disponível online, no servidor do RStudio para aplicativos Shiny, o Shinyapps, no link <https://ligialatorre.shinyapps.io/DCCA/>.

Cálculo do coeficiente DCCA

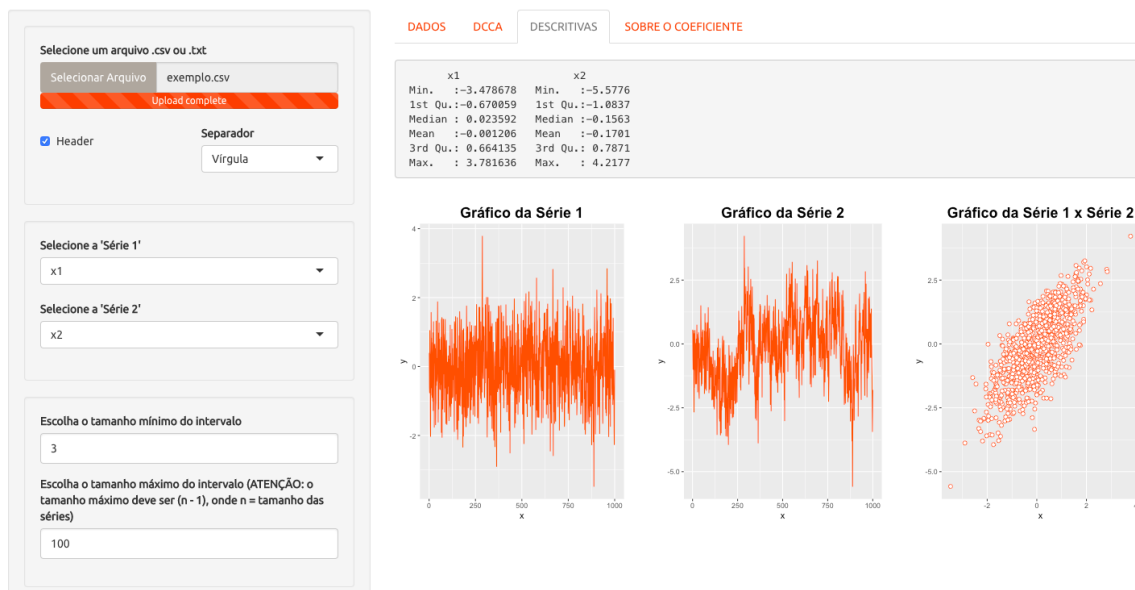


Figura 6.4: Aba “DESCRITIVAS” do aplicativo “Cálculo do coeficiente DCCA”, com o banco “exemplo.csv”

Cálculo do coeficiente DCCA

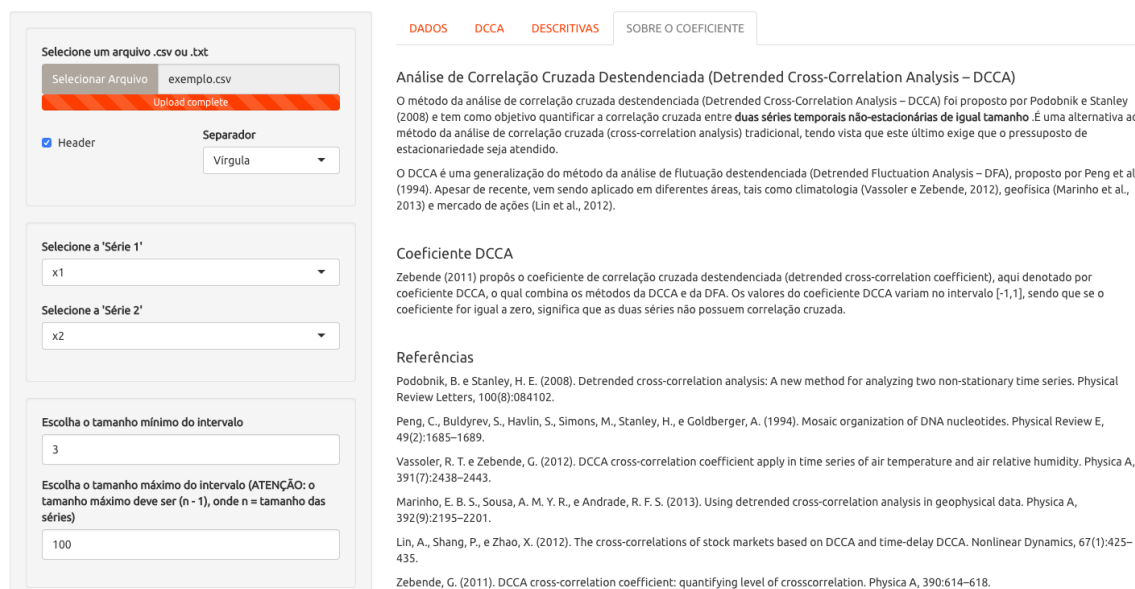


Figura 6.5: Aba “SOBRE O COEFICIENTE” do aplicativo “Cálculo do coeficiente DCCA”

Referências Bibliográficas

- Brockwell, P. J. e Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, 2^a edição.
- Gonçalves, L. F. (2018). O método da correlação cruzada destendenciada em séries temporais. Trabalho de Conclusão de Curso (Bacharelado em Estatística), Instituto de Matemática e Estatística, Departamento de Estatística, Universidade Federal do Rio Grande do Sul.
- Granger, C. W. J. e Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.
- Hosking, J. (1981). Fractional differencing. *Biometrika*, 68:165–167.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–799.
- Jun, W. e Da-Qing, Z. (2012). Detrended cross-correlation analysis of electroencephalogram. *Chinese Physics B*, 21(2):028703.
- Lin, A., Shang, P., e Zhao, X. (2012). The cross-correlations of stock markets based on DCCA and time-delay DCCA. *Nonlinear Dynamics*, 67(1):425–435.
- Machado, A., Silva, M. F., e Zebende, G. F. (2014). Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A*, 400:12–19.
- Marinho, E. B. S., Sousa, A. M. Y. R., e Andrade, R. F. S. (2013). Using detrended cross-correlation analysis in geophysical data. *Physica A*, 392(9):2195–2201.
- Morettin, P. A. e Toloi, C. M. C. (2004). *Análise de séries temporais*. Edgard Blücher, 2^a edição.
- 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.
- Podobnik, B. e Stanley, H. E. (2008). Detrended cross-correlation analysis: A new method for analyzing two non-stationary time series. *Physical Review Letters*, 100(8):084102.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2016). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Shumway, R. H. e Stoffer, D. S. (2006). *Time series analysis and its applications*. Springer-Verlag, 4^a edição.

Vassoler, R. T. e Zebende, G. (2012). DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A*, 391(7):2438–2443.

Zebende, G. (2011). DCCA cross-correlation coefficient: quantifying level of cross-correlation. *Physica A*, 390:614–618.