



UNIVERSIDADE D
COIMBRA

Beatriz de Sousa Palhinhas

Correlação entre Séries Temporais: Método DCCA

**Dissertação no âmbito do Mestrado em Métodos Quantitativos em Finanças,
orientada pelos Professores Doutores Pedro Bação e Joshua Duarte apresentada
ao Departamento de Matemática da Faculdade de Ciências e Tecnologia e à
Faculdade de Economia.**

Setembro de 2021

Correlação entre Séries Temporais: Método DCCA

Beatriz de Sousa Palhinas



UNIVERSIDADE D
COIMBRA

Mestrado em Métodos Quantitativos em Finanças

Master in Quantitative Methods in Finance

Dissertação de Mestrado | MSc Dissertation

Setembro 2021

Agradecimentos

Em primeiro lugar gostaria de agradecer ao meu orientador Professor Doutor Pedro Bação e ao coorientador Professor Doutor Joshua Duarte, por terem aceite acompanhar-me ao longo desta jornada tão importante da minha vida. Obrigada por toda a vossa disponibilização, dedicação e paciência que tiveram comigo nos piores e nos melhores momentos destes últimos meses. Obrigada do fundo do coração por tudo.

Obrigada aos meus pais, Sara Sousa e José Palhinhos, pela vossa compreensão nos dias em que tudo corria mal e eu não tinha a melhor atitude. É graças a vocês e ao vosso apoio que consegui atingir todos os meus objetivos até hoje.

Aos meus irmãos, Cláudia Palhinhos, Pedro Palhinhos e Susana Palhinhos quero agradecer por terem sido as pessoas que sempre me conseguiram dar motivação e força para continuar e nunca desistir.

Ao Sebastião Mateus por ter sido um grande apoio e por conseguir arrancar-me sempre um sorriso nos dias mais difíceis.

Às quatro raparigas que Coimbra me deu e que tenho a maior sorte em tê-las, sempre a apoiar-me e a responder a todos os meus pedidos de socorro, obrigada Ana Francisca Martins, Fátima Ramos, Patricia Fonseca e Sofia Ribeiro.

Quero agradecer aos meus alentejaninhos, com um especial agradecimento à Ana Rita Contreiras e à Susana Alves que estiveram mais perto neste momento mais emotivo da minha vida.

Por fim, obrigada a todas as pessoas que, de alguma forma, contribuíram para tornar a realização deste sonho possível. Muito obrigada a todos.

Resumo

Em 2008, B. Podobnik e H. E. Stanley, [48], propuseram um novo método para analisar duas séries temporais não estacionárias, Detrended Cross-Correlation Analysis (DCCA). Este método é uma generalização do método de Detrended Fluctuation Analysis (DFA), tomando como base o cálculo de uma "covariância sem tendência". Em 2011, G. F. Zebende, [65], propôs um novo coeficiente de correlação com o objetivo de quantificar o nível de correlação cruzada entre duas séries temporais não estacionárias. Este novo coeficiente de correlação (r_{DCCA}) é definido como uma função da variância calculada pelo método DFA e da covariância calculada pelo método DCCA. Também em 2011, B. Podobnik e co-autores, [47], apresentaram algumas propriedades do coeficiente de correlação obtido pelo método DCCA. L. Kristoufek, [37], mostrou, através de simulações, que o coeficiente de correlação calculado pelo método DCCA produz boas estimativas da correlação entre as inovações das séries temporais, pelo menos nos casos que estudou.

Este trabalho tem como principal objetivo estender a análise feita no artigo de Kristoufek a mais casos, prestando particular atenção à (não) estacionariedade das séries temporais usadas. Assim, é avaliado o desempenho do método DCCA (em amostras finitas), quando se fazem variar alguns parâmetros do modelo ARFIMA gerador das séries, tais como os parâmetros das componentes média móvel e autoregressiva, e a ordem de integração de cada variável.

Os resultados confirmam em geral a ideia, transmitida pelo trabalho de Kristoufek, de que o coeficiente de correlação calculado pelo método DCCA tem um bom desempenho para estimar a correlação entre as inovações das séries. Porém, também foram encontrados certos casos em que tal não sucede, sugerindo a necessidade de analisar mais detalhadamente as propriedades do coeficiente de correlação calculado pelo método DCCA.

Conteúdo

Lista de Figuras	ix
1 Introdução	1
2 Correlação entre Séries Temporais	3
2.1 Método de Monte Carlo	3
2.2 Séries Temporais	4
2.3 Análise de Correlação Cruzada	8
2.4 Análise de flutuações sem tendência-DFA	9
2.4.1 Método Multifractal DFA (MF-DFA)	10
2.5 Análise de Correlações Cruzadas sem Tendência-DCCA	11
2.6 Coeficiente de Correlação Cruzada sem Tendência	13
2.7 Medição da correlação entre duas séries temporais não estacionárias com r_{DCCA}	14
3 Desempenho do Coeficiente de Correlação Cruzada sem Tendência	17
3.1 Metodologia	17
3.2 1º Caso: Séries temporais com diferentes ordens de estacionariedade, ARFIMA(0,d,0)	18
3.3 2º Caso: Variação dos parâmetros AR/MA, f e q para a série x_t	21
3.3.1 1ªParte: $f_x \in [0;0.5]g, q_x \in [0;0.3;0.5;0.7]g; f_y = q_y = 0$	21
3.3.2 2ªParte: $f_x; q_x; f_y; q_y \in [0.9; 0.45; 0; 0.45; 0.9]g$	28
4 Conclusão	43
Bibliografia	45
Anexo A Gráficos adicionais	49

Lista de Figuras

3.1	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$	19
3.2	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$	20
3.3	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = f_y = q_y = 0$	22
3.4	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = f_y = q_y = 0$	23
3.5	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $q_x = 0.5; f_x = f_y = q_y = 0$	24
3.6	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $q_x = 0.5; f_x = f_y = q_y = 0$	25
3.7	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = 0.3; f_y = q_y = 0$	26
3.8	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = 0.3; f_y = q_y = 0$	27
3.9	$d_1 = 0.25; d_2 = 0.75$	30
3.10	$d_1 = 0.25; d_2 = 1.25$	31
3.11	$d_1 = 0.75; d_2 = 1.25$	32
3.12	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , ($f_x; q_x$). p_x corresponde a f_x e t_x corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série y_t $f_y = 0.9$ e $q_y = 0.9$	34

3.13	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e t_x corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série $f_y = 0.9$ e $q_y = 0.9$.	35
3.14	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e t_x corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série $f_y = 0$ e $q_y = 0.9$.	36
3.15	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e t_x corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série $f_y = 0$ e $q_y = 0.9$.	37
3.16	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série y_t , $(f_y; q_y)$. p_y corresponde a f_y e q_y corresponde a q_y . Resultados para as séries com parâmetros autorregressivo e média móvel da série x_t $f_x = q_x = 0$.	39
3.17	Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e q_x corresponde a q_x . Resultados para as séries com parâmetros autorregressivo e média móvel da série y_t $f_y = q_y = 0$.	39
3.18	Coeficiente de Correlação estimado pelo método DCCA para os parâmetros autorregressivo e média móvel da série x_t , $f_x = 0.45$, $q_x = 0.45$ e para y_t $f_y = q_y = 0$. p_x corresponde a f_x e t_x corresponde a q_x .	40
A.1	Gráficos 1º Caso para $T=1000$	50
A.2	Gráficos 1º Caso para $T=5000$	51
A.3	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$	52
A.4	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0; 0.5)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$	53
A.5	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0.3)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$	54
A.6	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$	55
A.7	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0; 0.5)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$	56
A.8	Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0.3)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$	57
A.9	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$	58
A.10	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.25; 0.75)$ e $(p_y; q_y) = (0.9; 0.9)$	59
A.11	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.25; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$	60
A.12	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$	61
A.13	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0; 0.9)$	62
A.14	Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0; 0.9)$	63
A.15	Gráficos do 2º Caso da 2ª Parte do Caso a)	64
A.16	Gráficos do 2º Caso da 2ª Parte do Caso b)	65

Capítulo 1

Introdução

Em muitos estudos, o aspecto mais importante é a análise da correlação cruzada entre diferentes séries. Tal acontece nas mais variadas áreas do saber, desde a geofísica ([51]) às finanças ([53], [57], [58], [33], [54]). Nomeadamente, as funções de correlação cruzada, em conjunto com as funções de auto-correlação, são muito utilizadas no estudo da dinâmica dos sistemas naturais.

Porém, as técnicas tradicionais apenas são bem sucedidas na presença de séries estacionárias. Uma vez que muitas das séries temporais estudadas nos sistemas da física, da biologia, da hidrologia, da sociologia, etc., são não estacionárias, torna-se difícil de estudar as propriedades estatísticas destes sistemas utilizando as técnicas desenvolvidas para séries estacionárias. Por forma a ultrapassar este problema, foi desenvolvido um método para estudar a relação entre a variância de séries temporais e a escala temporal da análise na presença de não estacionariedade: Detrended Fluctuation Analysis (DFA), [46]. O passo seguinte foi passar da quantificação da variância de uma série para a quantificação da covariância entre duas séries. Para tal foi proposto um método baseado na "covariância sem tendência", chamado Detrended Cross-Correlation Analysis (DCCA), [48]. A ideia chave deste método é caracterizar as relações entre as flutuações das séries temporais não estacionárias a partir da separação entre as flutuações relevantes e as flutuações resultantes das tendências locais, globais e periódicas que existam nas séries temporais, nomeadamente nas que abrangem um período temporal mais longo, [27].

Em 2011, Zebende, [65], introduziu um novo coeficiente de correlação que pretende ser adequado quando se trata de quantificar o nível de correlação cruzada entre duas séries temporais não estacionárias. Este coeficiente de correlação cruzada faz uso dos estimadores obtidos pelo método DFA (para a variância) e pelo método DCCA (para a covariância). No seu texto, Zebende demonstra a potencial utilidade deste coeficiente através de dois exemplos com dados reais, o primeiro relativo à análise de duas séries temporais climatológicas e o segundo envolvendo séries da bolsa de valores, associadas aos valores diários de abertura e de fecho de dois índices.

A presente dissertação foca-se na avaliação do comportamento deste coeficiente. Mais concretamente, analisamos o comportamento do coeficiente de autocorrelação obtido pelo método DCCA quando aplicado a amostras finitas de duas séries geradas por modelos ARFIMA. Consideramos diferentes casos para vários dos parâmetros dos modelos ARFIMA, tais como a ordem da componente autorregressiva (p , que poderá ser zero ou um), o coeficiente autorregressivo ($f \neq 0$ no caso em que $p = 1$), a ordem da componente média móvel (q , que poderá ser zero ou um), o coeficiente associado

a essa componente média móvel ($q \neq 0$ no caso em que $q = 1$), e o parâmetro de integração da série (d). Este estudo pretende assim complementar o estudo já feito por Kristoufek, [37], o qual mostrou que o coeficiente de correlação obtido pelo método DCCA tem um bom desempenho quando as séries têm o mesmo parâmetro de integração e não têm componentes autorregressivas nem média móvel.

Em termos gerais, os resultados apresentados nesta dissertação confirmam a ideia de que o método DCCA tem um bom desempenho. Porém, no caso em que as séries são não estacionárias, encontramos algumas parametrizações em que as estimativas produzidas pelo método DCCA não dão uma indicação fiável acerca da correlação existente entre as séries.

A dissertação está organizada da seguinte forma. No próximo capítulo introduzimos alguns conceitos, definições e métodos utilizados. No terceiro capítulo apresentamos a metodologia usada na simulação, os resultados e a discussão desses resultados. Por fim, no quarto capítulo são apresentadas as principais conclusões deste trabalho.

Capítulo 2

Correlação entre Séries Temporais

Neste capítulo serão apresentados os instrumentos e as definições utilizados no presente trabalho. Inicialmente é descrito o método de Monte Carlo, em seguida é feita uma introdução a alguns conceitos da área das Séries Temporais, nos subcapítulos seguintes temos a análise de correlação cruzada, a análise de flutuações sem tendência (DFA) e, na sequência, o método multifractal DFA (M-DFA), a análise de correlações cruzadas sem tendência (DCCA) e por fim o coeficiente de correlação cruzada sem tendência (r_{DCCA}). Terminamos este capítulo com uma pequena introdução ao estudo realizado por Ladislav Kristoufek em 2014, [37].

2.1 Método de Monte Carlo

O Método de Monte Carlo é um método computacional que utiliza números aleatórios e estatísticas para resolver problemas em áreas como a Matemática, a Engenharia e as Finanças. Segundo Johansen, [29], este método é utilizado principalmente em dois tipos de problemas estatísticos: otimização e integração. Segundo John Panos Najarian, [44], a definição clássica dos métodos de Monte Carlo diz-nos que este trata do cálculo de "números pseudoaleatórios" e a sua utilização em computação de integrais.

Um computador não é capaz de gerar números aleatórios puros porque, por definição, estes são uma construção abstrata: o resultado de um procedimento ideal que escolhe números de uma forma imprevisível, sem padrão. Em vez disso, os computadores criam sequências de números seguindo as instruções do código informático que for usado para gerar tais sequências. Um bom algoritmo para gerar números aleatórios produz números que aparentam não ter um padrão e que satisfazem certas condições estatísticas. Voltando a usar o mesmo algoritmo nas mesmas condições, os "números aleatórios" voltarão a ser os mesmos. São, assim, chamados "números pseudoaleatórios".

Num contexto mais geral, o termo "método de Monte Carlo" é sinónimo de "simulação probabilística". Na simulação probabilística é usado um código que duplica o comportamento de um sistema e é utilizado um gerador de números aleatórios para representar os aspetos probabilísticos desse sistema. Será este tipo de simulação que irá ser utilizada ao longo deste trabalho para chegar a conclusões sobre a utilização do método DCCA no cálculo do coeficiente de correlação cruzada.

Os métodos clássicos de Monte Carlo, referidos no início deste capítulo, são um subconjunto deste conceito mais amplo. Dentro deste conceito mais amplo de método de Monte Carlo, cabem

ainda subcategorias como o Modelo de Filas e a simulação do tipo determinístico. As simulações podem ainda ser classificadas como simulações de eventos discretos ou de eventos contínuos. Mais informações sobre as variedades de métodos de Monte Carlo podem ser obtidas em [44].

2.2 Séries Temporais

Hoje em dia são raras as áreas científicas que não são confrontadas com o estudo de fenómenos que evoluem ao longo do tempo, onde a identificação do tipo de evolução assume um papel fundamental [23]. Um contexto em que a identificação do tipo de evolução se torna especialmente importante é o da previsão de valores futuros.

À sucessão de observações (em datas, ou pontos no tempo, identificadas) do fenómeno sob estudo chamamos série temporal ou cronológica. Um aspeto importante na análise de séries temporais é a ordem de recolha das observações (identificação e ordenação das datas). Uma vez que o interesse está em analisar e modelar a dependência existente entre as observações vizinhas. Em certos casos, é necessário fazer a identificação de comportamentos sazonais e, possivelmente, a sua correção.

Até ao início dos anos 80 estes estudos eram, na sua maioria, baseados nos modelos lineares, com destaque para a classe dos processos auto-regressivos médias móveis (ARMA). A formulação retida para tais processos estocásticos permite exprimir o valor presente da série como função linear dos seus valores passados e dos valores presente e passados de um ruído que se interpreta como a inovação da série. No entanto, os modelos lineares são insuficientes, em particular para o tratamento de dados financeiros (nos quais frequentemente a variabilidade instantânea dos valores depende de modo significativo do passado) e ainda no estudo de fenómenos naturais susceptíveis de alterações bruscas e violentas (como fenómenos sismológicos, em que podem surgir sucessões de registos com picos repentinos seguidos de baixos valores em termos absolutos).

Com o objetivo de estudar este tipo de comportamentos, para os quais os modelos lineares devolvem uma representação insatisfatória, surgiram modelos com aspetos não lineares, entre os quais se podem destacar os modelos condicionalmente heteroscedásticos (Engle 1982, [14]) e os modelos bilineares (Granger e Andersen, 1978, [25]).

É, portanto, importante dizer que é uma série temporal, é uma realização de um processo estocástico, ou seja, de um conjunto de variáveis aleatórias ordenadas em intervalos regulares de tempo (cada dia, cada mês, cada ano, etc.).

É necessário começar por definir alguns conceitos básicos. As quatro definições seguintes foram retiradas do livro *Probabilidades: Princípios Teóricos* de E. Gonçalves e N. Lopes, [24].

Definição: Seja W um conjunto e A uma classe de subconjuntos de W . Diz-se que A é uma σ -álgebra, ou tribo, sobre W se verifica os seguintes axiomas:

a) $W \in A$.

b) $\delta A \in A$ (δA é o complemento de A no conjunto W).

c) Se $(A_n)_{n \in \mathbb{N}}$ é uma sucessão qualquer de acontecimentos de A então $\bigcup_{n=1}^{\infty} A_n \in A$.

Ao par $(W; A)$ chama-se espaço probabilizável ou mensurável,

Definição: Seja $(W; A)$ um espaço probabilizável; P é uma probabilidade sobre $(W; A)$ se P é uma aplicação definida sobre A com valores em $[0; 1]$ verificando os seguintes axiomas:

a) $P(W) = 1$:

b) Para toda a sucessão $(A_n)_{n \in \mathbb{N}}$ de elementos de A dois a dois incompatíveis, tem-se

$$P \left(\bigcup_{n=1}^{+\infty} A_n \right) = \sum_{n=1}^{+\infty} P(A_n):$$

Ao terno $(W; A; P)$ chama-se espaço de probabilidade.

Definição: Seja X uma variável aleatória real discreta de suporte $S = \{x_1; x_2; \dots; x_n; \dots; g\}$ e com função probabilidade $p(x_n) = P(X = x_n)$, $n \in \mathbb{N}$. Se $\sum_{n=1}^{+\infty} |x_n| P(X = x_n) < +\infty$ define-se a esperança matemática da variável aleatória real X como sendo

$$E(X) = \sum_{n=1}^{+\infty} x_n P(X = x_n):$$

Definição: Seja X uma v.a.r. absolutamente contínua com função densidade f . Se $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$ dizemos que a esperança matemática da v.a.r. X existe e é dada por

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx:$$

Consideremos então um espaço de probabilidade $(W; S; P)$, um espaço mensurável $(E; e)$ e um conjunto de índices T .

As definições seguintes foram retiradas dos textos de apoio disponibilizados pela Professora Doutora Esmeralda Gonçalves no âmbito da unidade curricular de Séries Temporais no ano letivo 2020/2021, [23].

Definição: Um processo estocástico (p.e.) indexado por T é uma família de variáveis aleatórias $(X_t; t \in T)$ definidas sobre $(W; S; P)$ e com valores em $(E; e)$. T é o espaço dos tempos, E é o espaço dos estados e, para cada $w \in W$ fixo, $(X_t(w); t \in T)$ define uma realização ou trajetória.

Definição: Um p.e. X é fortemente estacionário (ou estritamente estacionário) se $\forall k \in \mathbb{N}$, $\forall t_1, \dots, t_k \in \mathbb{Z}$, $\forall h \in \mathbb{Z}$, $(X_{t_1}, \dots, X_{t_k})$ e $(X_{t_1+h}, \dots, X_{t_k+h})$ têm a mesma lei.

Definição: Um p.e. $X = (X_t; t \in T)$ é de segunda ordem se $E \|X_t\|^2 < +\infty$, $\forall t \in T$, em que $E(\cdot)$ é a esperança matemática do argumento.

Um exemplo de um p.e. de segunda ordem é o Ruído Branco (univariado), que é uma sucessão de variáveis aleatórias reais $e = (e_t; t \in \mathbb{Z})$ com as seguintes características

- centradas: $E(e_t) = 0$

- com igual variância: $t(t;t) = Cov(e_t; e_t) = E(e_t^2) = s^2$
- não correlacionadas: $t(s;t) = Cov(e_s; e_t) = E(e_s e_t) = 0; s \neq t$.

Definição: Um p.e. de segunda ordem $X = (X_t; t \in \mathbb{Z})$ é estacionário no sentido fraco se

i) $E(X_t) = m; \forall t \in \mathbb{Z}$

ii) $Cov(X_t; X_{t+h}) = g(h); \forall t \in \mathbb{Z}$

Definição: Chamamos inovação do p.e. X ao p.e. $e = (e_t; t \in \mathbb{Z})$ formado pelos erros das previsões ao horizonte 1.

Definição: Um p.e. $X = (X_t; t \in \mathbb{Z})$ centrado e de segunda ordem (estacionário ou não) admite uma representação auto-regressiva de ordem p , $AR(p)$, se existirem números reais $f_1; f_2; \dots; f_p$ e um ruído branco $(e_t; t \in \mathbb{Z})$, tais que

$$X_t - f_1 X_{t-1} - \dots - f_p X_{t-p} = e_t;$$

com $f_p \neq 0$.

Definição: Um p.e. de segunda ordem $X = (X_t; t \in \mathbb{Z})$ admite uma representação média móvel de ordem q , $MA(q)$, quando

$$X_t = e_t - q_1 e_{t-1} - q_2 e_{t-2} - \dots - q_q e_{t-q}; \forall t \in \mathbb{Z};$$

onde $q_1; \dots; q_q \in \mathbb{R}$, com $q_q \neq 0$, $e = (e_t; t \in \mathbb{Z})$ um ruído branco.

Definição: Um p.e. estacionário $X = (X_t; t \in \mathbb{Z})$ admite uma representação $ARMA(p; q)$ se

$$X_t - f_1 X_{t-1} - \dots - f_p X_{t-p} = e_t - q_1 e_{t-1} - \dots - q_q e_{t-q}$$

com $f_p \neq 0; q_q \neq 0$.

A modelação $ARMA$ generaliza simultaneamente os modelos AR puros e as MA puras.

Para muitas séries temporais, a hipótese de estacionaridade não é plausível mas poderemos inverter esta situação considerando a hipótese de estacionaridade para a série das diferenças

$$X_t - X_{t-1} = (1 - L)X_t$$

em que L é o operador atraso (*lag*), ou mais geralmente, para a série das diferenças de ordem d , i.e.,

$$(1 - L)^d X_t; d \in \mathbb{N}$$

Os modelos $ARIMA(p; d; q)$ (AutoRegressive Integrated Moving Average) introduzidos por Box e Jenkins (1970), [4], incluem um parâmetro d . O parâmetro d é um número inteiro que estabelece o

nível de diferenciações necessárias para transformar a série temporal numa série estacionária.

Definição: Um processo $X = (X_t; t \geq 0)$ segue um modelo ARIMA($p; d; q$) (AutoRegressive Integrated Moving Average) se

$$F(L)(1 - L)^d X_t = Q(L)e_t; t \geq 0;$$

onde $e = (e_t; t \geq 0)$ é um ruído branco e $F(L) = 1 - f_1 L - \dots - f_p L^p$, $f_p \neq 0$, e $Q(L) = 1 - q_1 L - q_2 L^2 - \dots - q_q L^q$, $q_q \neq 0$, sendo ambos polinómios cujas raízes são de módulo superior a 1.

A partir dos anos 80, Granger e Joyeux, [26], e ainda Hosking (1981), [28], propuseram uma generalização deste tipo de modelos, na qual o parâmetro d pode tomar valores não só inteiros mas também fracionários. Introduziram assim os modelos que nos permitem estudar séries caracterizadas por longas dependências temporais. A este tipo de modelos foi dado o nome de modelos ARFIMA($p; d; q$), onde F significa "fractional".

Segundo [2], quando d é um valor não inteiro, o processo y_t dado por

$$F(L)(1 - L)^d (y_t - m) = Q(L)e_t; t \geq 0 \quad (2.1)$$

em que m é a sua média, L é o operador atraso, $F(L)$ e $Q(L)$ são os polinómios de ordem p e q , respetivamente, dados por

$$F(L) = 1 - f_1 L - \dots - f_p L^p$$

$$Q(L) = 1 - q_1 L - \dots - q_q L^q$$

com $f_i (i = 1, \dots, p)$ e $q_j (j = 1, \dots, q)$ constantes reais e $f_0 = q_0 = 1$, é então chamado AutoRegressive Fractional Integrated Moving Average, ARFIMA($p; d; q$), onde d é o grau de diferenciação. Segundo os apontamentos da unidade curricular de Análise de Séries Financeiras, dados pelo Professor Doutor Rui Pascoal no ano letivo 2019/2020, [45], este processo possui as seguintes propriedades:

1. se $d < \frac{1}{2}$, o processo é estacionário;
2. se $d \geq \frac{1}{2}$, o processo é não estacionário
 - (a) se $\frac{1}{2} < d < 1$ o processo é não estacionário mas reverte à média;
 - (b) se $d = 1$, o processo é não estacionário e não reverte à média (é explosivo);
3. se $0 < d < \frac{1}{2}$ temos um processo de memória longa positiva;
4. se $-1 < d < 0$ o processo é de memória longa negativa;
5. se $d = 0$ temos um processo ARMA, logo de memória curta.

Uma vez que este trabalho tem como principal objetivo avaliar a fiabilidade de um coeficiente de correlação estimado é necessário introduzir ainda algumas definições que foram retiradas dos apontamentos disponibilizados pela Professora Doutora Nazaré Lopes, no âmbito da disciplina de Estatística, [41].

Seja $(X_1; X_2; \dots; X_n)$ um conjunto de n amostras de uma v.a.r. X de lei Q pertencente a uma família de leis paramétricas $(Q_q; q \in Q)$ sobre $(R; B)$, onde Q é o espaço dos parâmetros e B a σ -álgebra de Borel.

Hipótese: O modelo é identificável, $\exists! q \in Q : Q_q = Q$:

Seja g uma função mensurável de Q em A , com $A \in B$, i.e.

$$\exists B \in B_A; g^{-1}(B) \in F;$$

com F uma σ -álgebra sobre Q e B_A a σ -álgebra dos borelianos de A . A estimação pontual de $g(q)$ a partir de uma amostra tem por base uma estatística T com valores em A .

Definição: a) Estimador de $g(q)$ é toda a estatística $T_n = T(X_1; \dots; X_n)$ com valores em A .

b) Estimativa de $g(q)$ é o valor do estimador associado a uma amostra particular $(x_1; \dots; x_n)$, $t_n = T(x_1; \dots; x_n)$.

Definição: Um estimador de T_n de $g(q)$ admitindo valor médio, para todo $q \in Q$, diz-se cêntrico ou sem viés se

$$E_q(T_n) = g(q):$$

No caso contrário, dir-se-á enviesado.

2.3 Análise de Correlação Cruzada

O cálculo de correlações cruzadas tem como objetivo o estudo das relações entre os movimentos de duas ou mais séries temporais.

Sejam $\{X_t\}_{t \in \mathbb{Z}}$ e $\{Y_t\}_{t \in \mathbb{Z}}$ duas séries temporais estacionárias com médias $m_X = E(X_t)$ e $m_Y = E(Y_t)$, variâncias $s_X^2 = Var(X_t)$ e $s_Y^2 = Var(Y_t)$, e funções de autocovariância $g_X(h)$ e $g_Y(h)$, com $h \in \mathbb{Z}$. As funções $g_{X;Y}(h)$ e $r_{X;Y}(h)$ que representam a função de covariância cruzada e a função de correlação cruzada, respetivamente, são dadas por

$$g_{X;Y}(h) = Cov(X_t; Y_{t+h}) = E[(X_t - m_X)(Y_{t+h} - m_Y)] \quad (2.2)$$

e

$$r_{X;Y}(h) = \frac{g_{X;Y}(h)}{g_X(0)g_Y(0)} = \frac{g_{X;Y}(h)}{s_X^2 s_Y^2} = \frac{g_{X;Y}(h)}{s_X s_Y} \quad (2.3)$$

Observação: Quando $h = 0$, $r_{X;Y}(0)$ representa o coeficiente de correlação de Pearson, que tem como principal objetivo medir a correlação linear entre duas variáveis aleatórias.

2.4 Análise de flutuações sem tendência-DFA

A análise de flutuações sem tendência, Detrended Fluctuation Analysis-DFA, foi proposta por Peng et.al, [46], e tem sido bastante utilizada para a determinação e detecção de correlações cruzadas de longo alcance entre duas séries temporais não estacionárias.

Segundo Kantelhardt e co-autores, [31], o DFA é um método bem estabelecido para determinar o comportamento em diferentes escalas de dados afectados pela presença tanto de ruído como de tendência de origem e forma desconhecidas. Ainda segundo Kantelhardt e co-autores, [31], podemos descrever o processo de cálculo do coeficiente de correlação cruzada na análise DFA através dos quatro passos seguintes, começando por supor que temos N observações de x_i e que apenas uma pequena fração destes valores são iguais a zero.

1ºPasso: Cálculo do sinal integrado da série $Y(l) = \sum_{k=1}^l x_k$.

Nota: podemos subtrair a média ao sinal integrado, mas não acrescentaria nada uma vez que no 3º passo será retirada a tendência da série.

2ºPasso: Dividir a série $Y(l)$ em $N_l = \frac{N}{l}$ intervalos não sobrepostos de igual comprimento (l).

Uma vez que o comprimento N não é necessariamente um múltiplo da escala de tempo considerada, l , pode acontecer que a parte final do sinal integrado não esteja em nenhum intervalo e por isso não seja usada. Para não desvalorizar esta parte da série, o mesmo procedimento é repetido a partir do outro extremo da série.

Assim, obtemos $2N_l$ segmentos.

3ºPasso: Calcular a tendência local para cada intervalo, ν , através do ajuste dos mínimos quadrados dos dados. São assim definidas as séries temporais sem tendência de cada intervalo, $Y_\nu(l)$, como a diferença entre a série temporal original e o ajuste

$$Y_\nu(l) = Y(l) - p_\nu(l);$$

onde $p_\nu(l)$ é o polinómio de ajuste em cada ν -ésimo intervalo. Podem ser utilizados, para este ajuste, polinómios lineares, quadráticos, cúbicos, ou de ordens superiores (DFA1, DFA2, DFA3, DFA).

Uma vez que a tendência das séries temporais é retirada através da subtração do polinómio de ajuste da série, este método destacou-se pela particularidade de eliminar as tendências das séries.

4ºPasso: É portanto calculada a variância das séries temporais sem tendência, para cada intervalo

$$F^2(\nu, l) = \frac{1}{l} \sum_{i=1}^l [Y((\nu-1)l + i) - p_\nu(i)]^2 \quad (2.4)$$

para cada segmento $\nu = 1; \dots; N_l$ e

$$F^2(\nu, l) = \frac{1}{l} \sum_{i=1}^l [Y[N - (\nu - N_l)l + i] - p_\nu(i)]^2 \quad (2.5)$$

para $v = N_l + 1; \dots; 2N_l$.

Por fim, é feita a média sobre todos os intervalos e tomamos a raiz quadrada desse valor para obter a função das flutuações DFA

$$F_{DFA}(l) = \frac{1}{2N_l} \sum_{v=1}^{2N_l} f_{DFA}^2(v)^{\# \frac{1}{2}} \quad (2.6)$$

Para diferentes ordens do polinômio de ajuste, n , obtemos diferentes funções de flutuações $F(l)$, que são representadas por $F^{(n)}(l)$. O autor faz a observação de que $F^{(n)}(l)$ irá aumentar com o aumento da escala l dos segmentos. Se a série x_i seguir uma lei de potência de longo alcance correlacionada (*long-range correlated power-law*), as funções de flutuações $F^{(n)}(l)$, para valores grandes de l , seguem uma lei de potência:

$$F^{(n)}(l) \propto l^\beta \quad (2.7)$$

2.4.1 Método Multifractal DFA (MF-DFA)

Em 2002, Kantelhardt e co-autores, [32], propuseram um novo método de análise multifractal baseado no método DFA e ao qual deram o nome de Análise Multifractal de Flutuações Sem Tendência (MF-DFA). Este método multifractal DFA é uma generalização do método do DFA referido anteriormente, consistindo em cinco passos, dos quais os primeiros três passos são idênticos aos do método DFA convencional. O ponto de partida é novamente uma série x_i de comprimento N , com $x_i = 0$ apenas numa fração insignificante da série.

1ºPasso: Determinar o "perfil" da série

$$Y(i) = \sum_{k=1}^i [x_k - \bar{x}]; \quad i = 1; \dots; N \quad (2.8)$$

onde $\bar{x} = \frac{1}{N} \sum_{k=1}^N x_k$.

Como já foi referido anteriormente, a subtração da média, \bar{x} , não é obrigatória, uma vez que será eliminada ao retirar a tendência no 3º Passo.

2ºPasso: Dividir o perfil $Y(i)$ em $N_l = \text{int} \frac{N}{l}$ segmentos não sobrepostos de igual tamanho l .

Tal como anteriormente, uma vez que o comprimento N não é necessariamente um múltiplo da escala de tempo considerada, l , pode acontecer que a parte final do perfil da série não esteja em nenhum intervalo e por isso não seja usada. Para não desvalorizar esta parte da série, o mesmo procedimento é repetido a partir do outro extremo da série. Vamos portanto obter $2N_l$ segmentos/intervalos.

3ºPasso: Calcular a tendência local para cada um dos N_l intervalos, através do ajuste dos mínimos quadrados da série, e determinar a variância

$$F^2(v; l) = \frac{1}{l} \sum_{i=1}^l fY[(v-1)l + i] - p_v(l) g^2 \quad (2.9)$$

para cada segmento $v = 1; \dots; N_l$ e

$$F^2(v; l) = \frac{1}{l} \sum_{i=1}^l \left(Y[N(v - N_l)l + i] - p_v(l) \right)^2 \quad (2.10)$$

para $v = N_l + 1; \dots; 2N_l$. Aqui, $p_v(l)$ é o polinómio de ajuste do segmento v . Os polinómios para o ajuste podem ser lineares, quadráticos, cúbicos ou de ordens superiores (DFA1, DFA2, DFA3, ...). No método (MF-)DFAm [(MF-)DFA de ordem m] a tendência estimada com um polinómio de ordem m é eliminada do perfil da série.

4º Passo: Fazer a média sobre todos os segmentos para obter a função de flutuações de ordem q

$$F_q(l) = \left(\frac{1}{2N_l} \sum_{v=1}^{2N_l} [F^2(v; l)]^{\frac{q}{2}} \right)^{\frac{1}{q}}; \quad (2.11)$$

onde, em geral, a variável q pode tomar qualquer valor real (para $q = 0$, ver o 5º Passo). Para $q = 2$, o procedimento convencional do DFA é recuperado. O objetivo é perceber como é que as funções das flutuações de ordem q , $F_q(l)$, dependem da escala de tempo l para valores diferentes de q . Assim deveremos repetir os passos 2-4 para várias escalas de tempo.

Conforme foi referido atrás, é previsível que $F_q(l)$ irá crescer com o aumento da escala de tempo l . $F_q(l)$ está definido apenas para $l \geq m + 2$.

5º Passo: Determinar o comportamento das flutuações em função da escala através da análise dos gráficos de dispersão de $F_q(l)$ versus l , numa escala log-log, para cada valor de q . Quando a correlação entre as séries segue uma lei de potência de longo alcance, $F_q(l)$ comporta-se, para valores grandes de l , de acordo com a lei de potência

$$F_q(l) \propto l^{h(q)}; \quad (2.12)$$

Para escalas de grande dimensão, $l > \frac{N}{4}$, $F_q(l)$ torna-se estatisticamente pouco fiável porque o número de segmentos N_l usados no procedimento de cálculo da média no 4º Passo torna-se muito pequeno. Por isso, normalmente, excluimos escalas em que $l > \frac{N}{4}$ no procedimento de ajuste para determinar $h(q)$. Também no caso de escalas muito pequenas, $l \leq 10$, podem ocorrer desvios sistemáticos em relação ao comportamento descrito pela equação 2.12.

Para séries temporais estacionárias, $h(2)$ é idêntico ao expoente de Hurst, H , [15]. Por isso, a função $h(q)$ é chamada de expoente de Hurst generalizado.

O método MF-DFA apenas consegue determinar expoentes de Hurst generalizados, $h(q)$, positivos e pode já ser impreciso quando $h(q)$ está próxima de zero.

2.5 Análise de Correlações Cruzadas sem Tendência-DCCA

Em 2008, Boris Podobnik e H. Eugene Stanley, [48], propuseram um novo método para analisar correlações cruzadas entre séries temporais não estacionárias quando essas relações seguem uma lei

de potência. Para mostrarem a utilidade deste novo método, os autores selecionaram alguns exemplos das áreas da física, da psicologia e das finanças. Desde então este é um método bastante utilizado em estudos/análises da correlação de séries temporais, nas quais, frequentemente, está presente a não estacionaridade. Este é portanto um método utilizado para identificar a correlação cruzada entre duas séries temporais não estacionárias, sendo uma generalização do método DFA.

Shen Chenhua, 2015, [7], refere que o facto de o método eliminar as tendências locais assegura que os resultados não são afetados pela tendência (incluindo tendências lineares, quadráticas, e mesmo de ordens superiores e tendências periódicas). Para analisar adequadamente as séries temporais, Horvatic e co-autores, 2011, [27], referem que é essencial retirar as tendências pelo menos por duas razões:

- ao retirarmos a tendência estamos a evitar que se encontre correlação entre as séries quando tal correlação não existe
- se existirem correlações, ao retirarmos a tendência, é revelada uma correlação genuína; no caso de as correlações seguirem uma lei de potência, obteremos um expoente de correlação genuíno.

Assim sendo, este é um método que pode revelar mais informação acerca das correlações do que outros tipos de análise tradicionais, mostrando-se assim útil na aplicação em diversas áreas.

No artigo escrito por Boris Podobnik e H. Eugene Stanley, [48], são consideradas duas séries temporais de longo alcance com correlação cruzada f_{X_i, Y_i} . As duas séries, $f_{Y_i, G}$, têm igual comprimento N . O método DCCA é composto pelos seguintes passos:

1º Passo São construídos dois sinais integrados,

$$X_k = \sum_{i=1}^k x_i \quad (2.13)$$

$$Y_k = \sum_{i=1}^k y_i \quad (2.14)$$

onde $k = 1; \dots; N$. Dividimos ambas as séries em N/n *boxes* (intervalos) sobrepostos, que irão conter $n+1$ valores cada.

2º Passo Para cada *box* que começa em i e termina em $i+n$, é definida a tendência local, \tilde{X}_k e \tilde{Y}_k ($i \leq k \leq i+n$) como sendo a ordenada de um ajuste linear dos mínimos quadrados. Esta tendência local é usada para definir o "detrended walk" (série sem tendência, "resíduos") como a diferença entre a série original e a tendência local.

3º Passo É então calculada a covariância dos resíduos para cada *box*

$$f_{DCCA}^2(n; i) = \frac{1}{n} \sum_{k=i}^{i+n} (\tilde{X}_k - \tilde{X}_i)(\tilde{Y}_k - \tilde{Y}_i) \quad (2.15)$$

4º Passo Por fim é calculada a covariância sem tendência usando todas as N/n *boxes* sobrepostas de tamanho n ,

$$F_{DCCA}^2(n) = \frac{1}{N} \sum_{i=1}^{N/n} f_{DCCA}^2(n; i) \quad (2.16)$$

Quando temos $X_k = Y_k$, a covariância sem tendência $F_{DCCA}^2(n)$ reduz-se à variância sem tendência $F_{DFA}^2(n)$ calculada através do método DFA.

Horvatic e co-autores referem no seu artigo sobre este mesmo método, [27], que na sua forma original o DCCA é baseado na eliminação da tendência local através de ajuste polinomial linear, ou seja, em cada janela, o método subtrai ao sinal o ajuste polinomial linear local por forma a obter um sinal local estacionário. No entanto, tanto o método DFA como o método DCCA na sua forma original, ou seja, com ajustes de polinómios lineares, podem não conseguir retirar todo o tipo de não estacionariedade, em particular as tendências sinusoidais; um ajuste linear é apenas capaz de imitar as tendências existentes em intervalos curtos, ou seja o ajuste não será tão eficiente quando utilizados intervalos de maiores dimensões. No entanto, se a tendência for retirada por um polinómio com uma ordem superior, é possível que sejam retiradas todas as tendências e daí a existência dos métodos DFA2, DFA3, ..., DCCA2, DCCA3, etc.

2.6 Coeficiente de Correlação Cruzada sem Tendência

O coeficiente de correlação cruzada sem tendência, r_{DCCA} , foi proposto por Zebende em 2011, [65], com o objetivo de quantificar o nível de correlação cruzada entre duas séries temporais não estacionárias, tendo como base o DFA e o DCCA. O expoente obtido através do método DCCA pode ser usado para estimar a auto-correlação no caso de a série seguir uma lei de potência, mas não a correlação cruzada entre duas séries.

O método do DFA permite usar a relação entre $F_{DFA}(l)$ (raiz quadrada da função das flutuações) e a escala l para avaliar se a série é caracterizada por uma lei de potência

$$F_{DFA}(l) \propto l^a;$$

Neste contexto, a é o expoente de escala da lei de potência com auto-correlação de longo alcance. Se $a = 0.5$, o sinal é não correlacionado; se $a < 0.5$, o sinal é anti-persistente; e se $a > 0.5$ o sinal é persistente. O método DCCA foi criado com o intuito de investigar a correlação cruzada entre duas séries temporais, no contexto de uma lei de potência e na presença de não estacionariedade, através da média das funções de covariância sem tendência $F_{DCCA}^2(l)$. Caso exista correlação cruzada de longo alcance entre essas duas séries temos

$$F_{DCCA} \propto l^{I_{DCCA}};$$

com $I_{DCCA} = (a_{DFA} + a_{DFA}^0) = 2$, sendo a_{DFA} e a_{DFA}^0 os expoentes de Hurst das duas séries. O expoente I_{DCCA} pode ser útil para detetar e analisar a correlação existente entre as séries, mesmo quando o comportamento destas é afetada por fenómenos como a sazonalidade, [66], mas não nos diz o grau de correlação cruzada.

Posto isto, Zebende decidiu propor então o coeficiente de correlação cruzada DCCA definido pelo quociente entre a função de covariância sem tendência e as funções de variância sem tendência F_{DFA} , i.e.,

$$r_{DCCA} = \frac{F_{DCCA}^2}{F_{DFAx} F_{DFAy}}; \quad (2.17)$$

Temos então um coeficiente adimensional (não tem nenhuma unidade física que o defina) que tem como intervalo de variação $-1 \leq r_{DCCA} \leq 1$, conforme foi provado teoricamente por [47]. Quando $r_{DCCA} = 0$, não há correlação cruzada entre as duas séries. Quando $r_{DCCA} = -1$ estamos perante uma anticorrelação perfeita (a variação de uma série implica a variação em sentido oposto da outra) e quando $r_{DCCA} = 1$ temos uma correlação perfeita (a variação de uma série implica a variação no mesmo sentido da outra).

Este coeficiente já foi usado em análises em diversas áreas da ciência, como no estudo do mercado financeiro, [50], em fenómenos climatológicos, [59], na análise de dados geofísicos, [43], em séries históricas de indicadores de criminalidade e violência, [42].

Segundo Silva Filho e co-autores, 2014, [11], uma das vantagens deste coeficiente em relação a outros coeficientes de correlação cruzada, como o coeficiente de correlação de Pearson, é a possibilidade de mensurar a correlação entre dois sinais em diferentes escalas de tempo, l , e ainda o facto de a análise da correlação entre os sinais ser realizada sem a componente tendência, que tende a mascarar as verdadeiras correlações, [21]. O que o mais o distingue do coeficiente de correlação de Pearson é o facto de ser capaz de analisar séries não lineares, [62].

2.7 Medição da correlação entre duas séries temporais não estacionárias com r_{DCCA}

Kristoufek em 2014, [37], realizou um estudo do coeficiente de correlação proposto por Zebende em 2011, [65], para testar a eficiência do método do DCCA quando utilizado para estimar o coeficiente de correlação cruzada, independentemente da força da não estacionariedade das séries temporais. Inicialmente Kristoufek faz uma breve introdução ao método DCCA e descreve como se calcula o coeficiente DCCA uma vez que é este o alvo em estudo.

Em seguida descreve as ferramentas que irá utilizar para as simulações realizadas no seu estudo. Como o objetivo é testar a habilidade do coeficiente DCCA em medir as correlações entre séries não estacionárias o autor simula uma quantidade considerável de processos fazendo variar o seu nível de correlação e o nível de (não) estacionariedade. Para tal simulou dois processos ARFIMA(0; d ; 0) em que as suas inovações estão correlacionadas

$$x_t = \sum_{n=0}^{\infty} a_n(d_1) e_{t-n} \quad (2.18)$$

$$y_t = \sum_{n=0}^{\infty} a_n(d_2) n_{t-n} \quad (2.19)$$

onde $a_n(d) = \frac{\Gamma(n+d)}{\Gamma(n+1)\Gamma(d)}$, $he_{t,i} = hn_{t,i} = 0$, $e_t^2 = n_t^2 = 1$ e $he_t n_{t,i} = r_{en}$. Para o seu estudo, o autor faz variar os parâmetros $d_1; d_2; r_{en}; l$ e T , onde T representa o comprimento das séries e l a escala utilizada no método DCCA.

Por forma a simplificar, Kristoufek apenas considera os casos em que $d_1 = d_2 = d$ e $r_{en} \in [0.9; 0.9]$ com um passo de 0.1. Utiliza dois comprimentos de séries temporais, $T = 1000$ e $T = 5000$, e quatro níveis diferentes para a escala: $l = \frac{T}{100}; \frac{T}{50}; \frac{T}{10}; \frac{T}{5}$.

Se o coeficiente de correlação estimado for 100% fiável então irá coincidir com r_{en} independentemente dos outros parâmetros, principalmente do parâmetro d . Como já referido no capítulo 2.2, o parâmetro d distingue as séries temporais estacionárias das não estacionárias, sendo $d = 0.5$ o ponto de referência. O autor realizou simulações para duas séries estacionárias ($d=0.1$ e $d=0.4$), duas séries não estacionárias com reversão à média ($d=0.6$ e $d=0.9$) e duas séries não estacionárias (explosivas) ($d=1.1$ e $d=1.4$).

Foi, portanto, realizado um estudo de simulação Monte Carlo com 1000 replicações das combinações dos parâmetros d , r_{en} , l e T de forma a analisar a influência destes parâmetros na performance do $r_{DCCA}(l)$. O autor optou por mostrar os resultados através de gráficos com o intuito de retirar mais facilmente as conclusões e para poder comparar melhor os seus resultados. Criou dois tipos de gráficos: os gráficos com os desvios padrão para os coeficientes de correlação DCCA, para os diferentes parâmetros fracionários de integração, d , e os gráficos com uma linha para a média dos valores estimados, uma linha com o valor r_{en} , e com os intervalos de confiança de 95% para cada valor l .

Através da observação dos gráficos dos coeficientes de correlação DCCA e dos desvios padrão dos coeficientes de correlação DCCA, o autor conclui que

- 1º o coeficiente de correlação DCCA é um estimador não enviesado do coeficiente de correlação verdadeiro, r_{en} , independentemente de todas as configurações de parâmetros;
- 2º As estimativas são mais precisas para valores absolutos mais altos do coeficiente de correlação verdadeiro;
- 3º A precisão (medida pelos desvios padrão) é aproximadamente simétrica em torno da correlação zero;
- 4º As estimativas são mais precisas para escalas menores;
- 5º A precisão das estimativas diminui com o aumento do d ;
- 6º A precisão das estimativas não varia muito com o aumento do comprimento da série temporal. O parâmetro dominante neste caso parece-lhe ser a escala l e a sua relação com a duração da série temporal, ao invés do T sozinho.

Com este estudo e estas observações o autor conclui que o coeficiente de correlação DCCA é capaz de estimar o coeficiente de correlação verdadeiro sem tendência, mesmo para séries fortemente não estacionárias.

Para finalizar o seu estudo, é ainda feita a mesma análise para o coeficiente de correlação de Pearson, tornando-se bastante claro que, em comparação com o coeficiente estimado através do método DCCA, não tem a mesma capacidade. Conclui então que o coeficiente de correlação de Pearson padrão é praticamente inútil para séries temporais não estacionárias.

Uma vez que o autor conseguiu provar que o coeficiente DCCA é capaz de estimar o coeficiente de correlação verdadeiro entre as séries com precisão independentemente da força da (não) estacionariedade das séries em causa, são muitos os artigos que o citam para dar ênfase ao facto de que este é um método fiável na análise das séries reais nas diferentes áreas científicas.

Na área da Física e Astronomia e da Matemática é onde surgem a maioria das citações deste artigo, tais como [13], [3], [6], [63], [30] (na área da Física e Astronomia), [12], [61], [38], [40], [52] (na área da Matemática), sendo que, quando falamos em utilizar um método de correlação cruzada num artigo, obviamente este faz parte da área da Matemática, pois estamos a utilizar um método matemático. Este é ainda um artigo bastante referido quando são feitos estudos das correlações entre mercados financeiros e acontecimentos na área da Economia, como nos estudos seguintes: [55], [56], [18], [9], [22], [19], [10], [17], [39], [1], [5], [20], [16], [36].

No entanto, ao estudar este artigo surgiram algumas dúvidas e principalmente duas questões:

1. o que será que acontece quando $d_1 \neq d_2$?
2. será que quando fazemos os parâmetros AR/MA variar obtemos os mesmo resultados?

Com o objetivo de responder a estas duas questões realizámos o estudo que é apresentado no próximo capítulo.

Capítulo 3

Desempenho do Coeficiente de Correlação Cruzada sem Tendência

Todas as simulações que serão apresentadas neste capítulo foram realizadas no software livre R, [49], versão 4.0.4. O código foi escrito na linguagem R, tendo sido programada uma função para o método DCCA (cálculo do r_{DCCA}), uma função para simulação das séries temporais da classe ARFIMA (usando o *package* `arfi ma`, [60]) e outra para a criação das suas inovações. A programação beneficiou da consulta do código disponibilizado por Javier E. Contreras-Reyes e Byron J. Idrovo-Aguirre (utilizado em [8]), a quem agradecemos.

3.1 Metodologia

O principal objetivo deste estudo será avaliar o desempenho da estimação do coeficiente de correlação cruzada de séries temporais (com várias hipóteses acerca da sua estacionaridade) através do método DCCA, complementando o estudo já feito por Ladislav Kristoufek em 2014, [37]. O nosso estudo foi realizado em duas partes, correspondendo ao estudo de dois casos.

No 1º caso realizámos um estudo de simulação de Monte Carlo, com 1000 replicações de duas séries, fazendo variar a correlação entre as inovações das duas, r_{en} , no intervalo de $[0;0.9]$ com um espaçamento de 0.1, considerando este como o valor real da correlação cruzada entre as duas séries temporais geradas. As séries foram criadas através de processos $ARFIMA(p; d; q)$, com o auxílio da função `arfi ma. sim`, que está presente na biblioteca `arfi ma` do software R. Nesta primeira simulação, para cada uma das séries $i \in \{1, 2\}$ (o índice 1 corresponde aos parâmetros da série x_t e o índice 2 corresponde aos da série y_t), considerámos $d_i \in \{0.1; 0.4; 0.6; 0.9; 1.1; 1.4\}$, $p_x = q_x = p_y = q_y = 0$, as inovações com variância 1 e as escalas do método DCCA analisadas foram $l = \frac{T}{5}; \frac{T}{10}; \frac{T}{50}; \frac{T}{100}$. A análise deste caso foi feita com o intuito de replicar o estudo já feito por Kristoufek, com a diferença de que considerámos também os casos em que $d_1 \neq d_2$. Tal como Kristoufek, considerámos os casos em que o comprimento das séries era de 1000 e de 5000 observações.

O estudo do 2º caso foi dividido em duas partes. No 2º caso: 1ª Parte voltámos a considerar $d_1, d_2 \in \{0.1; 0.4; 0.6; 0.9; 1.1; 1.4\}$, $T \in \{1000; 5000\}$, $l = \frac{T}{5}; \frac{T}{10}; \frac{T}{50}; \frac{T}{100}$ e a variância das inovações igual a 1. Uma diferença menor em relação ao 1º caso foi termos considerado $r_{en} \in \{0.9; 0.6; 0.3; 0; 0.3; 0.6; 0.9\}$. A diferença mais importante, com o objetivo de responder à

segunda questão referida no final do capítulo 2, foi termos criado pares de séries temporais ARFIMA em que uma delas podia ter componentes AR e/ou MA (no máximo, um desfasamento em cada componente), com os parâmetros $f_x \in f0;0.5g$, $q_x \in f0;0.3;0.5;0.7g$ (para a série com componentes AR/MA) e $f_y = q_y = 0$ (para a série sem componentes AR/MA).

Por último, no 2º caso: Parte quisemos estudar o que aconteceria se admitíssemos a possibilidade de haver componentes AR/MA nas duas séries. Portanto, considerámos $f_x; q_x; f_y; q_y \in f 0.9; 0.45; 0; 0.45; 0.9g$. Uma vez que só a variação destes parâmetros já gera muitos resultados, por forma a simplificar e baseando-nos nos resultados obtidos nos casos anteriores, considerámos apenas $T = 1000$, $d_1; d_2 \in f0.25; 0.75; 1.25g$ (ou seja, um caso em que a série é estacionária, um caso em que é não estacionária com reversão à média e outro caso em que é não estacionária explosiva), $r_{en} \in f0;0.3;0.6;0.9g$, a variância das inovações igual a 1 e $I = \frac{T}{5}; \frac{T}{10}; \frac{T}{50}; \frac{T}{100}$. Dentro desta 2ª parte, para clarificar a interpretação dos resultados obtidos, optámos ainda por dar ênfase a mais dois casos em particular: o Caso a) onde considerámos $f_x = q_x = 0$ e $f_y; q_y \in f 0.9; 0.45; 0; 0.45; 0.9g$; o Caso b) onde temos $f_y = q_y = 0$ e $f_x; q_x \in f 0.9; 0.45; 0; 0.45; 0.9g$, com a variância das inovações de ambas as séries igual a um e as escalas novamente a variarem no conjunto $I \in \frac{T}{100}; \frac{T}{50}; \frac{T}{10}; \frac{T}{5}$.

3.2 1º Caso: Séries temporais com diferentes ordens de estacionariedade, ARFIMA(0,d,0)

O primeiro passo foi verificar se realmente o coeficiente de correlação estimado pelo método do DCCA corresponde ao coeficiente de correlação entre as inovações dos processos, $r_{DCCA} = r_{en}$. Uma vez que Kristoufek, em 2014, [37], já tinha concluído que a precisão é simétrica em torno da correlação zero e que o processo de simulação é muito demorado, considerámos apenas as correlações positivas, $r_{en} \in [0;0.9]$ com um espaçamento de 0.1. Para além das variações dos parâmetros que o autor já tinha considerado quisemos verificar o que aconteceria nos casos em que $d_1 \neq d_2$. Uma vez que desta simulação resultaram 72 gráficos, apresentamos nas Figuras 3.1 e 3.2 apenas os mais representativos das conclusões finais, disponibilizando os restantes no apêndice.

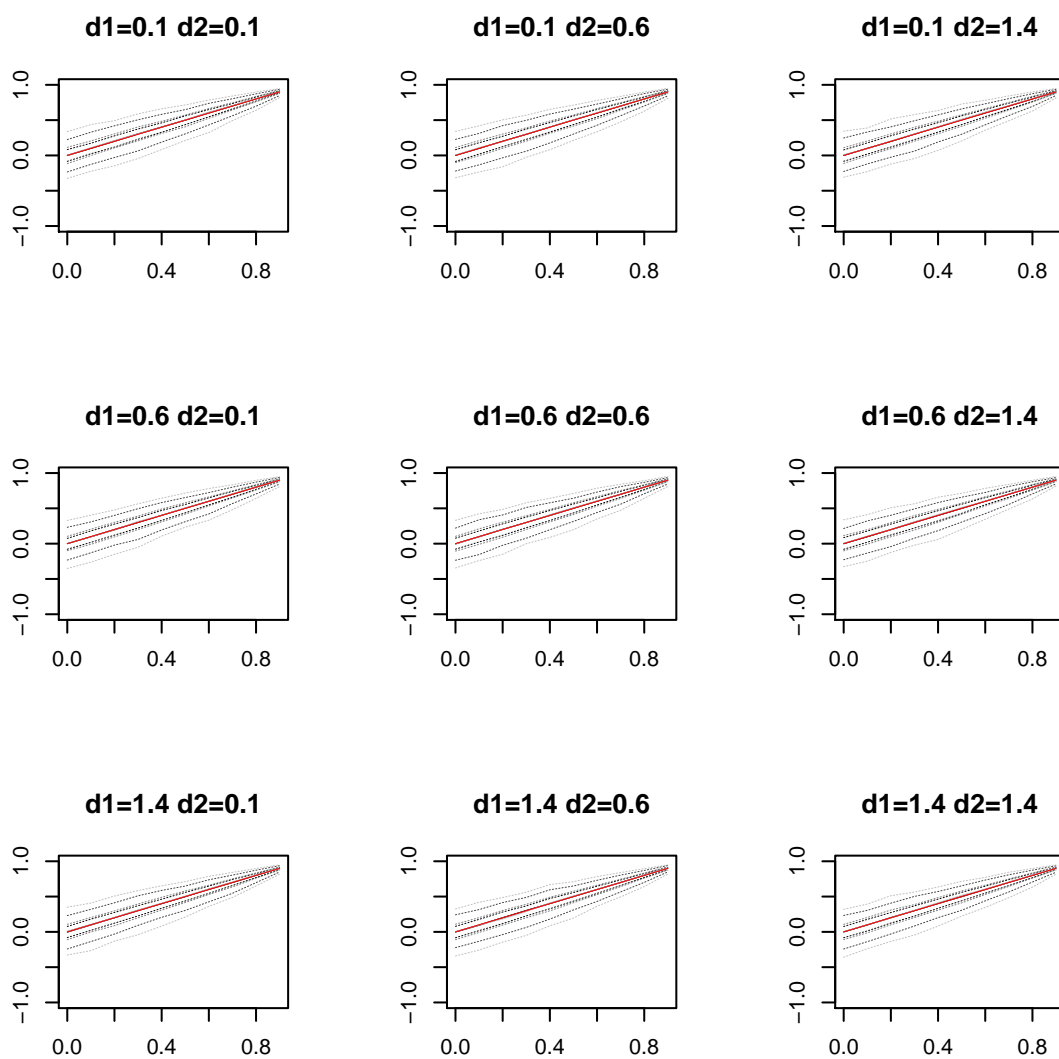


Fig. 3.1 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$.

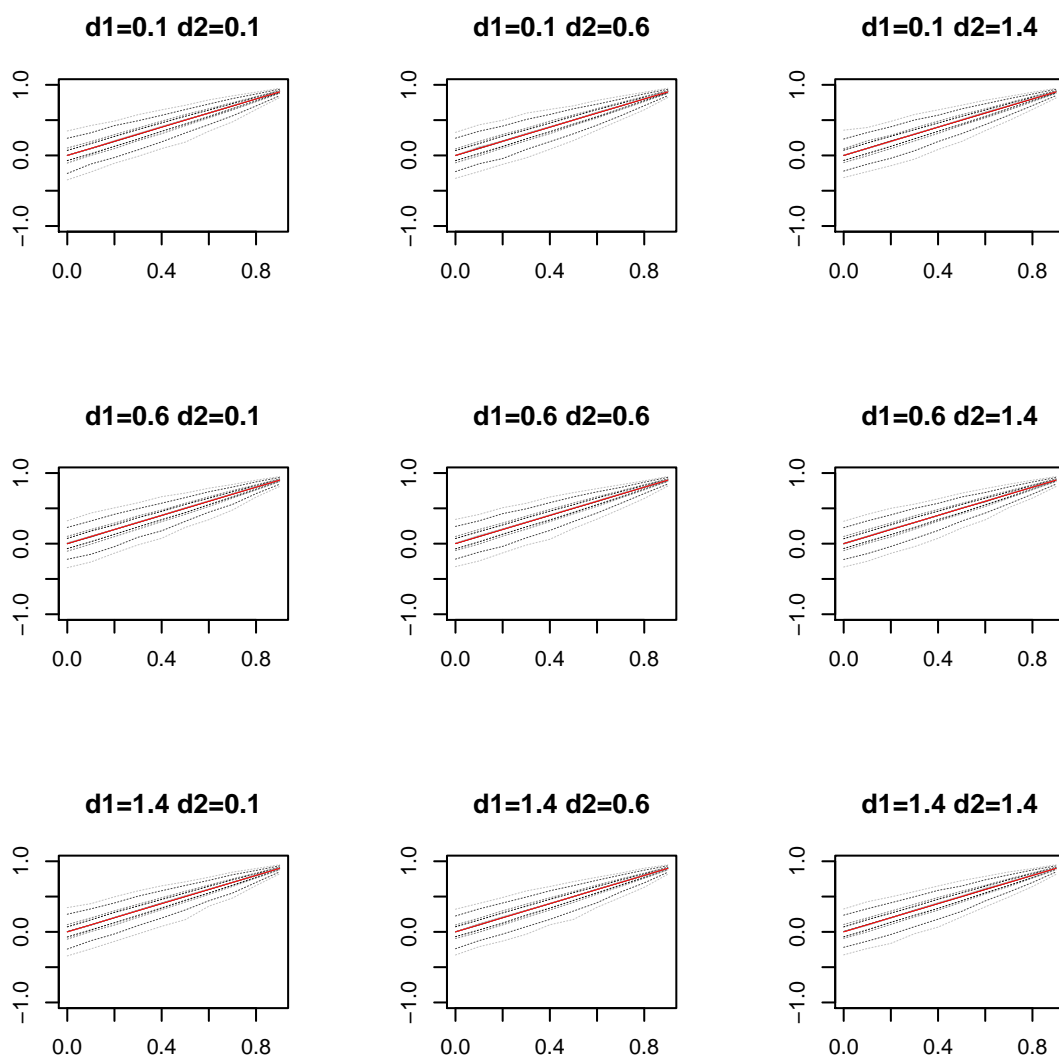


Fig. 3.2 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$.

Para ambas as figuras temos as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente (maioritariamente sobrepostas às linhas vermelhas). As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Estes resultados confirmam o que já tinha sido concluído por *Kristoufek Ladislav*, ou seja,

- 1º o coeficiente de correlação DCCA não tem viés relativamente ao coeficiente de correlação verdadeiro, r_{en} , independentemente de todas as configurações de parâmetros, incluindo de quando $d_1 \neq d_2$;
- 2º As estimativas são mais precisas para valores absolutos mais altos do coeficiente de correlação verdadeiro;
- 3º As estimativas são mais precisas para escalas menores;
- 4º A precisão das estimativas não varia muito com o aumento do comprimento da série temporal. Não existem diferenças visíveis da Figura 3.1 para a Figura 3.2.

Uma vez respondida a 1ª questão posto no final do capítulo anterior, ou seja, foi confirmada a invariância nas estimativas quando considerados os casos em que $d_1 \neq d_2$, seguimos para o 2º Caso.

3.3 2º Caso: Variação dos parâmetros AR/MA, f e q para a série x_t

Tentando responder à nossa segunda questão do final do capítulo 2, fizemos simulações em que os parâmetros autorregressivo e média móvel das séries temporais variam.

3.3.1 1ªParte: $f_x \in \{0, 0.5\}$, $q_x \in \{0, 0.3, 0.5, 0.7\}$; $f_y = q_y = 0$

Nesta 1ªParte optámos por fazer variar os parâmetros apenas para uma das séries, ou seja, gerámos uma serie ARFIMA com diferentes ordens de (não) estacionariedade e diferentes valores dos coeficientes AR/MA e outra série ARFIMA em que fazemos variar a ordem de (não) estacionariedade mas $f = q = 0$. Neste caso já iremos considerar valores negativos para r_{en} para obter um estudo mais completo, uma vez que é uma variação que ainda não foi estudada, pelo menos que seja do nosso conhecimento.

Vamos portanto considerar a seguinte gama de valores para os parâmetros: $r_{en} \in \{0.9, 0.6, 0.3, 0, 0.3, 0.6, 0.9\}$, $T \in \{1000, 5000\}$, $d_1, d_2 \in \{0.1, 0.4, 0.6, 0.9, 1.1, 1.4\}$, $l \in \{\frac{T}{100}, \frac{T}{50}, \frac{T}{10}, \frac{T}{5}\}$, $f_x \in \{0, 0.5\}$, $q_x \in \{0, 0.3, 0.5, 0.7\}$. Uma vez que se pode simplificar o modelo quando há factores comuns a ambas as componentes AR e MA, eliminamos o caso em que temos $f_x = q_x = 0.5$.

Para ser mais fácil analisar os resultados, construímos gráficos semelhantes aos apresentados no texto de Kristoufek, [37]. Tal como na Sub-Secção anterior, uma vez que são bastantes os gráficos produzidos pelas simulações, seleccionámos apenas alguns que representam as conclusões retiradas, e que estão nas Figuras 3.3 e 3.4, e apresentaremos os restantes gráficos no Apêndice.

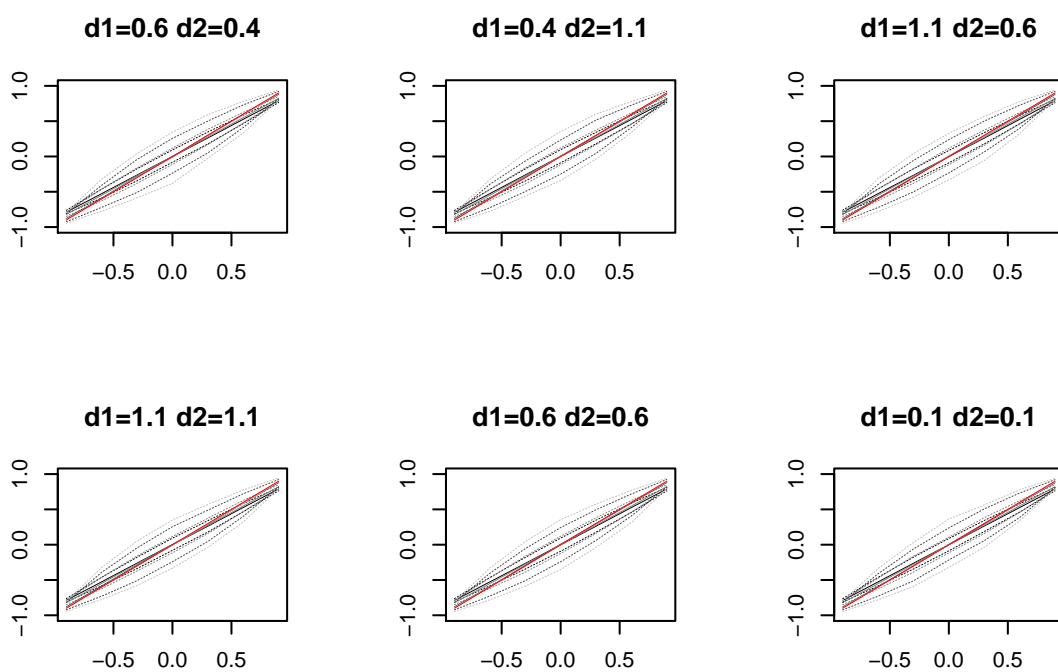


Fig. 3.3 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = f_y = q_y = 0$.

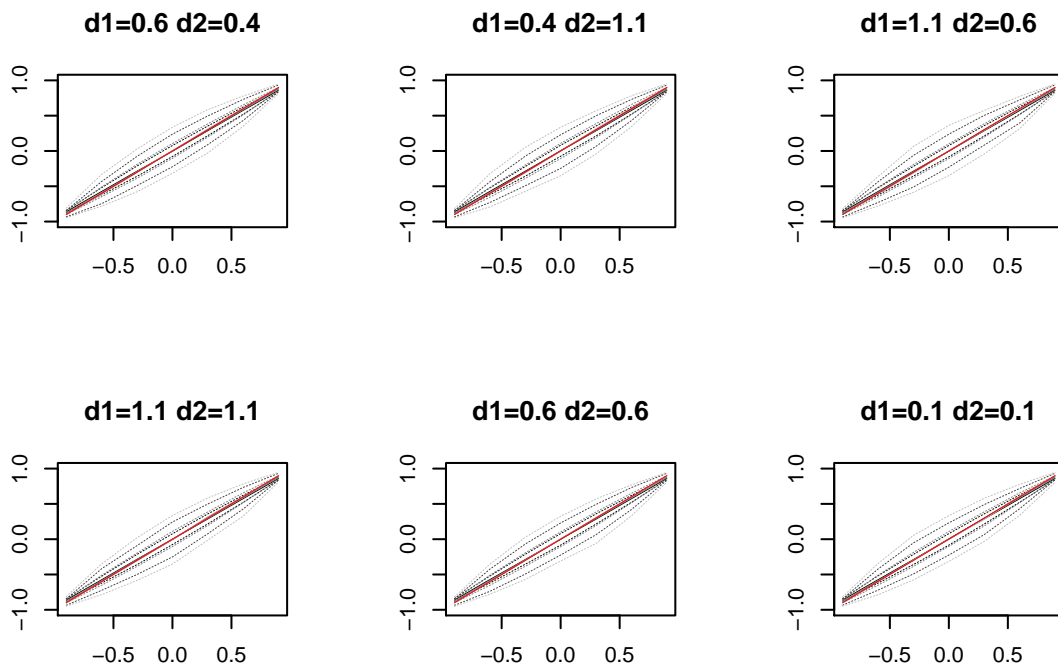


Fig. 3.4 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $\hat{f}_x = 0.5; q_x = \hat{f}_y = q_y = 0$.

Para ambas as Figuras anteriores, 3.3 e 3.4, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente (maioritariamente sobrepostas às linhas vermelhas). As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{50}$, o tom a seguir faz refere-se a $l = \frac{T}{5}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Das Figuras 3.3 e 3.4, que contém os gráficos da simulação em que temos uma série temporal gerada pelo modelo ARFIMA(1, d_1 ,0) com $\hat{f}_x = 0.5$ e outra pelo modelo ARFIMA(0, d_2 ,0), podemos retirar as seguintes conclusões:

- 1º Existe um enviesamento no estimador quando nos afastamos da correlação zero;
- 2º Parece que o enviesamento maior acontece para as escalas que têm também um intervalo de confiança mais preciso, $l = \frac{T}{100}$ e $l = \frac{T}{50}$, ou seja, para escalas menores este enviesamento é mais visível mas no entanto os seus intervalos de confiança têm uma amplitude menor;
- 3º Neste caso quando aumentamos o parâmetro T , parece que o enviesamento tende a ser menor. Ou seja, ao contrário do que acontece no Sub-Capítulo 3.3.1 parece existir uma diferença no comportamento do estimador quando aumentamos o comprimento na série.

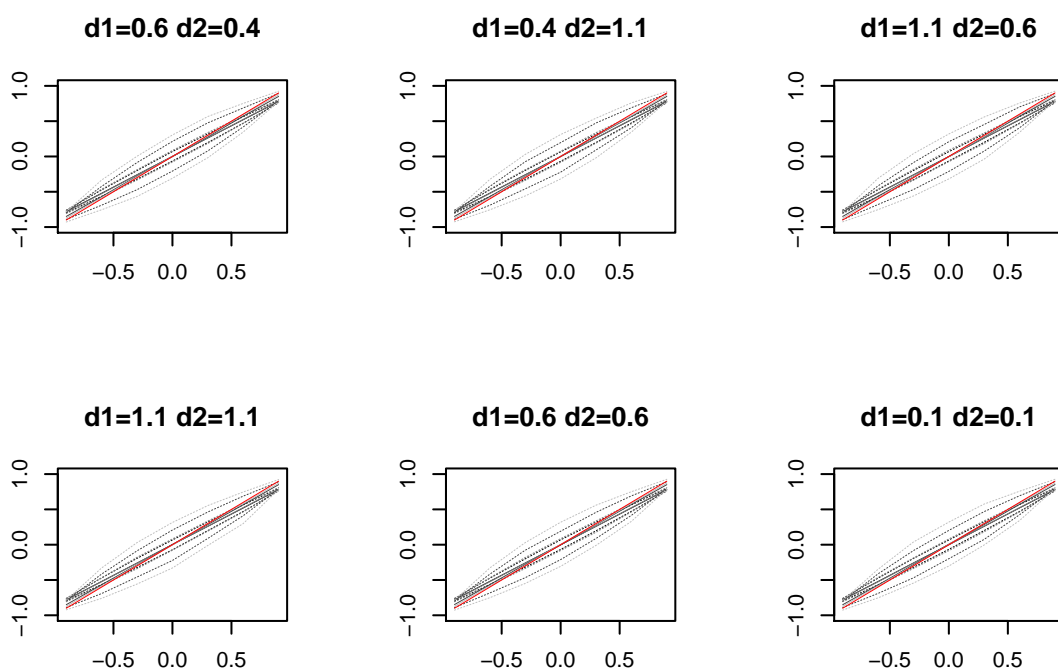


Fig. 3.5 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $q_x = 0.5; \hat{f}_x = \hat{f}_y = q_y = 0$.

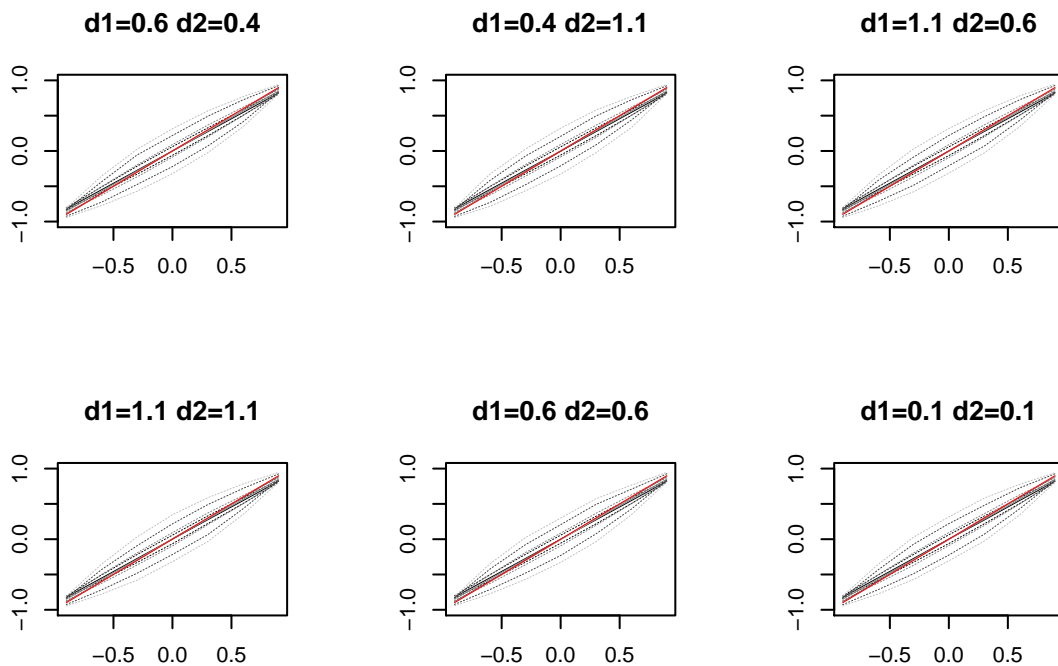


Fig. 3.6 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $q_x = 0.5$; $f_x = f_y = q_y = 0$.

Para ambas as Figuras 3.5 e 3.6, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente (maioritariamente sobrepostas às linhas vermelhas). As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Da observação das Figuras 3.5 e 3.6, que contém os gráficos da simulação em que temos uma série temporal gerada pelo modelo ARFIMA(0, d_1 , 1) com $q_x = 0.5$ e outra pelo modelo ARFIMA(0, d_2 , 0), podemos retirar as seguintes conclusões:

- 1º Existe novamente um enviesamento no estimador que vai aumentando à medida que nos afastamos da correlação zero e parece que o enviesamento maior acontece, novamente, para as escalas que têm um intervalo de confiança mais preciso, $l = \frac{T}{100}$ e $l = \frac{T}{50}$, ou seja para escalas menores;
- 2º Olhando para os intervalos de confiança, as estimativas parecem ser mais precisas para valores absolutos mais altos do coeficiente de correlação verdadeiro, caso em que o intervalo de confiança tem uma amplitude menor;
- 3º Tal como no caso anterior quando aumentamos o parâmetro T , parece que o enviesamento diminui mas nada muito significativo. Mas neste caso o enviesamento é visível para ambos os comprimentos de série.

As conclusões são bastante semelhantes às retiradas da análise das Figuras 3.3 e 3.4, tirando o facto de que neste caso quando aumentado o comprimento da série continua a ser visível o enviesamento para a escala menor, ou seja, para a linha mais escura que corresponde a $l = \frac{T}{100}$.

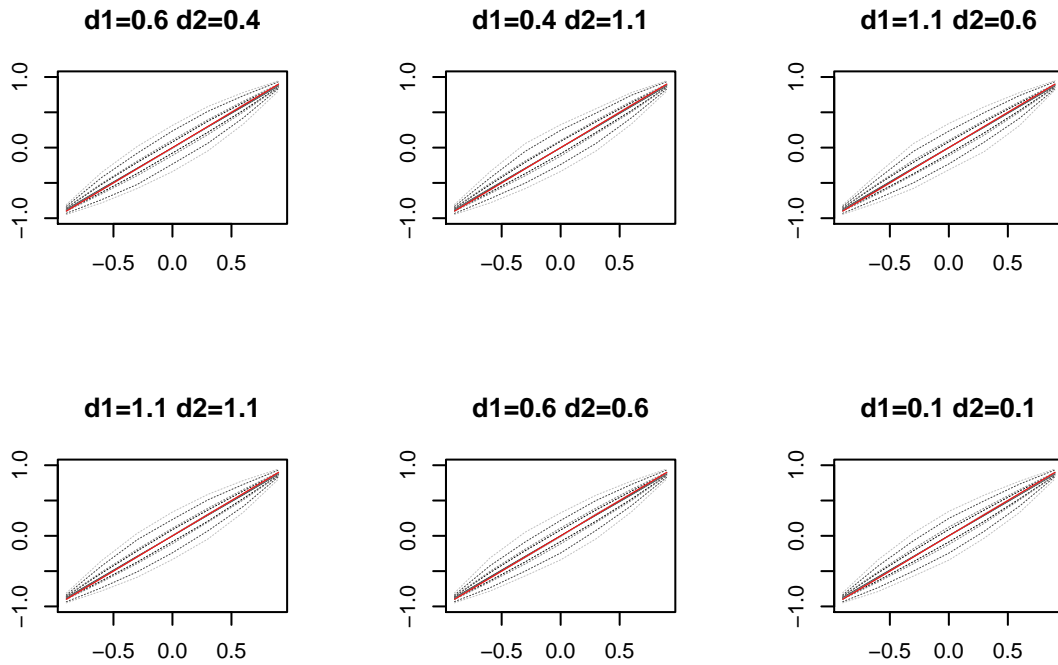


Fig. 3.7 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 1000$. Os parâmetros AR/MA são $f_x = 0.5; q_x = 0.3; f_y = q_y = 0$.

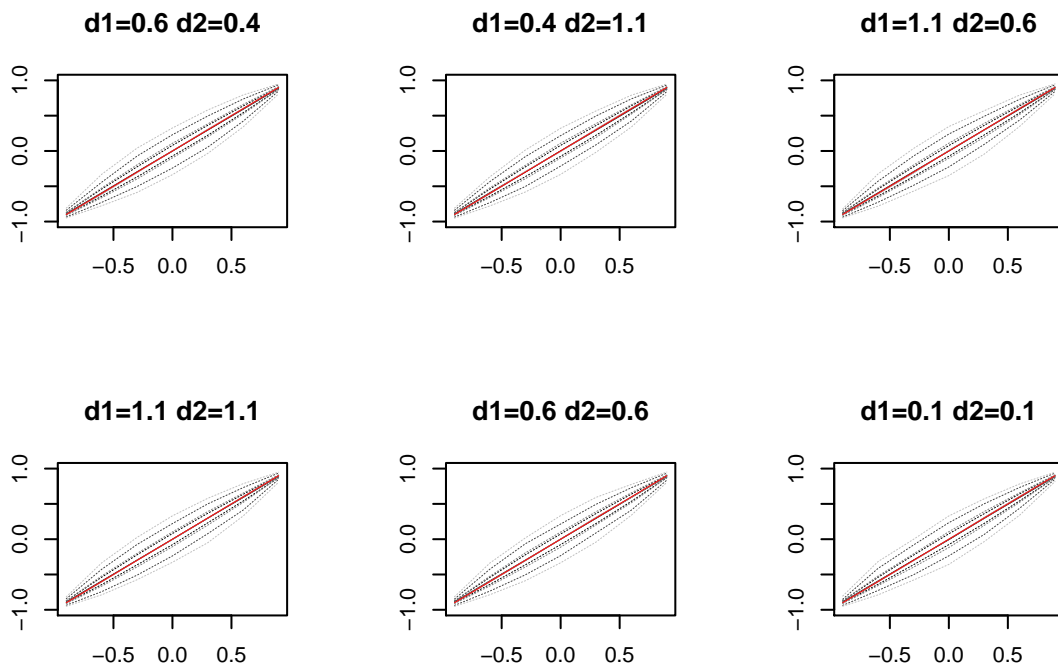


Fig. 3.8 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros de integração, d . Resultados para as séries com comprimento $T = 5000$. Os parâmetros AR/MA são $\bar{f}_x = 0.5; q_x = 0.3; \bar{f}_y = q_y = 0$.

Para ambas as Figuras 3.7 e 3.8, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente (maioritariamente sobrepostas às linhas vermelhas). As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Da observação das Figuras, 3.7 e 3.8, que contém os gráficos da simulação em que temos uma série temporal gerada pelo modelo ARFIMA(1, d_1 , 1) com $\bar{f}_x = 0.5$ e $q_x = 0.3$, e outra pelo modelo ARFIMA(0, d_2 , 0), podemos retirar as seguintes conclusões:

- 1º Neste caso destacamos duas escalas que têm um intervalo de confiança mais preciso, $l = \frac{T}{100}$ e $l = \frac{T}{50}$, ou seja, as escalas menores;
- 2º O coeficiente de correlação DCCA neste caso parece ser cêntrico para o coeficiente de correlação verdadeiro, r_{en} , independentemente da configuração dos parâmetros T , l , d_1 e d_2 , inclusive não parece existir mudança quando utilizamos séries com diferentes níveis de (não) estacionariedade;
- 3º Olhando para os intervalos de confiança, as estimativas parecem ser mais precisas para valores absolutos mais altos do coeficiente de correlação verdadeiro, caso em que o intervalo de confiança tem uma amplitude menor;

Quando comparamos todas as Figuras, 3.3, 3.4, 3.5, 3.6, 3.7 e 3.8, podemos retirar as seguintes conclusões:

- 1º O enviesamento no coeficiente de correlação DCCA é maior quando são utilizadas as séries de modelo ARFIMA(1, d_1 , 0) com $f_x = 0.5$ e ARFIMA(0, d_1 , 1) com $q_x = 0.5$ (ver Figuras 3.3 e 3.5 ou 3.4 e 3.6), pois quando observamos a Figura 3.7 ou 3.8, não é facilmente visível a existência do enviesamento, ou seja, é praticamente nulo. Enquanto nas duas primeiras figuras é algo um pouco mais visível, principalmente para os casos em que $T = 1000$;
- 2º Nos três casos o enviesamento, quando existe, vai aumentando à medida que nos afastamos da correlação zero. No último caso apenas existe um enviesamento, que embora insignificativo (só é visível fazendo um zoom às Figuras), apenas acontece para as correlações bastante próximas de um,
- 3º Embora, que para as duas primeiras Figuras deste Sub-Capítulo (3.3 e 3.4) obtenhamos resultados ligeiramente diferentes quando temos $T = 1000$ e $T = 5000$ não podemos concluir que essa diferença tenha uma relação direta apenas com a diferença do comprimento das séries uma vez que nos outros dois casos o mesmo já não se verifica. Assim sendo, concluímos que fazer o estudo para uma série com comprimento $T = 1000$ ou $T = 5000$ iremos chegar aos mesmos resultados e às mesmas conclusões.

Parece que, dependendo do tipo de modelo de série temporal com que estamos a trabalhar, o método DCCA poderá não ser eficaz para a estimação da correlação entre as séries temporais em causa.

Embora o enviesamento não seja muito significativo será interessante responder à questão: para que séries é que este método poderá não ser fiável e para que casos é que o método é realmente adequado? De acordo com os gráficos que vimos, parece haver uma influência da diferença entre os valores dos coeficientes autorregressivo e média móvel, uma vez que obtemos melhores resultados quando $jf_x - q_xj = j0.5 - 0.3j = 0.2$ do que quando $jf_x - q_xj = j0.5 - 0j = 0.5 = j0 - 0.5j$.

No entanto a única coisa que poderemos concluir através do estudo feito até agora, e que ficou bastante claro com estes gráficos, é que as escalas menores, $l = \frac{T}{100}$ e $l = \frac{T}{50}$, têm um intervalo de confiança mais preciso, mas é nessas escalas que é mais visível um enviesamento do coeficiente de correlação DCCA.

Mas será que é mesmo esta diferença que determina o enviesamento no estimador? Para tentar responder a esta questão iremos em seguida apresentar um estudo em que, para além de variarmos os parâmetros da série x_t iremos também variar os parâmetros da série y_t , com $f_x; f_y; q_x; q_y \in f \in 0.9; 0.45; 0; 0.45; 0.9g$.

3.3.2 2ª Parte: $f_x; q_x; f_y; q_y \in f \in 0.9; 0.45; 0; 0.45; 0.9g$

Com o objetivo de tentar perceber a razão do enviesamento do estimador, realizámos 1000 simulações de séries temporais geradas por um modelo ARFIMA($p; d; q$) e os seus coeficientes AR e MA. Este é um estudo em que há inúmeras combinações possíveis. Se continuássemos com as mesmas gamas de valores para os parâmetros que usámos anteriormente, o custo computacional seria bastante pesado e desse estudo resultaria uma enorme quantidade de resultados.

Dadas as limitações de tempo e com base nos estudos realizados anteriormente, iremos considerar apenas um dos comprimentos das séries, $T = 1000$. Uma vez que já foi concluído que existe simetria em relação à correlação zero, iremos considerar $r_{en} \in \{0; 0.3; 0.6; 0.9\}$. Considerámos os mesmos valores para a escala, $I \in \{\frac{T}{100}; \frac{T}{50}; \frac{T}{10}; \frac{T}{5}\}$, e para a variância das inovações (igual a um). Por fim, fazemos $d_1; d_2 \in \{0.25; 0.75; 1.25\}$.

Optámos novamente por basear a apresentação dos resultados em gráficos nos quais observamos as médias dos valores estimados e os intervalos de confiança de 95%.

Uma vez que, apesar de todas as simplificações nos intervalos de variação dos parâmetros, a análise ainda produziu muitos gráficos, seleccionámos apenas um conjunto deles que nos parece mais relevante para ilustrar as conclusões mais importantes, disponibilizando no apêndice as figuras com os gráficos adicionais. Para a seleção dos gráficos, primeiro foi feita uma análise apenas para as séries em que $f_y = 0.9$ e $q_y = 0.9$, procurando ver se existia alguma diferença significativa no comportamento do estimador quando se variavam os parâmetros d_1 e d_2 , como demonstrado nas Figuras 3.9, 3.10 e 3.11.

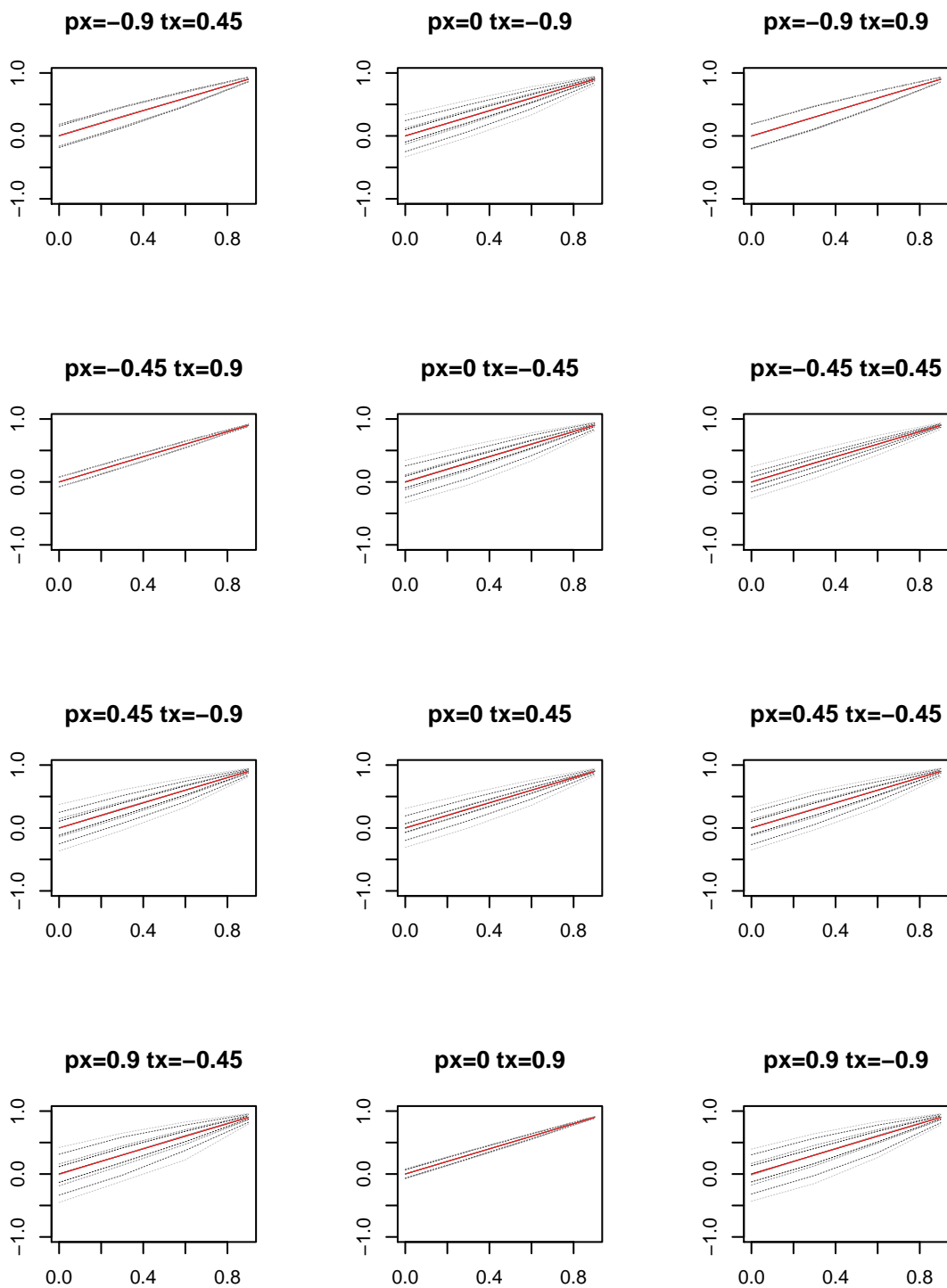


Fig. 3.9 $d_1 = 0.25; d_2 = 0.75$
 p_x corresponde a f_x e t_x corresponde a q_x .

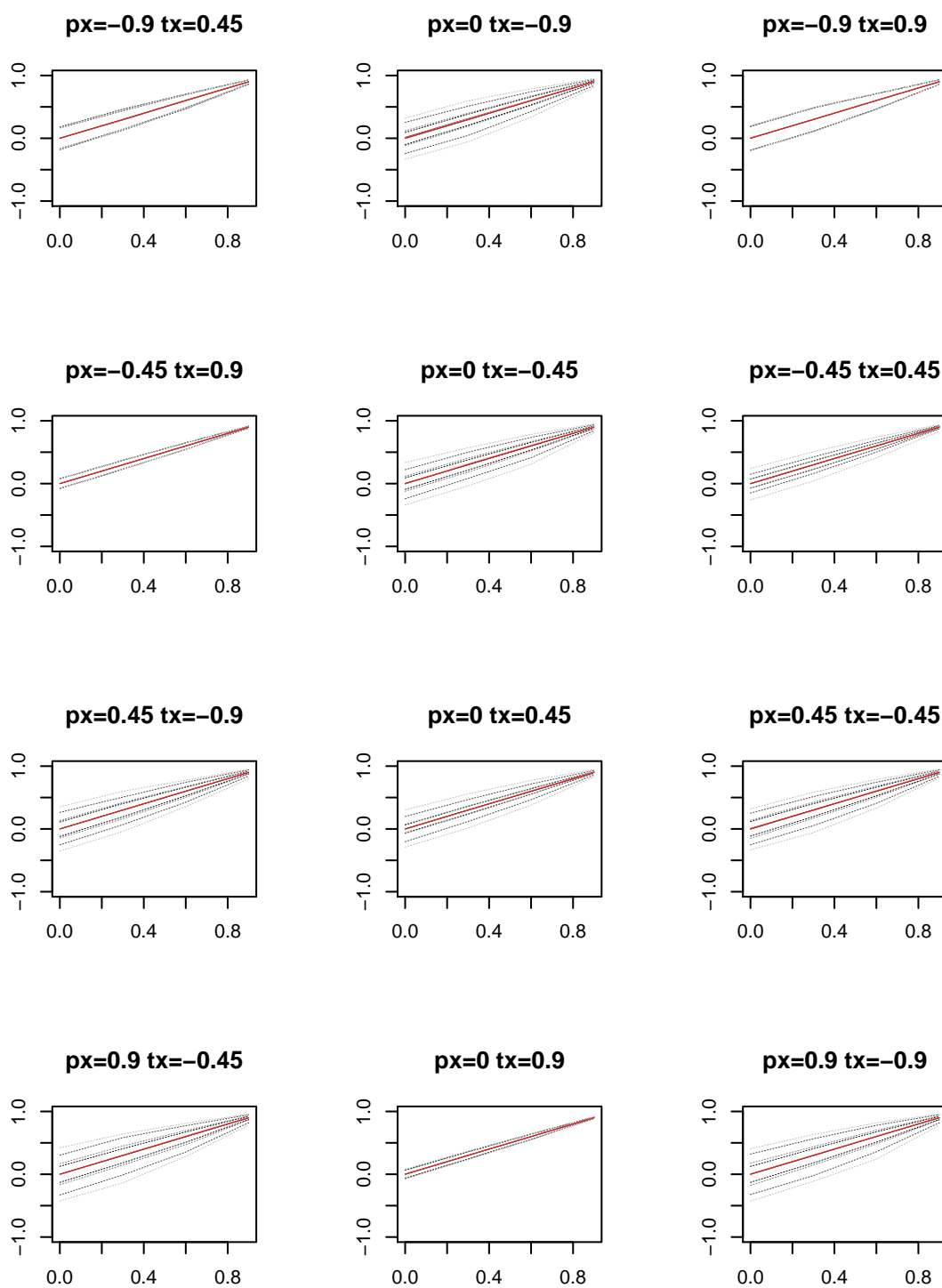


Fig. 3.10 $d_1 = 0.25; d_2 = 1.25$
 px corresponde a f_x e tx corresponde a q_x .

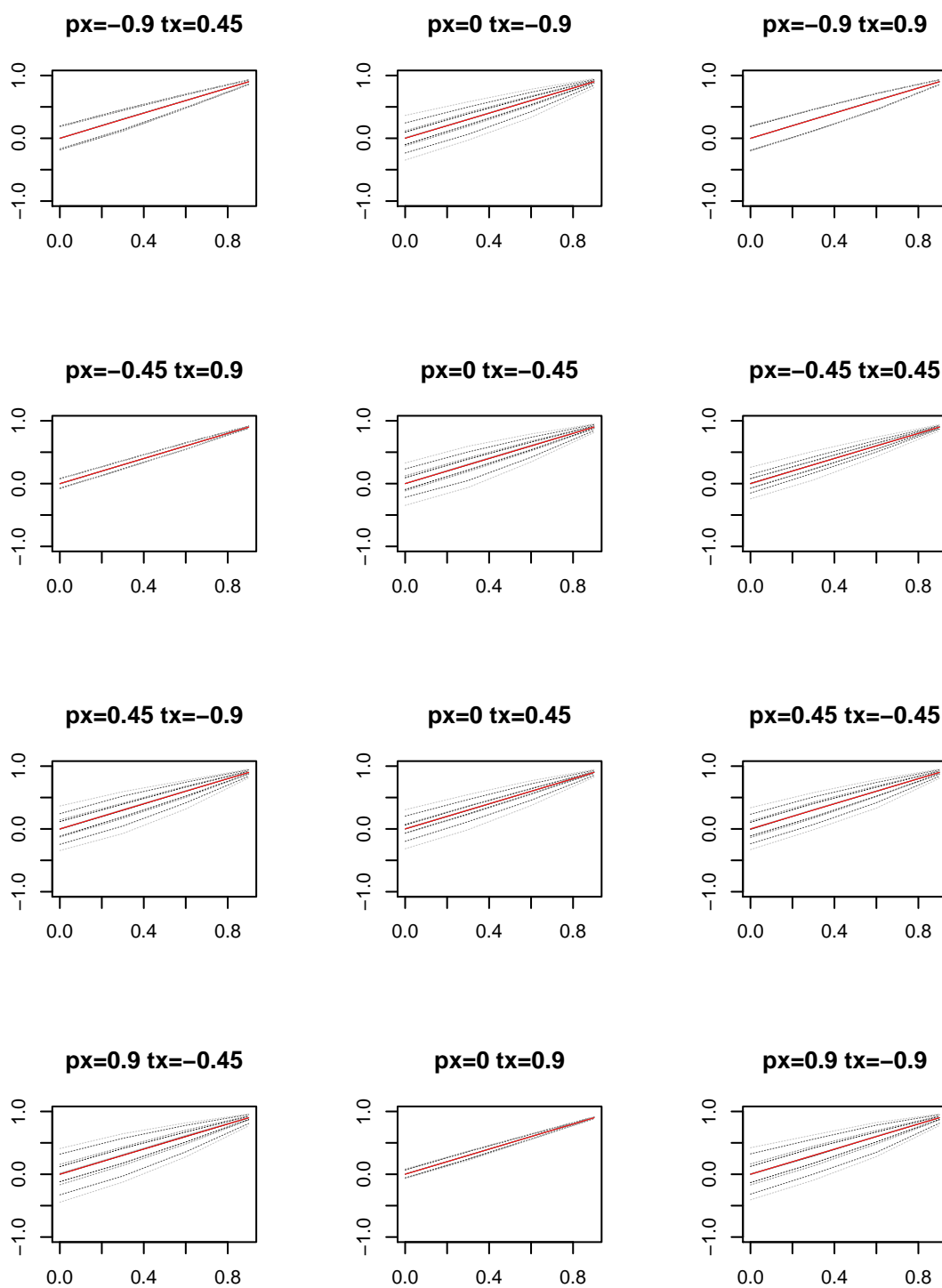


Fig. 3.11 $d_1 = 0.75; d_2 = 1.25$
 px corresponde a f_x e tx corresponde a q_x .

Para as três Figuras anteriores, 3.9, 3.10 e 3.11, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente. As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

É de lembrar que já tinha sido concluído na Sub-Secção 3.3.1 que utilizar $d_1 = d_2$ ou $d_1 \neq d_2$ não afeta o desempenho do coeficiente DCCA e por isso foram considerados apenas os três casos anteriores como suficientes para fazer a análise.

Uma vez que não é visível nenhuma diferença no comportamento dos gráficos da Figura 3.9 para a Figura 3.10 e para a Figura 3.11, iremos apresentar os gráficos apenas para um dos casos. Optámos por escolher a combinação $(d_1; d_2) = (0.75; 1.25)$, ou seja, a combinação de uma série não estacionária com reversão à média com uma série não estacionária explosiva. Isto porque o objetivo é avaliar o desempenho do método para séries não estacionárias. A escolha das combinações $(f_x; q_x)$ e $(f_y; q_y)$ é feita com o objetivo de demonstrar da melhor forma possível as conclusões retiradas.

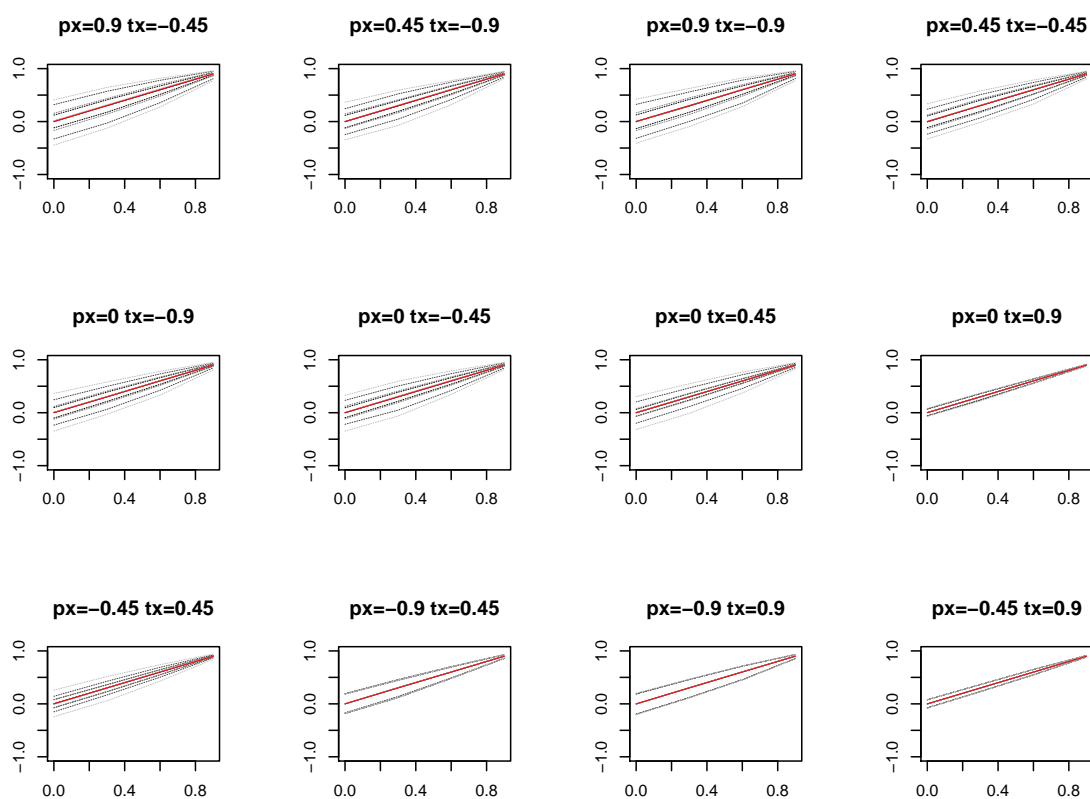


Fig. 3.12 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. px corresponde a f_x e tx corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série y_t $f_y = 0.9$ e $q_y = -0.9$.

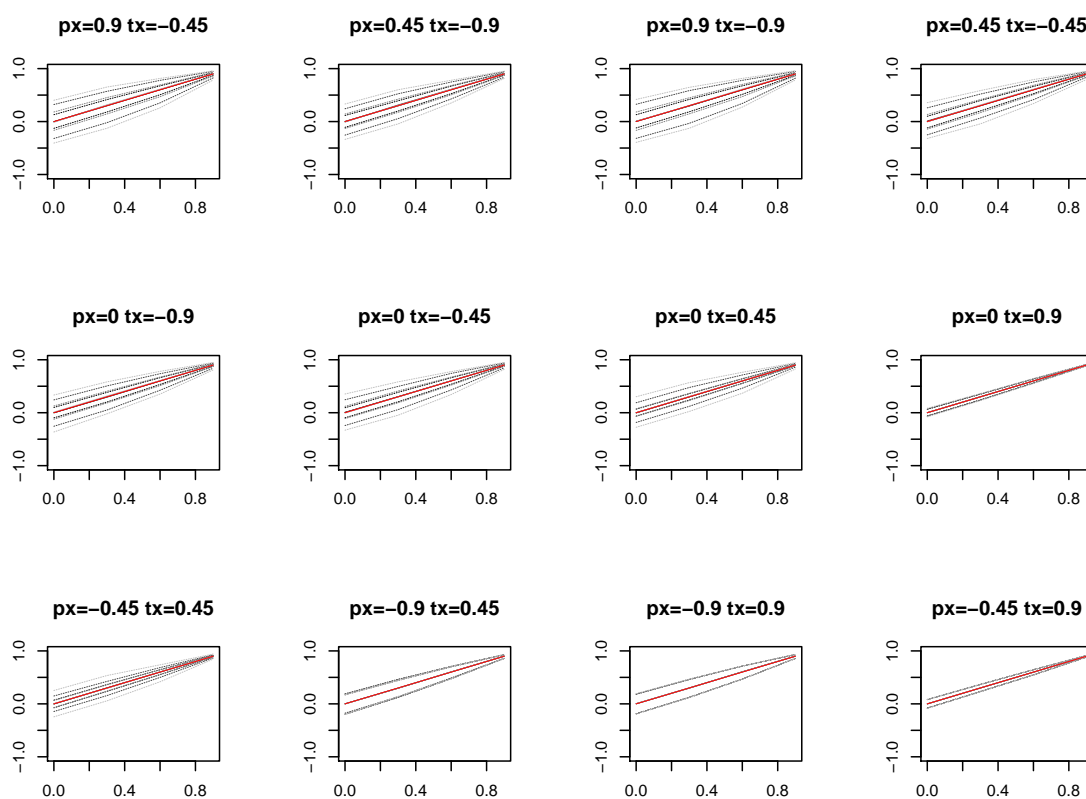


Fig. 3.13 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. px corresponde a f_x e tx corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série $f_y = 0.9$ e $q_y = 0.9$.

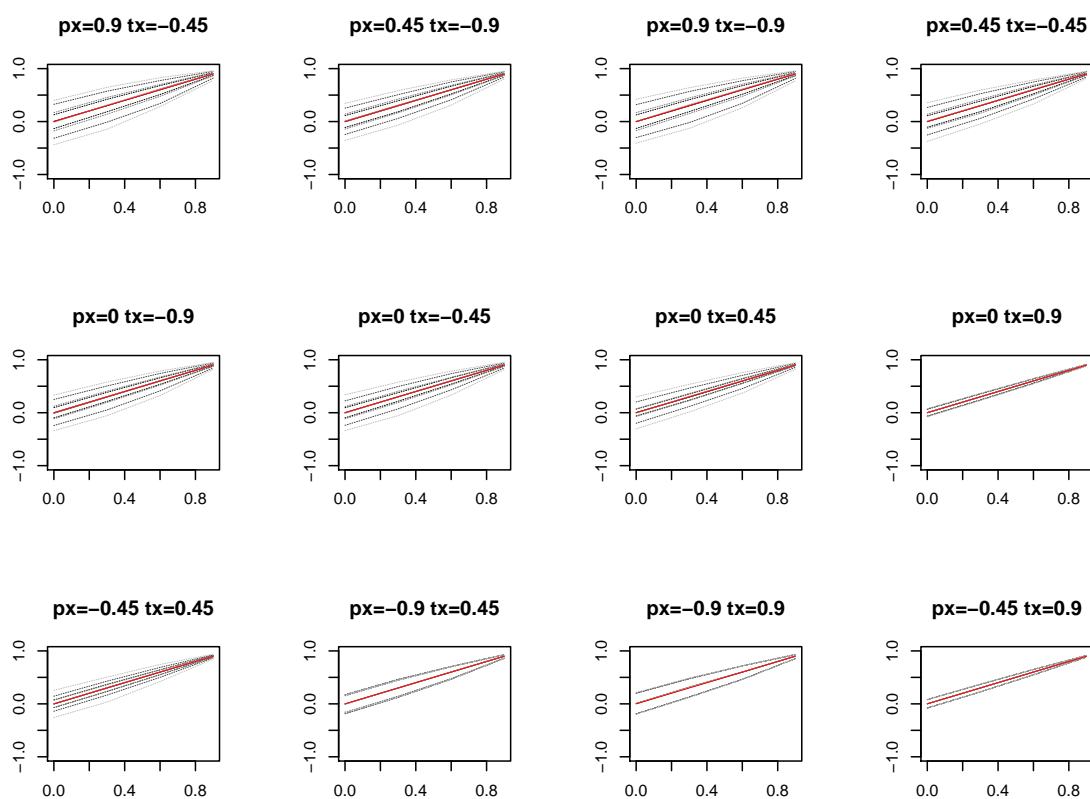


Fig. 3.14 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e t_x corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série y_t $f_y = 0$ e $q_y = 0.9$.

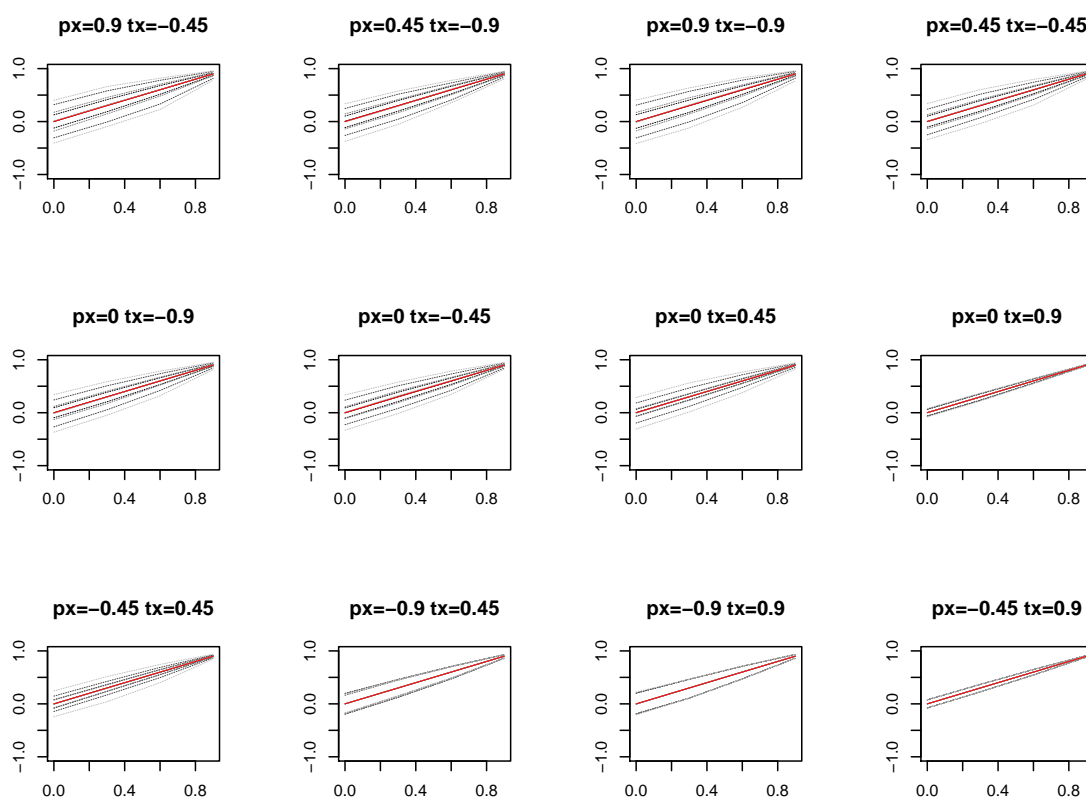


Fig. 3.15 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. px corresponde a f_x e tx corresponde a q_x . Resultados para as séries com parâmetros de integração $d_1 = 0.75$ e $d_2 = 1.25$ e parâmetros autorregressivo e média móvel da série y_t $f_y = 0$ e $q_y = 0.9$.

Para as quatro Figuras anteriores, 3.12, 3.13, 3.14 e 3.15, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente. As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Comparando as quatro Figuras, 3.12, 3.13, 3.14 e 3.15, podemos concluir que não existem grandes diferenças entre elas, ou seja, para cada par $(f_y; q_y)$ os pares $(f_x; q_x)$ têm um comportamento bastante semelhante. Existem apenas diferenças muito pouco significativas na largura dos intervalos de confiança. Posto isto, as conclusões serão as mesmas para estas quatro Figuras, que são um conjunto exemplificativo dos gráficos que resultaram desta simulação:

- 1º Ao observar os gráficos destaca-se logo a variação na largura dos intervalos de confiança consoante o par $(f_x; q_x)$. Podemos até dividi-los em dois conjuntos, em que as duas colunas mais à esquerda contêm os gráficos com os intervalos de confiança mais amplos, e as duas colunas mais à direita contêm os gráficos com os intervalos de confiança mais estreitos;
- 2º Parece ainda existir uma relação entre a largura do intervalo de confiança e a evolução do sinal do parâmetro f_x , sendo mais largo para valores positivos do parâmetro e mais justo para valores negativos. Ou seja, de cima para baixo existe um estreitamento dos intervalos de confiança;
- 3º De uma forma geral, para as duas menores escalas, $l = \frac{T}{100}$ e $l = \frac{T}{50}$, as estimativas parecem ser mais precisas, como já concluído anteriormente. Contudo, existem casos em que as linhas das quatro escalas estão praticamente sobrepostas umas às outras, como nos casos dos pares $(f_x; q_x) = f(0.45; 0.9); (0; 0.9); (0.9; 0.9); (0.9; 0.45)g$. Isto leva a uma indiferença quanto ao valor que é escolhido para escala nos casos referidos anteriormente.
- 4º Em nenhuma das figuras existe um gráfico em que seja visível um enviesamento do coeficiente de correlação DCCA. Ou seja, as linhas contínuas estão maioritariamente sobrepostas à linha vermelha que representa o valor verdadeiro da correlação entre as duas séries, r_{en} ;
- 5º Continua a existir uma maior precisão para as correlações mais próximas de um.

A conclusão mais importante a que se chega desta análise, em específico, é que não existe nenhum enviesamento significativo no estimador do coeficiente correlação do método DCCA, mas existem alguns modelos de séries temporais em que teremos uma estimativa mais precisa do que noutros casos possíveis. Porém, não será necessário modelar as séries antes de aplicar o método DCCA, pois este apresenta-nos um estimador para a correlação das inovações das séries bastante fiável independentemente da (não) estacionariedade e do tipo de modelo das séries temporais.

Uma vez que estes resultados não vão ao encontro das conclusões da Sub-Secção 3.3.1, iremos dar ênfase a dois casos em particular nos quais os parâmetros T , d_1 , d_2 e r tomam os seguintes valores: $T = 1000$, $d_1 = 0.75$, $d_2 = 1.25$, $r_{en} \in \{0.3; 0.6; 0.9\}g$. No 1º Caso, vamos chamar-lhe **Caso a)**, considerámos $f_x = q_x = 0$ e $f_y; q_y \in \{0.9; 0.45; 0; 0.45; 0.9\}g$. No 2º Caso, que será o **Caso b)**, fizemos $f_y = q_y = 0$ e $f_x; q_x \in \{0.9; 0.45; 0; 0.45; 0.9\}g$. Em todos os casos, a variância das inovações de ambas as séries é igual a um e as escalas l são $\frac{T}{100}; \frac{T}{50}; \frac{T}{10}; \frac{T}{5}$.

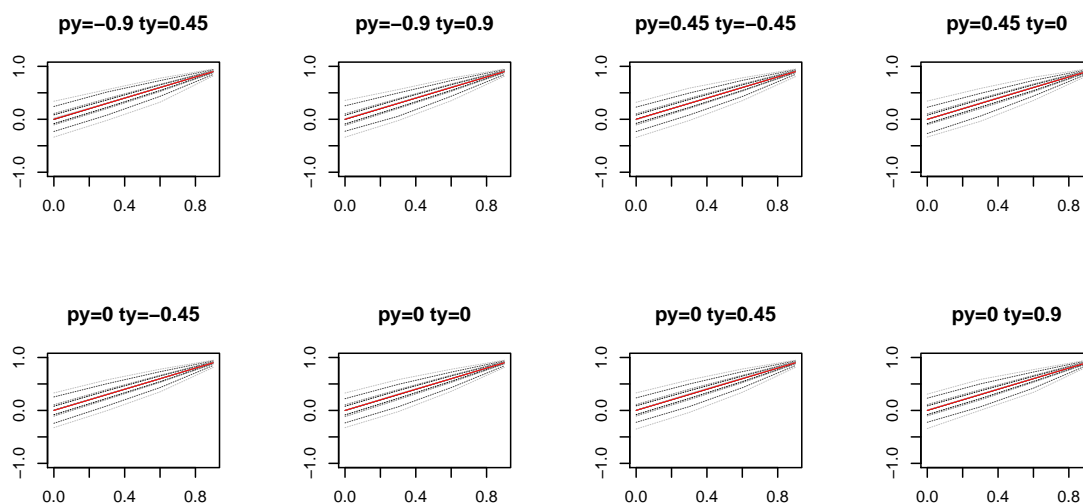


Fig. 3.16 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série y_t , $(f_y; q_y)$. p_y corresponde a f_y e q_y corresponde a q_y . Resultados para as séries com parâmetros autorregressivo e média móvel da série x_t $f_x = q_x = 0$.

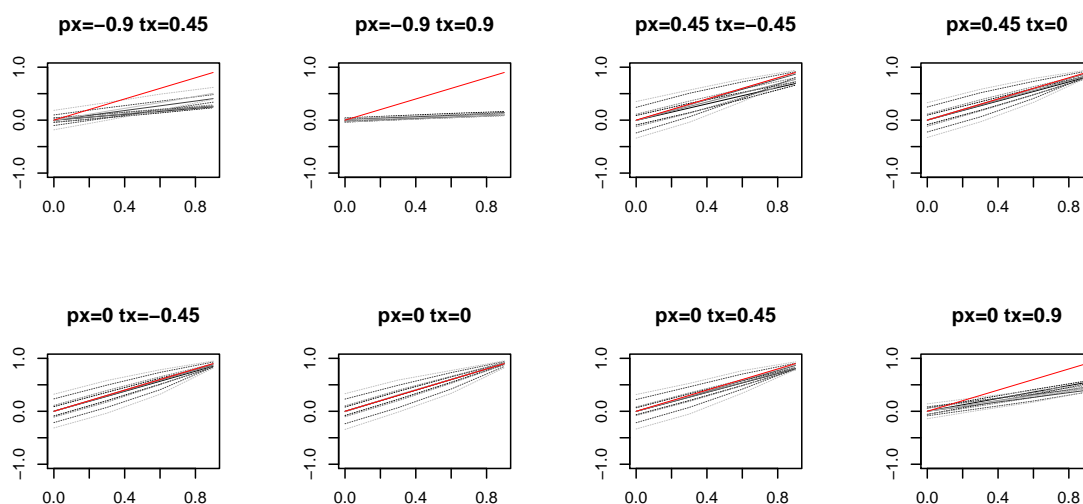


Fig. 3.17 Coeficiente de Correlação estimado pelo método DCCA para diferentes parâmetros autorregressivo e média móvel da série x_t , $(f_x; q_x)$. p_x corresponde a f_x e q_x corresponde a q_x . Resultados para as séries com parâmetros autorregressivo e média móvel da série y_t $f_y = q_y = 0$.

Para ambas as Figuras anteriores 3.16 e 3.17, as linhas vermelhas representam o valor verdadeiro de r_{en} . As linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações para os parâmetros descritos anteriormente. As linhas a tracejado representam os intervalos de confiança, 95%. Os diferentes tons de cinza representam diferentes valores de l , para $l = \frac{T}{100}$ temos o tom mais escuro, o segundo tom mais escuro refere-se a $l = \frac{T}{10}$, o tom a seguir faz refere-se a $l = \frac{T}{50}$ e $l = \frac{T}{5}$ é representada pelo tom mais claro.

Ao analisarmos as Figuras 3.16 e 3.17, podemos concluir que afinal sempre existe um enviesamento no estimador do coeficiente de correlação DCCA. Mas este enviesamento parece ser justificado pela relação da ordem de integração com as componentes AR/MA, ao invés de estar relacionado com a diferença em módulo entre os parâmetros autorregressivo e média móvel como tínhamos suposto no capítulo 3.3.1. Isto porque quando fazemos variar os parâmetros f e q para a série x_t que é a série com ordem de diferenciação $d_1 = 0.75$, não estacionária com reversão à média, Figura 3.17, é quando existe um enviesamento bastante visível enquanto que para a Figura 3.16 a variação dos parâmetros f e q é feita para a série que tem ordem de diferenciação $d_2 = 1.25$, série não estacionária explosiva e não é visível a olho nu nenhum enviesamento, o estimador neste caso parece ser centrado para r_{en} independentemente da variação dos parâmetros f e q . Parece que a causa deste enviesamento será a ordem de não estacionariedade da série para a qual fazemos variar os parâmetros f e q e não propriamente dos valores que estes parâmetros podem tomar.

Ainda assim existem casos em que, quando fazemos variar as componentes AR/MA na série com reversão à média, para os quais obtemos bons resultados. Tal é o caso quando temos $f_x = 0.45$ e $q_x = 0.45$, e tomamos em consideração as duas escalas maiores, $\frac{l}{10}$ e $\frac{l}{5}$, como mostram os gráficos na Figura 3.18.

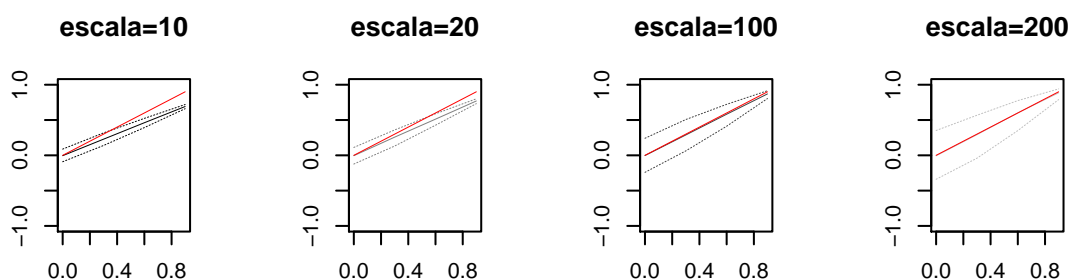


Fig. 3.18 Coeficiente de Correlação estimado pelo método DCCA para os parâmetros autorregressivo e média móvel da série x_t , $f_x = 0.45$, $q_x = 0.45$ e para y_t $f_y = q_y = 0$. px corresponde a f_x e tx corresponde a q_x .

Na Figura 3.18, as linhas vermelhas representam o valor verdadeiro de r_{en} . E as restantes linhas contínuas, com diferentes tons de cinza, representam a média dos valores das 1000 simulações e os seus intervalos de 95% confiança. Da esquerda para a direita temos as escalas $l = \frac{T}{100}$, $l = \frac{T}{50}$, $l = \frac{T}{10}$ e $l = \frac{T}{5}$.

Neste caso o método DCCA produz um estimador melhor quando são utilizadas escalas maiores, ou seja, para $\frac{T}{10}$ e $\frac{T}{5}$, apesar de os intervalos de confiança das escalas menores, $\frac{T}{100}$ e $\frac{T}{50}$, serem mais estreitos.

Para as escalas $\frac{T}{100}$ e $\frac{T}{50}$, os resultados não são fiáveis pois existe um enviesamento no estimador, que piora à medida que nos afastamos da correlação zero.

Posto isto, podemos concluir que para uma ordem de diferenciação pertencente aos intervalos $] -\infty; 0.5[$ e $[1; +\infty[$, ou seja para séries estacionárias e não estacionárias explosivas, o desempenho do estimador DCCA não é afetado. Contudo, é necessário ter em atenção que, para os casos em que temos séries temporais não estacionárias com reversão à média, $\frac{1}{2} < d < 1$ o método do DCCA não se mostra capaz de calcular as correlações verdadeiras para casos em que os parâmetros do AR/MA se afastam de zero e a outra série resulta de um modelo ARFIMA(0; d ; 0), não sendo um método fiável para estes casos. No caso de este método ser utilizado na análise de séries não estacionárias com reversão à média em que os parâmetros autoregressivo e média móvel estão próximos de zero e a outra série em estudo segue um modelo ARFIMA(0; d ; 0), dado serem os casos em que o método ainda mostrou bons resultados, deverão ser consideradas escalas maiores.

Este tipo de combinações de séries temporais, em que obtemos um enviesamento no estimador, merece algum estudo de um modo mais aprofundado no futuro.

Capítulo 4

Conclusão

Ao longo deste trabalho foram realizadas simulações de séries com diferentes características, dentro da classe ARFIMA, e foi utilizado o método DCCA na estimação do coeficiente de correlação entre essas séries. O objetivo foi avaliar o desempenho do método DCCA em circunstâncias variadas, nomeadamente nos casos em que as séries temporais são não estacionárias, situação que ocorre frequentemente em estudos aplicados a dados reais.

Os resultados apresentados apontam no sentido de se poder concluir que o método DCCA tem um bom desempenho na maioria dos casos analisados. Porém, foram detectados alguns casos em que as estimativas do coeficiente de correlação estavam enviesadas. Esses casos ocorreram com séries não estacionárias, nomeadamente quando estivemos a trabalhar com uma série com ordem de integração $\frac{1}{2}$ $d < 1$, ou seja, uma série não estacionária com reversão à média. Nesse caso, o desempenho do método DCCA parece depender dos valores dos parâmetros que determinam o comportamento das componentes autorregressiva e média móvel da série. Os resultados apresentados indicam que, nestas circunstâncias, existem combinações de valores dos parâmetros das componentes autorregressiva e média móvel para as quais os resultados do método DCCA não são fiáveis. Para estas combinações seria interessante fazer um estudo um pouco mais aprofundado, o que não nos foi permitido devido à falta de tempo.

Na maioria dos casos, as escalas com melhores resultados são as escalas menores. Para estas escalas obtemos um intervalo de confiança 95% mais preciso e não existe um enviesamento significativo no estimador DCCA. Assim, em termos gerais, a conclusão é que o método DCCA continua a ser um bom estimador da correlação entre séries temporais, mesmo quando passamos a trabalhar com séries com diferentes ordens de (não) estacionariedade, como havia já sido sugerido no estudo de Kristoufek, [37].

No entanto, o facto de termos encontrado alguns casos em que o método DCCA não é fiável torna aconselhável alguma prudência e alguma análise prévia quando se utiliza o método DCCA com séries não estacionárias. Tal poderá ser também o caso quando se utilizam outros métodos que foram desenvolvidos com base no DCCA, como o Detrended Partial-Cross-Correlation Analysis (DPCCA), [64], Detrending Moving-Average Cross-Correlation (DMCA), [35], Multifractal Height Cross-Correlation Analysis (MF-HXA), [34]. Uma análise semelhante à que foi realizada nesta dissertação poderá ajudar a perceber se estes novos métodos são mais robustos do que o método DCCA, ou se efectivamente o problema afecta também os métodos mais recentes.

Bibliografia

- [1] Aloui, C., Hkiri, B., Hammoudeh, S., and Shahbaz, M. (2018). A multiple and partial wavelet analysis of the oil price, inflation, exchange rate, and economic growth nexus in Saudi Arabia. *Emerging Markets Finance and Trade*.
- [2] Anesi, G. and Lopes, S. R. (2010). Inovações estáveis em processos arfima(p,d,q). <http://www2.ime.unicamp.br/sinape/sites/default/files/Gennaro%2520Anesi%2520-%2520Resumo%252019SINAPE.pdf>.
- [3] Baird, H., Shin, S., Oberprieler, R., Hulle, M., Vernon, P., Moon, K., Adams, R., McKenna, D., and Chown, S. (2021). Fifty million years of beetle evolution along the Antarctic polar front. *Proceedings of the National Academy of Sciences of the United States of America*.
- [4] Box, G. and Jenkins, G. (1970). Time series analysis forecasting and control. *Journal of Time Series Analysis*.
- [5] Cao, G., He, L.-Y., and Cao, J. (2018). *Multifractal detrended analysis method and its application in financial markets*.
- [6] Casa Nova, A., Ferreira, P., Almeida, D., Dionísio, A., and Quintino, D. (2021). Are mobility and COVID-19 related? A dynamic analysis for Portuguese districts. *Entropy*.
- [7] Chenhua, S. (2015). Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A*, 379:680–687.
- [8] Contreras-Reys, J. E. and Idrovo-Aguirre, B. J. (2020). Backcasting and forecasting time series using detrended cross-correlation analysis. *Physica A*, 560.
- [9] Costa, N., Silva, C., and Ferreira, P. (2019). Long-range behaviour and correlation in DFA and DCCA analysis of cryptocurrencies. *International Journal of Financial Studies*.
- [10] Da Cunha Cabral, I. and Ribeiro, P. and Nicolau, J. (2019). Tracking the relationship between euro area equities and sovereign bonds. *International Journal of Monetary Economics and Finance*.
- [11] da Silva Filho, A. M., Zebende, G. F., de Castro, A. P. N., and da Silva, M. F. (2014). Modelagem do teste estatístico do coeficiente de correlação cruzada sem tendência r_{DCCA} : teoria e aplicação. *Conference Paper*.
- [12] da Silva Filho, A.M. and Zebende, G. and de Castro, A.P.N. and Guedes, E. (2021). Statistical test for multiple detrended cross-correlation coefficient. *Physica A: Statistical Mechanics and its Applications*.
- [13] Ehelepola, N., Ariyaratne, K., Aththanayake, A., Samarakoon, K., and Thilakarathna, H. (2021). The correlation between three teleconnections and leptospirosis incidence in the Kandy district, Sri Lanka, 2004–2019. *Tropical Medicine and Health*.
- [14] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *The Econometric Society*, 50.

- [15] Feder, J. (1988). *Fractals*. Plenum Press, New York.
- [16] Ferreira, P. (2017). Portuguese and brazilian stock market integration: a non-linear and detrended approach. *Portuguese Economic Journal*.
- [17] Ferreira, P. (2018). What guides central and eastern european stock markets? a view from detrended methodologies. *Post-Communist Economies*.
- [18] Ferreira, P. (2020). Dynamic long-range dependences in the swiss stock market. *Empirical Economics*.
- [19] Ferreira, P., Da Silva, M., and De Santana, I. (2019). Detrended correlation coefficients between exchange rate (in dollars) and stock markets in the world's largest economies. *Economies*.
- [20] Ferreira, P. and Dionísio, A. (2017). Long range dependence in G7 stock markets' return rates using mutual information and detrended cross-correlation analysis. *Journal for Studies in Economics and Econometrics*.
- [21] Filho, A. M., da Silva, M., and Zebende, G. (2014). Autocorrelation and cross-correlation in time series of homicide and attempted homicide. *Physica A*, 400:12–19.
- [22] Ftiti, Z., Jawadi, F., Louhichi, W., and Madani, M. (2019). On the relationship between energy returns and trading volume: a multifractal analysis. *Applied Economics*.
- [23] Gonçalves, E. (2020). Processos auto-regrassivos médias móveis. Apontamentos de Séries Temporais.
- [24] Gonçalves, E. and Lopes, N. (2013). *Probabilidades: Princípios Teóricos*. ESCOLAR EDITORA.
- [25] Granger, C. and Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and their Applications* 8.
- [26] Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*.
- [27] Horvatic, D., Stanley, H. E., and Podobnik, B. (2011). Detrended cross-correlation analysis for non-stationary time series with periodic trends. *EPL (Europhysics Letters)*, 94.
- [28] Hosking, J. (1981). Fractional differencing. *Bometrika*, 68.
- [29] Johansen, A. (2010). Monte carlo methods. *International Encyclopedia of Education (Third Edition)*.
- [30] Junior, N., Santos, C., Zebende, G., and Filho, A. (2021). Geodetic sea level control in salvador: Analysis and correlations | [controle geodesico do nivel do mar em salvador: Analises e correlacoes]. *Revista Brasileira de Cartografia*.
- [31] Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H., Havlin, S., and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, (295):441–454.
- [32] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, (316):87–114.
- [33] Koscielny-Bunde, E., Bunde, A., Havlin, S., Roman, H. E., Goldreich, Y., and Schellnhuber, H. J. (1998). *Phys. Rev. Lett*, 81(729).
- [34] Kristoufek, L. (2011). Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations. *Europhysics Letters Association*, 95(6).

- [35] Kristoufek, L. (2014a). Detrending moving-average cross-correlation coefficient: Measuring cross-correlations between non-stationary series. *Physica A: Statistical Mechanics and its Applications*, 406:169–175.
- [36] Kristoufek, L. (2014b). Leverage effect in energy futures. *Energy Economics*.
- [37] Kristoufek, L. (2014c). Measuring correlations between non-stationary series with dcca coefficient. *Physica A*, (402):291–298.
- [38] Kristoufek, L. (2021). Power-law cross-correlations: Issues, solutions and future challenges. *Springer Proceedings in Complexity*.
- [39] Kristoufek, L. and Ferreira, P. (2018). Capital asset pricing model in portugal: Evidence from fractal regressions. *Portuguese Economic Journal*.
- [40] Kwapien, J., Oswwiecimka, P., and Drozd, S. (2015). Detrended fluctuation analysis made flexible to detect range of cross-correlated fluctuations. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*.
- [41] Lopes, M. d. N. M. (2020/2021). Estatística-capítulos iii e iv. Apontamentos de Estatística.
- [42] Machado, A., Silva, M., and Zebende, G. (2014). Autocorrelation and cross-correlation analysis in time series of homicide and attempted homicide. *Physica A, Amsterdam*, 400:12–19.
- [43] Marinho, E., Sousa, A., and Andrade, R. (2013). Using cross-correlation analysis in geophysical data. *Physica A, Amsterdam*, 392:2195–2201.
- [44] Najarian, J. P. (2020). Monte carlo techniques. *Salem Press Encyclopedia of Science*.
- [45] Pascoal, R. (2019/2020). Capítulo 5-processos de memória longa. Apontamentos de Análise de Séries Financeiras, Faculdade de Economia, Universidade de Coimbra.
- [46] Peng, C.-K., Buldyrev, S., Havlin, S., Simons, M., Stanley, H., and Goldberger, A. (1994). Mosaic organization of dna nucleotides. *Physical Review E*, 49.
- [47] Podobnik, B., Jiang, Z.-Q., Zhou, W.-X., and Standley, H. (2011). Statistical tests for power-law cross-correlated processes. *Physical Review*, 066118, 84.
- [48] Podobnik, B. and Standley, H. (2008). Detrended cross-correlation analysis: A new method for analysing two nonstationary time series. *Physical Review Letters*, 084102, 100.
- [49] R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [50] Reboredo, J., Rivera-Castro, M., and Zebende, G. (2014). Oil and us dollar exchange rate dependence: a detrended cross-correlation approach. *Energy Economics, Cambridge UK*, 42:132–139.
- [51] Robinson, P. A. (2003). *Phys. Rev.E*, 67(032902).
- [52] Shao, Y.-H., Gu, G.-F., Jiang, Z.-Q., and Zhou, W.-X. (2015). Effects of polynomial trends on detrending moving average analysis. *Fractals*.
- [53] Siwy, Z., Ausloos, M., and Ivanova, K. (2002). *Phys. Rev. E*, 65(031907).
- [54] Talkner, P. and Weber, R. O. (2000). *Phys. Rev. E*, 62(150).
- [55] Tilfani, O. and Ferreira, P. and El Boukfaoui, M. (2021). Dynamic cross-correlation and dynamic contagion of stock markets: a sliding windows approach with the dcca correlation coefficient. *Empirical Economics*.

- [56] Tilfani, O., Ferreira, P., and El Boukfaoui, M. (2020). Revisiting stock market integration in central and eastern european stock markets with a dynamic analysis. *Post-Communist Economies*.
- [57] Varotsos, P. A., Sarlis, N. V., and Skordas, E. S. (2003a). *Phys. Rev. E*, 67(021109).
- [58] Varotsos, P. A., Sarlis, N. V., and Skordas, E. S. (2003b). *Phys. Rev. E*, 68(031106).
- [59] Vassoler, R. and Zebende, G. (2011). Dcca cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A, Amesterdam*, 391:2438–2443.
- [60] Veenstra, J. Q. (2012). *Persistence and Anti-persistence: Theory and Software*. PhD thesis, Western University.
- [61] Wang, F. and Yang, Z. and Wang, L. (2016). Detecting and quantifying cross-correlations by analogous multifractal height cross-correlation analysis. *Physica A: Statistical Mechanics and its Applications*.
- [62] Wang, G.-J., Xie, C., Chen, Y.-J., and Chen, S. (2013). Statistical properties of the foreign exchange network at different time scales: Evidence from detrended cross-correlation coefficient and minimum spanning tree. *Entropy*, 15:1643–1662.
- [63] Wang, R., Xie, Y. and Chen, H., and Jia, G. (2021). Analyzing the impact of covid-19 on the cross-correlations between financial search engine data and movie box office. *Fluctuation and Noise Letters*.
- [64] Yuan, N., Fu, Z., Zhang, H., Piao, L., Xoplaki, E., and Luterbacher, J. (2015). Detrended partial-cross-correlation analysis: A new method for analyzing correlations in complex system. *Scientific Reports*, 5(8143).
- [65] Zebend, G. (2011). Dcca cross-correlation coefficiente:quantifying level of cross-correlation. *Physica A*, 390:614–618.
- [66] Zebende, G. and Filho, A. M. (2009). Cross-correlation between time series of vehicles and passengers. *Physica A*, 388:4863–4866.

Anexo A

Gráficos adicionais

Aqui serão apresentados os restantes gráficos que foram utilizados para retirar as conclusões do Capítulo 3. Para uma observação dos gráficos mais pormenorizada aconselhamos a que o leitor utilize a versão digital do documento.

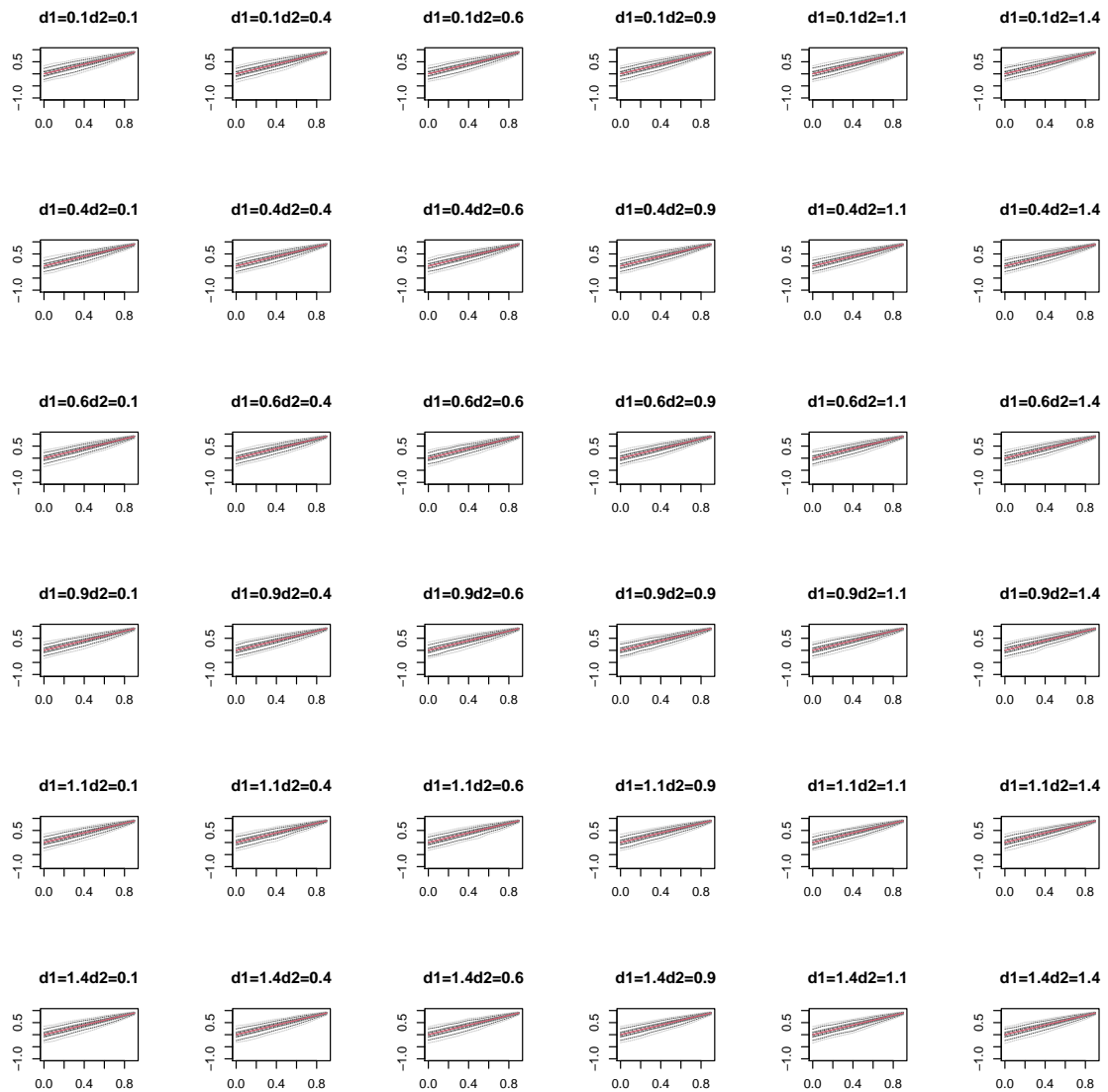


Fig. A.1 Gráficos 1º Caso para T=1000

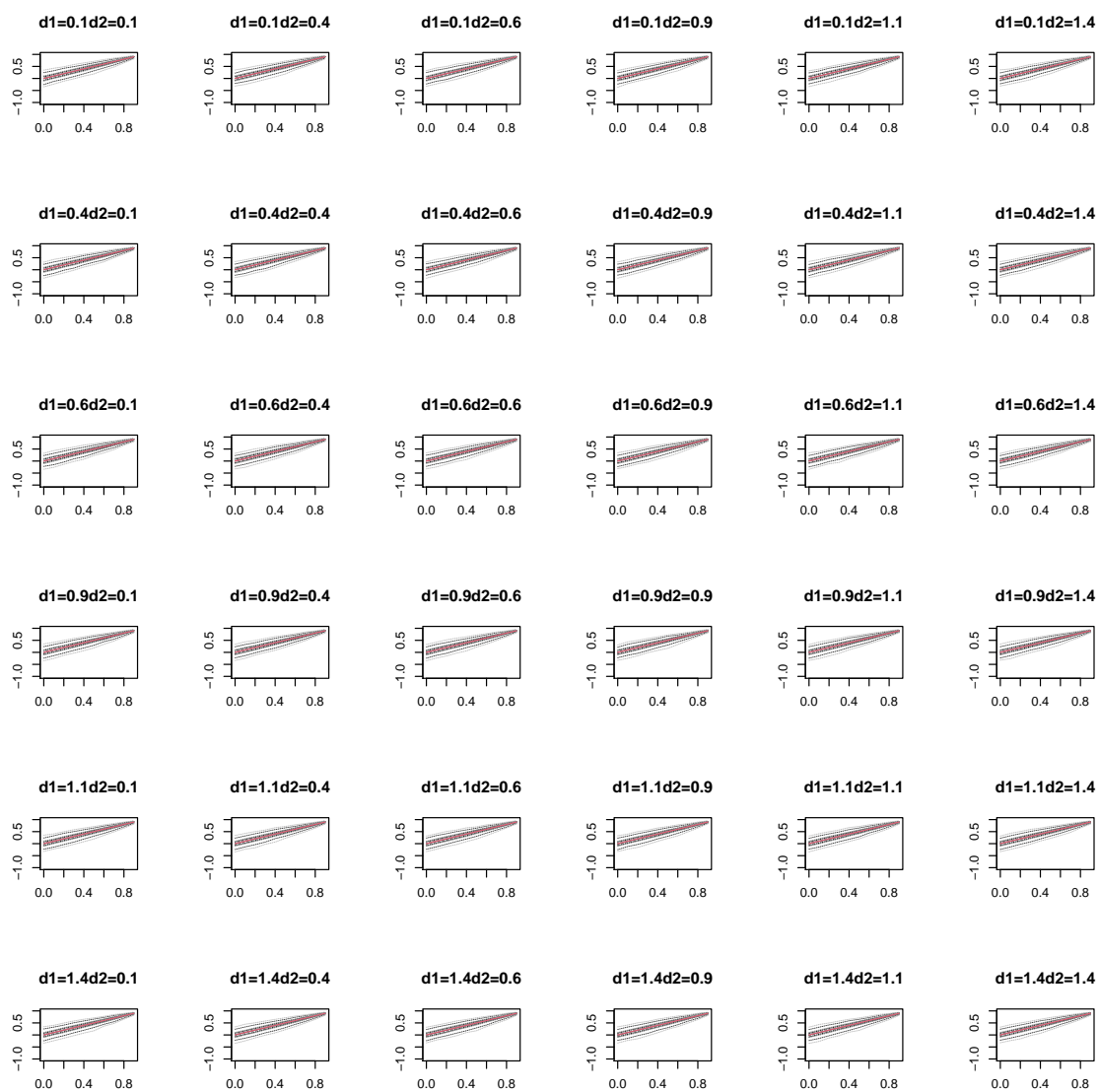


Fig. A.2 Gráficos 1º Caso para T=5000



Fig. A.3 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$



Fig. A.4 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0; 0.5)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$



Fig. A.5 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0.3)$ e $(f_y; q_y) = (0; 0)$ e $T = 1000$



Fig. A.6 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$



Fig. A.7 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0; 0.5)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$



Fig. A.8 Gráfico do 2º Caso 1ª Parte para os pares $(f_x; q_x) = (0.5; 0.3)$ e $(f_y; q_y) = (0; 0)$ e $T = 5000$

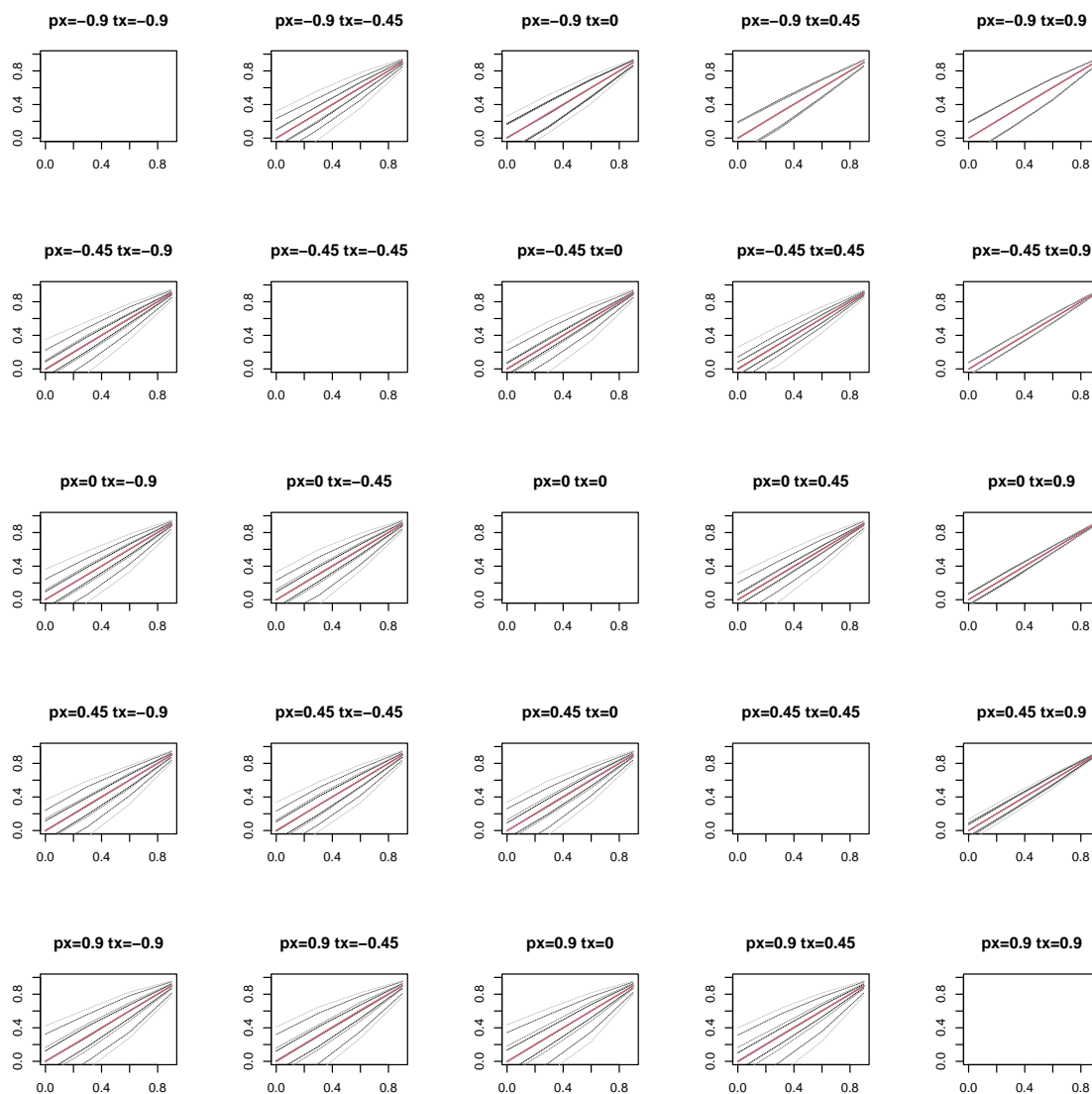


Fig. A.9 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$

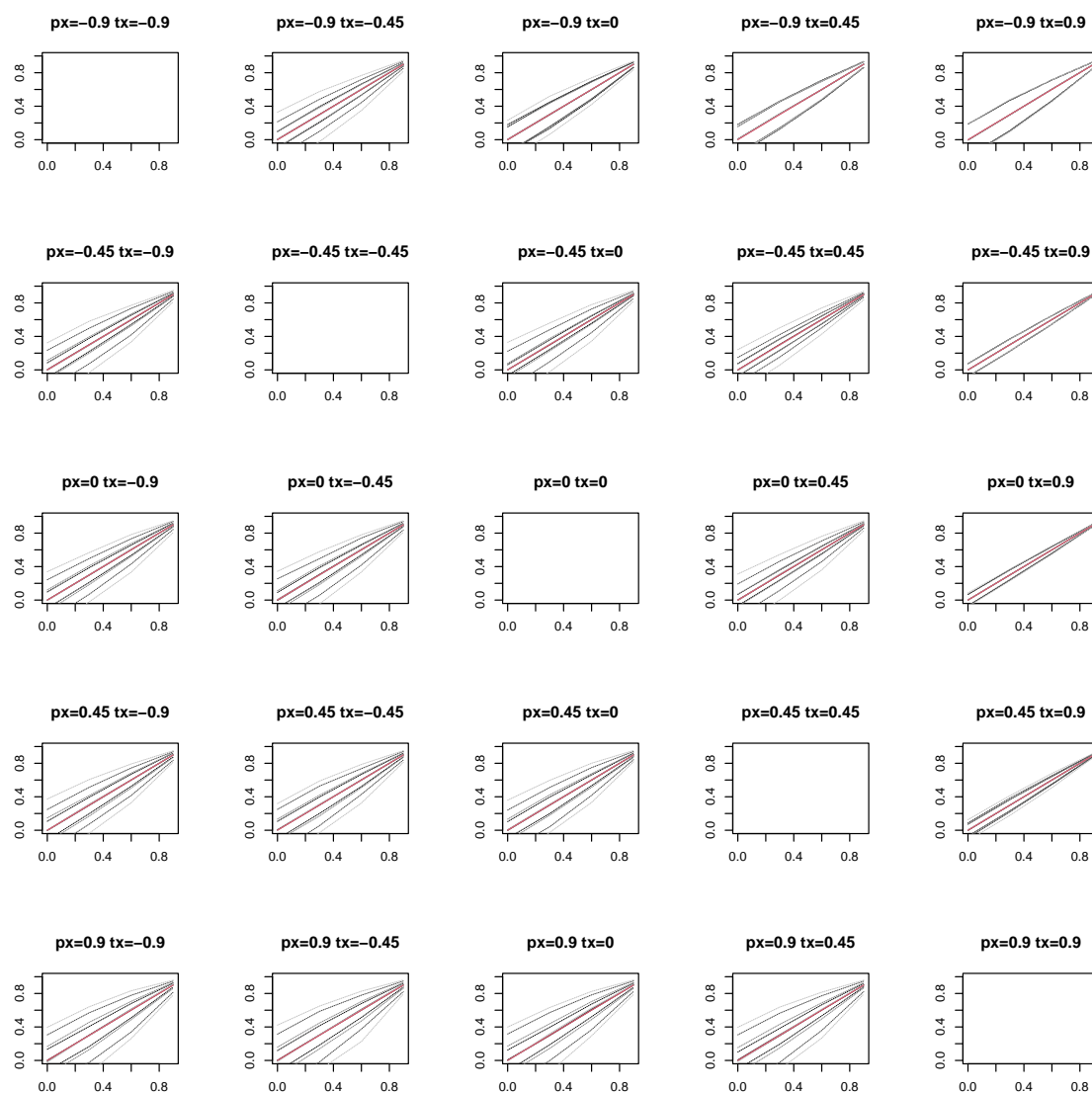


Fig. A.10 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.25; 0.75)$ e $(p_y; q_y) = (0.9; 0.9)$

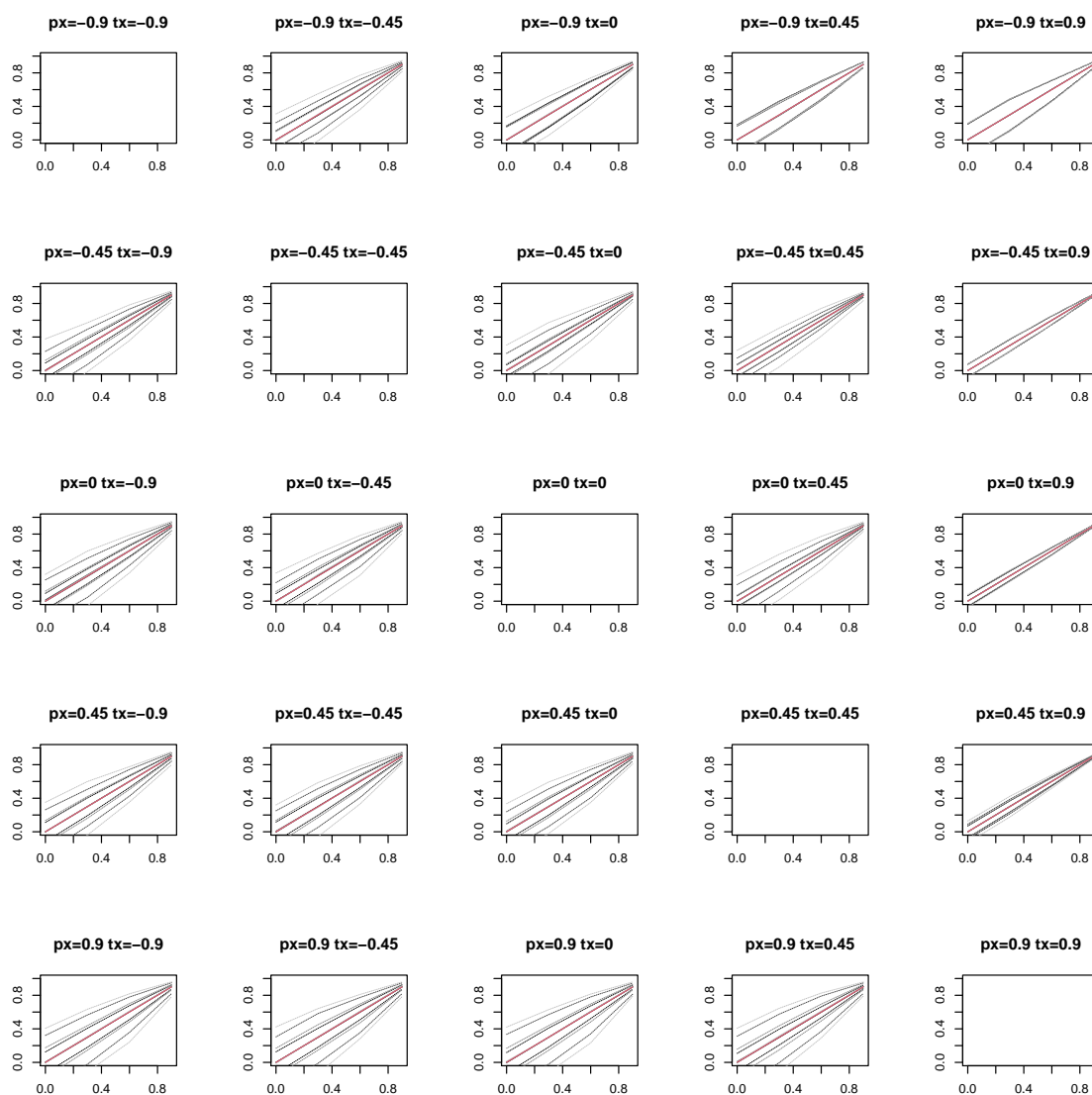


Fig. A.11 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.25; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$

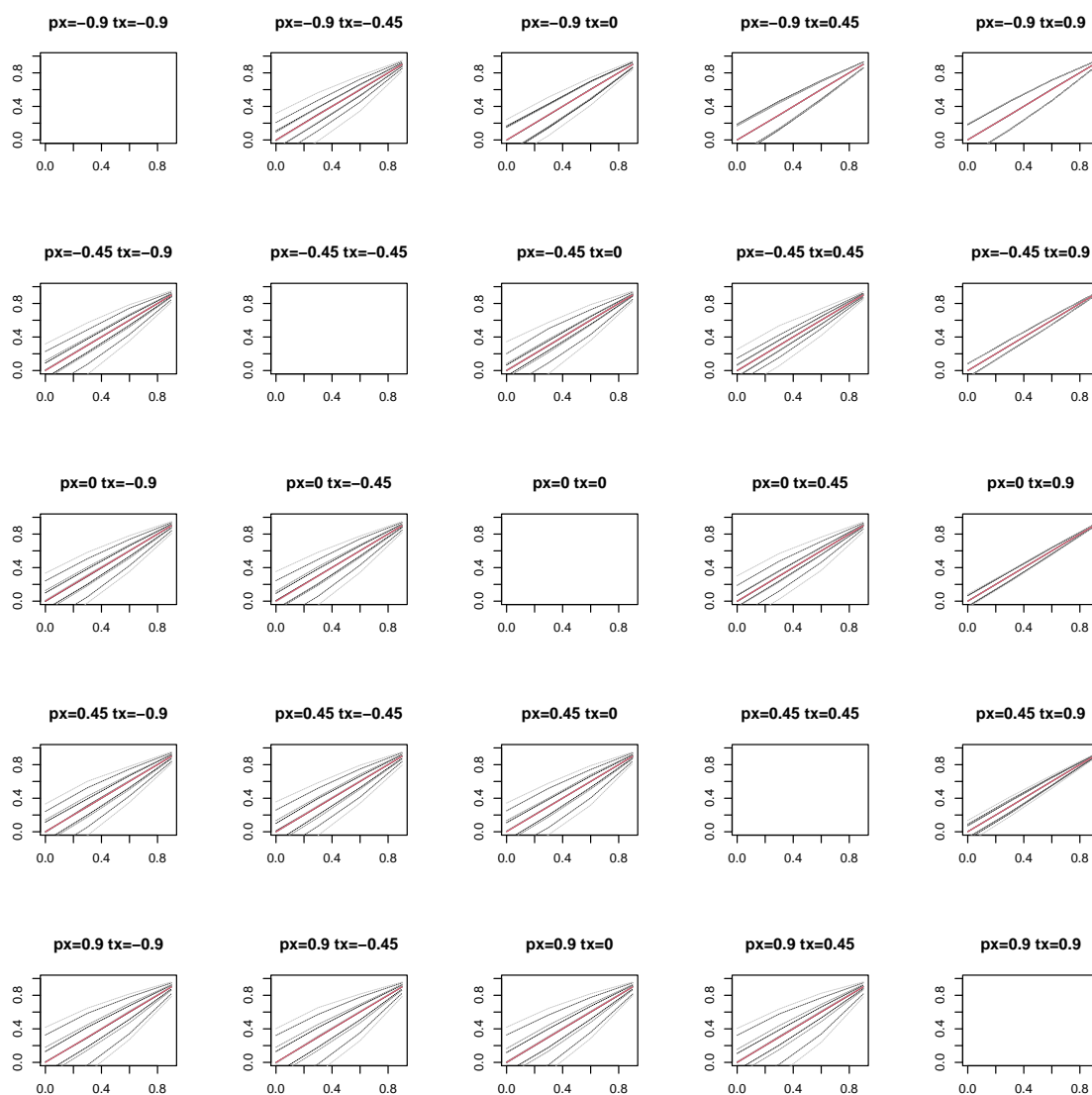


Fig. A.12 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0.9; 0.9)$

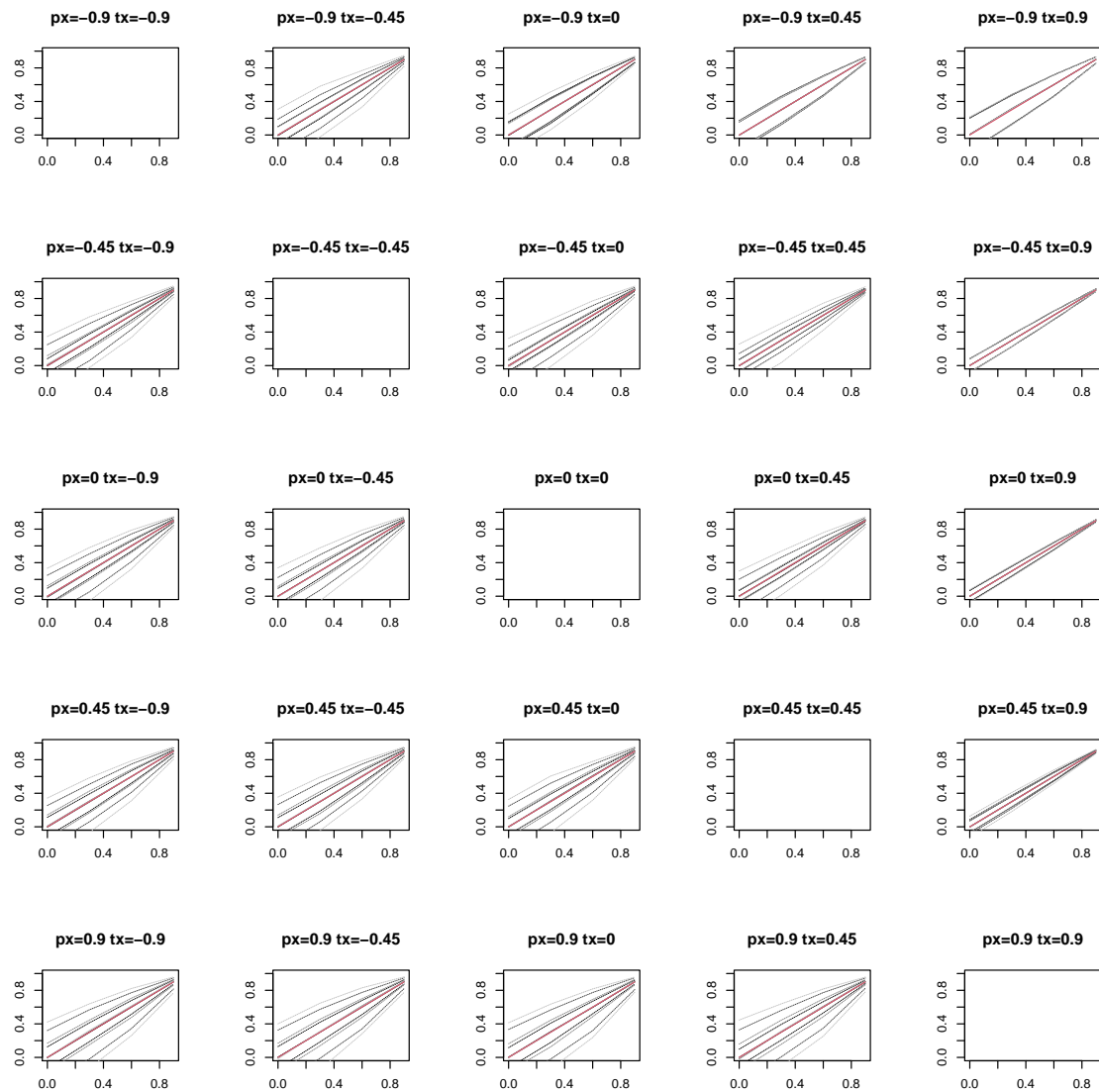


Fig. A.13 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0.75; 1.25)$ e $(p_y; q_y) = (0; 0.9)$

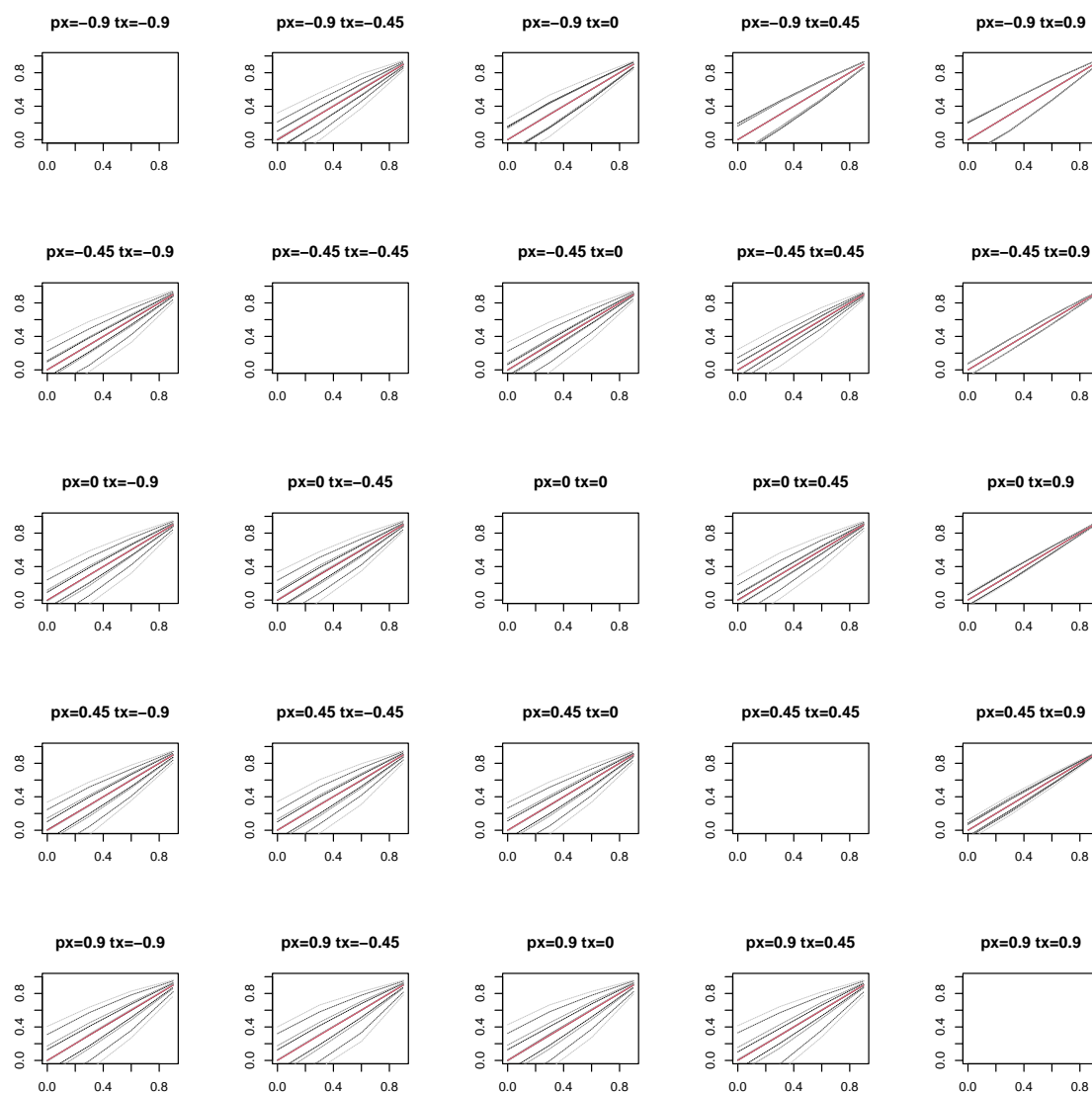


Fig. A.14 Gráfico do 2º Caso 2ª Parte para os pares $(d_1; d_2) = (0; 75; 1; 25)$ e $(p_y; q_y) = (0; 0; 9)$

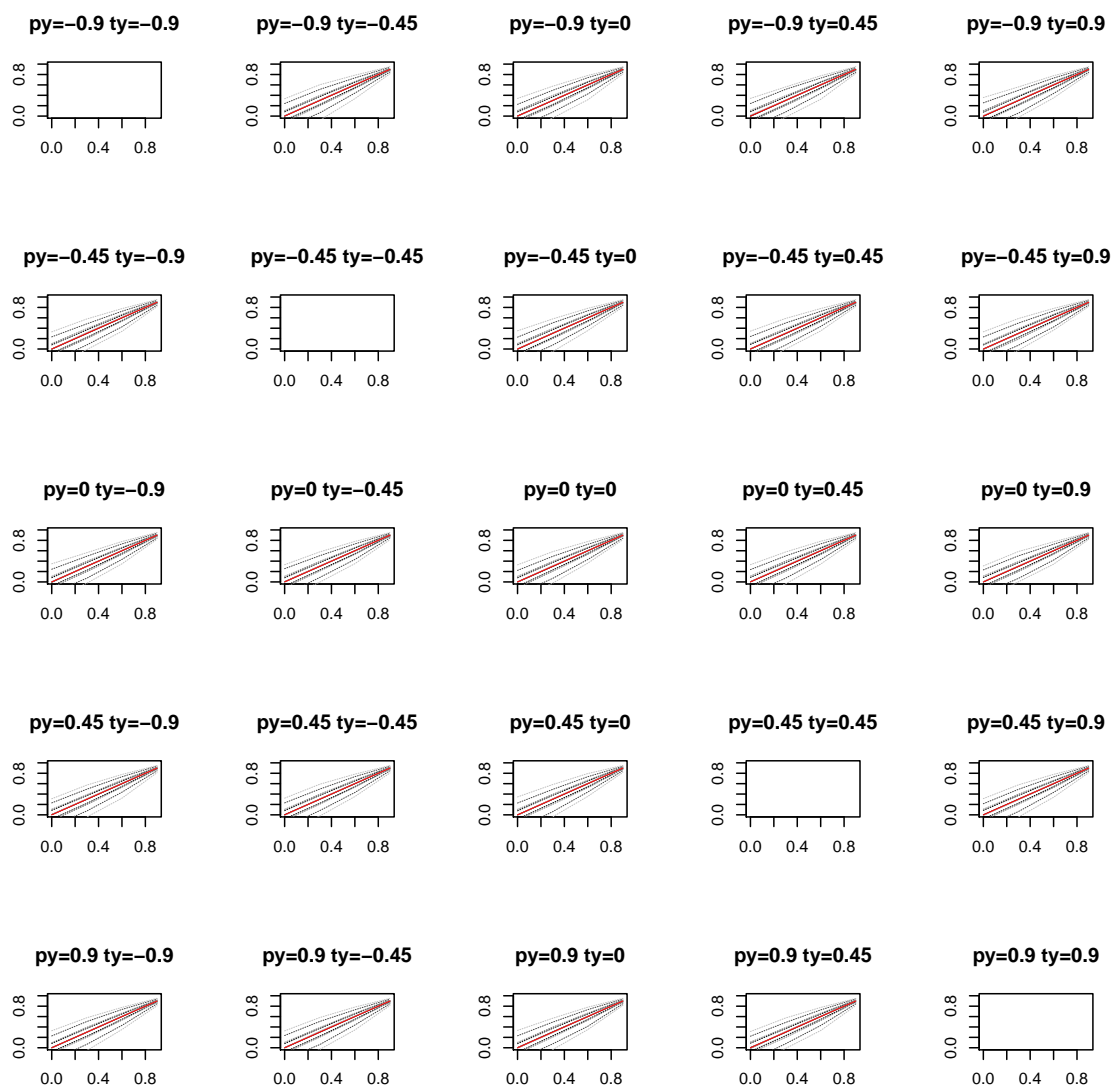


Fig. A.15 Gráficos do 2º Caso da 2ª Parte do Caso a)

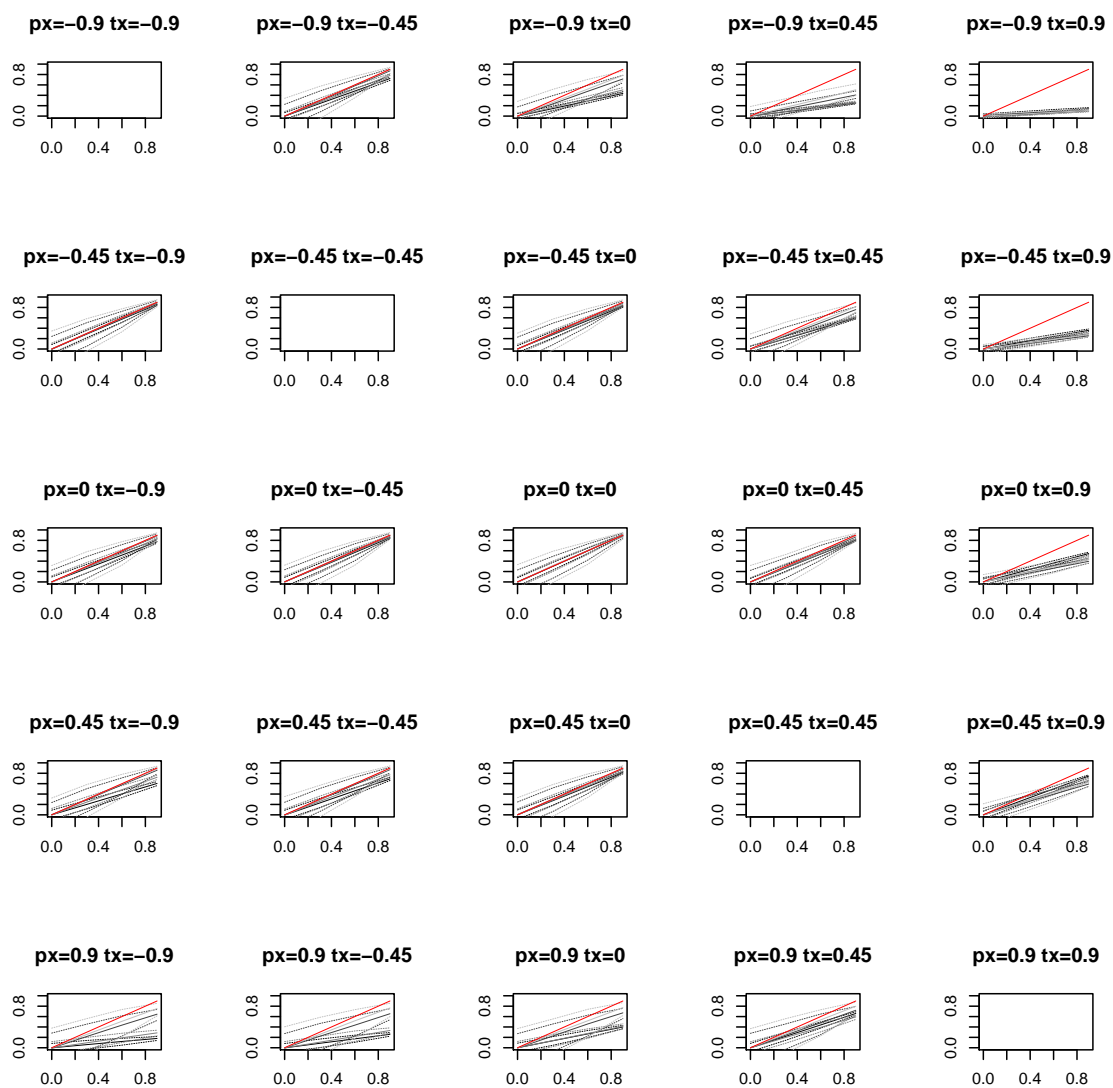


Fig. A.16 Gráficos do 2º Caso da 2ª Parte do Caso b)