Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Programa de Pós-Graduação em Matemática

Cópulas, Processos com Longa Dependência e Decaimento da Correlação em Processos Estocásticos

Guilherme Pumi

Porto Alegre, Janeiro de 2012.

Tese submetida por Guilherme Pumi[1] como requisito parcial para a obtenção do título de Doutor em Matemática pelo Programa de Pós-Graduação em Matemática da Universidade Federal do Rio Grande do Sul.

Orientadora:

Dra. Sílvia Regina Costa Lopes (PPGMAT - UFRGS)

Comissão Examinadora:

Dr. Artur Oscar Lopes (PPGMAT - UFRGS)

Dra. Chang Chung Yu Dorea (MAT - UnB)

Dr. Cleber Bisognin (Departamento de Estatística - UFRGS)

Dr. Luiz Koodi Hotta (IMECC - Unicamp)

Data de Apresentação: 26 de Janeiro de 2012

[1] Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

Agradecimentos

Primeiramente e em especial à Sílvia pela gentil e calorosa acolhida na nossa difícil volta dos Estados Unidos, pela confiança depositada em mim ao aceitar-me como aluno, pela atenciosa e dedicada orientação, pelos valiosos conselhos e incentivos, paciência, apoio e amizade.

Ao Prof. Artur por todo apoio e pelas diversas contribuições diretas e indiretas ao presente trabalho, aos membros da banca, pela disposição e valiosas sugestões ao trabalho e à Rosane pela atenção e carinho dedicados desde o nivelamento.

Ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo crucial apoio financeiro e ao CESUP - Centro Nacional de Supercomputação pelos recursos computacionais.

To fellow colleagues met at congresses and meetings throughout Brazil who contributed to the development of our work with their comments and suggestions. Special thanks to Roger B. Nelsen, Harry Joe, Aris Nikoloulopoulos, Matthias Scherer and Wembo V. Li.

Ao meu pai Juarez Campos Pumi pelo carinho, atenção, conselhos, brincadeiras, visitas e conversas que me dedicou nestes dois anos e meio de doutorado. À minha mãe Vani T. Kostaneski pelo carinho, apoio, incentivo e atenção desde a nossa volta dos Estados Unidos. Em especial também, agradeço à minha esposa Taiane, pelo enorme carinho, atenção, dedicação, paciência e amor que permitiram a conclusão de mais esta caminhada.

Resumo

Neste trabalho abordamos três assuntos importantes na área de Estatística Matemática, Análise de Séries Temporais e Processos Estocásticos, a saber, cópulas, processos com longa dependência e o decaimento da correlação em processos estocásticos. Nossa contribuição para a teoria de cópulas compreende a derivação e estudo das cópulas relacionadas com certos tipos de processos estocásticos obtidos a partir da iteração de uma transformação suave por partes à uma determinada variável aleatória inicial. Aplicações à estimação paramétrica em processos do tipo estudado são considerados e simulações de Monte Carlo são apresentadas.

Nossa contribuição à teoria de processos com longa dependência pode ser dividida em duas frentes. Primeiramente, o problema de estimação semiparamétrica em processos multivariados apresentando longa dependência é estudado. Duas classes de estimadores para o vetor de diferenciação fracionária são introduzidas e suas propriedades assintóticas estudadas. Simulações de Monte Carlo são realizadas para avaliar o desempenho dos estimadores na prática.

Em um segundo momento, estudamos a interdependência das coordenadas em processos VARFIMA$(0, \boldsymbol{d}, 0)$ bidimensionais sob o ponto de vista da distância de Mallows e do τ de Kendall sob diversas condições. O trabalho é baseado em simulações de Monte Carlo e foca em uma possível relação entre a distância de Mallows, o vetor de diferenciação fracionária $\boldsymbol{d}$, o tipo e grau de dependência induzido no ruído, bem como no comportamento das marginais do processo.

Como aplicações, um estimador do vetor de diferenciação fracionária e um teste para detectar a presença de coordenadas com longa dependência forte em processos VARFIMA$(0, \boldsymbol{d}, 0)$ de qualquer dimensão finita são introduzidos. Um estudo de Monte Carlo específico é realizado para avaliar o desempenho tanto do estimador quanto do teste. Propriedades assintóticas do estimador para a distância de Mallows utilizado no trabalho também são estudadas.

Finalmente, contribuímos para o estudo do decaimento da correlação em processos estocásticos investigando o problema de obter-se um determinado decaimento da correlação a partir da reparametrização de uma família de cópulas, dadas as marginais do processo. Como aplicações, uma metodologia geral de estimação de parâmetros, identificáveis pelo decaimento da autocovariância, em séries temporais é proposta e um método para a simulação de séries temporais com determinadas características é introduzido. A metodologia proposta ainda é aplicada à série temporal real do índice de ativos da S&P500.

Abstract

In this work we consider three major subjects in Statistics, Time Series Analysis and Stochastic Process, namely, copulas, long range dependence and the decay of correlation in stochastic processes. Our contribution to the theory of copulas is deriving and studying the copulas related to a class of chaotic processes obtained from the iteration of certain smooth piecewise transformations of the interval to an initial random variable. The theory is applied to the problem of parametric estimation in the class of stochastic processes studied. Monte Carlo simulations illustrating the methodology are also presented.

As for long range dependence, our contribution to the field is two-folded. Firstly, we consider semiparametric estimation in multivariate long range dependent processes. Two broad classes of estimators are introduced and their asymptotic properties are derived. A Monte Carlo study is also presented to assess the finite sample performance of the estimators.

Secondly, we analyze the dependence between the components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes under several different data generating methods through the Mallows distance and Kendall's τ point of view. The work is based on Monte Carlo simulations and the main goal is to investigate a possible relationship between the Mallows distance, the fractional differencing parameter $\boldsymbol{d}$, the type and level of dependence induced in the innovation process as well as its marginal behavior.

As applications, an estimator for the fractional differencing parameter $\boldsymbol{d}$ as well as a test to assess the presence of a strong long range dependent component in VARFIMA$(0, \boldsymbol{d}, 0)$ processes of any finite dimension are proposed. A Monte Carlo experiment is presented in order to assess the performance of both, the estimation procedure and the test. Asymptotic properties of the Mallows distance estimator introduced in the work are derived as well.

Lastly, we contribute to the study of the correlation decay in stochastic processes by investigating the problem of constructing stochastic processes with a predetermined decay of correlation with given marginals by reparameterizing a given parametric family of copulas. As applications, a general estimation procedure to estimate a given parameter identifiable through the decay of covariance in stochastic processes is proposed and the problem of simulating time series with some predetermined decay of covariance is studied. A Monte Carlo studied to exemplify and assess the methodology's performance is also presented. The methodology is further applied to the S&P500 US stock market index.

Índice

Capítulo 1

Introdução

Como o próprio título sugere, neste trabalho discutimos três grandes tópicos nas áreas de Estatística, Análise de Séries Temporais e Processos Estocásticos, a saber, cópulas, processos com longa dependência e o decaimento da correlação em processos estocásticos. Nossa intenção certamente não é ser exaustivo nesses tópicos mas, ao contrário, apresentar o trabalho realizado durante o período de pesquisas necessárias à obtenção do título de Doutor em Matemática pela Universidade Federal do Rio Grande do Sul. Os prerequisitos necessários ao entendimento deste trabalho são basicamente um curso em análise no espaço Euclideano, um curso em medida e integração, um curso de probabilidade e estatística matemática, um curso em processos estocásticos e análise de séries temporais. Alguns conceitos que possam não ser padrão em cursos nas áreas citadas são introduzidos no capítulo de preliminares e/ou no corpo do texto.

Conforme comentado no parágrafo anterior, o primeiro grande tópico abordado no trabalho são as cópulas. Informalmente, cópulas são funções que ligam uma função de distribuição m-dimensional às suas marginais de qualquer dimensão. O surgimento das cópulas está intimamente ligado ao desenvolvimento da teoria de espaços métricos probabilísticos, onde a necessidade de se obter um análogo probabilístico à desigualdade triangular levou ao desenvolvimento das cópulas. A teoria de cópulas e vários dos teoremas mais importantes dela, incluindo o Teorema de Sklar (Teorema 2.2), foram provados entre o final dos anos 50 até meados da década de 70. É bem verdade que a teoria de cópulas passou uma década praticamente esquecida, quando subitamente, na segunda metade dos anos 80, a teoria ganhou força novamente através do estudo de medidas de dependência não-paramétricas livres de escala. O grande impulso para a redescoberta de cópulas veio com o desenvolvimento de diversas aplicações em áreas como estatística, finanças, hidrologia, ciências atuariais, entre outros.

Atualmente, cópulas têm sido aplicadas com sucesso na construção e estudo de medidas de dependência livres de escala (veja por exemplo, Mari e Kotz, 2001 e Nelsen, 2006), na construção de distribuições multivariadas com determinadas características (veja por exemplo, Joe 1997 e Nelsen, 2006), na extensão de modelos univariados para um contexto multivariado (veja por exemplo, Lee e Long, 2009 e Palynchuk e Guo, 2011), no estudo de processos estocásticos (veja por exemplo, Darsow et al., 1992, Lagerås, 2010, Lopes e Pumi, 2011, Pumi e Lopes, 2011a), na estimação e estudo de séries temporais (veja por exemplo, Chen e Fan, 2006 e Pumi e Lopes, 2011c), como ferramenta para a simulação de processos estocásticos (veja por exemplo, Nelsen, 2006, Lopes et al., 2011 e Pumi e Lopes, 2011c) entre muitas outras áreas. Notavelmente, na área de finanças cópulas ganharam muita popularidade e métodos baseados em cópulas tornaram-se padrão em muitas áreas da análise financeira (veja por exemplo, Cherubini et al., 2004).

Neste trabalho, nossa contribuição para a teoria de cópulas é estudar uma classe particular delas relacionadas a certos processos caóticos. Mais especificamente, considere $T : [0,1] \to [0,1]$ uma transformação suave por partes e suponha que exista uma medida T-invariante μ_T. Seja U_0 uma variável aleatória com distribuição μ_T e seja $\varphi : [0,1] \to \mathbb{R}$ uma função μ_T-integrável. Podemos, desta forma, definir um processo estocástico através de

$$X_t := \varphi\big(T^t(U_0)\big), \quad \text{para todo } t = 0, 1, \cdots . \tag{1.1}$$

Na primeira parte do Capítulo 3, apresentamos os resultados contidos em Lopes e Pumi (2011). Neste, derivamos e estudamos as cópulas relacionadas a vetores aleatórios $(X_{t_1}, \cdots, X_{t_n})$ provindos de processos do tipo (1.1) quando a transformação T é a chamada transformação de Manneville-Pomeau (veja Definição 3.2). Essas cópulas são singulares e não possuem uma expressão fechada. Aspectos computacionais, aproximações, geração aleatória entre outros detalhes também são discutidos, além da estimação paramétrica em processos caóticos baseados na teoria desenvolvida. Na segunda parte do Capítulo 3, estudamos uma generalização dos resultados obtidos na primeira parte do capítulo. Tais resultados estão contidos em Pumi e Lopes (2011a).

O segundo grande tópico abordado no trabalho refere-se ao fenômeno da longa dependência. Longa dependência pode ser caracterizada pelo decaimento lento da autocovariância do processo ou pela presença de um singularidade na densidade espectral na frequência zero. O primeiro artigo relatando as características de longa dependência em uma série temporal (a saber, a famosa série dos níveis do Rio Nilo) foi Hurst (1951). Nesses 60 anos que se passaram desde o trabalho de Hurst, o estudo de séries temporais com longa dependência se desenvolveu e se expandiu muito, tanto do ponto de vista teórico quanto do ponto de vista de aplicações e hoje é parte importante da análise de séries temporais. Vários modelos acomodando características de longa dependência foram desenvolvidos na literatura ao longo dos anos, tanto no caso univariado, quanto no caso multivariado. Veja por exemplo Doukhan et al. (2003), Palma (2007) e Lopes (2008).

Uma das classes de processos com longa dependência mais populares na literatura é a chamada classe de processos ARFIMA, sigla em inglês para processos *autoregressivos de médias móveis com integração fracionária*. ARFIMA é uma classe de processos lineares bastante flexível, de simples aplicação e com excelentes propriedades. Para mais detalhes, veja por exemplo Brockwell e Davis (1991), Taqqu (2003), Palma (2007), Lopes (2008) e referências ali contidas. A extensão multivariada da classe de processos ARFIMA é chamada classe de processos VARFIMA, sigla em inglês para processos *vetoriais autoregressivos de médias móveis com integração fracionária*. A classe de processos VARFIMA foi introduzida em Sowell (1989) e aplicações incluem modelagem da volatilidade em séries financeiras, modelagem e previsão de dados de alta frequência entre outros (veja por exemplo, Chiriac e Voev, 2011, Diongue, 2010 e referências ali contidas). Nossa contribuição à área de processos com longa dependência, desenvolvida no Capítulo 4, se concentra basicamente no caso multivariado, embora diversos resultados são válidos no caso unidimensional também.

Na Seção 4.1, estudamos o problema da estimação semiparamétrica do vetor de diferenciação fracionária $\boldsymbol{d}$ em processos multivariados com longa dependência, no qual a classe de processos VARFIMA é um caso particular, resultados estes contidos em Pumi e Lopes (2011b). O ponto de partida do trabalho é o artigo de Shimotsu (2007). Nele o autor define uma classe de estimadores baseados na otimização de uma certa função objetiva a qual utiliza o periodograma como estimador da função densidade espectral. Nossa contribuição é estender os resultados contidos em Shimotsu (2007) considerando a substituição do periodograma na função objetiva

primeiramente por um estimador consistente arbitrário da função densidade espectral e, em um segundo momento, por um estimador arbitrário da função densidade espectral. Propriedades assintóticas dos estimadores propostos são derivadas e uma simulação de Monte Carlo é realizada para acessar as propriedades em amostras finitas destes, bem como para comparação com o estimador original de Shimotsu (2007).

Na Seção 4.2, o objetivo passa a ser estudar a interdependência entre as coordenadas de processos VARFIMA$(0, \boldsymbol{d}, 0)$ bidimensionais sob o ponto de vista da distância de Mallows. A distância de Mallows, embora tenha aparecido em trabalhos anteriores, recebeu este nome após o artigo de Mallows (1972), que a utilizou como ferramenta para provar a normalidade assintótica para somas de variáveis aleatórias independentes. Várias aplicações para a distância de Mallows apareceram depois disso. Por exemplo, Bickel e Freedman (1981) utilizam a distância de Mallows para obter resultados assintóticos para o método bootstrap.

O trabalho apresentado na Seção 4.2 está contido no artigo Lopes et al. (2011) e é baseado em simulações de Monte Carlo de processos VARFIMA$(0, \boldsymbol{d}, 0)$ bidimensionais para diversas combinações do vetor de diferenciação $\boldsymbol{d}$ e ruídos dos mais diversos, incluindo os casos Gaussiano e não Gaussiano, marginais de cauda pesada entre outros. Simulado o processo, a distância de Mallows entre as componentes do processo é estimada. Para efeitos de comparação, o τ de Kendall entre as coordenadas também é calculado. O objetivo é estudar o comportamento da distância de Mallows focalizando numa possível relação entre esta, o vetor de diferenciação $\boldsymbol{d}$, o tipo e grau de dependência induzido no ruído, bem como o comportamento das marginais do processo. O estimador da distância de Mallows utilizado é introduzido no próprio artigo e é baseado na distribuição empírica das coordenadas do processo.

Nossa contribuição nessa seção pode ser dividida em três frentes. Primeiro, apresentamos os resultados da simulação de Monte Carlo em si, visando estudar o comportamento da distância de Mallows e do τ de Kendall em processos VARFIMA$(0, \boldsymbol{d}, 0)$ bidimensionais, sob diversas condições de simulação. Segundo, baseado nos resultados da simulação, propomos um estimador simples para o vetor de diferenciação $\boldsymbol{d}$ em processos VARFIMA$(0, \boldsymbol{d}, 0)$ de qualquer dimensão finita (incluindo a classe de processos ARFIMA$(0, d, 0)$, cuja dimensão é 1). Terceiro, ainda baseado nos resultados da simulação, propomos um teste para identificar a existência de coordenadas apresentando longa dependência forte em processos VARFIMA$(0, \boldsymbol{d}, 0)$ de qualquer dimensão finita m, inclusive $m = 1$. Isto é, propomos um teste para identificar se $d_i > 0.3$, para algum $i = 1, \cdots, m$, onde $\boldsymbol{d} = (d_1, \cdots, d_m)$. O desempenho tanto do estimador quanto do teste é verificado através de estudos de Monte Carlo específicos. Ainda relativo ao trabalho, na Seção 4.3 provamos a consistência forte e derivamos a distribuição assintótica do estimador para a distância de Mallows introduzido no artigo.

O terceiro grande tópico abordado no trabalho é o decaimento da correlação em processos estocásticos. A importância do estudo do decaimento da correlação está no fato de que este traz informações importantes com relação à características de longo prazo em um processo. Um decaimento rápido é típico de processos com dependência fraca, nos quais a dependência entre variáveis aleatórias distantes (em termos de índice) uma da outra é quase inexistente enquanto um decaimento lento da correlação significa a presença de uma interdependência importante entre variáveis distantes uma das outras. Na classe de processos ARFIMA, por exemplo, o decaimento da correlação pode ser exponencialmente rápido ou hiperbolicamente lento, dependendo do parâmetro de diferenciação d.

Uma das consequências do Teorema de Sklar (Teorema 2.2) é que dado um processo estocástico $\{X_t\}_{t\in\mathbb{N}}$ com funções de distribuição marginais absolutamente contínuas $\{F_t\}_{t\in\mathbb{N}}$ e denotando por $C_{r,s}$ a cópula de X_r e X_s, para todo $r, s \in \mathbb{N}$, para calcularmos a autoco-

variância entre duas variáveis aleatórias X_r e X_s basta utilizar a terna $(F_r, F_s, C_{r,s})$ junto com o Lema de Hoeffding para cópulas (veja Lema 2.1). Desta forma, para cada $t \in \mathbb{N}$, podemos calcular o comportamento da autocovariância entre X_t e X_{t+h}, para h grande, através da terna $(F_t, F_{t+h}, C_{t,t+h})$. Assumindo que as cópulas $C_{t,t+h}$ dependem somente de h e não de t, isto é, $C_h := C_{0,h} = C_{t,t+h}$ para todo $t \in \mathbb{N}$, e que pertencem a uma única família $\{C_\theta\}_{\theta\in\Theta}$, então existe uma sequência $\{\theta_h\}_{h\in\mathbb{N}} \subset \Theta$ tal que $C_h = C_{\theta_h}$, para cada $h \in \mathbb{N}$. Isto implica que, para cada $t \in \mathbb{N}$ fixo, podemos calcular o comportamento da autocovariância entre X_t e X_{t+h}, para h grande, através da terna $(F_t, F_{t+h}, C_{\theta_h})$.

No Capítulo 5 estudamos o problema inverso, tratado em Pumi e Lopes (2011c), que pode ser posto da seguinte forma. Dada uma família de cópulas $\{C_\theta\}_{\theta\in\Theta}$, uma família de funções de distribuições absolutamente contínuas $\{F_t\}_{t\in\mathbb{N}}$ e uma função real $R(h)$, com $R(h) \to 0$, quando h tende ao infinito, será que existe e, em caso afirmativo, como podemos obter uma parametrização $\{\theta_h\}_{h\in\mathbb{N}} \subset \Theta$ tal que a autocovariância calculada a partir de $(F_t, F_{t+h}, C_{\theta_h})$ se comporte como $R(h)$, para h grande? Em outros termos, dada uma sequência de funções de distribuições, uma função $R(h)$ como anteriormente a partir de uma família de cópulas $\{C_\theta\}_{\theta\in\Theta}$, estudamos como obter uma parametrização $\{\theta_h\}_{h\in\mathbb{N}} \subset \Theta$ tal que a autocovariância tenha comportamento $R(h)$.

Em última instância, levando-se em conta o problema da compatibilidade de cópulas, estudamos como construir um processo com marginais dadas e com um certo decaimento da autocovariância a partir da parametrização adequada de uma família de cópulas bidimensionais. O foco do trabalho são processos fracamente estacionários com longa dependência. A teoria, porém, comporta qualquer tipo de decaimento da autocovariância e também processos não estacionários. Mostramos ainda como a teoria se encaixa com a classe de processos ARFIMA. A construção de séries temporais unidimensionais a partir de cópulas já foi abordado, por exemplo, em Joe (1997) no caso de dependência fraca. Outros trabalhos que podemos citar na área, sempre no caso de dependência fraca, são Chen e Fan (2006), Darsow (1992) e Lagerås (2010). Nossa abordagem, porém, é bastante diferente dos artigos recém citados e, como observado anteriormente, comporta decaimentos arbitrários da autocovariância bem como processos não estacionários.

Baseado nos resultados obtidos, apresentamos duas aplicações. A primeira é o desenvolvimento de uma metodologia geral de estimação de parâmetros identificáveis através do decaimento da autocovariância em séries temporais. Um estudo de Monte Carlo é conduzido para exemplificar a metodologia e acessar seu desempenho em amostra finitas. A segunda aplicação é relacionada com a simulação de séries temporais com decaimento da autocovariância dado. Nossas contribuições à área são, portanto, a abordagem diferenciada do problema permitindo grande flexibilidade das técnicas, a metodologia de estimação desenvolvida e na simulação de séries temporais com características dadas.

O trabalho se divide da seguinte maneira. No próximo capítulo introduzimos alguns conceitos preliminares necessários ao trabalho. No Capítulo 3 apresentamos os resultados sobre cópulas relacionadas à processos de Manneville-Pomeau e à funções monótonas por partes e processos associados. O Capítulo 4 trata de processos com longa dependência. Consideramos o problema da estimação semiparamétrica em processos multivariados com longa dependência e também apresentamos um estudo de Monte Carlo a respeito da interdependência entre as coordenadas de um processo VARFIMA$(0, \boldsymbol{d}, 0)$ sob o ponto de vista da distância de Mallows e do τ de Kendall. O Capítulo 5 discute o problema do decaimento da correlação em processos estocásticos e a sua relação com a parametrização de cópulas. Conclusões e futuros trabalhos são reservados ao Capítulo 6.

Capítulo 2

Preliminares

Este capítulo tem por objetivo introduzir a terminologia básica, bem como introduzir alguns conceitos que serão necessários ao longo do trabalho. Nem todos os conceitos utilizados no decorrer do trabalho, porém, serão mencionados em detalhes e muitos deles assumimos conhecidos. Entre eles, assumimos que o leitor está familiarizado com conceitos elementares e notações básicas em estatística matemática, análise, medida, probabilidade e processos estocásticos. Iniciamos com alguns conceitos de análise, medida e integração e probabilidade. A notação utilizada será tão padrão quanto possível.

2.1 Análise, Medida e Probabilidade

Neste trabalho convencionamos que o conjunto dos números naturais contém o zero e denotamos $\mathbb{N} := \{0, 1, \cdots\}$ e $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. Os símbolos $\mathbb{Z}$, $\mathbb{R}$ e $\mathbb{C}$ denotam o conjunto dos números inteiros, reais e complexos, respectivamente. Para um conjunto $A \subseteq \mathbb{R}$, denotamos $A^m := A \times \cdots \times A$ (m termos), o fecho de A é denotado por $\overline{A}$ e o conjunto de todos os pontos de acumulação de A por A', isto é,

$$A' := \big\{a \in \overline{A} : \text{ existe uma sequência não-constante } \{a_n\}_{n\in\mathbb{N}} \subset A \text{ tal que } a_n \to a\big\}.$$

No decorrer do trabalho, utilizamos alguns conceitos básicos de medida geralmente presentes em um primeiro curso no assunto. Particularmente, Royden (1988) contém os resultados básicos necessários ao trabalho. Conceitos tais como σ-álgebra, espaço mensurável, espaço de medida, continuidade absoluta, teoremas básicos de convergência, Teorema de Radon-Nikodym, etc, são assumidos conhecidos. Recapitulamos alguns conceitos que podem não constar em um curso básico de medida e que serão necessários.

- Para uma função f, denotamos o domínio de f por $\mathrm{Dom}(f)$ e a imagem de f por $\mathrm{Im}(f)$.
- Para uma transformação T, denotamos $T^m := T \circ \cdots \circ T$ (m termos).
- Seja $(\Omega, \mathcal{A})$ um espaço mensurável e $f : \Omega \to \Omega$ uma função mensurável. Uma medida μ definida em $(\Omega, \mathcal{A})$ é dita ser f-*invariante* se $\mu\big(f^{-1}(A)\big) = \mu(A)$, para todo $A \in \mathcal{A}$.
- Uma função real $f : A \to \mathbb{R}$ é dita ser α-*Hölder contínua* em A, para $\alpha \in (0, 1)$, se existe $K > 0$ tal que
$$|f(x) - f(y)| \leq K|x - y|^\alpha,$$
para todo $x, y \in A$.

- Uma função $f : A \to \mathbb{R}$ é dita ser de classe $C^{1+\alpha}$, para $\alpha \in (0,1)$, se f é diferenciável e sua derivada é α-Hölder contínua em A.

- Se $f : A \to \mathbb{R}$ é uma função diferenciável em A e tal que $|f'(x)| \geq \kappa > 1$, para todo $x \in A$, dizemos que f é *uniformemente expansiva* em A.

Neste trabalho, exceto menção explícita ao contrário, sempre que trabalharmos nos espaços $\mathbb{R}, \mathbb{R}^n$ ou qualquer subconjunto X deles (particularmente, $X = [0,1]$), o espaço de medida subentendido em conceitos como função mensurável, medida absolutamente contínua, etc e frases envolvendo/contendo conceitos valendo "quase toda a parte", será $(X, \mathscr{B}(X), \lambda)$ onde $\mathscr{B}(X)$ denota a σ-álgebra de Borel em X e λ é a medida de Lebesgue em (ou restrita a) $\big(X, \mathscr{B}(X)\big)$. Alguns conceitos elementares de convergência fraca de medidas de probabilidade serão necessários.

- Seja μ uma medida de probabilidade definida em um espaço mensurável $(X, \mathscr{B}(X))$. Um boreliano $A \in \mathscr{B}(X)$ é dito ser um *conjunto de μ-continuidade* (*μ-continuity set*), se $\mu(\partial A) = 0$, onde ∂A denota a borda (*boundary*) de A.

- Seja $\{\mu_n\}_{n\in\mathbb{N}}$ uma sequência de medidas de probabilidade definidas em $(X, \mathscr{B}(X))$. Dizemos que μ_n *converge fracamente* para uma medida de probabilidade μ se

$$\mu_n(A) \longrightarrow \mu(A),$$

  quando n tende ao infinito, para todo conjunto de μ-continuidade A. Note que se $X = \mathbb{R}$ e definindo $F_n(x) = \mu_n\big((-\infty, x]\big)$ e $F(x) = \mu\big((-\infty, x]\big)$, então μ_n convergir fracamente para μ é o mesmo que F_n convergir para F para todo ponto de continuidade de F.

- Uma caracterização útil de convergência fraca é a seguinte. Seja $\{\mu_n\}_{n\in\mathbb{N}}$ uma sequência de medidas de probabilidade em $(X, \mathscr{B}(X))$. Então, μ_n converge fracamente para μ se, e somente se, $\int_X f d\mu_n \to \int_X f d\mu$, para toda função real f contínua e limitada em X.

Para mais detalhes, referimos Billingsley (1999). Alguns fatos básicos a respeito de funções de variação lenta, que revisamos agora, serão necessários.

Definição 2.1. Uma função $L : S \to \mathbb{R}$, para $S \subseteq \mathbb{R}$ um conjunto não-vazio, é dita ser de *variação lenta em um ponto $a \in S'$*, se L é uma função mensurável, limitada em um intervalo limitado e satisfaz

$$\lim_{x\to a} \frac{L(cx)}{L(x)} = 1,$$

para todo $c > 0$.

Exemplos de funções de variação lenta no infinito são:

- Funções constantes;

- Funções limitadas, mensuráveis e com limite não-nulo no infinito;

- Logaritmos, potências de logaritmos e composições de logaritmos;

- Um exemplo fora da classe dos logaritmos é $L(x) = \exp\left(\frac{\ln(x)}{\ln(\ln(x))}\right)$;

- Funções de variação lenta no infinito podem exibir oscilação infinita, no sentido que

$$\liminf_{n\to\infty} L(n) = 0 \qquad \text{mas} \qquad \limsup_{n\to\infty} L(n) = \infty,$$

como, por exemplo, $L(x) = \exp\Big(\big(\ln(x)\cos\big(\ln(x)\big)\big)^{\frac{1}{3}}\Big)$.

Se L, L_1, L_2 são funções de variação lenta no infinito, então $L_1 + L_2$ e $L_1 \cdot L_2$ são de variação lenta no infinito também e o mesmo acontece para kL sempre que $k \neq 0$. Além disso, se $\beta > 0$, então $L(x)x^{-\beta} \to 0$, quando x tende ao infinito.

Uma maneira simples de se construir uma função de variação lenta no infinito é a partir do Teorema da Representação de Karamata, o qual estabelece que uma função L é de variação lenta no infinito se, e somente se, L pode ser escrita como

$$L(x) = f(x)\,\exp\left(\int_s^x \frac{g(u)}{u}\,\mathrm{d}u\right), \quad x \geq s, \tag{2.1}$$

para algum $s > 0$, onde f e g são funções mensuráveis satisfazendo $\lim_{x\to\infty} f(x) = k$, $k > 0$ e $\lim_{x\to\infty} g(x) = 0$. Mais detalhes sobre a teoria de funções de variação lenta, bem como a prova do Teorema da Representação de Karamata e das propriedades acima descritas podem ser encontrados em Bingham et al. (1987).

Neste trabalho estamos interessados apenas no conjunto de funções de variação lenta no infinito. Seja L uma função de variação lenta no infinito e $\{a_n\}_{n\in\mathbb{N}}$ uma sequência arbitrária de números reais. Considere a função real $\psi_{a_n}(\cdot) := \psi\big(\,\cdot\,;\{a_n\}_{n\in\mathbb{N}}\big)$ definida em $(0,\infty)$ por

$$\psi_{a_n}(x) := a_{\lceil x\rceil},$$

onde $\lceil\cdot\rceil$ denota a função maior inteiro. Se $a_n \to a \neq 0$, então ψ_{a_n} é sempre de variação lenta no infinito. De fato, se $a > 0$, basta tomar

$$f(x) = \psi_{a_n}(x)e^{\frac{1}{x}-1}, \quad g(x) = -\frac{1}{x} \quad \text{e} \quad s = 1$$

em (2.1), para obtermos o resultado. Se $a_n \to a < 0$, basta considerar $-a_n$ e o resultado segue. Se $a = 0$ então nem sempre ψ_{a_n} é de variação lenta no infinito. Por exemplo, $a_n = \big(\log(n)\big)^{-1}$ é de variação lenta no infinito mas $a_n = n^{-1}$ não o é. Dizemos que, para uma função L de variação lenta no infinito e uma sequência de números reais $\{a_n\}_{n\in\mathbb{N}}$, a_nL é de variação lenta no infinito sempre que $\psi_{a_n}L$ o for.

Adotamos as seguintes definições para ordens assintóticas. Sejam f e g duas funções reais. Dado $a \in \mathbb{R}\cup\{\pm\infty\}$, dizemos que

- $f(x) = O\big(g(x)\big)$, quando $x \to a$, se, e somente se, existe uma constante $k > 0$ tal que $|f(x)| \leq k|g(x)|$, quando $x \to a$;
- $f(x) = o\big(g(x)\big)$, quando $x \to a$, se, e somente se, $\lim_{x\to a} \dfrac{f(x)}{g(x)} = 0$;
- $f(x) \sim g(x)$, quando $x \to a$, se, e somente se, $\lim_{x\to a} \dfrac{f(x)}{g(x)} = 1$, ou, equivalentemente, quando $f(x) - g(x) = o\big(g(x)\big)$, quando $x \to a$.

Sejam $\{\boldsymbol{U}_n\}_{n\in\mathbb{N}}$ e $\{\boldsymbol{V}_n\}_{n\in\mathbb{N}}$ duas sequências de vetores aleatórios definidos num mesmo espaço de probabilidade $(X, \mathscr{B}(X), \mathbb{P})$. No caso de ordens assintóticas em probabilidade, dizemos que

- $\boldsymbol{U}_n = o_{\mathbb{P}}(1)$ se, e somente se, $\boldsymbol{U}_n \xrightarrow{\mathbb{P}} 0$, quando $n \to \infty$;
- $\boldsymbol{U}_n = O_{\mathbb{P}}(1)$ se, e somente se, para todo $\varepsilon > 0$, existe $M < \infty$ e $n_0 > 0$ tal que, se $n > n_0$, $\mathbb{P}\big(|\boldsymbol{U}_n| \geq M\big) \leq \varepsilon$;
- $\boldsymbol{U}_n = o_{\mathbb{P}}\big(\boldsymbol{V}_n\big)$ se, e somente se, $\frac{|\boldsymbol{U}_n|}{|\boldsymbol{V}_n|} = o_{\mathbb{P}}(1)$;
- $\boldsymbol{U}_n = O_{\mathbb{P}}\big(\boldsymbol{V}_n\big)$ se, e somente se, $\frac{|\boldsymbol{U}_n|}{|\boldsymbol{V}_n|} = O_{\mathbb{P}}(1)$.

Para $\alpha > 0$, denotamos por $\mathscr{F}_\alpha$ o espaço de todas as funções de distribuição satisfazendo $\int_{\mathbb{R}} |x|^\alpha dF < \infty$.

Definição 2.2. Seja $\alpha > 0$ e $F, G \in \mathscr{F}_\alpha$. Considere $A(F, G)$ o espaço de todos os vetores aleatórios (X, Y) tais que $X \sim F$ e $Y \sim G$. Definimos a *distância α de Mallows* entre F e G por

$$\mathscr{D}_\alpha(F, G) := \inf_{A(F,G)} \Big\{\mathbb{E}\big(|X - Y|^\alpha\big)^{\frac{1}{\alpha}}\Big\}. \tag{2.2}$$

A distância de Mallows recebeu este nome após o trabalho de Mallows (1972). Para $\alpha \geq 1$, $\mathscr{D}_\alpha$ é uma métrica em $\mathscr{F}_\alpha$ enquanto que se $\alpha < 1$, $\mathscr{D}_\alpha^\alpha$ é uma métrica em $\mathscr{F}_\alpha$ (cf. Bickel e Freedman, 1981). Uma forma alternativa para (2.2) que é bastante vantajosa do ponto de vista computacional é a seguinte (Major, 1978):

$$\mathscr{D}_\alpha(F, G) = \int_0^1 \big|F^{-1}(u) - G^{-1}(u)\big|^\alpha du.$$

Para mais detalhes, referimos ao leitor Major (1978), Bickel e Freedman (1981), Shao (2003) e referências ali contidas.

Uma medida de dependência que utilizamos adiante é o τ de Kendall. Para definí-la precisamos dos conceitos de concordância e discordância. Considere X e Y duas variáveis aleatórias contínuas definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e sejam (X_1, Y_1) e (X_2, Y_2) duas cópias independentes de (X, Y). Dado $\omega \in \Omega$, dizemos que os pares $\big(X_1(\omega), Y_1(\omega)\big)$ e $\big(X_2(\omega), Y_2(\omega)\big)$ são *concordantes* se $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) > 0$ e *discordantes* se $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) < 0$. Note que, como a igualdade ocorre com probabilidade 0, não precisamos nos preocupar com esse caso.

Definição 2.3. A *medida de dependência τ de Kendall* ou, simplesmente, o *τ de Kendall*, entre X e Y, denotado por $\tau_{X,Y}$ (ou simplesmente τ se não houver perigo de confusão) é definido como a probabilidade de concordância menos a probabilidade de discordância entre (X_1, Y_1) e (X_2, Y_2), ou seja

$$\tau = \tau_{X,Y} := \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) > 0\big) - \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) < 0\big).$$

Mais detalhes sobre o τ de Kendall podem ser encontrados em Joe (1997), Mari e Kotz (2001) e Nelsen (2006).

2.2 Cópulas

Nesta seção compilamos alguns conceitos e resultados sobre cópulas que facilitarão o entendimento do trabalho. Lembramos que cópulas são funções que interligam uma determinada função de distribuição conjunta multivariada à suas marginais univariadas. Antes de definirmos formalmente o que são cópulas, precisamos de alguns conceitos.

- Um retângulo em $\mathbb{R}^m$ é o produto cartesiano de m intervalos fechados $[u_1, v_1] \times \cdots \times [u_m, v_m]$.
- Dado um retângulo $R \subseteq \mathbb{R}^m$, $R = [u_1, v_1] \times \cdots \times [u_m, v_m]$, definimos o *conjunto de vértices* de R como sendo o conjunto

$$A_m(R) = \{\boldsymbol{w} = (w_1, \cdots, w_m) \in R \text{ tais que } w_k = u_k \text{ ou } w_k = v_k,\ k = 1, \cdots, m\}.$$

- A *função co-sinal* de um vetor $\boldsymbol{w} \in \mathbb{R}^m$, denotada por $\operatorname{csgn}(\cdot)$, é definida como

$$\operatorname{csgn}(\boldsymbol{w}) = \begin{cases} 1, & \text{se } w_i = u_i \text{ para um número par de } i\text{'s (ou nenhum } i), \\ -1, & \text{se } w_i = u_i \text{ para um número ímpar de } i\text{'s.} \end{cases}$$

- Seja $H : \mathbb{R}^m \to \mathbb{R}$ e seja $R = [u_1, v_1] \times \cdots \times [u_m, v_m]$ um retângulo em $\mathbb{R}^m$. Definimos o $H-volume$ de R por

$$V_H(R) = \sum_{\boldsymbol{w} \in A_m(R)} \operatorname{csgn}(\boldsymbol{w}) H(\boldsymbol{w}). \tag{2.3}$$

Note que se H é uma função de distribuição, (2.3) é simplesmente a fórmula de "inclusão/exclusão" (veja, por exemplo, Tucker, 1967). Agora estamos em condições de definir formalmente cópulas.

Definição 2.4. Seja $m \in \mathbb{N}$, $m \geq 2$. Uma *cópula* $m-dimensional$ é uma função C que satisfaz as seguintes condições:

(i). $\text{Dom}(C) = I^m$;

(ii). Para todo o retângulo $R \subseteq I^m$, $V_C(R) \geq 0$;

(iii). Dado um vetor $\boldsymbol{u} = (u_1, \cdots, u_m) \in I^m$, se $u_i = 0$ para algum $i \in \{1, \cdots, m\}$, então $C(\boldsymbol{u}) = 0$;

(iv). Para todo $k \in \mathbb{N}$, $k \leq m$, $C(1, \cdots, 1, u_k, 1, \cdots, 1) = u_k$.

O próximo teorema é um resumo de algumas propriedades importantes relacionadas com cópulas. A prova pode ser encontrada em Joe (1997), Schweizer e Sklar (2005), Nelsen (2006) e Pumi (2006).

Teorema 2.1. *Seja C uma $m-$cópula qualquer, $m \geq 2$. Então, C é uniformemente contínua, não-decrescente em cada coordenada e as derivadas parciais $\frac{\partial C(u_1, \cdots, u_m)}{\partial u_k}$, $k = 1, \cdots, m$, existem, tomam valores em I e são não-decrescentes em quase toda a parte. Além disso, dado $\boldsymbol{u} \in I^m$,*

$$W^m(\mathbf{u}) \leq C(\mathbf{u}) \leq M^m(\mathbf{u}),$$

onde $W^m(\boldsymbol{u}) := \max\{u_1 + \cdots + u_m - m + 1, 0\}$ e $M^m(\boldsymbol{u}) := \min\{u_1, \cdots, u_m\}$ são os chamados limite superior e limite inferior de Frèchet-Hoeffding, respectivamente.

Cabe observar que M^m é uma cópula, qualquer que seja $m \geq 2$, enquanto que W^m é cópula se, e somente se, $m = 2$, como pode ser facilmente verificado.

O próximo teorema é conhecido como Teorema de Sklar e é o teorema fundamental em cópulas. Um esboço da prova para o caso n-dimensional pode ser encontrado em Schweizer e Sklar (2005), uma prova alternativa mais detalhada pode ser encontrada em Carley e Taylor (2002) e uma prova relativamente simples para o caso bidimensional pode ser encontrada em Nelsen (2006). Enunciamos o teorema apenas para o caso em que as marginais envolvidas são funções contínuas, já que neste trabalho estudamos exclusivamente este caso.

Teorema 2.2. *Sejam $X_1, \cdots, X_m$ variáveis aleatórias com distribuição conjunta H e marginais contínuas $F_1, \cdots, F_m$, respectivamente. Então, existe uma única $m-$cópula C tal que,*

$$H(x_1, \cdots, x_m) = C\big(F_1(x_1), \cdots, F_m(x_m)\big).$$

Além disso, a cópula C é dada por

$$C(u_1, \cdots, u_m) = H\big(F_1^{-1}(u_1), \cdots, F_m^{-1}(u_m)\big), \tag{2.4}$$

para todo $(u_1, \cdots, u_m) \in I^m$.

Sejam $X_1, \cdots, X_n$ variáveis aleatórias contínuas com função de distribuição conjunta H e marginais contínuas $\{F_k\}_{k=1}^n$. Naturalmente, podemos associar ao conjunto $\{H, F_1, \cdots, F_n\}$ uma única cópula C a partir do Teorema de Sklar via (2.4). Neste caso, dizemos que a cópula C é a cópula associada a $X_1, \cdots, X_n$ e muitas vezes escrevemos $C_{X_1,\cdots,X_n}$ para deixar este fato evidenciado. Outra cópula importante é a chamada *cópula independência* dada por $\Pi(\boldsymbol{u}) := \prod_{k=1}^n u_k$. A cópula independência recebe este nome pois a cópula associada a um vetor aleatório $(X_1, \cdots, X_n)$ é Π se, e somente se, $X_1, \cdots, X_n$ são variáveis aleatórias independentes.

O próximo lema é a versão em termos de cópulas do celebrado Lema de Hoeffding. Uma prova detalhada pode ser encontrada em Pumi (2006) e referências ali contidas.

Lema 2.1. Sejam X e Y duas variáveis aleatórias contínuas com marginais F e G, respectivamente, e cópula associada C. Então,

$$\mathrm{Cov}(X, Y) = \iint_{I^2} \frac{C(u, v) - uv}{F'\big(F^{-1}(u)\big)G'\big(G^{-1}(v)\big)}\, dudv, \tag{2.5}$$

onde F' e G' denotam a derivada de F e G, respectivamente.

Um fato interessante e muito útil a respeito das cópulas é a invariância em relação à transformações crescentes em quase toda a parte e a forma simples que as cópulas assumem em relação à transformações decrescentes em quase toda a parte. Este é o conteúdo da próxima proposição cuja prova pode ser encontrada em Nelsen (2006) e Pumi (2006).

Proposição 2.1. *Seja $C_{X_1,\cdots,X_m}$ uma cópula $m-$dimensional associada a $X_1, \cdots, X_m$ e sejam $f_1, \cdots, f_m$ funções reais crescentes em quase toda a parte. Então, em todo I^m temos $C_{f_1(X_1),\cdots,f_m(X_m)} = C_{X_1,\cdots,X_m}$. Se f e g são funções decrescentes em quase toda a parte e $C_{X,Y}$ é uma cópula, então,*

$$C_{f(X),Y}(u, v) = v - C_{X,Y}(1 - u, v) \quad e \quad C_{f(X),g(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v),$$

para todo $(u, v) \in I^2$.

Para maiores detalhes sobre a teoria de cópulas, referenciamos o leitor a Joe (1997), Mari e Kotz (2001), Cherubini et al. (2004), Nelsen (2006) e Pumi (2006). Para maiores detalhes sobre a conexão entre cópulas e espaços métricos probabilísticos, referimos Schweizer e Sklar (2005).

2.3 Análise de Séries Temporais

Nesta seção introduzimos a terminologia associada às noções básicas em séries temporais necessárias para o desenvolvimento do trabalho. Em momentos diferentes do trabalho, precisamos tanto de conceitos sobre séries temporais univariadas quanto multivariadas. Tais conceitos são introduzidos separadamente para facilitar o entendimento. Ressaltamos, porém, que nossa intenção aqui não é ser exaustivo na exposição, por isso, ao leitor interessado em uma exposição mais completa, referências são sugeridas.

2.3.1 Séries Temporais Univariadas

Iniciamos com conceitos básicos na análise de séries temporais univariadas.

Definição 2.5. Seja $\{X_t\}_{t\in\mathbb{N}}$ um processo com variância finita (isto é, $\mathrm{Var}(X_t) < \infty$, para todo $t \in \mathbb{N}$) e tal que $\mathbb{E}(X_t) = \mu$, para todo $t \in \mathbb{N}$. Definimos a *função de autocovariância* de $\{X_t\}_{t\in\mathbb{N}}$ por

$$\gamma_X(r,s) := \mathrm{Cov}(X_r, X_s) = \mathbb{E}\big((X_r-\mu)(X_s-\mu)\big), \quad r,s \in \mathbb{N}.$$

Note que $\mathrm{Cov}(X_r, X_s)$ em geral depende de r e s. No caso especial em que $\mathrm{Cov}(X_t, X_{t+h})$ depende apenas de h para todo t, então $\{X_t\}_{t\in\mathbb{N}}$ é dito ser um processo *fracamente estacionário*. Neste caso, denotamos $\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h})$. Outra noção de estacionariedade importante é a seguinte. Um processo estocástico $\{X_t\}_{t\in\mathbb{N}}$ é dito ser *fortemente estacionário* se a distribuição de $(X_{t_1}, \cdots, X_{t_n})$ e $(X_{t_1+h}, \cdots, X_{t_n+h})$ é a mesma, para todo $n \in \mathbb{N}^*$, e para todo $t_1, \cdots, t_n, h \in \mathbb{N}$. Note que um processo fortemente estacionário com segundo momento finito é fracamente estacionário, mas a recíproca não é verdadeira, exceto quando o processo é Gaussiano (isto é, quando todas as distribuições finito-dimensionais do processo são Gaussianas).

Definição 2.6. Seja $\{X_t\}_{t\in\mathbb{N}}$ um processo com variância finita. Definimos a *função de autocorrelação* de $\{X_t\}_{t\in\mathbb{N}}$ por

$$\rho_X(r,s) := \mathrm{Corr}(X_r, X_s) = \frac{\mathrm{Cov}(X_r, X_s)}{\sqrt{\mathrm{Var}(X_r)}\sqrt{\mathrm{Var}(X_s)}}, \quad r,s \in \mathbb{N}.$$

Note que se $\{X_t\}_{t\in\mathbb{N}}$ for fracamente estacionário, então $\mathrm{Corr}(X_t, X_{t+h}) = \mathrm{Corr}(X_0, X_h)$ para todo $t, h \in \mathbb{N}$ e, neste caso, denotamos $\rho_X(h) = \mathrm{Corr}(X_t, X_{t+h})$. Para mais detalhes, veja, por exemplo, a seção 1.3 em Brockwell e Davis (1991) e também o capítulo 3 de Priestley (1981).

No que segue, recapitulamos alguns resultados importantes na análise espectral de processo fracamente estacionário $\{X_t\}_{t\in\mathbb{N}}$. O próximo teorema é chamado de Teorema de Herglotz e traz uma condição necessária e suficiente para que uma determinada função $\gamma(\cdot)$ possa ser uma função de autocovariância. Como consequência, fica também definida a função de distribuição espectral do processo $\{X_t\}_{t\in\mathbb{N}}$. Uma prova pode ser encontrada em Brockwell e Davis (1991).

Teorema 2.3. *Uma função $\gamma : \mathbb{C} \to \mathbb{Z}$ é não negativa-definida (e, portanto, pode representar uma função de autocovariância) se, e somente se,*

$$\gamma(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} dF(\lambda), \tag{2.6}$$

para todo $h \in \mathbb{Z}$, onde F é uma função limitada em $(-\pi, \pi]$, contínua à direita, não-decrescente e tal que $F(-\pi) = 0$.

A função F em (2.6) é de importância fundamental para a análise espectral de um processo $\{X_t\}_{t\in\mathbb{N}}$.

Definição 2.7. Seja $\{X_t\}_{t\in\mathbb{N}}$ um processo fracamente estacionário com função de autocovariância $\gamma_X(\cdot)$ e denote por F_X a função definida pela decomposição de γ_X em (2.6). A função F_X é chamada de *função de distribuição espectral* de $\{X_t\}_{t\in\mathbb{N}}$. Se F_X for absolutamente contínua, a função $f_X := dF_X$ é chamada de *função densidade espectral*, ou simplesmente, a *densidade espectral* de $\{X_t\}_{t\in\mathbb{N}}$. Além disso, nesse caso

$$f_X(\lambda) := \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} e^{i\lambda h} \gamma_X(h), \qquad \text{para todo } \lambda \in (-\pi, \pi].$$

Decorre imediatamente da definição que a densidade espectral é uma função contínua, par, positiva e integrável em $(-\pi, \pi]$. Mais detalhes podem ser encontrados no capítulo 4 de Brockwell e Davis (1991) e no capítulo 4 de Priestley (1991).

Consideramos agora dois estimadores para a função densidade espectral de um processo fracamente estacionário que serão úteis adiante. Lembrando que as *frequências de Fourier* entre $-\pi$ e π compreendem os números $\lambda_j := \frac{2\pi j}{n}$, para $j = \pm 1, \cdots \pm [n/2]$, onde $n \in \mathbb{N}^*$ é dado e $[x]$ denota a parte inteira de x.

Definição 2.8. Seja $X_1, \cdots, X_n$ uma série temporal obtida a partir de um processo estocástico fracamente estacionário $\{X_t\}_{t\in\mathbb{N}}$ (ou, abusando da linguagem, como fazemos repetidas vezes, simplesmente uma realização de um processo estocástico). Definimos a *função periodograma*, ou simplesmente, o *periodograma* de $X_1, \cdots, X_n$ nas frequências de Fourier $\lambda_j = 2\pi j/n$ por

$$I_n(\lambda_j) := \frac{1}{2\pi n} \Big| \sum_{t=1}^{n} X_t e^{\mathrm{i}t\lambda_j} \Big|^2.$$

Se $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty$, então o periodograma é assintóticamente não-viciado para a função densidade espectral mas ele não é consistente (veja, por exemplo, o capítulo 6 de Priestley, 1981 e a seção 10.3 de Brockwell e Davis, 1991). A hipótese de que a autocovariância do processo é absolutamente somável é crucial e não pode ser relaxada (cf. Hurvich e Beltrão, 1993).

Embora o periodograma pode ser considerado um estimador "natural" da função densidade espectral, alguns autores o consideram (em tradução livre) "um estimador extremamente pobre (se não inútil) da densidade espectral" (cf. Priestley, 1981, pag.420). A seção 6.2.2 em Priestley (1981) traz uma excelente discussão sobre o assunto. De qualquer forma, por ser um estimador simples e com algumas propriedades boas, o periodograma ainda é muito difundido e aplicado na prática e é especialmente útil para a identificação de processos com longa dependência. Muitos autores definem o periodograma sem a constante multiplicadora $\frac{1}{2\pi}$, mas sabemos da teoria de séries de Fourier que a constante é mera questão de gosto, fazendo diferença apenas na hora de

aplicar a fórmula de inversão. Aqui definimos o periodograma com a constante multiplicadora por coerência aos artigos que utilizamos como apoio para este trabalho.

Uma classe de estimadores com propriedades muito superiores aos do periodograma é obtida através de uma ponderação do periodograma em frequências adjacentes à frequência desejada.

Definição 2.9. Seja $X_1, \cdots, X_n$ uma realização de um processo $\{X_t\}_{t\in\mathbb{N}}$ fracamente estacionário. Definimos a *função periodograma suavizado*, ou simplesmente, o *periodograma suavizado* de $X_1, \cdots, X_n$ nas frequências de Fourier $\lambda_j = 2\pi j/n$ por

$$\hat{f}(\lambda_j) := \sum_{|k|\leq \ell(n)} W_n(k) I_n(\lambda_{k+j}),$$

onde $\{W_n(\cdot)\}$ é uma sequência de funções de ponderação satisfazendo $W_n(k) \geq 0$ e $W_n(k) = Wn(-k)$, para todo $k = \pm 1, \cdots, \pm\ell(n)$, e ainda $\sum_{|k|\leq\ell(n)} W_n(k) = 1$, onde $\ell(n) \in \mathbb{N}$, $\ell(n) < n$ é o ponto de truncamento. Se $\lambda_{j+k} \notin [-\pi, \pi]$, tomamos o periodograma como tendo período 2π.

Na Definição 2.9 definimos o periodograma suavizado apenas para as frequências de Fourier λ_j em $[-\pi, \pi]$. É vantajoso, porém, definí-lo para todos os números reais. Para isso utilizamos o seguinte artifício. Primeiro, para $\lambda \in [0, \pi]$, consideramos a função auxiliar

$$g(\lambda; n) := \begin{cases} \lambda_{k_0-1}, & \text{se} \quad \lambda - \lambda_{k_0-1} \leq \lambda_{k_0} - \lambda; \\ \lambda_{k_0}, & \text{se} \quad \lambda - \lambda_{k_0-1} > \lambda_{k_0} - \lambda. \end{cases}$$

Em seguida, se $\lambda \in [-\pi, 0)$, definimos $g(\lambda; n) := g(-\lambda; n)$. Então, se λ não é uma frequência de Fourier, estendemos o periodograma suavizado definindo

$$\hat{f}_n(\lambda) := \hat{f}_n\big(g(\lambda; n)\big).$$

Note que o periodograma também pode ser estendido dessa maneira, ou seja, definindo $I_n(\lambda) := I_n\big(g(\lambda; n)\big)$ quando λ não é uma frequência de Fourier. Na prática, utilizamos uma estimativa um pouco diferente para a frequência zero. Mais especificamente, utilizamos

$$\hat{f}_n(0) = W_n(0) I_n(\lambda_1) + 2 \sum_{k=1}^{\ell(n)} W_n(k) I_n(\lambda_{k+1}).$$

Observamos que os autores os quais definem o periodograma sem a constante multiplicadora $\frac{1}{2\pi}$, a incorporam na definição do periodograma suavizado.

Se além das condições da Definição 2.9, a função de autocovariância de um processo fracamente estacionário satisfaz $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty$ e $\sum_{|k|\leq\ell(n)} W_n^{ij}(k)^2 \longrightarrow 0$ e ainda $\frac{1}{\ell(n)} + \frac{\ell(n)}{n} \longrightarrow 0$, quando n tende ao infinito, então o periodograma suavizado é consistente para a densidade espectral. Veja, por exemplo, Priestley (1981) e Brockwell e Davis (1981). Novamente $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty$ é uma condição crucial e sem ela o resultado não é mais verdadeiro.

Existem diversas funções de ponderação populares na literatura. Na seção 6.2.3 de Priestley (1981), diversas funções de ponderação são discutidas e suas propriedades amostrais verificadas. Neste trabalho, utilizamos apenas a função de ponderação de Bartlett, dada por

$$W_n(k) := \frac{\sin^2(nk/2)}{2\pi n \sin^2(k/2)}.$$

2.3.2 Séries Temporais Multivariadas

Esta seção tem por objetivo introduzir a terminologia mínima relacionada a séries temporais multivariadas necessárias ao trabalho. Primeiramente, para uma matriz complexa M, denotamos por M' a transposta de M e por $\overline{M}'$ a transposta conjugada de M. Vetores aleatórios serão sempre denotados em negrito e as coordenadas de um processo m-dimensional serão denotadas por $\boldsymbol{X}_t = \big(X_t^{(1)}, \cdots, X_t^{(m)}\big)$, para todo $t \in \mathbb{N}$.

Definição 2.10. Seja $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ um processo m-dimensional com $\mathbb{E}\big((X_t^{(i)})^2\big) < \infty$, para todo $t \in \mathbb{N}$ e todo $i \in \{1, \cdots, m\}$. Definimos a *função de autocovariância* de $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ por

$$\boldsymbol{\Gamma}(r,s) := \mathrm{Cov}(\boldsymbol{X}_r, \boldsymbol{X}_s) := \mathbb{E}\big((\boldsymbol{X}_r - \mathbb{E}(\boldsymbol{X}_r))(\boldsymbol{X}_s - \mathbb{E}(\boldsymbol{X}_s))'\big), \quad r,s \in \mathbb{N}.$$

No caso multivariado, o conceito de estacionariedade fraca será suficiente para nossos propósitos.

Definição 2.11. Seja $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ um processo m-dimensional com $\mathbb{E}\big((X_t^{(i)})^2\big) < \infty$, para todo $t \in \mathbb{N}$ e todo $i \in \{1, \cdots, m\}$. Se $\mathbb{E}(\boldsymbol{X}_t) = \mathbb{E}(\boldsymbol{X}_0)$, para todo $t \in \mathbb{N}$, e $\mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t+h}) = \mathrm{Cov}(\boldsymbol{X}_0, \boldsymbol{X}_h)$, para todo $t \in \mathbb{N}$ e $h \in \mathbb{N}^*$, dizemos que o processo é *fracamente estacionário*. Neste caso, denotamos $\boldsymbol{\Gamma}(h) := \mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t+h})$.

No caso fracamente estacionário, a função de autocovariância $\boldsymbol{\Gamma}(h)$ será uma matriz real, finita, simétrica e positiva definida para todo h fixo. Além disso, escrevendo $\boldsymbol{\Gamma}(h)$ em forma matricial $\boldsymbol{\Gamma}(h) := (\gamma_{i,j}(h))_{i,j=1}^m$, então a função $\gamma_{r,s}(h)$, para $r, s \in 1, \cdots, m$ e $h \in \mathbb{N}$, nada mais é do que a função de *covariância cruzada* entre $X_t^{(r)}$ e $X_{t+h}^{(s)}$. Mais detalhes podem ser encontrados em Hannan (1970) e Brockwell e Davis (1991).

A definição das funções distribuição e densidade espectral para o caso multivariado é análogo ao caso univariado. A versão multivariada do Teorema de Herglotz nos diz que se $\boldsymbol{\Gamma}_X$ é a função matriz de autocovariância de um processo m-dimensional fracamente estacionário $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$, então $\boldsymbol{\Gamma}_X$ pode ser escrita como

$$\boldsymbol{\Gamma}_X(h) = \int_{(-\pi,\pi]} e^{\mathrm{i}h\lambda} d\boldsymbol{F}_X, \tag{2.7}$$

para todo $h \in \mathbb{Z}$, onde $\boldsymbol{F}_X$ é uma função limitada em $(-\pi, \pi]$, contínua à direita, não-decrescente e tal que $\boldsymbol{F}_X(-\pi) = 0$. Dizemos que $\boldsymbol{F}_X$, assim definida, é a *função matriz distribuição espectral* de $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$.

Definição 2.12. Seja $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ um processo m-dimensional fracamente estacionário com função de autocovariância $\boldsymbol{\Gamma}(h)$. Se $\boldsymbol{F}_X$ dada em (2.7) for absolutamente contínua, a função $\boldsymbol{f}_X := d\boldsymbol{F}_X$ é chamada de *função matriz densidade espectral*, ou simplesmente, a *matriz densidade espectral* de $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$. Além disso, nesse caso

$$\boldsymbol{f}_X(\lambda) := \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} e^{i\lambda h} \boldsymbol{\Gamma}_X(h), \quad \text{para todo } \lambda \in (-\pi, \pi].$$

O periodograma no caso multidimensional é meramente uma extensão matricial do caso unidimensional.

Definição 2.13. Seja $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ uma realização de um processo estocástico fracamente estacionário $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$. Definimos a *função periodograma*, ou simplesmente, o *periodograma*

de $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ nas frequências de Fourier $\lambda_j = 2\pi j/n$ por

$$\boldsymbol{I}_n(\lambda_j) := w_n(\lambda_j)\overline{w_n(\lambda_j)}', \quad \text{onde} \quad w_n(\lambda_j) := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{X}_t e^{\mathrm{i}t\lambda_j}.$$

O periodograma definido acima goza das mesmas propriedades assintóticas que seu análogo unidimensional não sendo, portanto, consistente. Para definirmos o periodograma suavizado no caso multivariado, lembramos que se $A = (a_{ij})_{ij}$ e $B = (b_{ij})_{ij}$ são duas matrizes $m \times n$ complexas, o *produto de Hadamard* entre A e B, denotado $A \odot B$, é definido como sendo a matriz $m \times n$ dada por $A \odot B := (a_{ij}b_{ij})_{i,j}$. Isto é, o produto de Hadamard entre duas matrizes nada mais é do que o produto entrada por entrada. O produto de Hadamard é comutativo, associativo e distributivo. Se A e B são duas matrizes positivas (semi-)definidas então (*Teorema de Shur*) $A \odot B$ é positiva (semi-)definida e (*desigualdade de Oppenheimer*) $\det(A \odot B) \geq \det(A)\det(B)$. A identidade do produto de Hadamard é a matriz cujas entradas são todas iguais a 1. Mais detalhes podem ser encontrados no capítulo 5 de Horn e Johnson (1991).

Definição 2.14. Seja $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ uma realização de um processo estocástico m-dimensional fracamente estacionário $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$. Seja $\boldsymbol{W}_n(\cdot) := \left(W_n^{i,j}(\cdot)\right)_{i,j=1}^{m}$ uma matriz de funções de ponderação satisfazendo $W_n^{ij}(k) = W_n^{ij}(-k)$, $W_n^{ij}(k) \geq 0$ e $\sum_{|k|\leq \ell(n)} W_n^{ij}(k) = 1$, para todo $i, j = 1, \cdots m$ e $k = \pm 1, \cdots, \pm\ell(n)$, onde $\ell(n) \in \mathbb{N}$, $\ell(n) < n$. Definimos a *função periodograma suavizado*, ou simplesmente, o *periodograma suavizado* de $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ por

$$\widehat{\boldsymbol{f}}_n(\lambda_j) := \sum_{|k|\leq \ell(n)} \boldsymbol{W}_n(k) \odot w_n(\lambda_{j+k})\overline{w_n(\lambda_{j+k})}', \tag{2.8}$$

onde tomamos w_n como tendo período 2π se $\lambda_{j+k} \notin (-\pi, \pi]$.

A extensão do periodograma suavizado multivariado para o caso em que $\lambda \in \mathbb{R}$ é idêntica ao caso unidimensional. A maioria dos autores define o periodograma suavizado multivariado como tendo uma única função de ponderação comum a todas as entradas. Note porém que em (2.8) permitimos que sejam usados funções de ponderação diferentes em cada entrada da matriz. Assim como no caso unidimensional, na frequência zero utilizamos a seguinte expressão para o periodograma suavizado multivariado:

$$\widehat{\boldsymbol{f}}_n(0) := \mathrm{Re}\left[\boldsymbol{W}_n(0) \odot w_n(\lambda_1)\overline{w_n(\lambda_1)}' + 2\sum_{k=1}^{\ell(n)} \boldsymbol{W}_n(k) \odot w_n(\lambda_{k+1})\overline{w_n(\lambda_{k+1})}'\right].$$

Se $\sum_{k\in\mathbb{Z}} |\gamma_{i,j}(k)| < \infty$, para todo i e j, e cada função de ponderação W_n^{ij}, $i, j = 1 \cdots, m$, satisfizer as mesmas condições necessárias para a consistência do periodograma suavizado no caso unidimensional, então o periodograma suavizado multivariado é consistente para a matriz densidade espectral. Veja Priestley (1981).

2.3.3 Longa Dependência

Desde sua "descoberta", reportada em Hurst (1951), o estudo de processos com longa dependência evoluiu muito e hoje estes representam uma parte importante da análise de séries temporais. O decaimento lento da autocovariância típico em processos com longa dependência produz efeitos dramáticos em estatísticas comumente utilizadas. Apenas para exemplificar,

lembramos que quando a autocovariância é absolutamente somável, sob certas condições nas funções de ponderação, o periodograma suavizado é $\sqrt{n}$-consistente para a densidade espectral (Priestley, 1981), mas no caso de longa dependência, essa propriedade não é mais verdadeira (Hurvich e Beltrão, 1993). Estes e outros problemas frequentes e, ao mesmo tempo, únicos na análise de séries temporais com longa dependência despertaram o interesse de muitos pesquisadores e fizeram com que houvesse um grande desenvolvimento na área nas últimas décadas, tanto em termos de teoria quanto em aplicações. Uma amostra deste desenvolvimento pode ser encontrada em Doukhan et al. (2003), Palma (2007), Lopes (2008) e referências ali contidas.

Embora tendo uma história relativamente longa, longa dependência ainda não possui uma definição globalmente aceita. Diversas maneiras de se definir e caracterizar longa dependência e condições para que as definições sejam equivalentes são discutidas na seção 4 de Taqqu (2003) e no capítulo 4 de Palma (2007). Neste trabalho adotamos a seguinte definição de processo com longa dependência.

Definição 2.15. Um processo fracamente estacionário $\{X_n\}_{n\in\mathbb{N}}$ é dito possuir *longa dependência* se

$$\gamma_X(h) \sim n^{-\beta}L(n), \tag{2.9}$$

quando n tende ao infinito, para algum $\beta \in (0,1)$ e alguma função L de variação lenta no infinito.

Observamos também que processos com longa dependência são comumente associados à presença de uma singularidade na função densidade espectral na origem e também à condição $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| = \infty$.

Uma das classes mais difundidas de modelos com longa dependência é a dos modelos ARFIMA, que revisamos a seguir. Denotamos por $\mathcal{B}$ o operador shift, isto é, $\mathcal{B}^k(X_t) = X_{t-k}$, $k \in \mathbb{N}^*$.

Definição 2.16. Um processo estocástico $\{X_t\}_{t\in\mathbb{N}}$ é dito ser um processo ARFIMA(p,d,q) se $\{X_t\}_{t\in\mathbb{N}}$ for uma solução fracamente estacionária de

$$\varphi(\mathcal{B})(1-\mathcal{B})^d X_t = \vartheta(\mathcal{B})Z_t, \tag{2.10}$$

onde $\varphi(z) := 1 - \varphi_1 z - \cdots - \varphi_p z^p$ e $\vartheta(z) := 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$ são polinômios que assumimos não ter raízes em comum, $\{Z_t\}_{t\in\mathbb{Z}}$ é um *ruído branco*, isto é, uma sequência de variáveis aleatórias não-correlacionadas com média 0 e segundo momento finito, e $(1-\mathcal{B})^d$ é definido pela sua expansão binomial $(1-\mathcal{B})^d := \sum_{k\in\mathbb{N}} \pi_k \mathcal{B}^k$, onde $\pi_0 := 1$ e $\pi_k := \prod_{j=1}^{k} \frac{j-1-d}{j}$, para todo $k \in \mathbb{N}^*$.

Pode ser mostrado que, para $d \in (-1, 0.5)$, se os polinômios φ e ϑ não têm raízes no disco unitário $\{z : |z| \leq 1\}$, então a solução de (2.10) é fracamente estacionária, causal e invertível. Neste caso, $\{X_t\}_{t\in\mathbb{N}}$ terá uma representação MA(∞) dada por $X_t = \sum_{k\in\mathbb{N}} c_k Z_{t-k}$, para todo $t \in \mathbb{N}$, onde a sequência $\{c_k\}_{k\in\mathbb{N}}$ é determinada através da expansão $(1-z)^{-d}\vartheta(z)/\varphi(z) = \sum_{k\in\mathbb{N}} c_k z^k$ e satisfaz $\sum_{k\in\mathbb{N}} c_k^2 < \infty$. Além disso, $\gamma_X(h) \sim Kh^{2d-1}$, para h grande. Desta forma, se $d \in (0, 0.5)$, $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| = \infty$ e o processo $\{X_t\}_{t\in\mathbb{N}}$ apresenta longa dependência no sentido da equação (2.9) com $\beta = 1-2d$. Se $d \in (-1, 0)$ então a função de autocovariância é absolutamente somável, $\sum_{k\in\mathbb{N}} |c_k| < \infty$ e dizemos que $\{X_t\}_{t\in\mathbb{N}}$ possui *dependência intermediária*. Se $d = 0$, então (2.10) se reduz a um processo ARMA(p,q), que possui *curta dependência*. Também pode-se mostrar que a densidade espectral de um processo ARFIMA(p,d,q)

satisfaz $f_X(\lambda) \sim K\lambda^{-2d}$, quando λ tende a 0, e, desta forma, na presença de longa dependência a densidade espectral é ilimitada na origem. Mais detalhes podem ser encontrados no capítulo 13 de Brockwell e Davis (1991), no capítulo 3 de Palma (2007) e em Lopes (2008).

A extensão multivariada "natural" dos processos ARFIMA(p, d, q) constitui a classe de processos conhecidos por VARFIMA$(p, \boldsymbol{d}, q)$.

Definição 2.17. Um processo estocástico m-dimensional $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$ com $\mathbb{E}(\boldsymbol{X}_t) = \boldsymbol{\mu}$, para todo $t \in \mathbb{N}$, é dito ser um VARFIMA$(p, \boldsymbol{d}, q)$ se for uma solução fracamente estacionária de

$$\boldsymbol{\Phi}(\mathcal{B})\,\mathrm{diag}\big\{(1-\mathcal{B})^{\boldsymbol{d}}\big\}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(\mathcal{B})\boldsymbol{\varepsilon}_t, \tag{2.11}$$

onde $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias m-dimensionais não-correlacionadas, com média 0 e variância finita, $\boldsymbol{\Phi}(\mathcal{B})$ e $\boldsymbol{\Theta}(\mathcal{B})$ são matrizes $m \times m$ em $\mathcal{B}$, dadas pelos polinômios

$$\boldsymbol{\Phi}(\mathcal{B}) = \sum_{\ell=0}^{p} \boldsymbol{\phi}_\ell \mathcal{B}^\ell \quad \text{e} \quad \boldsymbol{\Theta}(\mathcal{B}) = \sum_{\ell=0}^{q} \boldsymbol{\theta}_\ell \mathcal{B}^\ell,$$

os quais assumimos não conter raízes comuns, onde $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_q$ são matrizes reais $m \times m$ e $\boldsymbol{\phi}_0 = \boldsymbol{\theta}_0 = \boldsymbol{I}_{m\times m}$ é a matriz identidade $m \times m$.

Pode ser mostrado que se $\boldsymbol{d} \in (-0.5, 0.5)^m$, $\det\big(\boldsymbol{\Theta}(z)\big) \neq 0$ e $\det\big(\boldsymbol{\Phi}(z)\big) \neq 0$, para todo z no disco unitário, a solução fracamente estacionária de (2.11) é também causal e invertível. No caso de um processo VARFIMA, sempre que $d_i \in (0, 0.5)$, dizemos que a i-ésima componente $\{X_t^{(i)}\}_{t\in\mathbb{N}}$ do processo possui longa dependência. De fato, cada componente é um ARFIMA(p, d_i, q). Desta forma, as propriedades mencionadas para o caso do ARFIMA valem para cada componente do processo VARFIMA$(p, \boldsymbol{d}, q)$ observando-se as características determinadas pelo parâmetro de cada componente. Mais detalhes podem ser encontrados em Sowell (1989), Lutkepohl (1991), Reinsel (1993) e Tsay (2010).

Capítulo 3

Cópulas e Funções Monotônicas por Partes

Neste capítulo estudamos a dependência multidimensional entre variáveis provindas de processos caóticos associados a certos tipos de transformações monotônicas por partes sob a ótica de cópulas. Mais especificamente, seja $T : I \to I$ uma transformação suave do intervalo I e assuma que exista uma medida de probabilidade μ_T absolutamente contínua e T-invariante. Seja U_0 uma variável aleatória com distribuição $F_0(x) = \mu_T([0, x])$. Desta forma, dada uma função $\varphi : I \to \mathbb{R}$, μ_T-integrável, podemos definir o processo estocástico

$$X_t := (\varphi \circ T^h)(U_0) = \varphi\big(T^t(U_0)\big), \quad \text{para todo } t \in \mathbb{N}. \tag{3.1}$$

Processos deste tipo apresentam aplicações em diversas áreas como perfuração de rochas, evolução de organismos vivos e intermitência do ritmo cardíaco em humanos (cf. Lasota e Mackey, 1994 e referências ali contidas e Zebrowsky, 2001).

O processo $\{X_t\}_{t\in\mathbb{N}}$, definido por (3.1), reflete basicamente a evolução das iterações da transformação T consigo mesma. Note que o papel da função φ é reescalar as trajetórias do processo do intervalo unitário para a reta. Realizações de processos do tipo (3.1) tipicamente apresentam uma dinâmica complexa, com um comportamento instável, errático e com grande sensibilidade ao ponto inicial. Para exemplificar tal sensibilidade ao ponto inicial, na Figura 3.1(c) apresentamos a trajetória de um processo do tipo (3.1), com T dado por (3.2), para dois pontos iniciais distantes 10^{-3} um do outro. Note que, em menos de 10 passos, as trajetórias já estão completamente separadas e o comportamento de uma realização é completamente diferente da outra.

Em um primeiro momento, apresentamos os resultados do artigo Lopes e Pumi (2011) onde o caso particular em que T é uma transformação de Manneville-Pomeau é estudado. A partir das ideias e métodos apresentados neste artigo, os autores generalizam os resultados para uma classe maior de processos do tipo (3.1), os quais são discutidos no artigo Pumi e Lopes (2011a) que apresentamos na Seção 3.2.

3.1 Cópulas Relacionadas à Processos de Manneville-Pomeau

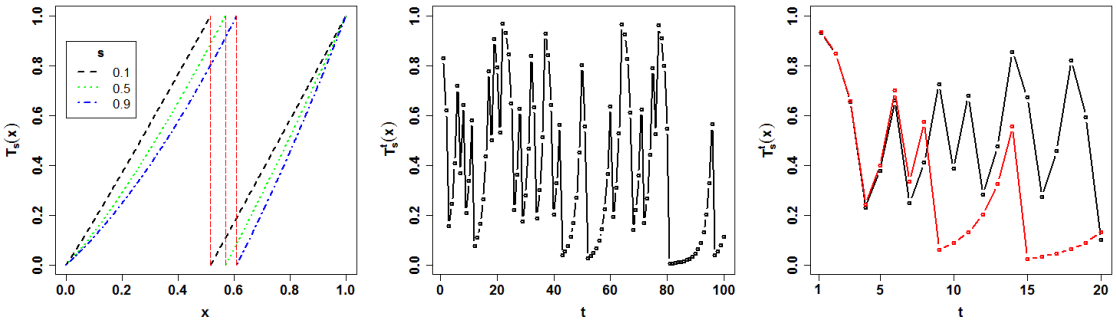
Nesta seção apresentamos os resultados do artigo Lopes e Pumi (2011) (veja Anexo A). Para $s > 0$, a função $T_s : [0,1] \to [0,1]$ dada por

$$T_s(x) := x + x^{1+s} (\mathrm{mod}\, 1), \tag{3.2}$$

é chamada de *transformação de Manneville-Pomeau*. A função T_s é uma função monótona por partes com dois ramos completos (no sentido que a transformação é sobrejetiva em cada ramo), de classe C^∞ e expansiva, exceto no ponto $x = 0$. Na Figura 3.1(a) mostramos o gráfico da transformação de Manneville-Pomeau para $s \in \{0.1, 0.5, 0.9\}$.

Pode-se mostrar que para a Transformação de Manneville-Pomeau sempre existe uma medida absolutamente contínua T_s invariante (denotada μ_s). Se $s \in (0,1)$, a medida será uma medida de probabilidade enquanto que para $s \geq 1$, esta medida não é mais finita, mas é σ-finita (cf. Pianigiani, 1980 e Fisher e Lopes, 2001). Por isso, sempre que $s \in (0,1)$, faz sentido considerar um processo do tipo (3.1) para $T = T_s$, ao qual chamamos de *processo de Manneville-Pomeau*. Na Figura 3.1(b) apresentamos uma realização típica de um processo Manneville-Pomeau para $s = 0.3$ tomando φ como a função identidade, enquanto que a Figura 3.1(c) ilustra a instabilidade com relação ao ponto inicial dos processos de Manneville-Pomeau.

Outras propriedades importantes do processo de Manneville-Pomeau são a estacionariedade forte, a ergodicidade e o decaimento hiperbólico da correlação. Para mais detalhes na teoria de transformações e processos de Manneville-Pomeau, referimos o leitor a Pianigiani (1980), Young (1999), Maes et al. (2000) e Fisher e Lopes (2001). Aplicações podem ser encontradas em Zebrowsky (2001), Olbermann et al. (2007) e Lopes e Lopes (1998).



(a) (b) (c)

Figura 3.1: (a) Transformação de Manneville-Pomeau para $s \in \{0.1, 0.5, 0.9\}$; (b) Realização de um Processo de Manneville-Pomeau para $s = 0.3$; (c) Instabilidade com relação ao ponto inicial de um Processo de Manneville-Pomeau.

Sendo $\{X_t\}_{t \in \mathbb{N}}$ um processo de Manneville-Pomeau com $s \in (0,1)$, é natural considerar o problema de caracterizar a dependência n-dimensional em vetores aleatórios $(X_{t_1}, \cdots, X_{t_n})$. Uma maneira bastante natural de se estudar tal dependência é a partir das cópulas relacionadas com o processo. Em Lopes e Pumi (2011) os autores derivam e estudam as cópulas relacionadas a vetores aleatórios provindos de processos de Manneville-Pomeau, chamadas de *cópulas de Manneville-Pomeau*. Estas cópulas são singulares em relação à medida de Lebesgue e seu suporte, no caso bidimensional, é uma função linear por partes passando pelos pontos de descontinuidade da transformação e suas potências. Os autores também mostram que a cópula

relacionada a (X_t, X_{t+h}) depende apenas do lag h.

Devido à sua complexa dinâmica, não existe uma fórmula fechada para a medida T_s-invariante μ_s e como as cópulas de Manneville-Pomeau são funções desta medida, estas também não possuem fórmula fechada. Para resolver este problema, Lopes e Pumi (2011) propõem uma aproximação para as cópulas de Manneville-Pomeau e mostram sua convergência uniforme para a cópula teórica. Também são discutidos aspectos computacionais desta aproximação como estabilidade em relação ao ponto inicial, T_s-invariância, pontos de truncamento, geração aleatória entre outros.

Por fim, é considerada uma aplicação dos resultados teóricos ao problema de estimação do parâmetro s em processos de Manneville-Pomeau. Um estimador computacionalmente simples e rápido é proposto. Para avaliar o comportamento do estimador na prática, uma simulação de Monte Carlo é realizada e os resultados mostram que além de rápido, o estimador também apresenta um bom desempenho.

3.2 Cópulas Relacionadas à Transformações Monotônicas por Partes

Na Seção anterior tratamos de cópulas relacionadas a processos de Manneville-Pomeau. Note que a transformação de Manneville-Pomeau dada em (3.2) é crescente em cada um de seus ramos. Os métodos empregados em Lopes e Pumi (2011), com algumas modificações, podem ser aplicados para derivar as cópulas para uma classe bem maior de transformações, que assim como a transformação de Manneville-Pomeau, também possuem ramos completos (sobrejetivos) e são monótonas em cada ramo. O artigo Pumi e Lopes (2011a) (veja Anexo B) apresenta esta generalização.

Mais precisamente, a classe de transformações consideradas em Pumi e Lopes (2011a) é composta por transformações $T : I \to I$ tal que T é uma função monótona por partes com um número finito de ramos, todos eles sendo bijetivos e de classe $C^{1+\alpha}$, $\alpha \in (0, 1)$, em seus domínios. Além disso, assume-se a existência de uma medida de probabilidade absolutamente contínua T-invariante. Esta última hipótese é necessária para garantir que o processo (3.1) está bem definido e terá propriedades boas, como estacionariedade. Ressaltamos, porém, que o problema de existência de uma medida invariante para uma dada transformação é um problema geralmente difícil e os resultados existentes geralmente envolvem condições fortes de suavidade para a transformação considerada. Veja, por exemplo, Rényi (1957), Lasota e Yorke (1973), Bowen (1979), Pianigiani (1980), Pianigiani (1981), Inoue e Ishitani (1991) e referências ali contidas.

Exemplos de transformações satisfazendo as condições acima são as seguintes:

1. A *transformação Tenda* com parâmetro $a \in (0, 1)$ é dada por

$$T_a(x) := \begin{cases} \dfrac{x}{a}, & \text{se} \quad 0 \leq x < a, \\ \dfrac{1-x}{1-a}, & \text{se} \quad a \leq x \leq 1. \end{cases}$$

   Pode-se mostrar que a medida de Lebesgue em I é T_a-invariante para todo a. A Figura 3.2(a) traz um gráfico típico da função Tenda enquanto que a Figura 3.2(d) traz uma realização típica de um processo do tipo (3.1) com $T = T_a$, $a = 0.8$.

2. A *transformação de Giampieri-Isola* com parâmetro $r \in (0,1)$, dada por

$$T_r(x) := \begin{cases} \frac{(2-r)x}{1-rx}, & \text{se} \quad 0 \leq x \leq \frac{1}{2}, \\ \frac{(2-r)(1-x)}{1-r(1-x)}, & \text{se} \quad \frac{1}{2} < x \leq 1. \end{cases}$$

Pode-se mostrar que $f_r(x) := -\frac{r}{\log(1-r)(1-r+rx)}$ é a densidade de uma medida absolutamente contínua e T_r-invariante (cf. Giampieri e Isola, 2005). A Figura 3.2(b) traz um gráfico típico da transformação Giampieri-Isola e a Figura 3.2(e) traz uma realização típica de um processo do tipo (3.1) com $T = T_r$, para $r = 0.8$. Note que T_r é bem definida para $r = 0$, recaindo na transformação Tenda com $a = 0.5$. Veja Giampieri e Isola (2005) para mais detalhes.

3. A transformação de Manneville-Pomeau já considerada na seção anterior.

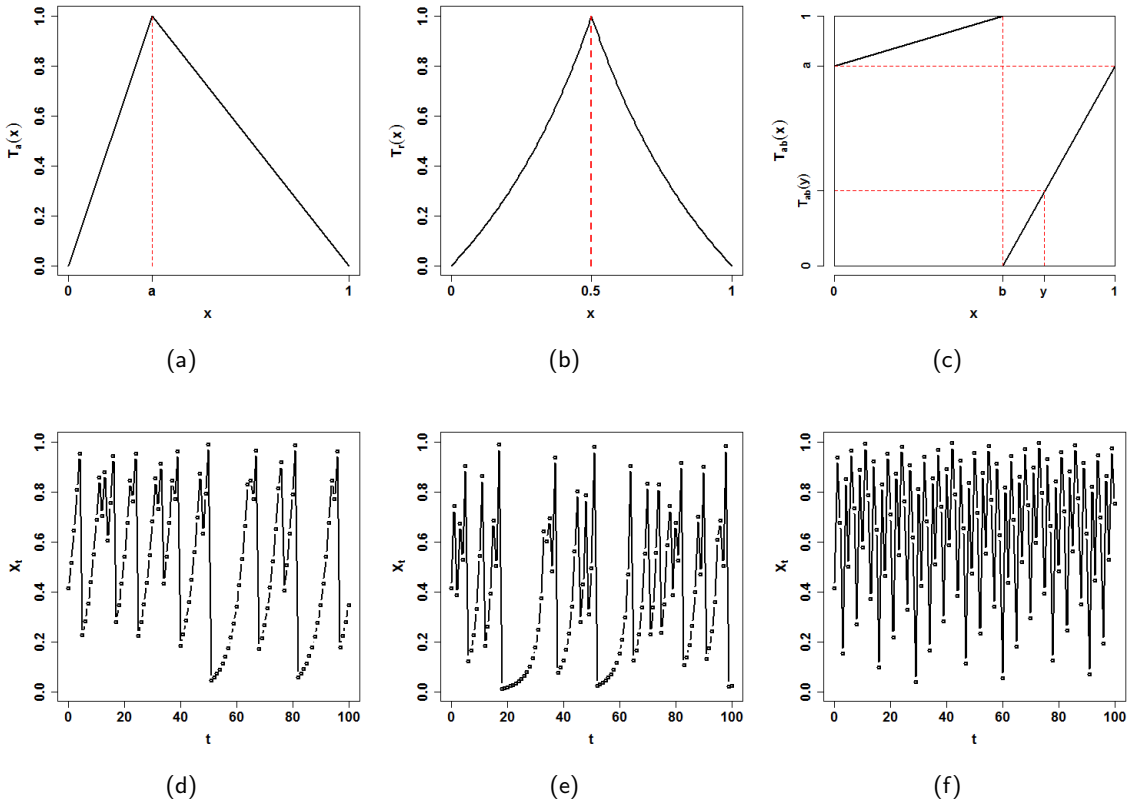


Figura 3.2: Gráfico típico da (a) transformação Tenda; (b) transformação Giampieri-Isola e (c) transformação (3.3). Realização típica do processo (3.1) com T (d) a transformação Tenda com parâmetro $a = 0.8$; (e) a transformação Giampieri-Isola com parâmetro $r = 0.8$ e (f) a transformação (3.3) com parâmetros $a = 0.8$ e $b = 0.6$. O ponto inicial em todos os casos é $\sqrt{2} (\mathrm{mod}\, 1)$.

Um exemplo de função suave do intervalo que não satisfaz as condições do artigo é a função dada por

$$T_{a,b}(x) := \begin{cases} a + \frac{(1-a)x}{b}, & \text{se} \quad 0 \leq x < b, \\ \frac{a(x-b)}{1-b}, & \text{se} \quad b \leq x < 1, \end{cases} \tag{3.3}$$

onde $a, b \in (0,1)$. A Figura 3.2(c) traz um gráfico típico da função (3.3) e a Figura 3.2(f) traz

uma realização típica de um processo do tipo (3.1) com $T = T_{a,b}$, $a = 0.8$ e $b = 0.6$. Pode ser mostrado que existe uma medida de probabilidade $T_{a,b}$-invariante (cf. Coelho et al., 2005). Note que a função não possui ramos completos.

Em Pumi e Lopes (2011a), as cópulas relacionadas a processos do tipo (3.1) para T como no parágrafo anterior são derivadas e estudadas. Estas famílias de cópulas compreendem cópulas singulares cujo suporte é uma função linear por partes com pontos de descontinuidade e orientação iguais aos da transformação original. Como no caso da transformação de Manneville-Pomeau, muitas vezes as cópulas estudadas não possuem fórmula fechada, por isso aproximações bastante gerais são discutidas. Sob certas condições, mostra-se a convergência uniforme destas para a cópula teórica. A geração de amostras aleatórias das cópulas derivadas no artigo também é discutida. Como exemplos, são apresentadas as cópulas relacionadas às transformações do tipo Tenda além, claro, da transformação de Manneville-Pomeau.

Como uma aplicação da teoria, os autores investigam a estimação paramétrica em processos do tipo (3.1) quando a transformação satisfaz às condições acima descritas assumindo que o parâmetro a ser estimado pode ser unicamente determinado pelo conhecimento dos pontos de descontinuidades da transformação. Este é o caso dos processos de Manneville-Pomeau. O método descrito em Pumi e Lopes (2011a) pode ser considerado uma generalização do método proposto em Lopes e Pumi (2011). Para exemplificar o método proposto e verificar suas propriedades, os autores conduzem uma simulação de Monte Carlo.

Capítulo 4

Processos com Longa Dependência

Este capítulo trata de processos multivariados com longa dependência sob dois pontos de vista diferentes, apresentados em Pumi e Lopes (2011b) e Lopes et al. (2011). Em um primeiro momento abordamos o problema da estimação semiparamétrica do vetor de diferenciação fracionária $\boldsymbol{d}$ em processos multivariados com longa dependência, e em um segundo momento, investigamos a interdependência entre as componentes de processos VARFIMA$(0, \boldsymbol{d}, 0)$ sob o ponto de vista da distância de Mallows.

4.1 Estimação Semiparamétrica em Processos Multivariados com Longa Dependência

Em meados da década de 80 e início da década de 90, a teoria assintótica dos estimadores paramétricos Gaussianos em processos univariados com longa dependência estava rigorosamente estabelecida com os trabalhos de Fox e Taqqu (1986), Dahlhaus (1989), Giraitis e Surgailis (1990), entre outros. Tais estimadores, que são baseados na função densidade espectral, com algumas condições de regularidade, apresentam excelentes propriedades como $\sqrt{n}$-consistência, normalidade assintótica e eficiência assintótica quando o processo for de fato Gaussiano. Estes estimadores, porém, dependem fortemente da Gaussianidade do processo (Giraitis e Surgailis, 1990 é exceção), requerem fortes condições distribucionais e não são robustos quanto à especificações incorretas do modelo, gerando, neste caso, estimativas inconsistentes.

Ainda no caso univariado, os primeiros estimadores da classe dos hoje conhecidos por estimadores semiparamétricos Gaussianos (GSE na sigla em inglês, que adotamos) apareceram no trabalho de Künsch (1987). Alguns anos depois Robinson (1995b) desenvolveu rigorosamente a teoria assintótica de tais estimadores. Algumas vantagens do estimador GSE em relação aos estimadores paramétricos é a maior eficiência, menores requerimentos distribucionais e, sobretudo, total independência quanto à Gaussianidade do processo.

No caso multivariado, a situação é um pouco diferente. O primeiro estimador para o vetor de diferenciação $\boldsymbol{d}$ em processos VARFIMA foi apresentado em Sowell (1989) que propôs o uso de um estimador de máxima verossimilhança. O custo computacional do método na época, porém, o tornou pouco atrativo. Alguns anos depois, Luceño (1996) apresentou uma alternativa computacionalmente menos intensiva baseadas em aproximações para a verossimilhança do processo. Recentemente, Tsay (2010) propôs um estimador baseado na verossimilhança condi-

cional exata para a estimação de $\boldsymbol{d}$. De qualquer forma, embora em geral métodos baseados no princípio da verossimilhança apresentem bom desempenho, seu custo computacional ainda é grande.

No caso multivariado, Robinson (1995a) apresentou o primeiro estudo assintótico rigoroso de um estimador GSE. Um estimador GSE multivariado a dois passos foi proposto em Lobato (1999), que mostrou a normalidade assintótica do estimador. Shimotsu (2007) introduziu um refinamento do estimador GSE a dois passos de Lobato (1999). Estendendo as técnicas apresentadas em Robinson (1995b) para o caso multivariado, Shimotsu (2007) provou a consistência e a normalidade assintótica de seu estimador além de provar a consistência do estimador GSE a dois passos de Lobato (1999). Recentemente, Nielsen (2011) estendeu o estimador GSE de Shimotsu (2007) para a região não-estacionária utilizando o chamado periodograma estendido.

O estimador GSE, introduzido em Shimotsu (2007), é baseado na minimização de uma função objetiva na qual a estimação da densidade espectral é feita utilizando-se o periodograma. Em Pumi e Lopes (2011b) (veja Anexo C), os autores estudam duas generalizações do estimador GSE proposto em Shimotsu (2007). Na primeira parte, consideram um novo estimador obtido substituindo-se o periodograma na função objetiva em Shimotsu (2007) por um estimador consistente arbitrário da densidade espectral. Os autores mostram que para $\boldsymbol{d} \in (0, 0.5)^m$, a consistência do estimador da densidade espectral sozinha é suficiente para garantir a consistência do novo estimador. Para a consistência do novo estimador quando $\boldsymbol{d} \in (-\beta/2, 0)^m$, para $\beta \in (0, 1)$, é necessário que o estimador para a densidade espectral seja $n^{-\beta}$-consistente. Na segunda parte, os autores consideram a função objetiva em Shimotsu (2007) substituindo-se o periodograma por um estimador arbitrário para a densidade espectral e derivam condições necessárias para que o novo estimador seja consistente e assintóticamente normal. Mostram ainda que matriz de variância-covariância da distribuição assintótica não depende do estimador para a densidade espectral.

Como exemplo da teoria, Pumi e Lopes (2011b) consideram substituir o periodograma na função objetiva pelo periodograma suavizado, mostram que o estimador GSE suavizado, assim definido, é consistente e apresentam uma simulação de Monte Carlo comparando o estimador original e o suavizado. Os autores concluem que o estimador GSE suavizado apresenta um desempenho consideravelmente melhor que o estimador GSE original.

4.2 Distância de Mallows em Processos VARFIMA$(0, \boldsymbol{d}, 0)$

Nesta seção, utilizamos a distância de Mallows como uma ferramenta no estudo da inter-dependência entre as coordenadas de processos VARFIMA$(0, \boldsymbol{d}, 0)$. Note que a distância de Mallows entre duas funções de distribuição F e G, pode ser vista como uma maneira de quantificar a relação entre elas. Com isso em mente, seja $\{X_t\}_{t\in\mathbb{N}}$ um processo bidimensional fracamente estacionário e assuma que as componentes do processo sejam identicamente distribuídas. Uma questão natural é a seguinte: a distância de Mallows calculada nas marginais do processo $\{X_t\}_{t\in\mathbb{N}}$ traz algum tipo de informação sobre o processo e, em caso afirmativo, esta informação pode ser útil, por exemplo, na estimação de algum parâmetro de interesse do processo? Esse é o caso, por exemplo, em processos nos quais as marginais são paramétricas e seus parâmetros são de interesse.

No artigo Lopes et al. (2011) (veja Anexo D), endereçamos a questão de como se comporta a distância de Mallows entre as componentes de um processo VARFIMA$(0, \boldsymbol{d}, 0)$ nos mais diversos cenários. O trabalho é baseado em simulações de Monte Carlo e tem por objetivo investigar uma possível relação da distância de Mallows com o vetor de diferenciação fracionária $\boldsymbol{d}$ e, em

caso de alguma relação existir, como esta relação se comporta em diferentes cenários.

No estudo de Monte Carlo conduzido no artigo, assumindo a notação da Definição 2.17, geram-se realizações de um processo VARFIMA $(0, \boldsymbol{d}, 0)$ bidimensional para diversas combinações do vetor de diferenciação $\boldsymbol{d}$ considerando-se vários cenários para o ruído $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ incluindo ruídos Gaussianos e não-Gaussianos, com diversos níveis de correlação, com marginais de caudas leves e pesadas e com marginais de variâncias iguais e diferentes. Em cada uma dessas situações, estima-se a distância de Mallows entre as componentes do processo. O estimador para a distância de Mallows é proposto no próprio artigo e é baseado na distribuição empírica das componentes da amostra. A partir daí, analisa-se uma possível relação entre o vetor de diferenciação $\boldsymbol{d}$, o tipo e a força da dependência no ruído e o comportamento das marginais do processo. Toda essa miscelânea de condições é geralmente induzida no processo diretamente em $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ utilizando-se ferramentas da teoria de cópulas, mais precisamente, via o Teorema de Sklar. Para efeito de comparação, os autores também apresentam o τ de Kendall para cada situação. Os autores concluem que o comportamento da distância de Mallows e do τ de Kendall é bastante diferente. Por exemplo:

1. Em todos os casos, a distância de Mallows só é sensível à diferenças na correlação induzida no ruído na presença de longa dependência forte, isto é, quando $\boldsymbol{d} \in [0.3, 0.5)^2$. Já o τ de Kendall é sempre muito sensível à correlação, mas apresenta um comportamento bastante suave em relação ao $\boldsymbol{d}$.

2. Quando as variâncias das marginais são diferentes, o comportamento global da distância de Mallows basicamente não muda, mas sua magnitude sofre alterações. Já o comportamento do τ de Kendall em função de $\boldsymbol{d}$ passa a ser errático.

3. Tanto no caso em que o ruído não é mais Gaussiano quanto no caso de marginais com caudas pesadas, o comportamento da distância de Mallows permanece estável com pequenas mudanças na magnitude. Dada a sua natureza, em ambos os casos não ocorreram diferenças significativas nem no comportamento nem na magnitude do τ de Kendall.

Como aplicação, Lopes et al. (2011) propõem um estimador semiparamétrico baseado na distância de Mallows e nos resultados empíricos obtidos. Algumas vantagens do estimador são sua versatilidade, pois pode ser utilizado para estimar o vetor de diferenciação $\boldsymbol{d}$ em processos VARFIMA$(0, \boldsymbol{d}, 0)$ de qualquer dimensão finita (inclusive dimensão 1) e sua agilidade, já que mesmo para dimensões altas, o estimador é computacionalmente simples e rápido, produzindo bons resultados.

Especialmente em dimensões altas, o estimador pode ser útil na obtenção de um vetor de inicialização a ser utilizado em estimadores do parâmetro de diferenciação baseados na otimização de alguma função objetiva, especialmente quando esta não é suave ou possui diversos pontos de mínimos e máximos locais. Um vetor de inicialização bom em tais condições pode significar uma redução drástica no tempo gasto com a otimização além de melhorar a precisão dos resultados desta, potencialmente evitando pontos de mínimos/máximos locais indesejáveis. Para exemplificar o uso e verificar o desempenho do estimador proposto, os autores conduzem uma simulação de Monte Carlo no contexto de processos VARFIMA$(0, \boldsymbol{d}, 0)$ Gaussianos.

Baseado nos resultados empíricos obtidos no trabalho, os autores propõem um teste para detectar a presença de longa dependência forte nas coordenadas de processos VARFIMA$(0, \boldsymbol{d}, 0)$ de dimensão finita, caracterizada pela presença de alguma coordenada para a qual $d_i > 0.3$. O desempenho do teste é verificada através de simulações em diversos casos, inclusive em casos marginais quando a evidência de ambas alternativas é fraca.

4.3 Teoria Assintótica

Considere $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$ duas amostras aleatórias de funções de distribuição F e G em $\mathscr{F}_\alpha$, respectivamente. Em Lopes et al. (2011), apresentado na seção anterior, os autores definem um estimador para a distância α de Mallows entre F e G baseado nas amostras $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$ através de

$$\widehat{\mathscr{D}}_\alpha(F,G) := \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \left(\int_0^1 \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_m^{(-1)}(u) \right|^\alpha \mathrm{d}u \right)^{1/\alpha}, \tag{4.1}$$

onde $\widehat{F}_n$ e $\widehat{G}_m$ denotam a distribuição empírica baseadas em $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$, respectivamente. Nesta seção o objetivo é mostrar que o estimador (4.1) é fortemente consistente e derivar sua distribuição assintótica no caso em que $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$ sejam amostras independentes e identicamente distribuídas (abreviado por $i.i.d.$) de duas variáveis aleatórias X e Y com distribuições F e G em $\mathscr{F}_\alpha$, respectivamente.

Seja $X_1, \cdots, X_n$ uma amostra i.i.d. de uma distribuição F. O Teorema de Glivenko-Cantelli garante que se $\widehat{F}_n$ é a distribuição empírica baseada em $X_1, \cdots, X_n$, então $\widehat{F}_n(x) \to F(x)$ uniformemente em x, exceto, talvez, em um conjunto de medida zero. Isto implica imediatamente que $\widehat{F}_n^{(-1)}(x) \to F^{(-1)}(x)$ uniformemente em x em quase toda a parte. Pode-se mostrar (Bickel e Freedman, 1981) que a convergência em distância α de Mallows é equivalente a convergência em distribuição junto com convergência do α-ésimo momento absoluto. Um argumento simples aliado à lei forte dos grandes números implica que $\mathscr{D}_\alpha(\widehat{F}_n, F) \longrightarrow 0$ em quase toda a parte (veja a Seção 5.1.1 em Shao, 1993, por exemplo). Uma aplicação deste resultado nos permite mostrar a consistência forte de $\widehat{\mathscr{D}}_\alpha$.

Teorema 4.1. *Seja $\alpha \geq 1$ e sejam $\widehat{F}_n$ e $\widehat{G}_m$ as distribuições empíricas baseadas em amostras i.i.d. $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$ de duas variáveis aleatórias X e Y com distribuições F e G em $\mathscr{F}_\alpha$, respectivamente. Então,*

$$\lim_{m,n\to\infty} \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \lim_{n\to\infty} \lim_{m\to\infty} \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \lim_{m\to\infty} \lim_{n\to\infty} \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \mathscr{D}_\alpha(F,G),$$

em quase toda a parte.

Prova: Primeiramente, é claro que $\widehat{F}_n$ e $\widehat{G}_m$ pertencem a $\mathscr{F}_\alpha$. Para $\alpha \geq 1$, a distância α de Mallows é uma métrica em $\mathscr{F}_\alpha$, desta forma, por um lado

$$\mathscr{D}_\alpha(F,G) \leq \mathscr{D}_\alpha(F, \widehat{F}_n) + \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) + \mathscr{D}_\alpha(\widehat{G}_m, G),$$

e por outro lado

$$\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) \leq \mathscr{D}_\alpha(\widehat{F}_n, F) + \mathscr{D}_\alpha(F,G) + \mathscr{D}_\alpha(G, \widehat{G}_m),$$

de onde segue que

$$\left| \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) - \mathscr{D}_\alpha(F,G) \right| \leq \mathscr{D}_\alpha(\widehat{F}_n, F) + \mathscr{D}_\alpha(\widehat{G}_m, G) \longrightarrow 0,$$

quase certamente tanto no sentido de limite duplo, quanto no sentido de limite iterado, já que $\mathscr{D}_\alpha(\widehat{F}_n, F) \to 0$ quando n tende ao infinito e $\mathscr{D}_\alpha(\widehat{G}_m, G) \to 0$ quando m tende ao infinito, independentemente um do outro. Isto completa a prova. ■

Na literatura podemos encontrar alguns resultados assintóticos para $\mathscr{D}_2(\widehat{F}_n, F)$. Por exemplo, distribuição assintótica, velocidade de convergência entre outros resultados para $\mathscr{D}_2(\widehat{F}_n, F)$ podem ser encontrados em Samworth e Johnson (2008) e referências ali contidas e também em Johnson e Samworth (2005). Procedemos a derivação da distribuição assintótica de (4.1). Focalizamos nossos esforços para o caso $\alpha = 2$, mas o caso geral para $\alpha \geq 1$ segue na mesma linha.

Para derivar a distribuição assintótica de (4.1) precisamos de alguns conceitos que revisamos aqui. Denotamos por $S := S[0,1]$ o espaço das funções reais definidas em $[0,1]$ contínuas à esquerda e possuindo limites laterais à direita em todos os pontos. Considere S com a norma uniforme,

$$\|f - g\|_\infty := \sup_{x\in[0,1]} \big\{\big|f(x) - g(x)\big|\big\}.$$

O espaço normado $\big(S, \|\cdot\|_\infty\big)$ é completo, mas não é separável, o que implica que a σ-álgebra gerada pelas bolas abertas em S, denotado por $\mathscr{B}_0$, é estritamente menor que a σ-álgebra de Borel em S, denotada por $\mathscr{B}$, gerada pelas bolas abertas em S. Este fato é inconveniente do ponto de vista da teoria da medida já que poderia acontecer de uma função real contínua em $\mathscr{B}$ não ser mensurável em $\mathscr{B}_0$. Para contornar esse problema, trabalhamos com o espaço $\big(S, \mathscr{B}_0, \|\cdot\|_\infty\big)$ e introduzimos um conceito diferente de convergência fraca conforme explicado, por exemplo, na seção 6 de Billingsley (1999).

Definição 4.1. Seja $\{X_t\}_{t\in\mathbb{N}}$ uma sequência de variáveis aleatórias em $\big(S, \mathscr{B}_0, \|\cdot\|_\infty\big)$. Definimos a *convergência em distribuição em* $\big(S, \mathscr{B}_0, \|\cdot\|_\infty\big)$, denotada por d°, como

$$X_n \xrightarrow[n\to\infty]{d^\circ} X \qquad \Longleftrightarrow \qquad \mathbb{E}\big(f(X_n)\big) \xrightarrow[n\to\infty]{} \mathbb{E}\big(f(X)\big),$$

para toda função $f : S \to \mathbb{R}$ contínua, limitada e $\mathscr{B}_0$-mensurável.

Além do conceito de convergência em distribuição em $\big(S, \mathscr{B}_0, \|\cdot\|_\infty\big)$, as distribuições assintóticas para funcionais da distribuição empírica geralmente envolvem os processos chamados pontes Brownianas, cuja definição lembramos em seguida.

Definição 4.2. Uma *ponte Browniana* é um processo Gaussiano $\{B(t)\}_{t\in[0,1]}$ com média 0,

$$\mathrm{Cov}\big(B(r), B(s)\big) = \min\{r, s\} - rs, \tag{4.2}$$

e tal que $B(0) = B(1) = 0$.

No próximo teorema, derivamos a distribuição assintótica de $\mathscr{D}_2(\widehat{F}_n, \widehat{G}_n)$.

Teorema 4.2. *Sejam $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_n$ amostras i.i.d. de duas variáveis aleatórias X e Y com distribuições F e G, respectivamente, em $\mathscr{F}_2$. Denote por $\widehat{F}_n$ e $\widehat{G}_n$ as distribuições empíricas baseadas nessas amostras, respectivamente. Suponha que F e G possuam densidades f e g, respectivamente, tais que $f\big(F^{-1}(t)\big)$ e $g\big(G^{-1}(t)\big)$ sejam positivas e contínuas para todo $t \in [0,1]$. Assuma também que os limites $\lim_{t\downarrow 0} F^{-1}(t)$, $\lim_{t\downarrow 0} G^{-1}(t)$, $\lim_{t\uparrow 1} F^{-1}(t)$ e $\lim_{t\uparrow 1} G^{-1}(t)$ existam em $\mathbb{R}$. Defina a variável aleatória*

$$Z := \int_0^1 \left(\frac{B_1(t)}{f\big(F^{-1}(t)\big)} + \frac{B_2(t)}{g\big(G^{-1}(t)\big)}\right) dt,$$

onde $\{B_i(t)\}_{t\in[0,1]}$ são pontes Brownianas independentes, para $i = 1, 2$. Se $\mathscr{D}_2(F, G) \neq 0$,

então

$$n^{\frac{1}{2}}\Big(\mathscr{D}_2(\widehat{F}_n,\widehat{G}_n)-\mathscr{D}_2(F,G)\Big)\xrightarrow[n\to\infty]{d}\operatorname{sign}(Z)\big|Z\big|^{-\frac{1}{2}},$$

onde $\operatorname{sign}(\cdot)$ *denota a função sinal, isto é* $\operatorname{sign}(x)=1$, *se* $x\geq 0$, *e* -1, *caso contrário. Se* $\mathscr{D}_2(F,G)=0$, *então* $F=G$ *e*

$$n^{\frac{1}{2}}\mathscr{D}_2(\widehat{F}_n,\widehat{G}_n)\xrightarrow[n\to\infty]{d}\left[2\int_0^1\left(\frac{B(t)}{f\big(F^{-1}(t)\big)}\right)^2\mathrm{d}t\right]^{\frac{1}{2}},$$

onde $\{B(t)\}_{t\in[0,1]}$ *é uma ponte Browniana.*

Prova: Como $f\big(F^{-1}(t)\big)$ e $g\big(G^{-1}(t)\big)$ são positivas por hipótese, o teorema 1 na pag.640 e o corolário 1 na pag.48 de Shorack e Wellner (1986) implicam

$$n^{\frac{1}{2}}\big(\widehat{F}_n^{-1}(t)-F^{-1}(t)\big)\xrightarrow[n\to\infty]{d^\circ}\frac{B_1(t)}{f\big(F^{-1}(t)\big)}\quad\text{e}\quad n^{\frac{1}{2}}\big(\widehat{G}_n^{-1}(t)-G^{-1}(t)\big)\xrightarrow[n\to\infty]{d^\circ}\frac{B_3(t)}{g\big(G^{-1}(t)\big)},$$

de onde segue que

$$n^{\frac{1}{2}}\Big(\widehat{F}_n^{-1}(t)-\widehat{G}_n^{-1}(t)-\big(F^{-1}(t)-G^{-1}(t)\big)\Big)\xrightarrow[n\to\infty]{d^\circ}\frac{B_1(t)}{f\big(F^{-1}(t)\big)}-\frac{B_3(t)}{g\big(G^{-1}(t)\big)},\tag{4.3}$$

em $\big(S,\mathcal{B}_0,\|\cdot\|_\infty\big)$, onde $B_1(\cdot)$ e $B_3(\cdot)$ são duas pontes Brownianas independentes. Observe que definindo $B_2(t):=-B_3(t)$, $\{B_2(t)\}_{t\in[0,1]}$ também é uma ponte Browniana. Observe também que, com probabilidade 1, $B_1(\cdot)$ e $B_2(\cdot)$ pertencem a $(\mathcal{C},\|\cdot\|_\infty)$, o espaço das funções reais contínuas em $[0,1]$ equipado com a norma uniforme que é completo e separável. Suponha que $\mathscr{D}_2(F,G)\neq 0$. Então, aplicando o delta método funcional (*functional delta method*) adaptado para convergência em d° para a função $h_1:\big(S,\mathscr{B}_0,\|\cdot\|_\infty\big)\to\big(S,\mathscr{B}_0,\|\cdot\|_\infty\big)$ dada por $h_1(\ell)(x)=\ell(x)^2$ (cf. teorema 20.8 em van der Vaart, 1998), segue que

$$n^{\frac{1}{2}}\Big[\big(\widehat{F}_n^{-1}(t)-\widehat{G}_n^{-1}(t)\big)^2-\big(F^{-1}(t)-G^{-1}(t)\big)^2\Big]\xrightarrow[n\to\infty]{d^\circ}2\bigg(\frac{B_1(t)}{f\big(F^{-1}(t)\big)}+\frac{B_2(t)}{g\big(G^{-1}(t)\big)}\bigg).$$

Agora pelo teorema da função contínua (*continuous mapping theorem*) adaptado para convergência em d° (cf. Billingsley, 1999) aplicado à função $h_2:\big(S,\mathscr{B}_0,\|\cdot\|_\infty\big)\to\mathbb{R}$ dada por $h_2(\ell)=\int_0^1\ell(t)dt$, segue que

$$n^{\frac{1}{2}}\Big(\mathscr{D}_2^2(\widehat{F}_n,\widehat{G}_n)-\mathscr{D}_2^2(F,G)\Big)\xrightarrow[n\to\infty]{d^\circ}2\int_0^1\bigg(\frac{B_1(t)}{f\big(F^{-1}(t)\big)}+\frac{B_2(t)}{g\big(G^{-1}(t)\big)}\bigg)dt.\tag{4.4}$$

Note que a integral em (4.4) é finita com probabilidade 1 já que B_1 e B_2 são limitadas pois são funções contínuas em quase toda a parte definidas no compacto $[0,1]$ e pela positividade de $f\big(F^{-1}(t)\big)$ e $g\big(G^{-1}(t)\big)$ em $[0,1]$. Observe também que a integral (4.4) depende do caminho das variáveis B_1 e B_2 e pode assumir valores positivos, negativos e nulos com probabilidade positiva. Seja Z como no enunciado e defina $S:=\{\omega:Z(\omega)\geq 0\}$. Em S, aplicando-se o delta método funcional à função $h_3:[0,\infty)\to[0,\infty)$ dada por $h_3(x):=x^{\frac{1}{2}}$ resulta

$$n^{\frac{1}{2}}\Big(\mathscr{D}_2(\widehat{F}_n,\widehat{G}_n)-\mathscr{D}_2(F,G)\Big)\xrightarrow[n\to\infty]{d^\circ}Z^{-\frac{1}{2}}.$$

Por outro lado, em S^c, o mesmo argumento com $h_4(x) = (-x)^{\frac{1}{2}}$ resulta

$$n^{\frac{1}{2}}\Big(\mathscr{D}_2(\widehat{F}_n, \widehat{G}_n) - \mathscr{D}_2(F, G)\Big) \xrightarrow[n\to\infty]{d^\circ} -(-Z)^{-\frac{1}{2}}.$$

Ambos os resultados podem ser escritos compactamente como

$$n^{\frac{1}{2}}\Big(\mathscr{D}_2(\widehat{F}_n, \widehat{G}_n) - \mathscr{D}_2(F, G)\Big) \xrightarrow[n\to\infty]{d^\circ} \operatorname{sign}(Z)\big|Z\big|^{-\frac{1}{2}}$$

e o resultado segue observando que qualquer função contínua e limitada de $\mathbb{R}$ em $\mathbb{R}$ é $\mathscr{B}$-mensurável (veja também pag.69 de Billingsley, 1999).

Considere agora o caso $\mathscr{D}_2(F, G) = 0$. Que $F = G$ segue do fato de $\mathscr{D}_2$ ser uma métrica em $\mathscr{F}_2$, e isto implica que $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_n$ são amostras (diferentes com probabilidade 1) da mesma distribuição. Neste caso, (4.3) se torna

$$n^{\frac{1}{2}}\big(\widehat{F}_n^{-1}(t) - \widehat{G}_n^{-1}(t)\big) \xrightarrow[n\to\infty]{d^\circ} \frac{B_1(t) - B_3(t)}{f\big(F^{-1}(t)\big)}.$$

Defina $B(t) = \big(B_1(t) - B_3(t)\big)/\sqrt{2}$, para todo $t \in [0, 1]$. Então $\{B_t\}_{t\in[0,1]}$ é uma ponte Browniana. De fato, segue diretamente da definição que $\{B_t\}_{t\in[0,1]}$ é Gaussiano, tem média 0 e $B(0) = B(1) = 0$. Falta mostrar que a covariância é da forma (4.2). Para $r, s \in [0, 1]$, temos

$$\begin{aligned} 2\operatorname{Cov}\big(B(r), B(s)\big) &= \operatorname{Cov}\big(B_1(r) - B_3(s), B_1(r) - B_3(s)\big) \\ &= \operatorname{Cov}\big(B_1(r), B_1(s)\big) - \operatorname{Cov}\big(B_1(r), B_3(s)\big) - \operatorname{Cov}\big(B_1(s), B_3(r)\big) + \operatorname{Cov}\big(B_3(r), B_3(s)\big) \\ &= 2\,(\min\{r, s\} - rs), \end{aligned}$$

já que B_1 e B_3 são pontes Brownianas independentes. Assim, segue que

$$n^{\frac{1}{2}}\big(\widehat{F}_n^{-1}(t) - \widehat{G}_n^{-1}(t)\big) \xrightarrow[n\to\infty]{d^\circ} \frac{\sqrt{2}B(t)}{f\big(F^{-1}(t)\big)}.$$

Aplicando-se um argumento semelhante ao caso anterior, podemos aplicar o teorema da função contínua para d° às funções $h_5 : \big(S, \mathscr{B}_0, \|\cdot\|_\infty\big) \to \big(S, \mathscr{B}_0, \|\cdot\|_\infty\big)$ dada por $h_5(\ell)(x) = \ell(x)^2$ e $h_6 : \big(S, \mathscr{B}_0, \|\cdot\|_\infty\big) \to \mathbb{R}$ dada por $h_6(\ell) = \int_0^1 \ell(t)dt$. Com isso obtemos

$$n\mathscr{D}_2^2(\widehat{F}_n, \widehat{G}_n) = n\int_0^1 \big(\widehat{F}_n^{-1}(t) - \widehat{G}_n^{-1}(t)\big)^2 dt \xrightarrow[n\to\infty]{d^\circ} 2\int_0^1 \left(\frac{B(t)}{f\big(F^{-1}(t)\big)}\right)^2 dt. \tag{4.5}$$

O mesmo argumento aplicado à (4.4) pode ser utilizado para mostrar que a integral em (4.5) é finita com probabilidade 1. Finalmente, como (4.5) é não-negativa, aplicando-se a teorema da função contínua para d° à função $h_7 : [0, \infty) \to [0, \infty)$ dada por $h_7(x) := x^{\frac{1}{2}}$, obtemos

$$n^{\frac{1}{2}}\mathscr{D}_2(\widehat{F}_n, \widehat{G}_n) \xrightarrow[n\to\infty]{d^\circ} \left[2\int_0^1 \left(\frac{B(t)}{f\big(F^{-1}(t)\big)}\right)^2 \mathrm{d}t\right]^{\frac{1}{2}}.$$

O resultado segue observando que qualquer função contínua e limitada de $\mathbb{R}$ em $\mathbb{R}$ é $\mathscr{B}$-mensurável, completando a prova do teorema. ■

Terminamos esta seção observando que as condições relativas à existência dos limites

$\lim_{t\downarrow 0} F^{-1}(t)$, $\lim_{t\downarrow 0} G^{-1}(t)$, $\lim_{t\uparrow 1} F^{-1}(t)$ e $\lim_{t\uparrow 1} G^{-1}(t)$ no Teorema 4.2 podem ser relaxadas assumindo condições de controle do decaimento das caudas das distribuições F e G, no mesmo espírito de Csörgő e Horváth (1990), del Barrio et al. (2000) e Samworth e Johnson (2008).

Capítulo 5

Parametrização de Cópulas e Decaimento da Covariância

Neste capítulo estudamos o decaimento da correlação em processos estocásticos e sua relação com a parametrizações de cópulas. No Capítulo 2 observamos que, pelo Teorema de Sklar, quaisquer que sejam as variáveis aleatórias contínuas $X_1, \cdots, X_n$ com distribuição conjunta H e marginais $F_1, \cdots, F_n$, respectivamente, podemos associar ao conjunto $\{F_1, \cdots, F_n, H\}$ uma única cópula C. Reciprocamente, ao conjunto $\{F_1, \cdots, F_n, C\}$, associamos uma única distribuição conjunta H. Desta forma, sendo $\{X_t\}_{t\in\mathbb{N}}$ um processo estocástico (unidimensional) tal que, para cada $t \in \mathbb{N}$, X_t seja contínua e tenha distribuição F_t, então para cada vetor aleatório (X_r, X_s) podemos associar uma única cópula $C_{r,s}$. Pelo Lema de Hoeffding, a tripla $(F_r, F_s, C_{r,s})$ nos permite calcular a autocovariância entre X_r e X_s. Assuma que todas as cópulas bidimensionais do processo $\{X_t\}_{t\in\mathbb{N}}$ pertencem a uma mesma família de cópulas $\{C_\theta\}_{\theta\in\Theta}$ e que a cópula de X_t e X_{t+h} não dependa de t, isto é, $C_h := C_{t,t+h}$ para todo $t \in \mathbb{N}$ e $h \in \mathbb{N}^*$ (isto ocorre, por exemplo, se o processo for fortemente estacionário). Assim, podemos associar uma sequência $\{\theta_n\}_{n\in\mathbb{N}^*} \subset \Theta$ de tal forma que $C_h := C_{\theta_h}$, para todo $h \in \mathbb{N}^*$. Com essa construção, para cada $h \in \mathbb{N}^*$ e todo $t \in \mathbb{N}$ podemos calcular a autocovariância entre X_t e X_{t+h} via Lema de Hoeffding a partir de $(F_t, F_{t+h}, C_{\theta_h})$. Em particular, isto nos permite determinar o decaimento da autocovariância entre X_t e X_{t+h}, para h grande. Note ainda que o decaimento da autocovariância pode ser determinado a partir das cópulas bidimensionais e das marginais do processo, não sendo necessário o conhecimento da estrutura de dependência de dimensões maiores do que 2.

Tendo em vista o parágrafo acima, quando se tratando de cópulas fica clara a vantagem e a naturalidade em se trabalhar com o decaimento da autocovariância do processo ao invés do decaimento da autocorrelação. Além disso, a autocorrelação pode ser obtida diretamente da autocovariância do processo. Com um mínimo de esforço, todos os resultados obtidos para o decaimento da autocovariância se traduzem para o decaimento da autocorrelação. Por isso, neste capítulo, tratamos exclusivamente do decaimento da autocovariância.

Neste capítulo estudamos o problema inverso apresentado em Pumi e Lopes (2011c) (veja Anexo E). Suponha que desejamos construir um processo fracamente estacionário $\{X_t\}_{t\in\mathbb{N}}$ com marginais absolutamente contínuas $\{F_n\}_{n\in\mathbb{N}}$ predefinidas e com um certo decaimento da autocovariância do processo. Para uma determinada família de cópulas paramétricas bidimensionais $\{C_\theta\}_{\theta\in\Theta}$ dada, investigamos como determinar uma parametrização $\{\theta_n\}_{n\in\mathbb{N}^*}$ de tal forma que, se a cópula associada a X_t e X_{t+h} for C_{θ_h}, então o decaimento da autocovariância desejado é obtido. O foco do trabalho é determinar parametrizações que produzam decaimento lento

da autocovariância típicos de processos com longa dependência, porém os resultados permitem obter qualquer tipo de decaimento. Além disso, embora o foco do trabalho seja obter processos estacionários, os métodos e resultados derivados se aplicam a processos não-estacionários com a devida interpretação.

A ideia de construir processos estocásticos a partir de cópulas não é novo. No caso multivariado, a utilização de cópulas para modelar a interdependência de duas ou mais variáveis na construção de modelos multivariados complexos tem se tornado rotina. Veja por exemplo os trabalhos recentes de Lee e Long (2009), Palynchuk e Guo (2011) e referências ali contidas. No caso univariado, o problema é um pouco mais delicado e a maioria dos resultados envolvem apenas processos com dependência fraca. Por exemplo, Darsow et al. (1992) desenvolvem a teoria de processos de Markov baseado em cópulas a partir de um certo produto introduzido no espaço das cópulas. Em Lagerås (2010), as propriedades das cadeias de Markov construídas desta maneira para várias famílias de cópulas são estudadas. Ainda no caso unidimensional, no capítulo 8 de Joe (1997) alguns métodos para a construção de séries temporais com curta dependência a partir de cópulas condicionais são desenvolvidos. O autor ainda discute a construção de processos Markovianos e séries temporais com curta dependência a partir da parametrização de distribuições pertencentes à classe de distribuições fechadas para a convolução e infinitamente divisíveis (*convolution-closed infinitely divisible class*). Outro trabalho interessante nessa linha é Chen e Fan (2006), onde as autoras introduzem uma classe de modelos lineares semiparamétricos generalizados baseado em cópulas com estrutura de dependência Markoviana e estudam propriedades, estimação e simulação nestes.

Nosso trabalho difere dos demais acima citados essencialmente por ser baseado na parametrização de cópulas focando na obtenção de um determinado decaimento ou estrutura de autocovariância dada. Além disso, nenhuma propriedade especial para as marginais do processo é assumida. A teoria e métodos propostos também possuem uma vantagem clara sobre métodos baseados diretamente na parameterização da covariância, a saber, a autonomia e flexibilidade na modelagem da dependência bidimensional do processo, aplicabilidade à simulação de séries temporais e generalidade dos resultados.

Vários exemplos são apresentados no trabalho englobando famílias de cópulas populares em aplicações, como por exemplo, as famílias Euclideana, de Valor Extremo e Gaussiana além de mostrar como a teoria se encaixa com a classe de modelos ARFIMA, entre outros. Duas aplicações da teoria geral são apresentadas. A primeira é relacionada com a estimação de parâmetros identificáveis pelo decaimento da autocovariância em séries temporais fracamente estacionárias. Um estudo de Monte Carlo é conduzido para acessar as propriedades do estimador proposto. A segunda aplicação é relacionada com a simulação de séries temporais com marginais dadas e decaimento predeterminado da autocovariância. As técnicas apresentadas permitem a simulação de séries temporais não tradicionais de forma relativamente simples e rápida. Para finalizar, a metodologia proposta é aplicada à série temporal real do índice de ativos da S&P500.

Capítulo 6

Conclusões e Trabalhos Futuros

Neste trabalho estudamos alguns aspectos da teoria de cópulas, de processos com longa dependência e do decaimento da correlação em processos estocásticos, refletindo o trabalho contido em Lopes e Pumi (2011), Pumi e Lopes (2011a,b,c) e Lopes et al. (2011). Na primeira parte do trabalho derivamos as cópulas relacionadas a vetores de variáveis aleatórias provindas de um certo tipo de processos caóticos, definidos a partir da iteração de uma transformação suave por partes a uma certa variável aleatória inicial.

Em um primeiro momento, consideramos a classe das transformações de Manneville-Pomeau e o processo caótico associado. Derivamos e estudamos as cópulas associadas a estes, que denominamos cópulas de Manneville-Pomeau. Como as cópulas derivadas não possuem fórmula fechada, consideramos alguns aspectos computacionais importantes, como por exemplo, aproximações às cópulas de Manneville-Pomeau, estabilidade e convergência dessas aproximações, geração aleatória de amostras, entre outros. Como aplicação, propomos um estimador para o parâmetro do processo baseado na teoria desenvolvida e conduzimos um experimento de Monte Carlo para determinar as propriedades do estimador.

As técnicas e resultados apresentados para a transformação de Manneville-Pomeau são, então, generalizadas para uma classe maior de transformações suaves por partes. Como aplicação, propomos uma metodologia para, em determinadas situações, estimar parâmetros relativos aos processos estudados. A metodologia é ilustrada e seu desempenho avaliado através de simulações de Monte Carlo.

No Capítulo 4, estudamos problemas relacionados com o fenômeno da longa dependência. Podemos dividir o capítulo em duas partes. Na primeira parte, estudamos a estimação semiparamétrica em processos multivariados com longa dependência. A ideia é baseada em modificações específicas do estimador proposto em Shimotsu (2007). Duas classes novas de estimadores são propostas e suas propriedades assintóticas analisadas. Em particular, encontramos condições para a consistência e a normalidade assintótica dos estimadores propostos. Para um caso particular dos estimadores considerados, uma simulação de Monte Carlo é conduzida para avaliar o seu desempenho em amostras finitas, bem como compará-lo ao estimador original em Shimotsu (2007). Concluímos que o estimador proposto apresenta boas propriedades assintóticas e em amostras finitas, tendo desempenho superior ao original em Shimotsu (2007).

Na segunda parte do Capítulo 4, analisamos, sob diversas condições, a interdependência entre as coordenadas de processos VARFIMA$(0, \boldsymbol{d}, 0)$ bidimensionais sob o ponto de vista da distância de Mallows e do τ de Kendall. O trabalho é baseado em simulações de Monte Carlo e tem por objetivo explorar uma possível relação entre a distância de Mallows, o vetor de diferenciação fracionária $\boldsymbol{d}$, o tipo e o nível de dependência induzido no ruído, bem como o comportamento

marginal do processo. Para fins de comparação com a distância de Mallows, o τ de Kendall também é calculado. Concluímos que o comportamento de ambos difere substancialmente. Por exemplo, o τ de Kendall, visto como função do parâmetro $\boldsymbol{d}$, é bastante sensível com relação à correlação induzida no processo através do ruído. Já a distância de Mallows só é sensível à correlação na presença de longa dependência em ambas as coordenadas do processo. A partir dos resultados empíricos obtidos, propomos tanto um estimador para o parâmetro $\boldsymbol{d}$ quanto um teste para acessar a existência de dependência forte em alguma coordenada de processos VARFIMA$(0, \boldsymbol{d}, 0)$ de qualquer dimensão finita. Simulações de Monte Carlo específicas são conduzidas para acessar o poder do teste e o desempenho do estimador em amostras finitas.

Um dos grandes atrativos do estimador proposto é sua facilidade de implementação, o baixo custo computacional, a relativa precisão e versatilidade, o que o torna uma alternativa atrativa para a estimação do parâmetro $\boldsymbol{d}$ em dimensões altas. Além disso, pode ser utilizado também para a obtenção de uma primeira estimativa para métodos mais refinados de estimação baseados na otimização de funções objetivas mal comportadas. Por fim, analisamos algumas propriedades assintóticas do estimador para a distância de Mallows proposto no trabalho, no contexto de amostras i.i.d.

O Capítulo 5 da tese trata do problema de se construir um processo estocástico com um determinado decaimento da correlação e marginais dadas, a partir da reparametrização de uma família de cópulas. As técnicas desenvolvidas e os resultados apresentados se aplicam para processos estacionários ou não, bem como para qualquer decaimento da correlação dado. O trabalho foca no decaimento lento da correlação, típico de processos com longa dependência. Todavia, para a maior abrangência dos resultados, a análise no capítulo é conduzida sempre em termos de autocovariâncias. Vários exemplos são apresentados, incluindo famílias de cópulas populares em aplicações como a Euclideana, a de Valores Extremos e a Gaussiana. Consideramos ainda como a teoria combina com a classe de modelos ARFIMA. Como aplicação da teoria, uma metodologia geral de estimação de parâmetros identificáveis através do decaimento da autocovariância em processos estacionários é proposta. Para avaliar a abrangência e o desempenho da metodologia tanto na presença de longa dependência, quanto no caso de curta dependência, simulações de Monte Carlo são realizadas. Os resultados mostram que a metodologia se adapta bem aos diversos tipos de dependência utilizados na simulação e apresenta um bom desempenho em amostras finitas. A metodologia ainda é aplicada à série temporal real do índice de ativos da S&P500.

Durante o período de pesquisas e todo o contato com o material estudado, nos deparamos com várias outras ideias para trabalhos futuros. Destacamos duas delas.

Primeiramente, em Lobato (1999) o autor introduz um estimador semiparamétrico Gaussiano baseado na otimização de uma determinada função objetiva, na qual o periodograma é utilizado como estimador da densidade espectral do processo. Em nosso próximo trabalho, pretendemos generalizar este estimador no mesmo espírito de Pumi e Lopes (2011b) em relação ao estimador introduzido em Shimotsu (2007). Em outras palavras, partindo da função objetiva em Lobato (1999), estudaremos primeiramente o que acontece quando substituímos o periodograma por um estimador consistente para a densidade espectral. Esperamos que as técnicas utilizadas em Pumi e Lopes (2011b) sejam adaptáveis a este problema e nos permitam obter conclusões parecidas, isto é, obter a consistência do estimador. Em seguida, consideramos a substituição do periodograma na função objetiva em Lobato (1999) por um estimador arbitrário da densidade espectral. Esperamos que, sob condições similares às consideradas em Pumi e Lopes (2011b), seja possível obter a consistência e a normalidade assintótica do estimador semiparamétrico modificado dessa maneira.

Em um segundo artigo, através de extensivas simulações de Monte Carlo, pretendemos comparar o desempenho dos estimadores em Shimotsu (2007), Pumi e Lopes (2011b), Lobato (1999) e a versão modificada deste conforme indicado no parágrafo anterior (e talvez outros estimadores), sob diversos cenários. Um trabalho assim é importante primeiro pelo fato de que não há na literatura tal comparação; segundo, porque para a versão modificada do estimador em Shimotsu (2007), e o mesmo provavelmente acontecerá para a versão modificada do estimador em Lobato (1999), a ideia "natural", dentro da teoria desenvolvida, é substituir o periodograma pelo periodograma suavizado. Nesse contexto, pontos de cortes e tipos de função de ponderação são uma fonte importante de discussão; terceiro, é de se esperar que não exista um estimador uniformemente melhor que os outros para todas as situações. Por isso, um estudo empírico apontando qual estimador é melhor em determinada situação, pode ser de grande valia em aplicações.

Referências Bibliográficas

[1] Bickel, P.J. e Freedman, D.A. (1981). "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics*, **9**, 1196 - 1217.

[2] Billingsley, P. (1999). *Convergence of Probability Measures*. New York: John Wiley. 2ª Edição.

[3] Bingham, N.H.; Goldie, C.M e Teugels, J.L. (1987). *Regular Variation*. Cambridge: Cambridge University Press.

[4] Bowen, R. (1979). "Invariant Measures for Markov Maps of the Interval". *Communications in Mathematical Physics*, **69**, 1 - 14.

[5] Brockwell, P.J. e Davis, R.A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag. 2ª Edição.

[6] Carley, H. e Taylor, M.D. (2002). "A New Proof of Sklar's Theorem". Em Cuadras, C.M.; Fortiana, J. e Rodríguez-Lallena, J.A. (editores). *Distributions with given marginals and statistical modelling*. Dordrecht: Kluwer Academic Publications, 29 - 34.

[7] Chen, X. e Fan, Y. (2006) "Estimation of Copula-Based Semiparametric Time Series Models". *Journal of Econometrics,* **130**, 307 - 335.

[8] Cherubini, U., Luciano; E. e Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester: Wiley.

[9] Chiriac, R. e Voev, V. (2011). "Modeling and Forecasting Multivariate Realized Volatility". *Journal of Applied Econometrics*, **26**(6), 922 - 947.

[10] Coelho, Z.; Lopes, A.O. e da Rocha, L.F. (1995). "Absolutely Continuous Invariant Measures for a Class of Affine Interval Exchange Maps". *Proceedings of the American Mathematical Society*, **123**(11), 3533 - 3542.

[11] Csörgő, M. e Horváth, L. (1990). "On the Distribution of L_p Norms of Weighted Quantile Processes". *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, **26**(1), 65 - 85.

[12] Dahlhaus, R. (1989). "Efficient Parameter Estimation for Self-Similar Processes." *The Annals of Statistics*, **17**, 1749 - 1766.

[13] del Barrio, E.; Cuesta-Albertos, J. A. e Matrán, C. (2000). "Contributions of Empirical and Quantile Processes to the Asymptotic Theory of Goodness-of-fit Tests". *Test*, **9**(1), 1 - 96.

[14] Dionge, A.K. (2010). "A Multivariate Generalized Long Memory Model". *Comptes Rendus de l'Académie des Sciences - Série I*, **348**, 327-330.

[15] Doukhan, P.; Oppenheim G. e Taqqu, M.S. (editores) (2003). *Theory and Applications of Long-Range Dependence.* Boston: Birkhäuser.

[16] Fisher, A.M. e Lopes, A. (2001). "Exact Bounds for the Polynomial Decay of Correlation, $1/f$ Noise and the CLT for the Equilibrium State of a Non-Hölder Potential". *Nonlinearity*, **14**, 1071 - 1104.

[17] Fox, R. e Taqqu, M.S. (1986). "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, **14**(2), 517 - 532.

[18] Giampieri, M. e Isola, S. (2005). "A One-Parameter Family of Analytic Markov Maps with an Intermittency Transition". *Discrete and Continuous Dynamical Systems*, **12**(1), 115 - 136.

[19] Giraitis, L. e Surgailis, D. (1990). "A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotic Normality of Whittle's Estimate". *Probability Theory and Related Fields*, **86**, 87 - 104.

[20] Hannan, E.J. (1970). *Multiple Time Series*. New York: John Wiley.

[21] Horn, R.A. e Johnson, C.R. (1991). *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

[22] Hurst, H.R. (1951). "Long-Term Storage in Reservoirs". *Transactions of the American Society of Civil Engineers*, **116**, 770 - 799.

[23] Hurvich, C.M e Beltrão, K.I. (1993). "Asymptotics for the Low-Frequency Ordinates of the Periodogram of a Long-Memory Time Series". *Journal of Time Series Analysis,* **14**(5), 455 - 472.

[24] Inoue, T. e Ishitani, H. (1991). "Asymptotic Periodicity of Densities and Ergodic Properties for Nonsingular Systems". *Hiroshima Mathematical Journal*, **21** (3), 597 - 620.

[25] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. New York: Chapman & Hall.

[26] Johnson, O. e Samworth, R. (2005). "Central Limit Theorem and Convergence to Stable Laws in Mallows Distance". *Bernoulli*, **11**(5), 829 - 845.

[27] Künsch, H. (1987). "Statistical Aspects of Self-Similar Processes". Em Prokhorov, Yu. e Sazanov, V.V. (editores), *Proceedings of the First World Congress of the Bernoulli Society*. Utrecht: VNU Science Press, 67 - 74.

[28] Lagerås, A.N. (2010). "Copulas for Markovian Dependence". *Bernoulli*, **16**(2), 331 - 342.

[29] Lasota, A. e Mackey, M.C. (1994). *Chaos, Fractals, and Noise. Stochastic Aspects of Dynamics.* New York: Springer-Verlag. 2ª Edição.

[30] Lasota, A. e Yorke, J.A. (1973). "On the Existence of Invariant Measures for Piecewise Monotonic Transformations". *Transaction of the American Mathematical Society*, **186**, 481 - 488.

[31] Lee, T-H. e Long, X. (2009). "Copula-Based Multivariate GARCH Model with Uncorrelated Dependent Errors". *Journal of Econometrics*, **150**(2), 207 - 218.

[32] Lobato, I.N. (1999). "A Semiparametric Two-Step Estimator in a Multivariate Long Memory Model". *Journal of Econometrics*, **90**, 129 - 153.

[33] Lopes, A. e Lopes, S.R.C. (1998). "Parametric Estimation and Spectral Analysis of Piecewise Linear Maps of the Interval". *Advances in Applied Probability*, **30**, 757 - 776.

[34] Lopes, S.R.C. (2008). "Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting". Em Sidoravicius, V. e Vares, M.E. (editores), *In and Out of Em quilibrium 2*. Basel: Birkhäuser, 497 - 525.

[35] Lopes, S.R.C. e Pumi, G. (2011). "Copulas Related to Manneville-Pomeau Processes". Em revisão.

[36] Lopes, S.R.C.; Pumi, G. e Zaniol, K. (2011). "Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes". *Communications in Statistics: Computations and Simulations*. In press.

[37] Luceño, A. (1996). "A Fast Likelihood Approximation for Vector General Linear Processes with Long Series: Application to Fractional Differencing". *Biometrika*, **83**, 603 - 614.

[38] Lutkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

[39] Maes, C.; Redig, F.; Takens, F.; Moffaert, A. e Verbitski, E. (2000). "Intermittency and Weak Gibbs States". *Nonlinearity*, **13**, 1681 - 1698.

[40] Major, P. (1978). "On the Invariance Principle for Sums of Independent Identically Distributed Random Variables". *Journal of Multivariate Analysis*, **8**(4), 487 - 517.

[41] Mallows, C.L. (1972). "A Note on Asymptotic Joint Normality". *The Annals of Mathematical Statistics,* **43**, 508 - 515.

[42] Mari, D.D. e Kotz, S. (2001). *Correlation and Dependence*. London: Imperial College Press.

[43] Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer Verlag. $2^{\underline{a}}$ Edição.

[44] Nielsen, F.S. (2011). "Local Whittle Estimation of Multi-Variate Fractionally Integrated Processes". *Journal of Time Series Analysis*, **32**(3), 317 - 335.

[45] Olberman, B.P; Lopes, S.R.C e Lopes, A.O. (2007). "Parameter Estimation in Manneville-Pomeau Processes". Manuscrito não publicado. Fonte: arXiv:0707.1600.

[46] Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. New Jersey: John Wiley.

[47] Palynchuk, B.A. e Guo, Y. (2011). "A Probabilistic Description of Rain Storms Incorporating Peak Intensities". *Journal of Hydrology*, **409**(1-2), 71 - 80.

[48] Pianigiani, G. (1980). "First Return Map and Invariant Measures". *Israel Journal of Mathematics*, **35**, 32 - 48.

[49] Pianigiani, G. (1981). "Existence of Invariant Measures for Piecewise Continuous Transformations". *Annales Polonici Mathematici*, **40**, 39 - 45.

[50] Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

[51] Pumi, G. (2006). Cópulas em Processos Estocásticos. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul. Disponível em http://hdl.handle.net/10183/8023.

[52] Pumi, G. e Lopes, S.R.C. (2011a). "Copulas Related to Piecewise Monotone Functions of the Interval and Associated Processes". Submetido.

[53] Pumi, G. e Lopes, S.R.C. (2011b). "Smoothed Semiparametric Estimation on Multivariate Long Memory Processes". Submetido.

[54] Pumi, G. e Lopes, S.R.C. (2011c). "Parameterization of Copulas and Covariance Decay of Stochastic Processes with Applications". Submetido.

[55] Reinsel, G.C. (1993). *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag.

[56] Rényi, A. (1957). "Representations for Real Numbers and their Ergodic Properties". *Acta Mathematica Academiae Scientiarum Hungaricae*, **8**, 477 - 493.

[57] Robinson, P.M. (1995a). "Log-Periodogram Regression of Time Series with Long Range Dependence". *Annals of Statistics*, **23**(3), 1048 - 1072.

[58] Robinson, P.M. (1995b). "Gaussian Semiparametric Estimation of Long Range Dependence". *Annals of Statistics*, **23**(5), 1630 - 1661.

[59] Royden, H.L. (1988). *Real Analysis.* New York: Macmillan. 3ª Edição.

[60] Samworth, R. e Johnson, O. (2008). "Convergence of the Empirical Process in Mallows Distance, with an Application to Bootstrap Performance". Manuscrito não publicado. Fonte: arXiv:math/0406603v1.

[61] Schweizer, B. e Sklar, A. (2005). *Probabilistic Metric Spaces*. Mineola: Dover Publications.

[62] Shao, J. (2003). *Mathematical Statistics*. New York: Springer-Verlag.

[63] Shimotsu, K. (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes". *Journal of Econometrics*, **137**, 277 - 310.

[64] Shorack, G.R. e Wellner, J.A. (1986). *Empirical Processes With Applications to Statistics*. New York: John Wiley.

[65] Sowell, F. (1989). "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models". Working Paper, Carnegie-Mellon University.

[66] Taqqu, M.S. (2003). "Fractional Brownian Motion and Long-Range Dependence". Em Doukhan, P.; Oppenheim, G. e Taqqu, M.S. (editores), *Theory and Applications of Long-Range Dependence*. Boston: Birkhäuser, 5 - 38.

[67] Tsay, W.-J. (2010). "Maximum Likelihood Estimation of Stationary Multivariate ARFIMA Processes". *Journal of Statistical Computation and Simulation*, **80**, 729 - 745.

[68] Tucker, H.G. (1967). *A Graduate Course in Probability*. New York: Academic Press.

[69] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: University Press.

[70] Young, L-S. (1999). "Recurrence Times and Rates of Mixing". *Israel Journal of Mathematics*, **110**, 153 - 188.

[71] Zebrowsky, J.J. (2001). "Intermittency in Human Heart Rate Variability". *Acta Physica Polonica B*, **32**, 1531 - 1540.

Anexo A

Artigo Lopes e Pumi (2011)

Copulas Related to Manneville-Pomeau Processes

Sílvia R.C. Lopes and Guilherme Pumi
Mathematics Institute
Federal University of Rio Grande do Sul

This version: October, 10th, 2011

Abstract

In this work we derive the copulas related to Manneville-Pomeau processes. We examine both bidimensional and multidimensional cases and derive some properties for the related copulas. Computational issues, approximations and random variate generation problems are addressed and simple numerical experiments to test the approximations developed are also performed. In particular, we propose an approximation to the derived copulas which we show to converge uniformly to the true copula. To illustrate the usefulness of the theory, we derive a fast procedure to estimate the underlying parameter in Manneville-Pomeau processes.

Keywords. Copulas; Manneville-Pomeau Processes; Invariant Measures; Parametric Estimation.

1 Introduction

The statistics of stochastic processes derived from dynamical systems has seen a grown attention in the last decade or so (see Chazottes et al. (2005) and references therein). The relationship between copulas and areas such as ergodic theory and dynamical systems also have seen some development, especially in the last few years (see, for instance, Kolesárová et al. (2008)). In this work our aim is to contribute with the area by identifying and studying the copulas related to random vectors coming from the so-called Manneville-Pomeau processes, which are obtained as iterations of the Manneville-Pomeau transformation to a specific chosen random variable (see Definitions 2.1 and 2.2). We cover both, bidimensional and n-dimensional cases, which share a lot more in common than one could expect.

The copulas derived here depend on a probability measure which has no closed formula. In order to minimize this deficiency, we propose an approximation to the copula which we show to converge uniformly to the true copula. The copula also depend on several functions which have to be approximated as well, so the approximation depends on several intermediate steps. The results related to the convergence of the proposed approximation presented here are far more general than we need and actually allows one to change these intermediate approximations and still obtain the uniform convergence result for the approximated copula. We also address problems related to random variate generation of the copula and present the results of some simple numerical experiments in order to assess the stability and precision of the intermediate approximations. The usefulness of the theory is illustrated by a simple application to the problem of estimating the underlying parameter in Manneville-Pomeau processes.

The paper is organized as follows: in the next section, we briefly review some concepts and results on Manneville-Pomeau transformations and processes and on copulas. Section 3 is devoted to determine the copulas related to any pair (X_t, X_{t+h}) from a Manneville-Pomeau process and to explore some consequences. In Section 4, the multidimensional extensions are shown. In Section 5 an approximation to the copulas derived in Section 3 is proposed. This

approximation, which is shown to converge uniformly to the true copula, is then applied to exploit some characteristics of the copulas related to Manneville-Pomeau process through statistical and graphical analysis. Some computational and random variate generation problems are also addressed. In Section 6 we illustrate the usefulness of the theory by deriving a fast procedure to estimate the underlying parameter in Manneville-Pomeau processes. Conclusions are reserved to Section 7.

2 Some Background

In this section we shall briefly review some basic results on Manneville-Pomeau transformations and related processes as well as some concepts on copulas needed later. We start with the definition of the Manneville-Pomeau transformation.

Definition 2.1. The map $T_s : [0,1] \longrightarrow [0,1]$, given by

$$T_s(x) = x + x^{1+s} (\mathrm{mod}\, 1),$$

for $s > 0$, is called the *Manneville-Pomeau transformation* (MP *transformation*, for short).

In what follows, λ shall denote the Lebesgue measure in $I := [0,1]$ and the k-fold composition will be denoted, as usual, by $T_s^k = T_s \circ \cdots \circ T_s$. Figure 1 shows the plot of the MP transformation for the values of $s \in \{0.5, 1, 10\}$. The plots show the usual behavior of the MP transformations: for any s, they are increasing and differentiable functions by parts in I. Furthermore, for any $s > 0$, the function T_s^k will have exactly 2^k parts.

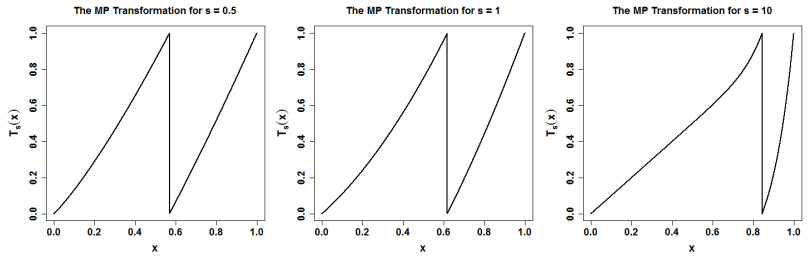


Figure 2.1: Plot of the Manneville-Pomeau transformation for different values of $s \in \{0.5, 1, 10, 10\}$.

Pianigiani (1980) shows the existence of a T_s-invariant and absolutely continuous measure with respect to the Lebesgue measure in I which will be denoted henceforth by μ_s. However, the proof uses Perron-Frobenius operator theory and is, for practical purposes, non-constructive so that an explicit form for a T_s-invariant measure is unknown. However, this measure will be a Sinai-Bowen-Ruelle (SBR) measure in the sense that the weak convergence

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T_s^k(x)}(A) \longrightarrow \mu_s(A) \tag{2.1}$$

holds for almost all $x \in I$ and all μ_s-continuity sets[1] A, where $\delta_a(\cdot)$ is the Dirac measure at a.

As a dynamical system, the triple (I, μ_s, T_s) is exact (that is, $\lim_{k\to\infty}(\mu_s \circ T_s^k)(A) = 1$, for all positive μ_s-measurable sets A) which implies ergodicity and strong-mixing. When $s < 1$,

[1]Recall that a set A is a μ-continuity set if $\mu(\partial A) = 0$, where ∂A denotes the boundary of A. The measure theoretical results applied here can be found, for instance, in Royden (1988). A good reference in weak convergence of probability measures is Billingsley (1999) and for ergodic theoretical related results, see Pollicott and Yuri (1998).

μ_s is a probability measure, while if $s \geq 1$, μ_s is no longer finite, but σ-finite (see Fisher and Lopes (2001)). Furthermore, it can be shown that μ_s has a positive, bounded continuous Radon-Nikodym derivative $\mathrm{d}\mu_s = h_s(x)\mathrm{d}x$, fact that will be useful later. For further details in the theory of MP transformations and related results, we refer to Pianigiani (1980), Young (1999), Maes et al. (2000) and Fisher and Lopes (2001). For applications, see Zebrowsky (2001), Olbermann et al. (2007) and Lopes and Lopes (1998).

Definition 2.2. Let $s \in (0,1)$ and let U_0 be a random variable distributed according to (the probability measure) μ_s. Let $\varphi : [0,1] \longrightarrow \mathbb{R}$ be a function in $\mathcal{L}^1(\mu_s)$. The stochastic process given by

$$X_t = (\varphi \circ T_s^t)(U_0), \quad \text{for all} \quad t \in \mathbb{N},$$

is called a *Manneville-Pomeau process* (or MP *process*, for short).

The MP process, as defined above, is stationary since μ_s is a T_s-invariant measure and $\mu_s \ll \lambda$. It is also ergodic since μ_s is ergodic for T_s. By its turn, an n-dimensional copula is a distribution function whose marginals are uniformly distributed on I. The copula literature has grown enormously in the last decade, especially in terms of empirical applications and have become standard tools in financial data analysis (see Nelsen (2006) and references therein). The next theorem, known as Sklar's theorem, is the key result for copulas and elucidates the role played by them. See Schweizer and Sklar (2005) for a proof.

Theorem 2.1 (Sklar's Theorem)**.** *Let $X_1, \cdots, X_n$ be random variables with marginals $F_1, \cdots, F_n$, respectively, and joint distribution function H. Then, there exists a copula C such that,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all} \quad (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

If the F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_n)$*, where for a function f,* $\mathrm{Ran}(f)$ *denotes the range of f. The converse also holds. Furthermore,*

$$C(u_1, \cdots, u_n) = H\big(F_1^{(-1)}(u_1), \cdots, F_n^{(-1)}(u_n)\big), \quad \textit{for all} \quad (u_1, \cdots, u_n) \in I^n,$$

where for a function F, $F^{(-1)}$ denotes its pseudo-inverse given by $F^{(-1)}(x) := \inf\big\{u \in \mathrm{Ran}(F) : F(u) \geq x\big\}$.

The next theorem, whose proof can be found, for instance, throughout Nelsen (2006), shall prove very useful in what follows. Except stated otherwise, the measure implicit to phrases like "almost sure", "almost everywhere" and so on will be the (appropriate) Lebesgue measure.

Theorem 2.2. *Let X and Y be continuous random variables with copula C. If f is an almost everywhere decreasing function then $C_{f(X),Y}(u,v) = u - C_{X,Y}(u, 1-v)$. Furthermore, if f_1 and f_2 are functions increasing almost everywhere, then $C_{f_1(X),f_2(Y)}(u,v) = C_{X,Y}(u,v)$.*

For an introduction to copulas, we refer the reader to Nelsen (2006). For more details and extensions to the multivariate case with emphasis in modeling and dependence concepts, see Joe (1997). The theory of copulas is also intimately related to the theory of probabilistic metric spaces (see Schweizer and Sklar (2005) for more details in this matter).

3 Copulas and MP Processes: Bidimensional Case

In this section we shall investigate the bidimensional copulas associated to pairs of random variables coming from MP processes which we shall call MP *copulas*. As we will see later, the

multidimensional case is very similar to the bidimensional case, so we shall give special attention to the latter.

First, let $\{X_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s \in (0,1)$ and $\varphi \in \mathcal{L}^1(\mu_s)$ be an increasing almost everywhere function. Throughout this section and in the rest of the paper, we shall treat $s \in (0,1)$ as a given fixed number. Let

$$F_0(x) := \mathbb{P}(U_0 \leq x) = \mu_s\big([0,x]\big).$$

Since $\mu_s \ll \lambda$, μ_s is non-atomic and, therefore, F_0 is (uniformly) continuous. The existence of a positive Radon-Nikodym density for μ_s also shows that F_0 is increasing and its inverse is well defined. Let F_t be the distribution function of $T_s^t(U_0)$, for all $t \in \mathbb{N}$. For $x \in I$, notice that

$$F_t(x) := \mathbb{P}\big(T_s^t(U_0) \leq x\big) = \mu_s\big((T_s^t)^{-1}\big([0,x]\big)\big) = \mu_s\big([0,x]\big) = F_0(x), \tag{3.1}$$

since μ_s is a T_s-invariant measure.

In what follows, we shall need the solution for the inequality $T_s^t(X) \leq y$, $y \in (0,1)$, in X, for X a random variable taking values in I. Now, since each of the 2^t parts of T_s^t is one-to-one in its domain, the inverse of T_s^t will also be continuous by parts and each part will also be a one-to-one function in its domain. Let $0 = a_{t,0}, \cdots, a_{t,2^t} = 1$ be the end points of each part of T_s^t. We shall call each interval $[a_{t,k}, a_{t,k+1})$ a *node* of T_s^t, for $k = 0, \cdots, 2^t - 1$ and $t > 0$. The (piecewise) inverse of T_s^t can be conveniently written as

$$\begin{aligned} (T_s^t)^{-1} : I &\longrightarrow I^{2^t} \\ y &\longmapsto \big(\mathcal{T}_{t,0}(y), \cdots, \mathcal{T}_{t,2^t-1}(y)\big), \end{aligned} \tag{3.2}$$

where $\mathcal{T}_{t,k}(y)$ denotes the inverse of T_s^t restricted to its k-th node, for all $k \in \{0, \cdots, 2^t - 1\}$. Notice that both $\mathcal{T}_{t,k}$ and $a_{t,k}$ depend on s for each k, but since no confusion will arise, and for the sake of simplicity, we shall omit this dependence from the notation as we shall do in several other occasions. Now, the solution of the inequality $T_s^t(X) \leq y$ in X can be determined and is given by $X \in A_{t,0}(y) \bigcup \cdots \bigcup A_{t,2^t-1}(y)$, where

$$A_{t,k}(y) = \big[a_{t,k}, \mathcal{T}_{t,k}(y)\big], \tag{3.3}$$

which is a proper closed subinterval of $[a_{t,k}, a_{t,k+1})$, for each $k = 0, \cdots, 2^t - 1$. Notice that $A_{t,k}(y)$ (whose dependence on s was omitted from the notation) is just the inverse image of $[0,y]$ by the transformation T_s^t restricted to the node $[a_{t,k}, a_{t,k+1})$. We can now use this result to prove the following useful lemma.

Lemma 3.1. *Let X be a random variable taking values in I and let T_s be the* MP *transformation with parameter $s > 0$. Then, for any $t \in \mathbb{N}$ and $x \in I$,*

$$\mathbb{P}\big(T_s^t(X) \leq x\big) = \mathbb{P}\big(X \in \textstyle\bigcup_{k=0}^{2^t-1} A_{t,k}(x)\big) = \displaystyle\sum_{k=0}^{2^t-1} \mathbb{P}\big(X \in A_{t,k}(x)\big),$$

where $A_{t,k}$'s are given by (3.3).

Proof: The result follows easily from what was just discussed and from the fact that the intervals $A_{t,k}$'s are (pairwise) disjoints. ■

As for the copulas related to MP processes, in view of the stationarity of the MP process, the following result follows easily.

Proposition 3.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an* MP *process with parameter $s \in (0,1)$ and $\varphi \in \mathcal{L}^1(\mu_s)$ be an almost everywhere increasing function. Then, for any $t, h \in \mathbb{N}$,*

$$C_{X_t, X_{t+h}}(u,v) = C_{X_0, X_h}(u,v),$$

everywhere in I^2.

Proof: As consequence of the stationarity of $\{X_t\}_{t\in\mathbb{N}}$, if we let the joint distribution of the pair (X_p, X_q), for any $p, q \in \mathbb{N}$, $p \neq q$, be denoted by $\widetilde{H}_{p,q}(\cdot,\cdot)$, it follows that for all $x, y \in (0,1)$, $t \in \mathbb{N}$ and $h \in \mathbb{N}^* := \mathbb{N}\backslash\{0\}$, $\widetilde{H}_{t,t+h}(x,y) = \widetilde{H}_{0,h}(x,y)$. Now, upon applying Sklar's Theorem and (3.1), it follows that

$$C_{X_t,X_{t+h}}(u,v) = \widetilde{H}_{t,t+h}\big(F_t^{-1}(u), F_{t+h}^{-1}(v)\big) = \widetilde{H}_{0,h}\big(F_0^{-1}(u), F_h^{-1}(v)\big) = C_{X_0,X_h}(u,v),$$

for all $(u,v) \in I^2$. ■

Corollary 3.1. *Let T_s be the* MP *transformation for some $s \in (0,1)$, μ_s be a T_s-invariant probability measure and let U_0 be distributed as μ_s. Then, for any $t, h \in \mathbb{N}$, $h \neq 0$,*

$$C_{T_s^t(U_0),T_s^{t+h}(U_0)}(u,v) = C_{U_0,T_s^h(U_0)}(u,v)$$

everywhere in I^2.

Proof: Immediate from Theorem 2.2 applied to Proposition 3.1. ■

Now we turn our attention to determine the copula associated to any pair of random variables (X_p, X_q), $p, q \in \mathbb{N}$, obtained from an MP process with φ an increasing almost everywhere function. For the sake of simplicity, let us introduce the following functions: let h be a positive integer and for $k = 0, \cdots, 2^h - 1$, let $\mathcal{F}_{h,k} : I \to \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big]$ be given by

$$\mathcal{F}_{h,k}(x) := F_0\big(\mathcal{T}_{h,k}\big(F_0^{-1}(x)\big)\big).$$

Notice that for each k, $\mathcal{F}_{h,k}(0) = F_0(a_{h,k})$ and $\mathcal{F}_{h,k}(1) = F_0(a_{h,k+1})$ and $\mathcal{F}_{h,k}$ is a one to one, increasing and uniformly continuous function.

Proposition 3.2. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an* MP *process with parameter $s \in (0,1)$, $\varphi \in \mathcal{L}^1(\mu_s)$ be an increasing almost everywhere function and let F_0 be the distribution function of U_0. Then, for any $t, h \in \mathbb{N}$, $h \neq 0$ and $(u,v) \in I^2$,*

$$C_{X_t,X_{t+h}}(u,v) = \left(\sum_{k=0}^{n_0-1} \mathcal{F}_{h,k}(v) - F_0(a_{h,k})\right)\delta_{\mathbb{N}^*}(n_0) + \min\big\{u, \mathcal{F}_{h,n_0}(v)\big\} - F_0(a_{h,n_0}), \tag{3.4}$$

where $\delta_{\mathbb{N}^*}(x)$ equals 1, if $x \in \mathbb{N}^*$ and 0, otherwise, $\{a_{h,k}\}_{k=0}^{2^h}$ are the end points of the nodes of T_s^h and $n_0 := n_0(u;h) = \big\{k : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\} \in \{0, \cdots, 2^h - 1\}$.

Proof: By Propositions 3.1 and 2.2, it suffices to derive the copula of the pair $\big(U_0, T_s^h(U_0)\big)$. So let again $\{X_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s \in (0,1)$ and $\varphi \in \mathcal{L}^1(\mu_s)$ be an increasing almost everywhere function and let $H_{0,h}(\cdot,\cdot)$ denote the distribution function of the pair $\big(U_0, T_s^h(U_0)\big)$. Notice that

$$\begin{aligned} H_{0,h}(x,y) =& \mathbb{P}(U_0 \le x, T_s^h(U_0) \le y) = \mathbb{P}\big(U_0 \le x, U_0 \in \textstyle\bigcup_{k=0}^{2^h-1} A_{h,k}(y)\big) \\ &= \mathbb{P}\big(U_0 \in [0,x] \textstyle\bigcap \bigcup_{k=0}^{2^h-1} A_{h,k}(y)\big) = \mathbb{P}\big(U_0 \in \textstyle\bigcup_{k=0}^{2^h-1} \big[[0,x] \bigcap A_{h,k}(y)\big]\big) \\ &= \sum_{k=0}^{2^h-1} \mathbb{P}\big(U_0 \in [0,x] \textstyle\bigcap A_{h,k}(y)\big), \end{aligned}$$

for any $x, y \in (0,1)$. Now let $n_1 := n_1(x;h) = \big\{k : x \in [a_{h,k}, a_{h,k+1})\big\} \in \{0, \cdots, 2^h - 1\}$ and assume for the moment that $n_1 \ge 1$. Since $A_{h,k}(y) = \big[a_{h,k}, \mathcal{T}_{h,k}(y)\big]$, it follows

$$H_{0,h}(x,y) = \sum_{k=0}^{n_1-1} \mathbb{P}\big(U_0 \in A_{h,k}(y)\big) + \mathbb{P}\big(U_0 \in A_{h,n_1}(y) \textstyle\bigcap [a_{h,n_1}, x]\big)$$

$$= \sum_{k=0}^{n_1-1} \mu_s\big(A_{h,k}(y)\big) + \mu_s\big(\big[a_{h,n_1}, \mathcal{T}_{h,n_1}(y)\big] \bigcap [a_{h,n_1}, x]\big)$$
$$= \sum_{k=0}^{n_1-1} \mu_s\big(\big[a_{h,k}, \mathcal{T}_{h,k}(y)\big]\big) + \mu_s\big(\big[a_{h,n_1}, \min\{x, \mathcal{T}_{h,n_1}(y)\}\big]\big),$$

which can be written, since $F_0(x) = \mu_s([0,x])$ is increasing, as

$$H_{0,h}(x,y) = \sum_{k=0}^{n_1-1} \big[F_0\big(\mathcal{T}_{h,k}(y)\big) - F_0(a_{h,k})\big] + \min\big\{F_0(x), F_0\big(\mathcal{T}_{h,n_1}(y)\big)\big\} - F_0(a_{h,n_1}).$$

If $n_1 = 0$, the summation is absent of the formula and we have

$$H_{0,h}(x,y) = \min\big\{F_0(x), F_0\big(\mathcal{T}_{h,0}(y)\big)\big\} - F_0(a_{h,0}),$$

so that, in any case, we have

$$H_{0,h}(x,y) = \left(\sum_{k=0}^{n_1-1} \big[F_0\big(\mathcal{T}_{h,k}(y)\big) - F_0(a_{h,k})\big]\right)\delta_{\mathbb{N}^*}(n_1) + \min\big\{F_0(x), F_0\big(\mathcal{T}_{h,n_1}(y)\big)\big\} - F_0(a_{h,n_1}).$$

Now upon applying Sklar's Theorem, it follows that

$$C_{U_0,T_s^h(U_0)}(u,v) = H_{0,h}\big(F_0^{-1}(u), F_h^{-1}(v)\big) = H_{0,h}\big(F_0^{-1}(u), F_0^{-1}(v)\big)$$
$$= \left(\sum_{k=0}^{n_0-1} \mathcal{F}_{h,k}(v) - F_0(a_{h,k})\right)\delta_{\mathbb{N}^*}(n_0) + \min\big\{u, \mathcal{F}_{h,n_0}(v)\big\} - F_0(a_{h,n_0}),$$

where $n_0 := n_0(u;h) = n_1\big(F_0^{-1}(u);h\big) = \big\{k : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\}$. The result now follows from Proposition 3.1. ∎

Remark 3.1. Notice that the copula (3.4) can be expressed in terms of μ_s as

$$C_{X_t,X_{t+h}}(u,v) = \left(\sum_{k=0}^{n_0-1} \mu_s\Big(\big[a_{h,k}, \mathcal{T}_{h,k}\big(F_0^{-1}(v)\big)\big]\Big)\right)\delta_{\mathbb{N}^*}(n_0) +$$
$$+ \mu_s\Big(\big[a_{h,n_0}, \min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}\big]\Big), \tag{3.5}$$

which will prove useful in Section 5. Also, expression (3.5) is helpful if one desires to verify directly that the marginals of (3.4) are indeed uniform.

In the next proposition we address the case where φ is an almost everywhere decreasing function. In view of Theorem 2.2, one could, at first glance, think that a result like $C_{X_0,X_h} = C_{X_t,X_{t+h}}$ would not hold anymore, but in fact it still does, as it is shown in the next proposition.

Proposition 3.3. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an* MP *process with parameter $s \in (0,1)$, $\varphi \in \mathcal{L}^1(\mu_s)$ be an almost everywhere decreasing function and let F_0 be the distribution function of U_0. Then, $C_{X_0,X_h}(u,v) = C_{X_t,X_{t+h}}(u,v)$ everywhere in I^2 and, for any $t, h \in \mathbb{N}$ and $h \neq 0$,*

$$C_{X_t,X_{t+h}}(u,v) = u + v - 1 + \left(\sum_{k=0}^{n_0-1} [\mathcal{F}_{h,k}(1-v) - F_0(a_{h,k})]\right)\delta_{\mathbb{N}^*}(n_0) +$$
$$+ \min\big\{1-u, \mathcal{F}_{h,n_0}(1-v)\big\} - F_0(a_{h,n_0}), \tag{3.6}$$

for all $(u,v) \in I^2$, where $\{a_{h,k}\}_{k=0}^{2^h}$ are the end points of the nodes of T_s^h and $n_0 := n_0(u;h) = \big\{k : u \in \big(1 - F_0(a_{h,k+1}), 1 - F_0(a_{h,k})\big]\big\}$.

Proof: Since the inverse of an almost everywhere decreasing function is still decreasing almost everywhere and $X_t = \varphi\big(T_s^t(U_0)\big)$, upon applying Theorem 2.2 twice, it follows that

$$\begin{aligned} C_{T_s^t(U_0),\,T_s^{t+h}(U_0)}(u,v) &= C_{\varphi^{-1}(X_t),\,\varphi^{-1}(X_{t+h})}(u,v) = u - C_{X_t,\varphi^{-1}(X_{t+h})}(u,1-v) \\ &= u - \big(1 - v - C_{X_t,\,X_{t+h}}(1-u,1-v)\big), \end{aligned}$$

or, equivalently (changing u by $1-u$ and v by $1-v$),

$$C_{X_t,\,X_{t+h}}(u,v) = u + v - 1 + C_{T_s^t(U_0),\,T_s^{t+h}(U_0)}(1-u,1-v). \tag{3.7}$$

Now (3.6) follows upon applying Proposition 3.2 with the identity map and substituting equation (3.4) into (3.7). As for the equality $C_{X_0,X_h}(u,v) = C_{X_t,X_{t+h}}(u,v)$, Corollary 3.1 and Theorem 2.2 applied to (3.7) yield

$$\begin{aligned} C_{X_t,\,X_{t+h}}(u,v) &= u + v - 1 + C_{U_0,\,T_s^h(U_0)}(1-u,1-v) \\ &= u + v - 1 + C_{\varphi^{-1}(\varphi(U_0)),\varphi^{-1}(\varphi(T_s^h(U_0)))}(1-u,1-v) \\ &= C_{\varphi(U_0),\varphi(T_s^h(U_0))}(u,v) \;=\; C_{X_0,X_h}(u,v), \end{aligned}$$

everywhere in I^2, as desired. ■

Remark 3.2. In view of the "stationarity" results of Theorems 3.1 and 3.3, a copula associated to a pair (X_t, X_{t+h}) from an MP process will be referred to as *lag h* MP *copula*.

The copulas in (3.4) and (3.6) are both singular, as it can be readily verified. So the question that naturally arises is, for each h, what is the support of $C_{X_t,X_{t+h}}$? The question is addressed in the next proposition, which will be useful in Sections 5 and 6. For simplicity, for a given MP process and $h > 0$, let $\ell_{h,k}^+, \ell_{h,k}^- : \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big) \to I$ be functions defined by

$$\ell_{h,k}^+(x) = \frac{x - F_0(a_{h,k})}{F_0(a_{h,k+1}) - F_0(a_{h,k})} \quad \text{and} \quad \ell_{h,k}^-(x) = \frac{F_0(a_{h,k+1}) - x}{F_0(a_{h,k+1}) - F_0(a_{h,k})},$$

for all $k = 0, \cdots, 2^h - 1$. Notice that, for each k, $\ell_{h,k}^+$ is the linear function connecting the points $\big(F_0(a_{h,k}), 0\big)$ and $\big(F_0(a_{h,k+1}), 1\big)$, while $\ell_{h,k}^-$ connects the points $\big(F_0(a_{h,k}), 1\big)$ and $\big(F_0(a_{h,k+1}), 0\big)$.

Proposition 3.4. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an* MP *process with parameter $s \in (0,1)$, for $\varphi_1 \in \mathcal{L}^1(\mu_s)$ an almost everywhere increasing function and let $\{Y_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s \in (0,1)$, for $\varphi_2 \in \mathcal{L}^1(\mu_s)$ an almost everywhere decreasing function. Also let F_0 be the distribution function of U_0. Then, for any $t, h \in \mathbb{N}$, $h > 0$,*

$$\operatorname{supp}\{C_{X_t,X_{t+h}}\} = \bigcup_{k=0}^{2^h-1} \big\{\big(u, \ell_{h,k}^+(u)\big) : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\} \tag{3.8}$$

and

$$\operatorname{supp}\{C_{Y_t,Y_{t+h}}\} = \bigcup_{k=0}^{2^h-1} \big\{\big(u, \ell_{h,k}^-(u)\big) : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\}. \tag{3.9}$$

Proof: Let $R = [u_1, u_2] \times [v_1, v_2]$ be a rectangle in I^2 and let its $C_{X_t,X_{t+h}}$-volume be denoted by $V_{C_{\boldsymbol{X}}}(R)$. Let $k \in \{0, \cdots, 2^h - 1\}$ be fixed and suppose that $u_i \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big]$. This implies that $n_0 = k$ for all four terms in $V_{C_{\boldsymbol{X}}}(R)$, hence the summands and constants on the copula cancel out so that we have

$$\begin{aligned} V_{C_{\boldsymbol{X}}}(R) &= \min\big\{u_1, \mathcal{F}_{h,k}(v_1)\big\} + \min\big\{u_2, \mathcal{F}_{h,k}(v_2)\big\} - \min\big\{u_1, \mathcal{F}_{h,k}(v_2)\big\} - \min\big\{u_2, \mathcal{F}_{h,k}(v_1)\big\} \\ &= V_M\big([u_1, u_2] \times [\mathcal{F}_{h,k}(v_1), \mathcal{F}_{h,k}(v_1)]\big), \end{aligned}$$

where $M(u,v) = \min\{u,v\}$ is the Frechèt upper bound copula whose support is the main diagonal in I^2. Since $[u_1, u_2] \times [\mathcal{F}_{h,k}(v_1), \mathcal{F}_{h,k}(v_1)] \subset [F_0(a_{h,k}), F_0(a_{h,k+1})]^2$, $V_{C_{\boldsymbol{X}}}(R) > 0$ if, and only if, $R \bigcap \big\{\big(u, \ell_{h,k}^+(u)\big) : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\} \neq \emptyset$.

Analogously, denoting the $C_{Y_t,Y_{t+h}}$-volume of R by $V_{C_{\boldsymbol{Y}}}(R)$, if $u_i \in \big[1-F_0(a_{h,k}), 1-F_0(a_{h,k+1})\big]$, we have

$$\begin{aligned} V_{C_{\boldsymbol{Y}}}(R) &= \min\big\{1-u_1, \mathcal{F}_{h,k}(1-v_1)\big\} + \min\big\{1-u_2, \mathcal{F}_{h,k}(1-v_2)\big\} - \\ &\quad - \min\big\{1-u_1, \mathcal{F}_{h,k}(1-v_2)\big\} - \min\big\{1-u_2, \mathcal{F}_{h,k}(1-v_1)\big\} \\ &= V_M\big([1-u_1, 1-u_2]\times[\mathcal{F}_{h,k}(1-v_2), \mathcal{F}_{h,k}(1-v_1)]\big). \end{aligned} \tag{3.10}$$

Since $[1-u_1,1-u_2]\times[\mathcal{F}_{h,k}(1-v_1),\mathcal{F}_{h,k}(1-v_2)] \subset [F_0(a_{h,k}), F_0(a_{h,k+1})]^2$, $V_{C_{\boldsymbol{Y}}}(R)$ is positive if, and only if, $R \bigcap \big\{\big(u, \ell^-_{h,k}(u)\big) : u \in \big[F_0(a_{h,k}), F_0(a_{h,k+1})\big)\big\} \neq \emptyset$ (notice the terms $1-v_i$ in expression (3.10), for $i=1,2$). Now (3.8) and (3.9) follow by observing that $I = \bigcup_{k=0}^{2^h-1}\big[F_0(a_{h,k}), F_0(a_{h,k+1})\big] = \bigcup_{k=0}^{2^h-1}\big[1-F_0(a_{h,k+1}), 1-F_0(a_{h,k})\big]$. ∎

Remark 3.3. We close this section by noticing that as an application of Propositions 3.1 and 3.3, together with the so-called copula version of Hoeffding's lemma (see Nelsen (2006)), we can show in a rather different way that an MP process is weakly stationary. Let F_{X_t} be the distribution function of X_t and notice that $F_{X_t}(x) = F_{X_0}(x)$, for all $t \in \mathbb{N}$, by the stationarity of $\{X_t\}_{t\in\mathbb{N}}$ and since $C_{X_t,X_{t+h}}(u,v) = C_{X_0,X_h}(u,v)$, the result follows immediately.

4 Multidimensional Case

In this section we are interested in extending the results from the previous section to the multidimensional case, that is, in this section we are interested in deriving the copulas associated to n-dimensional vectors $(X_{t_1},\cdots,X_{t_n})$, $t_1,\cdots,t_n \in \mathbb{N}$, coming from an MP process with φ an increasing almost everywhere function. In view of Theorem 2.2, it suffices to derive the copula associated to the vector $\big(T_s^{t_1}(U_0),\cdots,T_s^{t_n}(U_0)\big)$. It turns out that there are more similarities between the bidimensional and multidimensional cases than one could expect. In fact, an expression very similar in form to (3.4) holds for the multidimensional case as well.

Let $\{X_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s \in (0,1)$ and $\varphi \in \mathcal{L}^1(\mu_s)$ be an almost everywhere increasing function. For the sake of simplicity, we shall use the following notation: let $a,b \in \mathbb{N}$, $a<b$, we shall write $x_{a:b} := (x_a,\cdots,x_b)$ and for a function f, $f(x_{a:b}) := \big(f(x_a),\cdots,f(x_b)\big)$. Again we shall denote the distribution function of U_0 by F_0.

Theorem 4.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an* MP *process with parameter $s\in(0,1)$, with $\varphi \in \mathcal{L}^1(\mu_s)$ an almost everywhere increasing function. Let $t_1,\cdots,t_n \in \mathbb{N}$ and set $h_i := t_i - t_1$. Then, for all $(u_1,\cdots,u_n) \in I^n$,*

$$\begin{aligned} C_{X_{t_1},\cdots,X_{t_n}}(u_1,\cdots,u_n) &= \left(\sum_{k=0}^{n_0-1} F_0\Big(b_{h_n,k}\big(F_0^{-1}(u_{2:n})\big)\Big) - F_0(a_{h_n,k})\right)\delta_{\mathbb{N}^*}(n_0) + \\ &\quad + \min\big\{u_1, F_0\big(b_{h_n,n_0}\big(F_0^{-1}(u_{2:n})\big)\big)\big\} - F_0(a_{h_n,n_0}), \end{aligned} \tag{4.1}$$

where $n_0 := n_0(u_1,n) = \big\{k : u_1 \in \big[F_0(a_{h_n,k}), F_0(a_{h_n,k+1})\big)\big\}$, $\{a_{h_n,k}\}_{k=0}^{2^h}$ are the end points of the nodes of $T_s^{h_n}$, for $i=2,\cdots,n$, $j=0,\cdots 2^{h_i}-1$, $\mathcal{T}_{h_i,j}$ is given by (3.2) *and for a vector $(x_2,\cdots,x_n)\in I^{n-1}$, $b_{h_n,k}(x_{2:n}) = \min\limits_{i=2,\cdots,n}\big\{c_i(x_i;h_n,k)\big\}$, with*

$$c_i(x_i;h_n,k) = \begin{cases} a_{h_n,k}, & \text{if } \ B_i(x_i;h_n,k) = \emptyset; \\ B_i(x_i;h_n,k), & \text{otherwise.} \end{cases}$$

and

$$B_i(x_i;h_n,k) = \min_{j=0,\cdots,2^{h_i}-1}\big\{\mathcal{T}_{h_i,j}(x_i) : \mathcal{T}_{h_i,j}(x_i) > a_{h_n,k} \text{ and } a_{h_i,j} < a_{h_n,k+1}\big\}.$$

Proof: Let $\{X_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s\in(0,1)$ and $\varphi\in\mathcal{L}^1(\mu_s)$ be an almost everywhere increasing function. Without loss of generality, we can assume that $0<t_1<\cdots<t_n$. In view of Theorem 2.2, it suffices to work with the vector $\big(T_s^{t_1}(U_0),\cdots,T_s^{t_n}(U_0)\big)$. Let $H_{t_1,\cdots,t_n}$ be the distribution function of $\big(T_s^{t_1}(U_0),\cdots,T_s^{t_n}(U_0)\big)$. Let $h_i=t_i-t_1$, for each $i=1,\cdots,n$, and notice that $h_i>0$ since $t_1<t_i$, for all $i=2,\cdots,n$. Let $(x_1,\cdots,x_n)\in(0,1)^n$ and for the sake of simplicity, let $Y_{t_1}:=T_s^{t_1}(U_0)$, so that we have

$$\begin{aligned}
H_{t_1,\cdots,t_n}(x_1,\cdots,x_n)&=\mathbb{P}\big(T_s^{t_1}(U_0)\le x_1,\cdots,T_s^{t_n}(U_0)\le x_n\big)\\
&=\mathbb{P}\big(Y_{t_1}\le x_1,T_s^{h_2}(Y_{t_1})\le x_2,\cdots,T_s^{h_n}(Y_{t_1})\le x_n\big)\\
&=\mathbb{P}\Big(Y_{t_1}\in[0,x_1],Y_{t_1}\in\bigcup_{k=0}^{2^{h_2}-1}A_{h_2,k}(x_2),\cdots,Y_{t_1}\in\bigcup_{k=0}^{2^{h_n}-1}A_{h_n,k}(x_n)\Big)\\
&=\mathbb{P}\Big(Y_{t_1}\in[0,x_1]\bigcap_{i=2}^n\big[\bigcup_{k=0}^{2^{h_i}-1}A_{h_i,k}(x_i)\big]\Big)\\
&=\mathbb{P}\Big(U_0\in\bigcap_{i=2}^n\bigcup_{k=0}^{2^{h_i}-1}\big[[0,x_1]\bigcap A_{h_i,k}(x_i)\big]\Big),
\end{aligned}\tag{4.2}$$

where $A_{h_i,k}$'s are given by (3.3) and the last equality is a consequence of the T_s-invariance of μ_s. For $k=0,\cdots,2^{h_n-1}$, let

$$\widetilde{A}_{h_n,k}(x_{2:n})=A_{h_n,k}(x_n)\bigcap_{i=2}^{n-1}\big[\bigcup_{j=0}^{2^{h_i}-1}A_{h_i,j}(x_i)\big].$$

In order to simplify the notation, for $i=2,\cdots,n$ and $k=0,\cdots,2^{h_n}-1$, let $B_i(x_i;h_n,k)$ be as in the enunciate. Notice that, for each k and i, $B_i(x_i;h_n,k)$ is either empty or the smallest $\mathcal{T}_{h_i,j}(x_i)$ which is greater than $a_{h_n,k}$ and such that the correspondent $A_{h_i,j}(x_i)$ has non-empty intersection with $A_{h_n,k}(x_n)$. Let

$$c_i(x_i;h_n,k)=\begin{cases}a_{h_n,k}, & \text{if}\quad B_i(x_i;h_n,k)=\emptyset;\\ B_i(x_i;h_n,k), & \text{otherwise.}\end{cases}$$

Then, for each $k=1,\cdots,2^{h_n}-1$, setting $b_{h_n,k}(x_{2:n})=\min\limits_{i=2,\cdots,n}\big\{c_i(x_i;h_n,k)\big\}$, it follows that

$$\widetilde{A}_{h_n,k}(x_{2:n})=\big[a_{h_n,k},b_{h_n,k}(x_{2:n})\big],$$

which is a closed subset of $[a_{h_n,k},a_{h_n,k+1}]$. Also notice that, from the definition of $b_{h_n,k}(x_{2:n})$, we could have $\widetilde{A}_{h_n,k}(x_{2:n})=\{a_{h_n,k}\}$, in which case we set $\widetilde{A}_{h_n,k}(x_{2:n})=\emptyset$ (although from a measure-theoretical point of view, this correction makes no difference). Again we are omitting the dependence in s from the notation on both, $b_{h_n,k}$ and $\widetilde{A}_{h_n,k}$. Each $b_{h_n,k}(x_{2:n})$ determines the smallest $\mathcal{T}_{h_i,j}(x_i)$ that lies on the k-th node of $T_s^{h_n}$ (which has the smallest nodes among all $T_s^{h_i}$'s), so that $\widetilde{A}_{h_n,k}$'s are just the intersection of all $A_{h_i,k}(x_i)$'s with end point in the k-th node of $T_s^{h_n}$. Also notice that the $\widetilde{A}_{h_n,k}$'s are pairwise disjoints. One can rewrite (4.2) as

$$H_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=\mathbb{P}\Big(U_0\in\bigcup_{k=0}^{2^{h_n}-1}\big[\widetilde{A}_{h_n,k}(x_{2:n})\bigcap[0,x_1]\big]\Big).\tag{4.3}$$

Now, let $n_1:=n_1(x_1;n)=\big\{k:x_1\in[a_{h_n,k},a_{h_n,k+1})\big\}\in\{0,\cdots,2^{h_n}-1\}$, and assume for the moment that $n_1\ge1$. Then (4.3) becomes

$$\begin{aligned}
H_{t_1,\cdots,t_n}(x_1,\cdots,x_n)&=\sum_{k=0}^{n_1-1}\mathbb{P}\big(U_0\in\widetilde{A}_{h_n,k}(x_{2:n})\big)+\mathbb{P}\big(U_0\in\widetilde{A}_{h_n,n_1}(x_{2:n})\bigcap[a_{h_n,n_1},x_1]\big)\\
&=\sum_{k=0}^{n_1-1}\mu_s\big([a_{h_n,k},b_{h_n,k}(x_{2:n})]\big)+\mu_s\big([a_{h_n,n_1},\min\{x_1,b_{h_n,n_1}(x_{2:n})\}]\big)\\
&=\sum_{k=0}^{n_1-1}\big[F_0\big(b_{h_n,k}(x_{2:n})\big)-F_0(a_{h_n,k})\big]+\min\big\{F_0(x_1),F_0(b_{h_n,n_1}(x_{2:n}))\big\}-F_0(a_{h_n,n_1}).
\end{aligned}$$

If $n_1 = 0$, then

$$H_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = \min\big\{F_0(x_1), F_0(b_{h_n,0}(x_{2:n}))\big\} - F_0(a_{h_n,0}).$$

In any case, we can write

$$\begin{aligned} H_{t_1,\cdots,t_n}(x_1,\cdots,x_n) &= \left(\sum_{k=0}^{n_1-1} F_0\big(b_{h_n,k}(x_{2:n})\big) - F_0(a_{h_n,k})\right)\delta_{\mathbb{N}^*}(n_1) + \\ &\quad + \min\big\{F_0(x_1), F_0(b_{h_n,n_1}(x_{2:n}))\big\} - F_0(a_{h_n,n_1}). \end{aligned}$$

Recall that the distribution function of $T_s^t(U_0)$ is also F_0 by the T_s-invariance of μ_s. Now applying Sklar's Theorem, it follows that,

$$\begin{aligned} C_{X_{t_1},\cdots,X_{t_n}}(u_1,\cdots,u_n) &= H_{t_1,\cdots,t_n}\big(F_0^{-1}(u_1),\cdots,F_0^{-1}(u_n)\big) \\ &= \left(\sum_{k=0}^{n_0-1} F_0\Big(b_{h_n,k}\big(F_0^{-1}(u_{2:n})\big)\Big) - F_0(a_{h_n,k})\right)\delta_{\mathbb{N}^*}(n_1) + \\ &\quad + \min\big\{u_1, F_0\big(b_{h_n,n_0}\big(F_0^{-1}(u_{2:n})\big)\big)\big\} - F_0(a_{h_n,n_0}), \end{aligned}$$

where $n_0 := n_1\big(F_0^{-1}(u_1), n\big) = \big\{k : u_1 \in \big[F_0(a_{h_n,k}), F_0(a_{h_n,k+1})\big)\big\}$, which is the desired formula. ■

Remark 4.1. Notice that the proof of Theorem 4.1 from equation (4.3) on is exactly the same as the one in Proposition 3.2 with the obvious notational adaptations.

Now we turn our attention to the case where φ is an almost everywhere decreasing function. In view of Theorem 2.2, one cannot expect a simple expression for the copula. What happens is that the copula in this case will be the sum of the lower dimensions copulas related to the iterations $T_s^k(U_0)$, as the next proposition shows.

Proposition 4.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an MP process with parameter $s \in (0,1)$, and $\varphi \in \mathcal{L}^1(\mu_s)$ be an almost everywhere decreasing function. Let $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$ and set $Y_0 := U_0$ and $Y_k := T_s^{h_k}(U_0)$. Denote the copula associated to the random vector $(X_t, X_{t+h_1}, \cdots, X_{t+h_n})$ by $C_{\boldsymbol{t}}$. Then the following relation holds*

$$\begin{aligned} C_{\boldsymbol{t}}(u_0,\cdots,u_n) &= 1 - n + \sum_{i=0}^{n} u_i + \sum_{i=0}^{n}\sum_{j=i+1}^{n} C_{Y_i,Y_j}(1-u_i, 1-u_j) + \cdots + \\ &\quad + (-1)^{n-1}\sum_{k_1=0}^{n}\sum_{k_2=k_1+1}^{n}\cdots\sum_{k_{n-1}=k_{n-2}+1}^{n} C_{Y_{k_1},\cdots,Y_{k_{n-1}}}(1-u_{k_1},\cdots,1-u_{k_{n-1}}) + \\ &\quad + (-1)^n C_{U_0,Y_1,\cdots,Y_n}(1-u_0,\cdots,1-u_n), \end{aligned} \tag{4.4}$$

everywhere in I^{n+1}.

Proof: Let $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$, $t \neq 0$. Set $Y_0 := U_0$, $Y_k := T_s^{h_k}(U_0)$ and $y_k := \varphi(x_k)$. We have

$$\begin{aligned} H_{X_0,X_{h_1},\cdots,X_{h_n}}(x_0,x_1,\cdots,x_n) &= \mathbb{P}\big(U_0 \geq y_0, Y_1 \geq y_1, \cdots, Y_n \geq y_n\big) \\ &= \mathbb{P}\Big(U_0 \geq y_0 \big| Y_1 \geq y_1, Y_2 \geq y_2, \cdots, Y_n \geq y_n\Big)\mathbb{P}\Big(Y_1 \geq y_1, \cdots, Y_n \geq y_n\Big) \\ &= \mathbb{P}\Big(Y_1 \geq y_1, \cdots, Y_n \geq y_n\Big) - \mathbb{P}\big(U_0 \leq y_0, Y_1 \geq y_1, \cdots, Y_n \geq y_n\big). \end{aligned} \tag{4.5}$$

Upon applying a long chain of a conditioning argument on both terms in (4.5), we arrive at

$$H_{X_0,X_{h_1},\cdots,X_{h_n}}(x_0,x_1,\cdots,x_n) = 1 - \sum_{i=0}^{n} F_0(y_i) + \sum_{i=0}^{n}\sum_{j=i+1}^{n} H_{Y_i,Y_j}(y_i,y_j) +$$

$$+\cdots+(-1)^{n-1}\sum_{k_1=0}^{n}\sum_{k_2=k_1+1}^{n}\cdots\sum_{k_{n-1}=k_{n-2}+1}^{n}H_{Y_{k_1},\cdots,Y_{k_{n-1}}}(y_{k_1},\cdots,y_{k_{n-1}})+$$
$$+(-1)^n H_{U_0,Y_1,\cdots,Y_n}(y_0,\cdots,y_n). \tag{4.6}$$

A simple calculation (using the T_s-invariance of μ_s) shows that, for all $t\in\mathbb{N}^*$ and $x\in(0,1)$,

$$F_{X_t}(x)=1-F_0\big(\varphi(x)\big)\quad\text{ and }\quad F_{X_t}^{-1}(x)=\varphi^{-1}\big(F_0^{-1}(1-x)\big),$$

so that, the result follows upon applying Sklar's Theorem to (4.6) (recall that $y_k=\varphi(x_k)$). ■

Remark 4.2. Notice that the copula in Proposition 4.1 can be explicitly calculated since (4.4) is written as sums of the copulas of vectors containing U_0 and $T^t(U_0)$ for different t's, so that the desired formulas can be deduced in terms of the copulas in Theorem 4.1.

5 Numerical Approximations to the MP copulas

The MP copulas derived in the last sections do not have readily computable formulas, especially because μ_s does not have explicit expression and because even apparently simple tasks like determining the discontinuity points of T_s^h or to compute explicit formulas for the branches of T_s^h can be highly complex ones. However, one can still study the copulas derived in the last sections by using appropriate approximations to the functions appearing in the copula expression. Besides the invariant measure μ_s, computation of the bidimensional copulas so far discussed also involves computation of the quantile function F_0^{-1}, the inverse of T_s^h and the end points $\{a_{h,k}\}_{k=0}^{2^h}$ of the nodes of T_s^h.

In this section our goal is to derive simple approximations to these functions in order to obtain an approximation to the copula itself, which we shall prove to converge uniformly in its arguments to the true copula. The approximations presented here are simple ones, usually a linear interpolation based on a grid of values, but the technique and results we shall use and prove here are stronger and cover a wide range of approximations, for instance, all results hold if we use some type of spline interpolation instead of a linear one. This is so because the functions to be approximated are generally very smooth. We also evaluate the stability and performance of the approximations by simple numerical experiments.

Approximation to μ_s

We start with an approximation to μ_s. In this direction there are at least two ways to compute approximations to μ_s. One way is by using the ideas and results outlined in Dellnitz and Junge (1999), which are based on a discretization of the Perron-Frobenius operator by means of a Garlekin projection type approximation in order to compute the eigenvectors of the discretized operator corresponding to the eingenvalue 1. Although it can be used to approximate any SBR measure, the method is especially suited to approximate and study (almost) cyclical behavior of dynamical systems. However, its complexity makes the efficient implementation troublesome. A much simpler idea, which we shall adopt here, is to approximate the measure by truncating equation (2.1) for a reasonably large value of n. That is, we consider the approximating measure

$$\mu_n(A;s,x_0)=\frac{1}{n}\sum_{k=0}^{n-1}\delta_{T_s^k(x_0)}(A) \tag{5.1}$$

which converges in a weak sense to μ_s as n tends to infinity, for almost all initial points $x_0\in I$ and all μ_s-continuity sets A. The iterations of T_s are known to be unstable with respect to the initial point in the sense that, given a small $\varepsilon>0$ and a point $x\in(0,1)$, the trajectories $T_s^k(x)$

and $T_s^k(x+\varepsilon)$ become far apart exponentially fast. The approximation (5.1), however, is quite stable with respect to the initial point x_0 for large n. For instance, in Figure 5.1 we show the measure of the sets $[0.1, 0.2]$ and $[0.4, 0.6]$ obtained by using $\mu_n(\cdot; s, x_0)$ with $s = 0.5$, for 50 different initial points x_0 and 3 different truncation points $n \in \{300,000; 1,000,000; 3,000,000\}$. All plots are in the same scale (within set) in order to make comparison possible. In Table 5.1 we show basic statistics related to Figure 5.1. Notice that, in average, the 1,000,000 and 3,000,000 iteration cases are very similar and all cases are fairly stable with respect to the initial points (observe the scale).

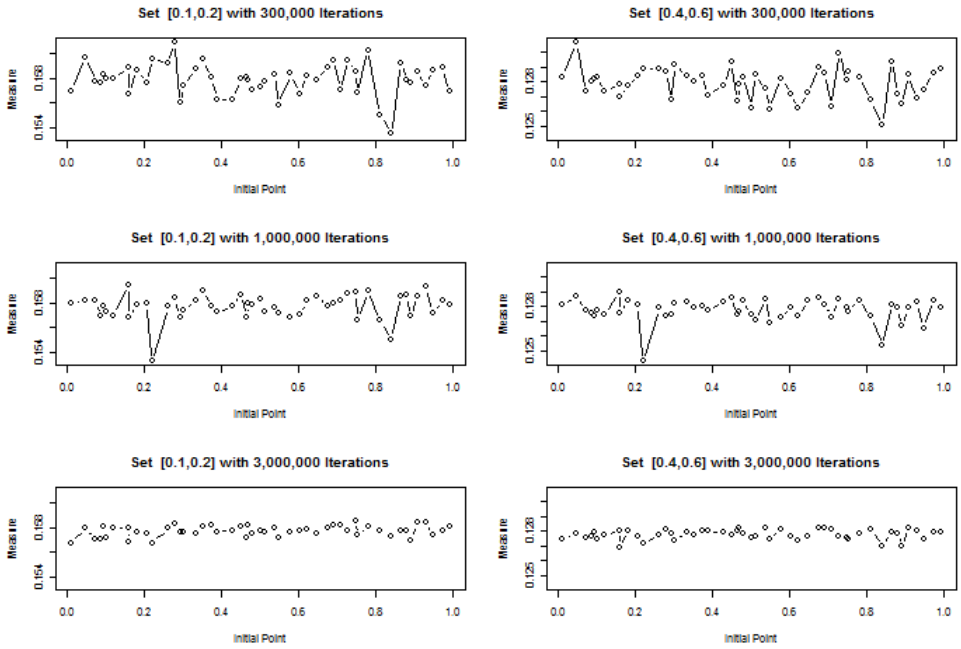


Figure 5.1: Performance of the approximation (5.1) for truncation points $n \in \{300,000; 1,000,000; 3,000,000\}$ (top, middle and bottom, respectively) and 50 different initial points for $s = 0.5$. The measured sets are (left) $[0.1, 0.2]$ and (right) $[0.4, 0.6]$. All plots within the same set are in the same scale.

Table 5.1: Summary statistics for the data presented in Figure 5.1.

Set	n	300,000	1,000,000	3,000,000
[0.2,0.3]	[min, max]	[0.12511,0.13067]	[0.12431,0.12901]	[0.12688,0.12825]
	range	0.00556	0.00470	0.00137
	mean	0.12790	0.12775	0.12777
[0.4,0.6]	[min, max]	[0.15349,0.16092]	[0.15326,0.15944]	[0.15676,0.15857]
	range	0.00743	0.00618	0.00181
	mean	0.15792	0.15771	0.15771

Next question is how good is the approximation (5.1)? One way to test this is by testing whether the approximation is invariant under T_s. For given initial points, say $x_1, \cdots, x_k$ and some interval $[a, b]$, we calculate $\mu_n\big([a, b]; s, x_i\big)$ and $\mu_n\big(T_s^{-1}([a, b]); s, x_j\big)$. If the difference between the two quantities is small for different pairs (x_i, x_j), one can conclude that the approximation is reasonably good. In Table 5.2 we present the difference $\big|\mu_n\big([a, b]; s, x_i\big) - \mu_n\big(T_s^{-1}([a, b]); s, x_j\big)\big|$ for 7 different initial points and 3 different sets $[a, b]$. The truncation point was taken to be 3,000,000 and $s = 0.5$. From Table 5.2 we conclude that the approxi-

mation (5.1) performs very well in all cases and that it can be taken to be T_s-invariant. As expected, when $x_i = x_j$ the differences are the smallest ($< 10^{-8}$ in all cases).

Table 5.2: Difference $\left|\mu_n\big([a,b];s,x_i\big) - \mu_n\big(T_s^{-1}([a,b]);s,x_j\big)\right|$ for different values of x_0 and sets $[a,b]$. The truncation point was taken to be $n = 3,000,000$ and $s = 0.5$. The initial points are $(x_1,\cdots,x_7) = \big(\pi, \pi/(\sqrt{2}+1), \pi\sqrt{2}, \pi+\sqrt{2}, \sqrt{7}, \pi+\sqrt{7}, \sqrt{11}+\sqrt{7}\big)(\text{mod } 1)$.

	initial	x_1	x_2	x_3	x_4	x_5	x_6	x_7
[0.05,0.2]	x_1	0.00000	0.00019	0.00040	0.00008	0.00004	0.00062	0.00022
	x_2	0.00019	0.00000	0.00020	0.00027	0.00024	0.00043	0.00042
	x_3	0.00040	0.00000	0.00000	0.00047	0.00044	0.00022	0.00062
	x_4	0.00008	0.00030	0.00047	0.00000	0.00003	0.00070	0.00015
	x_5	0.00004	0.00020	0.00044	0.00003	0.00000	0.00066	0.00018
	x_6	0.00062	0.00043	0.00022	0.00070	0.00066	0.00000	0.00084
	x_7	0.00022	0.0004	0.00062	0.00015	0.00018	0.00084	0.00000
	initial	x_1	x_2	x_3	x_4	x_5	x_6	x_7
[0.3,0.8]	x_1	0.00000	0.00019	0.00011	0.00009	0.00052	0.00036	0.00155
	x_2	0.00019	0.00000	0.00008	0.00028	0.00033	0.00016	0.00136
	x_3	0.00011	0.00008	0.00000	0.00020	0.00041	0.00024	0.00144
	x_4	0.00009	0.00028	0.00020	0.00000	0.00061	0.00045	0.00164
	x_5	0.00052	0.00033	0.00041	0.00061	0.00000	0.00016	0.00103
	x_6	0.00036	0.00016	0.00024	0.00045	0.00016	0.00000	0.00119
	x_7	0.00155	0.00136	0.00144	0.00164	0.00103	0.00119	0.00000
	initial	x_1	x_2	x_3	x_4	x_5	x_6	x_7
[0.7,0.95]	x_1	0.00000	0.00011	0.00005	0.00012	0.00003	0.00012	0.00089
	x_2	0.00011	0.00000	0.00016	0.00022	0.00013	0.00022	0.00078
	x_3	0.00005	0.00016	0.00000	0.00006	0.00003	0.00006	0.00094
	x_4	0.00012	0.00022	0.00006	0.00000	0.00009	0.00000	0.00100
	x_5	0.00003	0.00013	0.00003	0.00009	0.00000	0.00009	0.00091
	x_6	0.00012	0.00022	0.00006	0.00000	0.00009	0.00000	0.00101
	x_7	0.00089	0.00078	0.00094	0.00100	0.00091	0.00101	0.00000

In the remaining of this section we shall assume that $s \in (0,1)$ has been fixed and $x_0 \in (0,1)$ has been chosen so that the approximation (5.1) converges to μ_s. Since no confusion will arise, we shall drop s and x_0 from the notation and write the approximation (5.1), based on a size n iteration vector, just by $\mu_n(\cdot)$.

Approximating F_0^{-1} and the nodes of T_s^h

In order to approximate F_0^{-1}, one can use an empirical version based on the same iteration vector from which μ_n is derived. First we need to define an approximation to F_0 from which an approximation to F_0^{-1} will be derived. Let $\widehat{F}_n$ be the empirical distribution based on a size n iteration vector $\big(x_0, T_s(x_0), \cdots, T_s^{n-1}(x_0)\big)$ and let $x_1,\cdots,x_n$ be the jump points[2] of $\widehat{F}_n$. Consider the set $L_n := \{0 = x_0, x_1, \cdots, x_n, x_{n+1} = 1\}$. Given $x \in I\backslash L_n$, there exists a $k \in \{0,\cdots,n\}$ such that $x \in (x_k, x_{k+1})$. We define the approximate value of $F_0(x)$, denoted by $F_n(x)$, as the linear interpolation of x between the points $\big(x_k, \widehat{F}_n(x_k)\big)$ and $\big(x_{k+1}, \widehat{F}_n(x_{k+1})\big)$, that is, we set

$$F_n(x) := \left(\frac{\widehat{F}_n(x_{k+1}) - \widehat{F}_n(x_k)}{x_{k+1} - x_k}\right) x + \frac{\widehat{F}_n(x_k)x_{k+1} - \widehat{F}_n(x_{k+1})x_k}{x_{k+1} - x_k}. \tag{5.2}$$

If $x \in L_n$, we simply define $F_n(x) := \widehat{F}_n(x)$. Notice that, for each n, $F_n : I \to I$ is a one-to-one, increasing and uniformly continuous function, so that its inverse, F_n^{-1}, is well defined and is also one-to-one and uniformly continuous. In the next proposition, we show that $F_n(x) \to F_0(x)$ and $F_n^{-1}(x) \to F_0^{-1}(x)$, both limits being uniform in x.

[2] by the choice of x_0, there will be exactly n jump points.

Proposition 5.1. *Let $\widehat{F}_n$ be the empirical distribution based on an iteration vector $\big(x_0, T_s(x_0), \cdots, T_s^{n-1}(x_0)\big)$ and let $x_1, \cdots, x_n$ be the jump points of $\widehat{F}_n$. Let F_n be the approximation (5.2) based on $\{x_1, \cdots, x_n\}$ and F_n^{-1} be its inverse. Then,*

$$F_n(x) \longrightarrow F_0(x) \quad \text{and} \quad F_n^{-1}(x) \longrightarrow F_0^{-1}(x),$$

uniformly in x.

Proof: By the Glivenko-Cantelli theorem, $\widehat{F}_n(x) \to F_0(x)$ uniformly in $x \in [0,1]$, so that, given $\varepsilon > 0$, one can find $n_0 := n_0(\varepsilon) > 0$ such that if $n > n_0$, then $\big|\widehat{F}_n(x) - F_0(x)\big| < \varepsilon$ uniformly in x. Now, for $x \in (0,1)$ (if x equals 0 or 1, the result is trivial), there exists a $k \in \{1, \cdots, n\}$ such that $x \in [x_k, x_{k+1})$. Hence, if $n > n_0$

$$\begin{aligned}
\big|F_n(x) - F_0(x)\big| &\leq \big|F_n(x) - \widehat{F}_n(x)\big| + \big|\widehat{F}_n(x) - F_0(x)\big| \\
&< \big|\widehat{F}_n(x_{k+1}) - \widehat{F}_n(x_k)\big| + \varepsilon \\
&\leq \sup_{i=1,\cdots,n-1} \big\{\big|\widehat{F}_n(x_{i+1}) - \widehat{F}_n(x_i)\big|\big\} + \varepsilon \\
&\leq \frac{1}{n} + \varepsilon,
\end{aligned}$$

uniformly in x. To show the convergence of the inverse, let $y \in [0,1]$ and $\varepsilon > 0$ be given and notice that F_n^{-1} being uniformly continuous, one can find a $\delta := \delta(\varepsilon) > 0$ such that

$$|x - y| < \delta \implies |F_n^{-1}(x) - F_n^{-1}(y)| < \varepsilon.$$

Now, since F_n converges uniformly to F_0, there exists $n_1 := n_1(\varepsilon) > 0$ such that,

$$n > n_1 \implies \big|F_n(x) - F_0(x)\big| < \delta,$$

for all $x \in I$. Also, since F_0 is one to one, there exists $v_0 \in [0,1]$ such that $y = F_0(v_0)$. Therefore, if $n > n_1$

$$\big|F_n^{-1}(y) - F_0^{-1}(y)\big| = \big|F_n^{-1}\big(F_0(v_0)\big) - v_0\big| = \big|F_n^{-1}\big(F_0(v_0)\big) - F_n^{-1}\big(F_n(v_0)\big)\big| < \varepsilon$$

and since n_1 is independent of y, the desired convergence follows. ∎

As for the end points $\{a_{h,k}\}_{k=0}^{2^h}$ of the nodes of T_s^h, let $\{x_1, \cdots, x_m\} \in (0,1)$, $x_i \neq x_j$ and consider the set $\{T_s^h(x_1), \cdots, T_s^h(x_m)\}$, for $m > 0$ sufficiently large[3]. Note that $a_{h,0} = 0$ and $a_{h,2^h} = 1$, for any h. Let $D = \big\{i : T_s^h(x_i) > T_s^h(x_{i+1})\big\} \subset \{1, \cdots, m\}$. The set D contains the indexes $i \in \{1, \cdots, m\}$ for which the interval $[x_i, x_{i+1}]$ contains a discontinuity of T_s^h. Let $\{d_j\}_{j=1}^{2^h-1}$ denote the ordered elements of D, so that the interval $[x_{d_j}, x_{d_{j+1}}]$ contains the j-th discontinuity of T_s^h. Now consider the function $T_{i,h;s}^* : [x_{d_i}, x_{d_i+1}] \to [0,2]$ given by $T_{i,h}^*(x;s) := T_s^{h-1}(x) + \big(T_s^{h-1}(x)\big)^{1+s}$ and notice that we can write

$$T_s^h(x) = T_{i,h}^*(x;s) - \delta_{[1,2]}\big(T_{i,h}^*(x;s)\big).$$

Since there is a discontinuity of T_s^h in the interval $[x_{d_i}, x_{d_i+1}]$, we have $T_{i,h}^*(x_{d_i};s) \leq 1$ and $T_{i,h}^*(x_{d_i+1};s) \geq 1$ and since $T_{i,h}^*$ is continuous and increasing, there exists a point $x \in [x_{d_i}, x_{d_i+1}]$

[3]By "sufficiently large" we mean that m should be at least large enough to guarantee that the set $\{T_s^h(x_1), \cdots, T_s^h(x_m)\}$ reflects the $2^h - 1$ discontinuities of T_s^h, or, in other words, $m \geq 2^h$. The limits in m taken for an approximation are understood to be in terms of partitions, that is, we start with a sufficiently large set of points, say $I_m = \{x_1, \cdots, x_m\}$ and consider refinements of the form $I_{m+1} = I_m \bigcup \{x_{m+1}\}, \cdots, I_{m+k} = I_{m+k-1} \bigcup \{x_{m+k}\}$. Suppose that $R_m := R(I_m)$ is an approximation based on I_m. For a sequence of refinements $\{I_k\}_{k=m+1}^{\infty}$ we consider the sequence $\{R(I_k)\}_{k=m+1}^{\infty}$. Whenever the last limit exists, we set $\lim\limits_{m\to\infty} R_m = \lim\limits_{k\to\infty} R(I_k)$.

such that $T_s^*(x;s) = 1$, which is precisely $a_{h,i}$. With this in mind, let $a_{h,i}^m$ denote the approximation to $a_{h,i}$ obtained from $\{x_1, \cdots, x_m\}$ by using a linear interpolation between the points $\big(x_{d_i}, T_{i,h}^*(x_{d_i};s)\big)$ and $\big(x_{d_i+1}, T_{i,h}^*(x_{d_i+1};s)\big)$. That is, $a_{h,i}^m$ is given by

$$a_{h,i}^m = x_{d_i} + \frac{x_{d_i+1} - x_{d_i}}{T_{i,h}^*(x_{d_i+1};s) - T_{i,h}^*(x_{d_i};s)} \big(1 - T_{i,h}^*(x_{d_i};s)\big), \tag{5.3}$$

for all $d_i \in D$. Clearly $a_{h,i}^m \underset{m\to\infty}{\longrightarrow} a_{h,i}$, since $|x_{d_i+1} - x_{d_i}| \underset{m\to\infty}{\longrightarrow} 0$ and by the continuity of $T_{i,h}^*$, for each $i \in \{1, \cdots, 2^h - 1\}$.

Approximating $\mathcal{T}_{h,k}$

Concerning the approximation of $\mathcal{T}_{h,k}$, we shall use an argument based on an empirical inverse and linear interpolation, but we shall also need a doubling argument in order to improve accuracy of the approximation near the discontinuities and guarantee the uniform convergence of the approximation to its target. So let $\{0 = x_1, \cdots, x_m = 1\} \in I$, $x_i < x_j$ and consider the set $\big\{T_s^h(x_1), \cdots, T_s^h(x_m)\big\}$, for $m > 0$ sufficiently large. Given $y \in [0,1]$, recall that the inverse image of y by T_s^h is a size 2^h vector denoted by $\big(\mathcal{T}_{h,0}, \cdots, \mathcal{T}_{h,2^h-1}\big)$. Let again $D = \big\{i : T_s^h(x_i) > T_s^h(x_{i+1})\big\} \subset \{1, \cdots, m\}$ and $\{d_i\}_{i=1}^{2^h-1}$ be the ordered points in D. Suppose that we know exactly or have good estimates for the nodes $\{a_{h,k}\}_{k=0}^{2^h}$ of T_s^h (for instance, we could use $\{a_{h,k}^m\}_{k=0}^{2^h}$, as described before, based on the same set $\{x_1, \cdots, x_m\}$ considered here). For $i = 0, \cdots, 2^h - 1$, let $p_i = d_{i+1} - d_i + 2$ and let

$$R_{h,i} = \{x_{h,i}^{(1)}, \cdots, x_{h,i}^{(p_i)}\} := \{a_{h,i}^m, x_{d_i+1}, \cdots, x_{d_{i+1}}, a_{h,i+1}^m\}$$

and

$$I_{h,i} = \{y_{h,i}^{(1)}, \cdots, y_{h,i}^{(p_i)}\} := \big\{0, T_s^h(x_{d_i+1}), \cdots, T_s^h(x_{d_{i+1}}), 1\big\}.$$

Given $y \in [0,1]$, for each $i = 0, \cdots, 2^h - 1$, there exists a $y_{h,i}^{(k)} \in I_{h,i}$ such that $y \in \big[y_{h,i}^{(k)}, y_{h,i}^{(k+1)}\big)$. We define the approximation $\mathcal{T}_{h,i}^m(y)$ of $\mathcal{T}_{h,i}(y)$, as being the linear interpolation of y between the points $\big(x_{h,i}^{(k)}, y_{h,i}^{(k)}\big)$ and $\big(x_{h,i}^{(k+1)}, y_{h,i}^{(k+1)}\big)$. That is, for each $i = 0, \cdots, 2^h - 1$,

$$\mathcal{T}_{h,i}^m(y) = x_{h,i}^{(k)} + \frac{x_{h,i}^{(k+1)} - x_{h,i}^{(k)}}{y_{h,i}^{(k+1)} - y_{h,i}^{(k)}} \big(y - y_{h,i}^{(k)}\big). \tag{5.4}$$

Notice that if y equals 0 or 1, we have $\mathcal{T}_{h,i}^m(y) = \mathcal{T}_{h,i}(y)$. Also, as the partition $\{x_1, \cdots, x_m\}$ increases, $\big|x_{k+1} - x_k\big| \underset{m\to\infty}{\longrightarrow} 0$ and the uniform continuity of T_s^h clearly implies $\mathcal{T}_{h,i}^m(y) \underset{m\to\infty}{\longrightarrow} \mathcal{T}_{h,i}(y)$, for each $y \in [0,1]$, for $i = 0, \cdots, 2^h - 1$. More is true: the convergence is actually uniform in y, as we show in the next proposition.

Proposition 5.2. *Let $\mathcal{T}_{h,k}^m$ be the approximation of $\mathcal{T}_{h,k}$ given by* (5.4) *based on a partition R_m. Then,*

$$\mathcal{T}_{h,k}^m(y) \longrightarrow \mathcal{T}_{h,k}(y),$$

for each $k = 0, \cdots, 2^h - 1$, as m goes to infinity (that is, as the partition gets thinner). Moreover, the convergence is uniform in $y \in [0,1]$.

Proof: Given $\varepsilon > 0$, the uniform continuity of $\mathcal{T}_{h,k}$ implies the existence of a $\delta := \delta(\varepsilon) > 0$ such that

$$|x - y| < \delta \Longrightarrow \big|\mathcal{T}_{h,k}(x) - \mathcal{T}_{h,k}(y)\big| < \varepsilon,$$

for all $x \in [0,1]$. Let $R_0 = \{0 = x_1, \cdots, x_{m_0} = 1\} \in I$ for a sufficiently large $m_0 \in \mathbb{N}^*$ such that

$$\sup_{i=1,\cdots,m_0-1} \big\{|x_{i+1} - x_i|\big\} < \delta.$$

For $m > m_0$, let $R_m = \{x_1^*, \cdots, x_m^*\} \supset R_0$ be a size m refinement of R_0. Given $y \in (0,1)$, for each $i = 0, \cdots, 2^h - 1$, let $\mathcal{T}_{h,i}^m$ be the approximation (5.4) based on R_m. By construction and since $y \in (0,1)$, it follows that

$$\mathcal{T}_{h,i}\big(x_{h,i}^{(k)}\big) \le \mathcal{T}_{h,i}^m(y) < \mathcal{T}_{h,i}\big(x_{h,i}^{(k+1)}\big) \quad \text{and} \quad \mathcal{T}_{h,i}\big(x_{h,i}^{(k)}\big) \le \mathcal{T}_{h,i}(y) < \mathcal{T}_{h,i}\big(x_{h,i}^{(k+1)}\big),$$

so that

$$\begin{aligned} \big|\mathcal{T}_{h,i}^m(y) - \mathcal{T}_{h,i}(y)\big| &\le \big|\mathcal{T}_{h,i}\big(x_{h,i}^{(k+1)}\big) - \mathcal{T}_{h,i}\big(x_{h,i}^{(k)}\big)\big| \\ &\le \sup_{j=1,\cdots,m-1} \big\{\big|\mathcal{T}_{h,i}(x_{j+1}) - \mathcal{T}_{h,i}(x_j)\big|\big\} < \varepsilon, \end{aligned}$$

for all $y \in (0,1)$. If $y \in \{0,1\}$, by construction $\mathcal{T}_{h,i}(y) = \mathcal{T}_{h,i}^m(y)$, so that the result follows uniformly for all $y \in [0,1]$, as desired. ■

5.1 Approximating the lag h MP copula

With these approximations in hand, we can now define the approximation for the copula $C_{X_t,X_{t+h}}$ when φ is almost everywhere increasing given in Proposition 3.2 but in the form (3.5). For $(u,v) \in I^2$, $n > 0$ and $m \ge 2^h$, we set

$$\begin{aligned} C_{m,n}(u,v;h) = \Bigg(\sum_{k=0}^{n_0^*-1} \mu_n\Big(\big[a_{h,k}^m, \mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big)\big]\Big)\Bigg)\delta_{\mathbb{N}^*}(n_0^*) + \\ + \mu_n\Big(\big[a_{h,n_0}^m, \min\big\{F_n^{-1}(u), \mathcal{T}_{h,n_0}^m\big(F_n^{-1}(v)\big)\big\}\big]\Big), \end{aligned} \tag{5.5}$$

where $n_0^* := n_0(m,n) = \big\{k : u \in \big[F_n(a_{h,k}^m), F_n(a_{h,k+1}^m)\big)\big\}$ and $\lim_{m,n\to\infty} n_0^* = n_0$ since F_n converges uniformly to F_0 and $a_{h,k}^m$ converges to $a_{h,k}$. In the next theorem we establish the convergence of the approximation (5.5) to the true copula.

Theorem 5.1. *Let $C_{m,n}(u,v;h)$ be given by (5.5). Then, for all $(u,v) \in I^2$, $t \ge 0$ and $h > 0$*

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}(u,v;h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}(u,v;h) = \lim_{m,n\to\infty} C_{m,n}(u,v;h)$$

and the common limit is $C_{X_t,X_{t+h}}(u,v)$ (given by (3.4)). Furthermore, the above limits are uniform in $(u,v) \in I^2$.

The proof of Theorem 5.1, is a consequence of the following stronger lemma.

Lemma 5.1. *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures defined in I such that $\mu_n \xrightarrow{w} \mu$. Let $f_n : I \to I$ be a sequence of continuous functions converging uniformly to a function $f : I \to I$. Let $\{a_m\}_{m\in\mathbb{N}}$ be a sequence of real numbers such that $a_m \in [0,1]$ for all m and $a_m \to a$. Also let $g_m : [a_m, 1] \to I$ be a sequence of continuous functions converging uniformly to a function $g : I \to I$, $S_{m,n}(v) := \big[a_m, g_m\big(f_n(v)\big)\big]$ and $S(v) := \big[a, g\big(f(v)\big)\big]$. Then,*

$$\lim_{m\to\infty}\lim_{n\to\infty}\mu_n\big(S_{m,n}(v)\big) = \lim_{n\to\infty}\lim_{m\to\infty}\mu_n\big(S_{m,n}(v)\big) = \lim_{m,n\to\infty}\mu_n\big(S_{m,n}(v)\big) = \mu\big(S(v)\big)$$

uniformly in $v \in I$.

Proof: For all $m, n > 0$ and $v \in [0,1]$, let $S_{m,n}(v)$ and $S(v)$ be as in the enunciate and let

$$S_n(v) := \big[a, g\big(f_n(v)\big)\big] \quad \text{and} \quad S_m(v) := \big[a_m, g_m\big(f(v)\big)\big].$$

Notice that all sets just defined are μ-continuity sets for all m, n and v. Since the convergence of f_n to f is uniform, we have

$$\lim_{m,n\to\infty} g_m\big(f_n(v)\big) = \lim_{n\to\infty}\lim_{m\to\infty} g_m\big(f_n(v)\big) = \lim_{m\to\infty}\lim_{n\to\infty} g_m\big(f_n(v)\big) = g\big(f(v)\big)$$

for all v, so that, both, the iterated and the double limits exist and $S_{m,n}(v) \to S(v)$, for all $v \in [0,1]$. Also notice that we have $\delta_{S_{m,n}}(x) \leq \delta_I(x)$ uniformly in m, n and x, and since μ_n converges weakly to μ and I is a μ-continuity set, it follows that

$$\int \delta_I(x)\mathrm{d}\mu_n \longrightarrow \int \delta_I(x)\mathrm{d}\mu.$$

Now, in one hand, since $S_{m,n}(v) \to S_m(v)$ for all v and $\delta_{S_{m,n}} \leq \delta_I$, by the Lebesgue convergence theorem, it follows that

$$\mu_n\big(S_{m,n}(v)\big) = \int \delta_{S_{m,n}}(x)\mathrm{d}\mu_n \underset{n\to\infty}{\longrightarrow} \int \delta_{S_m}(x)\mathrm{d}\mu,$$

and, since $\delta_{S_m} \leq \delta_I$ and $\int \delta_I \mathrm{d}\mu < \infty$, by the Lebesgue dominated theorem, we conclude that

$$\int \delta_{S_m}(x)\mathrm{d}\mu \underset{m\to\infty}{\longrightarrow} \int \delta_S(x)\mathrm{d}\mu = \mu\big(S(v)\big),$$

which shows that $\lim_{m\to\infty}\lim_{n\to\infty}\mu_n\big(S_{m,n}(v)\big) = \mu\big(S(v)\big)$ and the convergence holds uniformly in $v \in (0,1)$. On the other hand, since $\delta_{S_{m,n}} \leq \delta_I$ and $\int \delta_I \mathrm{d}\mu_n < \infty$, by the Lebesgue dominated theorem, it follows that

$$\mu_n\big(S_{m,n}(v)\big) = \int \delta_{S_{m,n}}(x)\mathrm{d}\mu_n \underset{m\to\infty}{\longrightarrow} \int \delta_{S_n}(x)\mathrm{d}\mu_n,$$

and, since $\delta_{S_n} \leq \delta_I$ and $\int \delta_I \mathrm{d}\mu_n \to \int \delta_I \mathrm{d}\mu$, by the Lebesgue convergence theorem we conclude that,

$$\int \delta_{S_n}(x)\mathrm{d}\mu_n \underset{n\to\infty}{\longrightarrow} \int \delta_S(x)\mathrm{d}\mu = \mu\big(S(v)\big),$$

that is, $\lim_{n\to\infty}\lim_{m\to\infty}\mu_n\big(S_{m,n}(v)\big) = \mu\big(S(v)\big)$, which also holds uniformly in v. Since the iterated limits are established, in order to finish the proof we need to show that the double limit exists and is equal to the iterated ones. Let $\varepsilon > 0$ be given. Since $\mu \ll \lambda$, the Radon-Nikodym theorem implies the existence of a non-negative continuous function h, which will be bounded since we are restricted to the interval I, such that, for any $A \in \mathcal{B}(I)$,

$$\mu(A) = \int_A h(x)\mathrm{d}\lambda \leq M\lambda(A),$$

where $M = \sup_{x\in I}\{h(x)\} < \infty$. Now, since $a_m \to a$, one can find $m_1 := m_1(\varepsilon) > 0$ such that, if $m > m_1$,

$$a_m \in K_1(\varepsilon) := \Big[a - \frac{\varepsilon}{10M}, a + \frac{\varepsilon}{10M}\Big]$$

and

$$\mu\big(K_1(\varepsilon)\big) \leq M\lambda\left(\Big[a - \frac{\varepsilon}{10M}, a + \frac{\varepsilon}{10M}\Big]\right) = \frac{\varepsilon}{5}.$$

The uniform convergence of g_m to g implies the existence of $m_2 := m_2(\varepsilon) > 0$ such that, if $m > m_2$, $|g_m(x) - g(x)| < \varepsilon/20M$, for all $x \in I$, or equivalently, taking $x = f_n(v)$, if $m > m_2$

$$g_m\big(f_n(v)\big) \in \Big[g\big(f_n(v)\big) - \frac{\varepsilon}{20M}, g\big(f_n(v)\big) + \frac{\varepsilon}{20M}\Big].$$

Now, the uniform continuity of g implies the existence of a $\delta := \delta(\varepsilon) > 0$ such that

$$\big|x - f_n(v)\big| < \delta \implies \big|g(x) - g\big(f_n(v)\big)\big| < \frac{\varepsilon}{20M}.$$

But since f_n converges to f uniformly, there exists a $n_1 = n_1(\delta) > 0$ such that

$$n > n_1 \implies \big|f_n(v) - f(v)\big| < \delta,$$

for all v so that, taking $x = f(v)$, for $n > n_1$, we have

$$g\big(f_n(v)\big) \in \Big[g\big(f(v)\big) - \frac{\varepsilon}{20M}, g\big(f(v)\big) + \frac{\varepsilon}{20M}\Big],$$

for all $v \in I$. Hence, if we take $m > m_2$ and $n > n_1$,

$$g\big(f_n(v)\big) - \frac{\varepsilon}{20M} \in \Big[g\big(f(v)\big) - \frac{\varepsilon}{10M}, g\big(f(v)\big)\Big]$$

and

$$g\big(f_n(v)\big) + \frac{\varepsilon}{20M} \in \Big[g\big(f(v)\big), g\big(f(v)\big) + \frac{\varepsilon}{10M}\Big]$$

so that, setting

$$K_2(\varepsilon) := \Big[g\big(f(v)\big) - \frac{\varepsilon}{10M}, g\big(f(v)\big) + \frac{\varepsilon}{10M}\Big],$$

for $m > m_2$ and $n > n_1$, it follows that

$$g_m\big(f_n(v)\big) \in \Big[g\big(f_n(v)\big) - \frac{\varepsilon}{20M}, g\big(f_n(v)\big) + \frac{\varepsilon}{20M}\Big] \subseteq K_2(\varepsilon),$$

for all $v \in I$. Also observe that

$$\mu\big(K_2(\varepsilon)\big) \leq M\lambda\Big(\Big[g\big(f(v)\big) - \frac{\varepsilon}{10M}, g\big(f(v)\big) + \frac{\varepsilon}{10M}\Big]\Big) \leq \frac{\varepsilon}{5}.$$

The convergence of μ_n to μ implies the existence of $n_2 := n_2(\varepsilon) > 0$ such that if $n > n_2$ ($K_i(\varepsilon)$ is a $\mu-$continuity set)

$$\big|\mu_n\big(K_i(\varepsilon)\big) - \mu\big(K_i(\varepsilon)\big)\big| < \frac{\varepsilon}{5},$$

for $i = 1, 2$. Also, if we set $F_n(x) = \mu_n\big([0, x]\big)$ and $F_0(x) = \mu\big([0, x]\big)$, then F_0 is continuous (since $\mu \ll \lambda$), $F_n \to F_0$, and, by Pólya's theorem, there exists a $n_3 := n_3(\varepsilon) > 0$ such that, if $n > n_3$

$$\sup_{x \in I}\big\{\big|F_n(x) - F_0(x)\big|\big\} < \frac{\varepsilon}{10}.$$

Now, notice that, if $n > n_3$

$$\begin{aligned}\big|\mu_n\big(S(v)\big) - \mu\big(S(v)\big)\big| &\leq \Big|F_n\big(g\big(f(v)\big)\big) - F_0\big(g\big(f(v)\big)\big)\Big| + |F_n(a) - F_0(a)| \\ &\leq 2\sup_{x \in I}\big\{\big|F_n(x) - F_0(x)\big|\big\} < \frac{\varepsilon}{5},\end{aligned}$$

for all $v \in I$. Observe further that, by construction, if $m > \max\{m_1, m_2\}$ and $n > n_1$,

$$S_{m,n}(v) \backslash S(v) \subset K_1(\varepsilon) \bigcup K_2(\varepsilon),$$

for all v so that, setting $n_0 = n_0(\varepsilon) := \max\{m_1, m_2, n_1, n_2, n_3\}$, if $m, n > n_0$, we have

$$\begin{aligned}&\big|\mu_n\big(S_{m,n}(v)\big) - \mu\big(S(v)\big)\big| \leq \\ &\quad\leq \big|\mu_n\big(S_{m,n}(v)\big) - \mu_n\big(S(v)\big)\big| + \big|\mu_n\big(S(v)\big) - \mu\big(S(v)\big)\big| \\ &\quad< \big|\mu_n\big(K_1(\varepsilon)\big) + \mu_n\big(K_2(\varepsilon)\big)\big| + \frac{\varepsilon}{5} \\ &\quad\leq \big|\mu_n\big(K_1(\varepsilon)\big) - \mu\big(K_1(\varepsilon)\big)\big| + \mu\big(K_1(\varepsilon)\big) + \mu\big(K_2(\varepsilon)\big) + \big|\mu\big(K_2(\varepsilon)\big) - \mu_n\big(K_2(\varepsilon)\big)\big| + \frac{\varepsilon}{5} \\ &\quad< \varepsilon,\end{aligned}$$

for all v, which implies the existence of the double limit, equality with the iterated ones and the desired uniform convergence. ■

Proof of Theorem 5.1: First notice that taking $f_n = F_n^{-1}$, $g_m = \mathcal{T}_{h,k}^m$, $a_m = a_{h,k}^m$, it follows from Lemma 5.1 that

$$\mu_n\Big(\big[a_{h,k}^m, \mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big)\big]\Big) \underset{m,n \to \infty}{\longrightarrow} \mu\Big(\big[a_{h,k}, \mathcal{T}_{h,k}\big(F_0^{-1}(v)\big)\big]\Big),$$

for each $k = 0, \cdots, n_0 - 1$. It remains to show that

$$\lim_{m,n\to\infty} \mu_n\Big(\big[a^m_{h,n_0}, \min\big\{F_n^{-1}(u), \mathcal{T}^m_{h,n_0}\big(F_n^{-1}(v)\big)\big\}\big]\Big) = \mu\Big(\big[a_{h,n_0}, \min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}\big]\Big),$$

and that the iterated limits exist and are equal to the double limit. First, since we can write $\min\{u, v\} = \frac{u+v}{2} - \frac{|u-v|}{2}$, it is routine to show that if $f_n \to f$ uniformly, with f_n and f uniformly continuous and $g_m \to g$ uniformly, with g_m and g uniformly continuous, we have $\min\big\{f_n(u), g_m\big(f_n(v)\big)\big\}$ converging uniformly to $\min\big\{f(u), g\big(f(v)\big)\big\}$ in n, m, u and v. So, the problem simplifies to show that if $a_m \to a$, $g_{m,n}(u, v)$ is a sequence of functions such that $g_{m,n}(u, v) \to g(u, v)$ uniformly in u, v, n, m and $a_m \le g_{m,n}(u, v)$ for all u, v, n, m and $\mu_n \xrightarrow{w} \mu$, then

$$\lim_{m,n\to\infty} \mu_n\big([a_m, g_{m,n}(u, v)]\big) = \mu\big([a, g(u, v)]\big),$$

uniformly in u and v and the double limit above is equal to the iterated limits. A similar argument to the one used in Lemma 5.1 to establish the existence and equality of the iterated limits can be used to show the existence and equality of the iterated limits in this case. As for the double limit, let M be as in the proof of Lemma 5.1. By the uniform convergence of $g_{m,n}(u, v)$ to $g(u, v)$ and since $g_{m,n}$ and g are uniformly continuous for all m, n, it follows that there exists $m_1 := m_1(\varepsilon) > 0$, depending on ε only, such that, if $m, n > m_1$,

$$g_{m,n}(u, v) \in K(\varepsilon) := \Big[g(u, v) - \frac{\varepsilon}{10M}, g(u, v) + \frac{\varepsilon}{10M}\Big],$$

for all u and v and $\mu\big(K(\varepsilon)\big) \le \varepsilon/5$. The rest of the proof is carried out by mimicking the proof of Lemma 5.1 with the obvious adaptations. Identification of $g_{m,n}(u, v)$, $g(u, v)$, a_m and a with $\min\big\{F_n^{-1}(u), \mathcal{T}^m_{h,n_0}\big(F_n^{-1}(v)\big)\big\}$, $\min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}$, a^m_{h,n_0} and a_{h,n_0}, respectively, completes the proof. ∎

Remark 5.1. Notice that neither the convergence proved in Lemma 5.1 nor the one in Theorem 5.1 is uniform in m and n.

As for the case when φ is almost everywhere decreasing, we observe that, in view of (3.7), the function

$$C^*_{m,n}(u, v; h) = u + v - 1 + C_{m,n}(1 - u, 1 - v; h)$$

is an approximation to the copula in (3.6). Clearly $C^*_{m,n}$ converges to the true copula as m and n tends to infinity (view either as an iterated or a double limit) and the convergence is uniform in (u, v).

Implementation and Random Variate Generation

The implementation of the approximations so far discussed is routine. All the approximations we mentioned can share the same iteration vector, which further improves the efficiency and precision of the task and greatly reduces the computational burden. In the top panel of Figure 5.2 we show the three dimensional plot of the lag 1 and 2 MP copula for values of $s \in \{0.1, 0.4\}$. The respective level plots are shown in the bottom panel of Figure 5.2. Notice the non-exchangeability of the copulas in all cases.

Obtaining random samples from an MP copulas is a trivial task in view of Proposition 3.4. There we show that the support of an MP copula is the union of graphs of certain linear functions. The following algorithm can be used to generate a pair of variates from a bidimensional MP copula for φ an almost everywhere increasing function.

1. Generate an uniform $(0, 1)$ variate u.

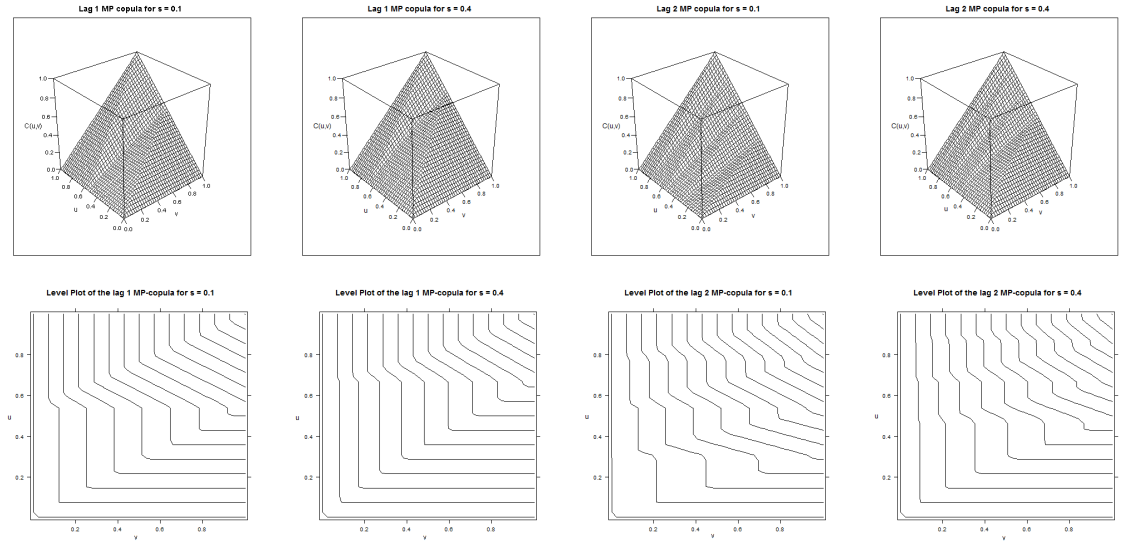


Figure 5.2: From left to right, three dimensional plots of the lag 1 MP copula for $s \in \{0.1, 0.4\}$ and lag 2 MP copula for the same parameters (top panel) and respective level sets (bottom panel) obtained from approximation (5.5).

2. Let κ_0 denote the index for which $u \in \left[F_0(a_{h,\kappa_0}), F_0(a_{h,\kappa_0+1})\right]$ and set $v = \ell^+_{h,\kappa_0}(u)$.

3. The desired pair is (u, v).

In practice the T_s-invariant probability measure is unknown and F_0 has to be approximated. Furthermore, most of times the nodes related to T_s^h, for $h > 0$, $s \in (0, 1)$ cannot be analytically obtained. However, we can apply the approximations developed in this section together with the algorithm above to obtain approximated samples from MP copulas. In Figure 5.3 we show 500 approximated sample points from a lag 1 and 2 MP copula for $s \in \{0.1, 0.4\}$ and φ an almost everywhere. Obvious modifications in the algorithm, allow handling the case where φ is an almost everywhere decreasing function.

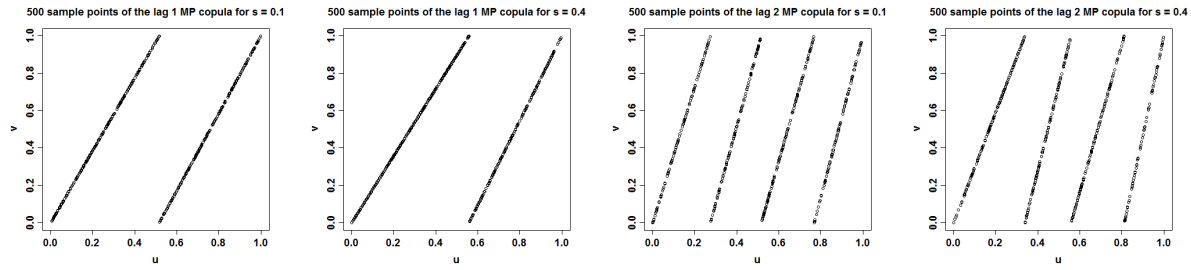


Figure 5.3: Left to right: 500 approximated sample points from a lag 1 MP copula for $s \in \{0.1, 0.4\}$ and lag 2 MP copula for the same parameters.

Remark 5.2. For small values of the lag, the resemblance of the sample to a piecewise continuous function is very clear, but this is not always the case as it can be seen in Figure 5.4, where we show 500 approximated sample points of the lag 4, 5 and 7 MP copulas for $s = 0.2$. This is a general principle, for a fixed sample size the higher the lag, the harder to distinguish the support of the copula based on the sample, since the number of branches of T_s^h grow as fast as 2^h. For instance, for $h = 7$ in Figure 5.4 is difficult to say that the sample came from a singular copula at all.

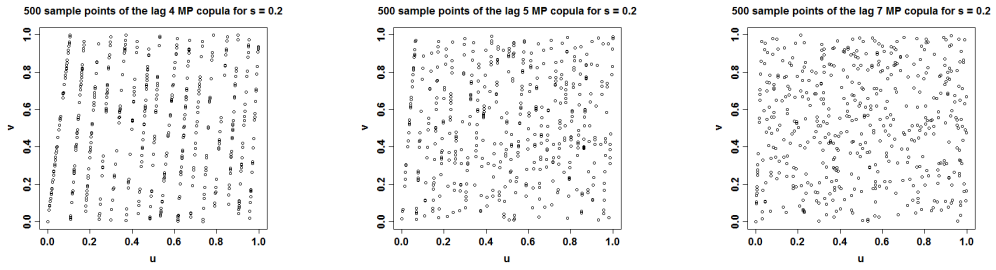


Figure 5.4: Left to right: 500 approximated sample points from the lag 4, 5 and 7 MP copulas for $s = 0.2$.

6 Application

In this section we apply the theory developed in Section 3 to the problem of estimating the parameter s in MP processes. This problem have been studied before in Olbermann et al. (2007), where the authors adapt and apply several estimation methods from the classical theory of long-range dependence to the problem of estimating the parameter s. In this section we propose an estimator for the parameter s based on the ideas developed in Section 3, which is both, precise and fast.

The mathematical framework is as follows. Let $s \in (0,1)$ and consider the associated MP process $\{X_n\}_{n\in\mathbb{N}}$ for φ the identity map. Suppose we observe a realization $x_1, \cdots, x_N$ from X_n and our goal is to estimate the unknown parameter s. Let $a := a(s) \in \left(\frac{1}{2}, \frac{\sqrt{5}-1}{2}\right)$ denote the discontinuity point of the MP transformation and notice that s and a are related by

$$a + a^{1+s} = 1 \quad \Longleftrightarrow \quad s = \frac{\log(1-a)}{\log(a)} - 1.$$

Hence, the problem of estimating s is equivalent to the problem of estimating a.

To define the proposed estimator, we start by observing that Proposition 3.4 for $h = 1$ implies that the lag 1 MP copula's support is given by the graph of the piecewise linear function

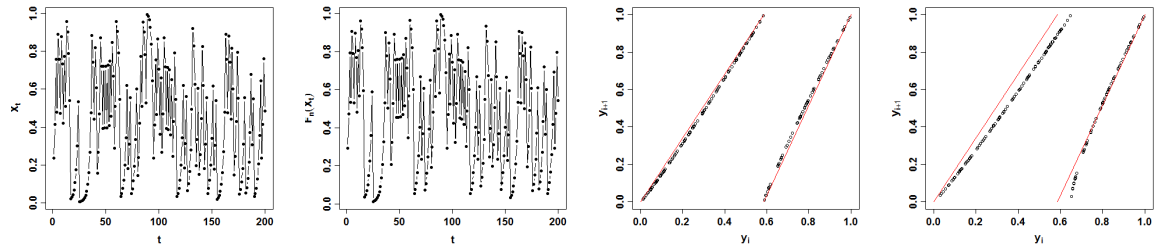
$$\ell(x) := \begin{cases} \frac{x}{F_0(a)}, & \text{if } x \in \big[0, F_0(a)\big) \\ \frac{x - F_0(a)}{1 - F_0(a)}, & \text{if } x \in \big[F_0(a), 1\big], \end{cases}$$

so that, any (independent or correlated) sample from a lag 1 MP copula consists of points scattered through the lines defined by ℓ (see Figure 5.3). The discontinuity point of the function ℓ is precisely $F_0(a)$. Let $y_i = F_0(x_i)$, for $i = 1, \cdots, N$, and consider the series $\{u_i := (y_i, y_{i+1})\}_{i=1}^{N-1}$. By Sklar's Theorem, $\{u_i\}_{i=1}^{N-1}$ is a (correlated) sample from the lag 1 MP copula, so all points should lie in the graph of the function ℓ.

These considerations suggest the following procedure to obtain s based on a path $x_1, \cdots, x_N$ of X_n within a given accuracy $\varepsilon > 0$. We choose $s_0 \in (0,1)$ as an initial guess for s and calculate $\hat{y}_i = F_n(x_i; s_0)$, $i = 1, \cdots, N$, where F_n is the approximation of F_0 given in (5.2). Next we define $\{\hat{u}_i := (\hat{y}_i, \hat{y}_{i+1})\}_{i=1}^{N-1}$, from which we estimate the slope of the two branches of the approximated sample from the lag 1 MP copula obtained by this way. The discontinuity point (and hence s) can then be easily calculated. In this manner we obtain an estimative $\tilde{s}$ which can be compared to s_0. If s_0 is close to the true value s, then the difference between $\tilde{s}$ and s_0 should be small. If not, we choose another starting value and repeat the operation until obtain the desired accuracy. This leads to an optimization procedure to obtain s within a predefined accuracy.

To illustrate the procedure, Figure 6.1(a) shows a sample path of an MP process for $s = 0.2$, with $N = 200$ while Figure 6.1(b) shows the sample path $y_i = F_n(x_i; 0.2)$, $i = 1, \cdots, N$. From $\{y_i\}_{i=1}^{N}$, we construct the sequence $\{u_i\}_{i=1}^{N-1}$, where $u_i = \big(F_n(y_i; s), F_n(y_{i+1}; s)\big)$, $i = 1, \cdots, N -$

1, for the correctly specified $s = 0.2$ and for $s = 0.3$. Figure 6.1(c) presents the graph of $\{u_i\}_{i=1}^{N-1}$ obtained from the correct specification of s, while Figure 6.1(d) shows the graph of the misspecified one. In Figures 6.1(c) and 6.1(d), the solid lines represent the respective theoretical support of the copula given in Proposition 3.4. Some distortion in the points can be seen given to the use of the approximation F_n instead of the theoretical F_0, especially in lower quantiles. From Figure 6.1(d) it is clear that the line obtained from the sequence $\{u_i\}_{i=1}^{N-1}$ and the theoretical one for the chosen value of s_0, namely, 0.3, do not match, while for the correct specified one in Figure 6.1(c), they do.



(a) (b) (c) (d)

Figure 6.1: (a) Sample path $x_1, \cdots, x_{200}$ of an MP process with $s = 0.2$ starting at $\sqrt{5}(\text{mod}1)$. (b) Sample path $y_i = F_n(x_i)$. Plot of $u_i = (y_i, y_{i+1})$ for the (c) correct and (d) misspecified s. The solid lines correspond to the theoretical support of the respective lag 1 MP copula.

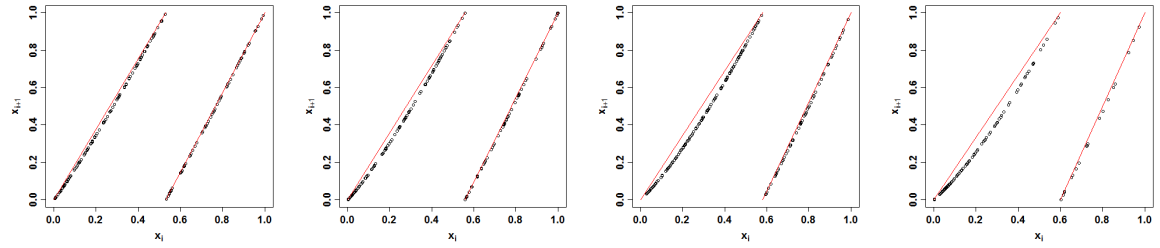
The procedure just outlined is, however, computationally expensive given the fact that to calculate the approximation F_n with reasonable stability and accuracy, for each s, it requires the construction of an iteration vector of large size (see Figure 5.1 and Table 5.1). Such an optimization procedure can easily take hundreds of evaluations, depending on the desired accuracy, and hence, can be a very time consuming task.

To overcome this difficulty, observe that in Figures 6.1(a) and 6.1(b), little differences can be seen between them. In fact, since F_0 is a smooth distribution, an alternative is to apply the previous argument to the points $\hat{v}_i := (x_i, x_{i+1})$, $i = 1, \cdots, N$. There will certainly be some distortion in the lines due to the absence of F_0, but we expect to be able to estimate the discontinuity point a based on v_i by similar idea as before.

As an illustration, Figure 6.2 shows the plots of $v_i = (x_i, x_{i+1})$, $i = 1, \cdots, 199$, based on MP processes with $s \in \{0.2, 0.4, 0.6, 0.8\}$ all starting at $\sqrt{5}(\text{mod}1)$. The solid lines correspond to the lines joining the points $(0, 0)$ and $(a, 1)$ and joining $(a, 0)$ and $(1, 1)$, where a denotes the correct discontinuity point of the respective MP transformation. From the graphs in Figure 6.2 we see the identification of the line based on v_i with the correct line, especially in the second branch of the graph. That is so because $a \in \left(\frac{1}{2}, \frac{\sqrt{5}-1}{2}\right)$, so that the second branch, being smaller, is less affected by the distortion due to the absence of F_0.

In order to assess the performance of the estimation procedure, we perform the following experiment. We randomly select 100 initial points[4] in $(0, 1)$ and for each initial point we generate a path (of size $N = 200$) of an MP process for $s \in \{0.1, 0.15, \cdots, 0.95\}$. For each path, say $x_1, \cdots, x_{200}$, we perform the proposed estimation procedure. In order to estimate a, we applied two methods: the first one is a simple least squares method applied to the points lying in the second branch of (x_i, x_{i+1}). The second method is the following: let (x_{m_0}, x_{m_0+1}) and (x_{m_1}, x_{m_1+1}) denote the points among the ones lying on the second branch of $\{(x_i, x_{i+1})\}_{i=1}^{N-1}$

[4]Tables with the initial values applied in our experiments and the complete simulation results are available upon request.



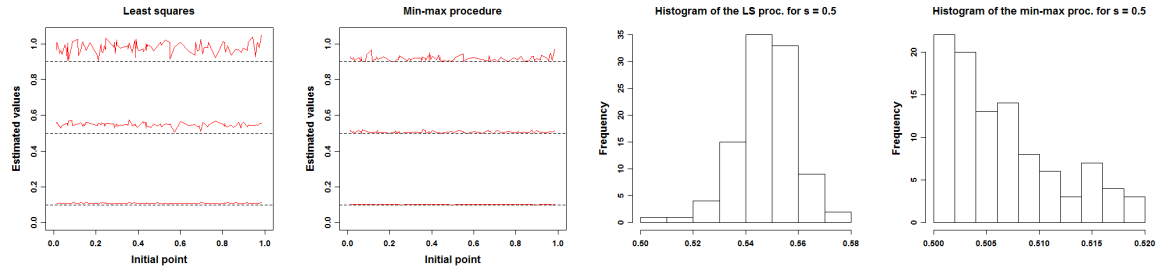
(a) (b) (c) (d)

Figure 6.2: Plot of $v_i = (x_i, x_{i+1})$, $i = 1, \cdots, 199$ from a sample path of an MP process with (a) $s = 0.2$, (b) $s = 0.4$, (c) $s = 0.6$ and (d) $s = 0.8$. The solid lines correspond to the lines joining the points $(0, 0)$ and $(a, 1)$ and joining $(a, 0)$ and $(1, 1)$, where a denotes the correct discontinuity point of the respective MP transformation.

for which x_{m_0} is minimum and x_{m_1} is maximum. We define the estimator of a, say $\hat{a}$, as

$$\hat{a} = -\frac{B}{A}, \quad \text{where} \quad A := \frac{x_{m_1+1} - x_{m_0+1}}{x_{m_1} - x_{m_0}} \quad \text{and} \quad B := x_{m_0+1} - A x_{m_0}. \tag{6.6}$$

For reference, in the subsequent we shall call this the *min-max procedure*. Geometrically, $\hat{a}$ is the inverse image of 0 by the linear function joining (x_{m_0}, x_{m_0+1}) and (x_{m_1}, x_{m_1+1}).



(a) (b) (c) (d)

Figure 6.3: Plot of the estimated values for $s \in \{0.1, 0.5, 0.9\}$ for 100 random initial points by using (a) the least squares procedure and (b) the min-max procedure. The dashed lines correspond to the correct value of s. Also shown the histogram of the estimated values for $s = 0.5$ by using (c) the least squares procedure and (d) the min-max procedure.

Table 6.1 summarizes the experiment results by presenting the mean, range, standard deviation (st.d.) and mean square error (mse) of the results. Figures 6.3(a) and 6.3(b) present graphically the results for both methods for $s \in \{0.1, 0.5, 0.9\}$ while in Figures 6.3(c) and 6.3(d), the histogram of the results for $s = 0.5$ are presented. From Table 6.1 and Figure 6.3, we see that the min-max procedure (MM) outperforms the least squares estimates (LS) obtained. Some bias can be seen for both estimates, especially when s increases.

The min-max procedure can be carried out even for time series of sample size as small as 20, as long as the second branch of $\{(x_i, x_{i+1})\}_{i=1}^{N-1}$ contains at least 2 points, which does not always happen (for instance, for $N = 110$, a sample path of an MP process with $s = 0.8$ starting at $\sqrt{74}(\text{mod}1)$ has only one point in the second branch). In such a situation, a straightforward adaptation of the min-max procedure can be applied to the first branch and still yields reasonable estimates. The closer to 0 and 1 the points x_{m_0} and x_{m_1} in (6.6) are, respectively, the better the estimation performance.

Table 6.1: Summary statistics of the experiment results. Presented are the mean estimate ($\hat{s}$), the range, the standard deviation (st.d.) and the mean square error values (mse) of the estimates. The min-max procedure is denoted by MM while LS denotes the least squares.

Proc.	s	$\hat{s}$	range	st.d.	mse	s	$\hat{s}$	range	st.d.	mse
MM	0.10	0.1008	[0.1000, 0.1024]	0.0006	0*	0.55	0.5581	[0.5500, 0.5888]	0.0069	0.0001
LS		0.1087	[0.1056, 0.1128]	0.0017	0.0001		0.6036	[0.5501, 0.6287]	0.0142	0.0031
MM	0.15	0.1516	[0.1500, 0.1597]	0.0015	0*	0.60	0.6091	[0.6000, 0.6451]	0.0090	0.0002
LS		0.1632	[0.1573, 0.1710]	0.0026	0.0002		0.6545	[0.6012, 0.7023]	0.0186	0.0033
MM	0.20	0.2023	[0.2001, 0.2101]	0.0021	0*	0.65	0.6600	[0.6501, 0.6927]	0.0087	0.0002
LS		0.2179	[0.2098, 0.2315]	0.0038	0.0003		0.7089	[0.6579, 0.7584]	0.0197	0.0039
MM	0.25	0.2534	[0.2501, 0.2632]	0.0027	0*	0.70	0.7125	[0.7001, 0.7726]	0.0119	0.0003
LS		0.2730	[0.2636, 0.2875]	0.0049	0.0006		0.7646	[0.7038, 0.8170]	0.0226	0.0047
MM	0.30	0.3036	[0.3000, 0.3128]	0.0028	0*	0.75	0.7621	[0.7502, 0.8019]	0.0110	0.0003
LS		0.3272	[0.3128, 0.3410]	0.0052	0.0008		0.8177	[0.7505, 0.8612]	0.0246	0.0052
MM	0.35	0.3544	[0.3500, 0.3669]	0.0039	0*	0.80	0.8165	[0.8001, 0.8659]	0.0131	0.0004
LS		0.3835	[0.3550, 0.4024]	0.0078	0.0012		0.8781	[0.8005, 0.9449]	0.0277	0.0069
MM	0.40	0.4050	[0.4000, 0.4214]	0.0049	0*	0.85	0.8677	[0.8500, 0.9428]	0.0151	0.0005
LS		0.4367	[0.4082, 0.4570]	0.0078	0.0014		0.9307	[0.8507, 1.0111]	0.0297	0.0074
MM	0.45	0.4556	[0.4501, 0.4703]	0.0046	0.0001	0.90	0.9172	[0.9002, 0.9704]	0.0141	0.0005
LS		0.4909	[0.4702, 0.5174]	0.0102	0.0018		0.9774	[0.9011, 1.0477]	0.0306	0.0069
MM	0.50	0.5065	[0.5001, 0.5189]	0.0051	0.0001	0.95	0.9706	[0.9500, 1.0517]	0.0189	0.0008
LS		0.5475	[0.5059, 0.5746]	0.0111	0.0024		1.0371	[0.9500, 1.1641]	0.0384	0.0090

Note: 0* means that the mse is smaller than 5×10^{-5}.

7 Conclusions

In this work we derive the copulas related to Manneville-Pomeau processes for almost everywhere monotonic functions φ. In the bidimensional case, we find that the copulas of any random pair (X_t, X_{t+h}) depend only on the lag h and are singular. The support of the copulas is derived as well.

As for the multidimensional case, when φ is increasing almost everywhere, the functional form of the copulas are very similar to the ones derived in the bidimensional case. We conclude that the copulas of vectors $(X_{t_1}, \cdots, X_{t_n})$ and $\left(U_0, T_s^{t_2-t_1}(U_0), \cdots, T_s^{t_n-t_1}(U_0)\right)$ are the same. When φ is decreasing almost everywhere, we find that the copulas of an n-dimensional random vector from an MP process can be deduced from the ones derived for the increasing case.

The copulas derived here depend on the T_s-invariant measure μ_s which has no explicit formula. For the bidimensional case, we propose an approximation to the copula which is shown to converge uniformly to the true copula. From this approximation, we are able to present plots of the copulas for different parameters and lags and to present a simple algorithm to generate approximated samples from the copulas. Some simple numerical calculation are presented to test the steps of the approximation. To illustrate the usefulness of the theory, we derive a fast estimation procedure of the underlying parameter s in Manneville-Pomeau processes.

Acknowledgements

Sílvia R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by INCT *em Matemática* and also by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1.

Guilherme Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil.

References

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd ed., New York: John Wiley. MR1700749

Chazottes, J.-R., P. Collet and B. Schmitt (2005). "Statistical Consequences of the Devroye Inequality for Processes. Applications to a Class of Non-Uniformly Hyperbolic Dynamical Systems". *Nonlinearity*, **18(5)**, 2341-2364. MR2165706

Dellnitz, M. and Junge, O. (1999). "On the Approximation of Complicated Dynamical Behavior". *SIAM Journal of Numerical Analysis*, Vol. **36(2)**, 491-515. MR1668207

Fisher, A.M. and Lopes, A. (2001). "Exact Bounds for the Polynomial Decay of Correlation, $1/f$ Noise and the CLT for the Equilibrium State of a Non-Hölder Potential". *Nonlinearity*, Vol. **14**, 1071-1104. MR1862813

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability, 73. London: Chapman & Hall. MR1462613

Kolesárová, A.; Mesiar, R.; Sempi, C. (2008). "Measure-Preserving Transformations, Copulæ and Compatibility". *Mediterranean Journal of Mathematics*, **5(3)**, 325-339. MR2465579

Lopes, A. and Lopes, S.R.C. (1998). "Parametric Estimation and Spectral Analysis of Piecewise Linear Maps of the Interval". *Advances in Applied Probability*, Vol. **30**, 757-776. MR1663557

Maes, C.; Redig, F.; Takens, F.; Moffaert, A. and Verbitski, E. (2000). "Intermittency and Weak Gibbs States". *Nonlinearity*, Vol. **13**, 1681-1698. MR1781814

Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag. 2nd Edition. MR2197664

Olberman, B.P.; Lopes, S.R.C. and Lopes, A.O. (2007). "Parameter Estimation in Manneville-Pomeau Processes". Unpublished manuscript. Source: arXiv:0707.1600.

Pianigiani, G. (1980). "First Return Map and Invariant Measures". *Israel Journal of Mathematics*, Vol. **35**, 32-48. MR0576460

Pollicott, M. and Yuri, M. (1998). *Dynamical Systems and Ergodic Theory*. Cambridge: Cambridge University Press. MR1627681

Royden, H.L. (1988). *Real Analysis*. New York: Macmillan. 3nd Edition. MR1013117

Schweizer, B. and Sklar, A. (2005). *Probabilistic Metric Spaces*. Mineola: Dover Publications. MR0790314

Young, L-S. (1999). "Recurrence Times and Rates of Mixing". *Israel Journal of Mathematics*, Vol. **110**, 153-188. MR1750438

Zebrowsky, J.J. (2001). "Intermittency in Human Heart Rate Variability". *Acta Physica Polonica B*, Vol. **32**, 1531-1540.

Anexo B

Artigo Pumi e Lopes (2011**a**)

Copulas Related to Piecewise Monotone Functions of the Interval and Associated Processes

Guilherme Pumi and Sílvia R.C. Lopes
Mathematics Institute
Federal University of Rio Grande do Sul

This version: 12/14/2011

Abstract

In this work we derive the copulas related to vectors coming from stochastic processes defined by the iteration of certain smooth piecewise monotone functions of the interval $[0, 1]$ to some initial random variable. We study some of its properties and present some examples. Since often these type of copulas do not have closed formulas, we provide a general method of approximation which converges uniformly to the true copula. Our results cover a wide class of processes, including the so-called Manneville-Pomeau processes. The general theory is applied to the parametric estimation in certain chaotic processes and we also present a Monte Carlo study.

Keywords. Chaotic Processes, Copulas, Interval Maps, Invariant Measures, Parametric Estimation.

1 Introduction

Let $T : [0, 1] \to [0, 1]$ be some smooth function of the interval $[0, 1]$ and suppose there exists an absolutely continuous T-invariant probability measure, say, μ_T. Let U_0 be distributed as μ_T. We can define a stochastic process by setting $X_t := \varphi\big(T^t(U_0)\big)$, $t \in \mathbb{N}$, for a given μ_T-integrable function φ. This type of stochastic process has been applied in a variety of problems from rock drilling (see Lasota and Mackey, 1994 and references therein) to intermittency in human cardiac rate (see Zebrowsky, 2001). Realizations of this type of process usually present complex dynamics, chaotic behavior and instability with respect to the initial point X_0.

In this work we consider the problem of characterizing the copulas related to random vectors obtained from these type of stochastic process. More specifically, we are interested in the case where the transformation T is a smooth piecewise monotone function of the interval and φ is a strictly monotone function. Our results cover a wide class of stochastic process, such as the so-called Manneville-Pomeau process (studied in Lopes and Pumi, 2011) and the process related to the tent transformation (see Example 6.2) among many others (see, for instance, Lasota and Mackey, 1994).

The present work greatly generalizes the results in Lopes and Pumi (2011), where the authors derive and study the copulas related to Manneville-Pomeau processes, by considering T belonging to a certain general class of smooth piecewise transformations which can have either increasing, decreasing, or monotone branches. We consider the important bidimensional case as well as the multidimensional case. As the problem of existence of an invariant measure for a given transformation is usually a difficult one, often the copulas derived do not have closed formulas. In that case, we have to rely on approximations to study these copulas. In this direction, we develop a somewhat general approximation to the copulas which is shown to converge uniformly to the theoretical one. The problem of random variate generation of the copulas presented here is also addressed. As an application of the general theory, the problem of parametric estimation in certain chaotic process is discussed and an estimation procedure motivated by the results obtained in this paper is proposed. To assess the finite sample performance of the proposed

estimator, a Monte Carlo study is performed. Examples are also provided.

The paper is organized as follows: in the next section, we briefly review some concepts and results on copulas and introduce the class of stochastic processes we will be interested in this work. Section 3 is devoted to determine the copulas related to any pair (X_t, X_{t+h}) coming from the aforementioned processes and to explore some consequences. In section 4, some multidimensional extensions are shown. In Section 5 we develop a general approximation to the copulas derived in Section 3 and prove its uniform convergence to the true copula. Random variate generation is also addressed. Section 6 brings two examples. In Section 7 we apply the general theory of Section 3 to the problem of parametric estimation in certain chaotic processes. A Monte Carlo study is also performed. The conclusions are reserved to Section 8.

2 Preliminaries

We begin by establishing the necessary mathematical set up. We say that a map $T : I \to I$, where $I := [0, 1]$, is of class $C^{1+\alpha}_{1\text{-}1}(I)$ if it is a one-to-one function of class $C^{1+\alpha}(I)$ (the class of C^1 functions whose derivative is α-Hölder continuous), for $\alpha \in (0, 1)$. A map $T : I \to I$, is said to be *finitely piecewise* $C^{1+\alpha}_{1\text{-}1}(I)$ if there exists a partition $\{I_k\}_{k=1}^n$ of I such that the restriction of T to the interior of each I_k is a $C^{1+\alpha}_{1\text{-}1}(I)$ function. Each I_k will be called a *node* of T. If, in addition, the restriction of T to the interior of each I_k is monotone (increasing or decreasing), then we call T a *finitely piecewise monotone* (*increasing* or *decreasing*) $C^{1+\alpha}_{1\text{-}1}(I)$ function.

For consistency, when we call for a partition relative to a map T, we shall always mean the *maximal partition* $\{I_k\}$, in the sense that if $\{\tilde{I}_k\}$ is another partition such that T restricted to the interior of each $\tilde{I}_k$ is a $C^{1+\alpha}_{1\text{-}1}(I)$ function, then, for each k, there exists j such that $\tilde{I}_k \subseteq I_j$. We shall refer as the *net* relative to the nodes of a (finitely piecewise $C^{1+\alpha}_{1\text{-}1}(I)$) map T, the collection of all endpoints of each I_k in the maximal partition $\{I_k\}$. Notice that the net of the nodes of T are discontinuity points of T, except, perhaps, the points 0 and 1. For the sake of simplicity, but without loss of generality, we shall assume that $I_k = [a_{k-1}, a_k)$ and that T is right continuous, except at 1 where we assume that T is left continuous. The reason why this convention is advantageous will be clear later.

Given an arbitrary function $T : I \to I$ one can ask whether there exists a smooth T-invariant probability measure. The problem of establishing the existence of such measure is often a hard one. The literature on the subject is relatively extense and has a long history. Rényi (1957) shows that, for transformation of the form $T_a(x) = ax \pmod 1$, a unique absolutely continuous invariant probability measure always exists. Remember that a map $T : I \to I$ is said to be *uniformly expanding* in I if $|T'(x)| \geq \kappa > 1$, for all $x \in I$, whenever T' is defined. Lasota and Yorke (1973) show that for piecewise smooth uniformly expanding function there always exists an absolutely continuous invariant measure. Under more stringent, but easily verifiable conditions, Pianigiani (1980) show that this measure is unique and it is a probability measure. However, many non-expanding functions are known to possess such a measure. Sufficient conditions for this to happen for non-expanding functions can be found, for instance, in Bowen (1979), Pianigiani (1980) and Pianigiani (1981). We shall denote the space of all transformations $T : I \to I$ for which an absolutely continuous invariant probability measure exists by $\mathscr{S}$. More specifically, let us define the following spaces:

$$\mathcal{T}^{\updownarrow} := \left\{T \in \mathscr{S} \text{ and } T \text{ is finitely piecewise monotone } C^{1+\alpha}_{1\text{-}1}(I) \text{ function}\right\};$$
$$\mathcal{T}^{\downarrow} := \left\{T \in \mathscr{S} \text{ and } T \text{ is finitely piecewise decreasing } C^{1+\alpha}_{1\text{-}1}(I) \text{ function}\right\};$$
$$\mathcal{T}^{\uparrow} := \left\{T \in \mathscr{S} \text{ and } T \text{ is finitely piecewise increasing } C^{1+\alpha}_{1\text{-}1}(I) \text{ function}\right\}.$$

Certainly $\mathcal{T}^{\uparrow} \bigcap \mathcal{T}^{\downarrow} = \emptyset$ and $\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow} \subsetneq \mathcal{T}^{\updownarrow}$. When $T \in \mathcal{T}^{\updownarrow}$, for any $t \geq 1$ we shall denote by

$$K_t^{\uparrow} := \{k : \ T^t|_{I_k} \text{ is increasing}\} \quad \text{and} \quad K_t^{\downarrow} := \{k : \ T^t|_{I_k} \text{ is decreasing}\}, \tag{2.1}$$

where, as usual, $T^k = T \circ T^{k-1}$ denote the k-fold composition. Clearly, $K_t^{\uparrow} \bigcap K_t^{\downarrow} = \emptyset$, $K_t^{\uparrow} = \emptyset$, if $T \in \mathcal{T}^{\downarrow}$ and $K_t^{\downarrow} = \emptyset$, if $T \in \mathcal{T}^{\uparrow}$.

In this paper we are interested in study the copulas related to the following class of stochastic process.

Definition 2.1. Let $T \in \mathcal{T}^{\updownarrow}$ and let μ_T be a T-invariant probability measure. Let U_0 be a random variable distributed according to μ_T and $\varphi : I \to \mathbb{R}$ be a function in $\mathcal{L}^1(\mu_T)$. The stochastic process given by

$$X_t := (\varphi \circ T^t)(U_0), \quad \text{for all } t \in \mathbb{N},$$

is called a $\mathcal{T}_{\varphi}^{\updownarrow}$*-induced process* (or $\mathcal{T}_{\varphi}^{\updownarrow}$ process, for short).

By taking, in Definition 2.1, $T \in \mathcal{T}^{\uparrow}$ or $T \in \mathcal{T}^{\downarrow}$ instead, we define a $\mathcal{T}_{\varphi}^{\uparrow}$*-induced process* and a $\mathcal{T}_{\varphi}^{\downarrow}$*-induced process*, respectively, or, for short, $\mathcal{T}_{\varphi}^{\uparrow}$ and $\mathcal{T}_{\varphi}^{\downarrow}$ process. The processes just defined are stationary since μ_T is T-invariant and absolutely continuous.

Regarding copulas, a n-dimensional copula is a distribution function whose marginals are uniformly distributed on I and whose support is the n-dimensional cube I^n. The usefulness of copulas lies on its ability to model dependence independently of the marginals and vice-versa. The literature on the subject has grown enormously, especially in the last decade, due to the discover of several applications of copulas on many areas such as finance, actuarial science, time series, hydrology, among others. For instance, an interesting application of copulas in finances can be found in Wang et al. (2009), where the authors present one-factor models to pricing credit default index swap tranches and collateralized debt obligations, based on heavy tailed copulas specially designed to allow a continuous time tail-fatness control. We refer the reader to Cherubini et al. (2004) (and references therein) for applications in finances, Frees and Valdez (1998) for a review on the use of copula on actuarial sciences, Chen and Fan (2006) for an interesting copula based on a unidimensional time series model, and to Genest and Favre (2007) for a step-by-step guide to the dependence study through copulas with application to hydrology.

Among others, the invariance by almost everywhere increasing function and the simple functional form the copula takes when the transformation is decreasing almost everywhere are some of the properties we shall use very often in what follows. In the next theorem, we summarize these properties. The proof can be found in Nelsen (2006). In this work the measure implicit to phrases like "almost everywhere" and "almost sure" will be the appropriated Lebesgue measure.

Theorem 2.1. *Let C be any copula and let $f_1, \cdots, f_n$ be almost everywhere increasing functions. Then $C_{f_1(X_1),\cdots,f_n(X_n)}(u_1, \cdots, u_n) = C_{X_1,\cdots,X_n}(u_1, \cdots, u_n)$. Moreover, if f and g are two almost everywhere decreasing functions instead, then $C_{f(X),g(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v)$.*

The next theorem, the so-called Sklar's theorem, is the key result for copulas. See Schweizer and Sklar (2005) for a sketch of the proof in the n-dimensional case and Nelsen (2006) for a more detailed proof in the bidimensional case.

Theorem 2.2 (Sklar's Theorem)**.** *Let $X_1, \cdots, X_n$ be random variables with joint distribution function H and marginals $F_1, \cdots, F_n$, respectively. Then, there exists a copula C such that,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all } \ (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

If the F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_n)$*. The converse also holds. Furthermore,*

$$C(u_1, \cdots, u_n) = H\big(F_1^{(-1)}(u_1), \cdots, F_n^{(-1)}(u_n)\big), \quad \textit{for all } \ (u_1, \cdots, u_n) \in I^n,$$

where for a function F, $F^{(-1)}$ denotes its pseudo-inverse given by $F^{(-1)}(x) := \inf\big\{u \in \mathrm{Ran}(F) : F(u) \geq x\big\}$.

For more details on the theory of copulas we refer the reader to Nelsen (2006) and Joe (1997). Copulas are also in close connection to probabilistic metric spaces. See Schweizer and Sklar (2005) for details on this matter.

3 Bidimensional case

In this section we shall investigate the bidimensional copulas associated to $\mathcal{T}_\varphi^\uparrow$, $\mathcal{T}_\varphi^\downarrow$ and $\mathcal{T}_\varphi^\updownarrow$ processes where φ will be taken to be an almost surely monotone function. As we will see later, the multidimensional case is very similar to the bidimensional case, so we shall give special attention to the latter. Let us start with the following lemma of general interest.

Lemma 3.1. *Let $T \in \mathcal{T}^\updownarrow$ and suppose that T has $s \geq 1$ nodes. Then, $T^t \in \mathcal{T}^\updownarrow$, for $t \geq 1$, and it will have s^t nodes.*

Proof: We shall use induction on the number of nodes. For $t = 1$, the result is clear. Suppose that for $t > 1$, T^t has s^t nodes. Let $\{I_k\}_{k=1}^{s^t}$ be the partition relative to T^t. Since $T \in \mathcal{T}^\updownarrow$, it is clear that $T^t \in \mathcal{T}^\updownarrow$ and $T^{t+1} \in \mathcal{T}^\updownarrow$ as well. Now $T^{t+1} = T\big(T^t\big)$ and since T^t restricted to each I_k is surjective, it implies that, for an arbitrary node I_k,

$$T^{t+1}(I_k) = T\big(T^t(I_k)\big) = T(I),$$

and since $T(I)$ has s nodes, it follows that T^{t+1} will part each node of T^t into s new nodes, and since by assumption there are s^t such nodes for T^t, T^{t+1} will have $s \times s^t = s^{t+1}$ nodes as desired. ■

Given $T \in \mathcal{T}^\updownarrow$, we shall always fix an absolutely continuous T-invariant probability measure and denote it by μ_T. Now let $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost surely increasing function and consider $\{X_t\}_{t\in\mathbb{N}}$ the $\mathcal{T}^\updownarrow$ process associated to T. For all $t \in \mathbb{N}$, let $F_t(\cdot)$ be the distribution function of X_t. By definition, for all $x \in I$

$$F_0(x) := \mathbb{P}(U_0 \leq x) = \mu_T\big([0, x]\big).$$

Observe that the T-invariance of μ_T, implies that, for any $t \in \mathbb{N}$, $t > 0$, and $x \in I$,

$$F_t(x) := \mathbb{P}\big(T^t(U_0) \leq x\big) = \mu_T\left((T^t)^{-1}\big([0, x]\big)\right) = \mu_T\big([0, x]\big) = F_0(x). \tag{3.1}$$

Also, $\mu_T \ll \lambda$ implies that μ_T is non-atomic and since $T \in C^{1+\alpha}(I)$ implies the existence of a continuous positive density for μ_T, F_t is continuous, increasing and its inverse is well defined.

Remark 3.1. Notice that for $t > 0$ and $T \in \mathcal{T}^\updownarrow$, the restriction of T^t to each of its s^t nodes (say $\{I_k\}_{k=1}^{s^t}$) is a one-to-one function, so that on each I_k its inverse is locally well defined. Now, we can conveniently define the inverse of T^t at $y \in (0, 1)$ ($y \in \{0, 1\}$ is trivial) as a piecewise function by setting $\mathcal{T}_{t,k}(y) := (T^t)^{-1}|_{I_k}(y)$, so that

$$(T^t)^{-1}(y) = \big((T^t)^{-1}|_{I_1}(y), \cdots, (T^t)^{-1}|_{I_{s^t}}(y)\big) = \big(\mathcal{T}_{t,1}(y), \cdots, \mathcal{T}_{t,s^t}(y)\big).$$

This is just a simple way of writing the inverse image of the singleton $\{y\}$ by T^t. With this in mind, let $y \in (0, 1)$ ($y \in \{0, 1\}$ is trivial), $t > 0$, X be a random variable taking values in I and $\{a_{t,k}\}_{k=0}^{s^t}$ be the net associated to the nodes of T^t. For $T \in \mathcal{T}^\uparrow \bigcup \mathcal{T}^\downarrow$, the solution of the inequality $T^t(X) \leq y$ in X can be written as $X \in A_{t,1}(y) \bigcup \cdots \bigcup A_{t,s^t}(y)$, where

$$A_{t,k}(y) := \begin{cases} [a_{t,k-1}, \mathcal{T}_{t,k}(y)], & \text{if} \quad T \in \mathcal{T}^\uparrow, \\ [\mathcal{T}_{t,k}(y), a_{t,k}], & \text{if} \quad T \in \mathcal{T}^\downarrow, \end{cases} \tag{3.2}$$

will be a proper closed subinterval of $[a_{t,k-1}, a_{t,k}]$, for each $k = 1, \cdots, s^t$. Notice that $A_{t,k}(y)$ is just the inverse image of $[0, y]$ by the transformation T^t restricted to its k-th node, that is, $A_{t,k}(y) = (T^t)^{-1}\big([0, y]\big) \bigcap I_k$.

The next result will be used several times during the work.

Lemma 3.2. *Let X be a random variable taking values in I and let T belong to either $\mathcal{T}^{\uparrow}$ or $\mathcal{T}^{\downarrow}$ and suppose that T has s nodes. Then, for any $t \in \mathbb{N}$ and $x \in I$,*

$$\mathbb{P}\big(T^t(X) \leq x\big) = \mathbb{P}\big(X \in \bigcup_{k=1}^{s^t} A_{t,k}(x)\big) = \sum_{k=1}^{s^t} \mathbb{P}\big(X \in A_{t,k}(x)\big),$$

where the $A_{t,k}$'s are given by (3.2).

Proof: The result follows from Remark 3.1 and from the fact that the $A_{t,k}$'s are (pairwise) disjoint. ■

The case where $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ is treated in the next lemma. As one could expect, it will be a mix of the results in Lemma 3.2.

Lemma 3.3. *Let X be a random variable taking values in I and distributed according to a non-atomic distribution. Let T belong to $\mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ and suppose that T has s nodes. Then, for any $t \in \mathbb{N}$ and $x \in I$,*

$$\mathbb{P}\big(T^t(X) \leq x\big) = \mathbb{P}\Big(X \in \bigcup_{k=1}^{s^t} \big(A^{\uparrow}_{t,k}(x) \bigcup A^{\downarrow}_{t,k}(x)\big)\Big) = \sum_{k=1}^{s^t} \big(\mathbb{P}(X \in A^{\uparrow}_{t,k}(x)) + \mathbb{P}(X \in A^{\downarrow}_{t,k}(x))\big),$$

where

$$A^{\uparrow}_{t,k}(x) := \begin{cases} [a_{t,k-1}, \mathcal{T}_{t,k}(x)], & k \in K^{\uparrow}_t, \\ \emptyset, & \textit{otherwise,} \end{cases} \quad \textit{and} \quad A^{\downarrow}_{t,k}(x) := \begin{cases} [\mathcal{T}_{t,k}(x), a_{t,k}], & k \in K^{\downarrow}_t, \\ \emptyset, & \textit{otherwise.} \end{cases} \tag{3.3}$$

Proof: Let $A^{\uparrow}_{t,k}$ and $A^{\downarrow}_{t,k}$ be as in expression (3.3). Since $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, $K^{\uparrow}_t \subsetneq \{1, \cdots, s^t\}$ and $K^{\downarrow}_t \subsetneq \{1, \cdots, s^t\}$ are non-empty. To prove the first equality, notice that

$$\mathbb{P}\big(T^t(X) \leq x\big) = \mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) \bigcup \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big) = \mathbb{P}\Big(X \in \bigcup_{k=1}^{s^t} \big(A^{\uparrow}_{t,k}(x) \bigcup A^{\downarrow}_{t,k}(x)\big)\Big).$$

As for the second, one can write

$$\begin{aligned} \mathbb{P}\big(T^t(X) \leq x\big) &= \mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) \bigcup \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big) \\ &= \mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big)\Big) + \mathbb{P}\Big(X \in \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big) - \\ &\qquad - \mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) \bigcap \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big). \end{aligned} \tag{3.4}$$

Now upon noticing that $A^{\uparrow}_{t,i}(x) \bigcap A^{\uparrow}_{t,j}(x) = \emptyset$ and $A^{\downarrow}_{t,i}(x) \bigcap A^{\downarrow}_{t,j}(x) = \emptyset$ whenever $i \neq j$ and by the definition of $A^{\uparrow}_{t,i}$ and $A^{\downarrow}_{t,i}$, it follows that

$$\mathbb{P}\big(X \in \bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) = \mathbb{P}\big(X \in \bigcup_{k=1}^{s^t} A^{\uparrow}_{t,i}(x)\big) = \sum_{k=1}^{s^t} \mathbb{P}\big(X \in A^{\uparrow}_{t,k}(x)\big)$$

and similarly $\mathbb{P}\big(X \in \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\big) = \sum_{k=1}^{s^t} \mathbb{P}\big(X \in A^{\downarrow}_{t,k}(x)\big)$. At this point, upon substituting these two equalities into (3.4), the lemma will be proved if we show that

$$\mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) \bigcap \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big) = 0.$$

This follows upon observing that

$$\begin{aligned} \mathbb{P}\Big(X \in \big(\bigcup_{i \in K^{\uparrow}_t} A^{\uparrow}_{t,i}(x)\big) \bigcap \big(\bigcup_{j \in K^{\downarrow}_t} A^{\downarrow}_{t,j}(x)\big)\Big) &= \mathbb{P}\Big(X \in \bigcup_{k=1}^{s^t} \big(A^{\uparrow}_{t,k}(x) \bigcap A^{\downarrow}_{t,k}(x)\big)\Big) \\ &\leq \mathbb{P}\Big(X \in \big(\bigcup_{k=0}^{s^t} \{a_{t,k}\}\big) \bigcup \big(\bigcup_{k=1}^{s^t} \{\mathcal{T}_{t,k}(x)\}\big)\Big) = 0, \end{aligned}$$

since, by assumption, X has a non-atomic distribution. This completes the proof. ■

Remark 3.2. Lemma 3.3 is the analogous of Lemma 3.2 in the case where $T \in \mathcal{T}^{\updownarrow} \backslash (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, but with an extra condition: we had to impose the condition of non-atomicity in the distribution of X. The non-atomicity condition is necessary because the intersection in (3.4) may be non-empty, in which case it will contain only a finite number of isolated points. However, in our work we will always assume that the underline distribution is absolutely continuous, so the extra imposed condition has no impact in the development of the work.

Proposition 3.1. *Let $T \in \mathcal{T}^{\updownarrow}$, μ_T be a T-invariant probability measure and let U_0 be distributed as μ_T. Then, for any $t, h \in \mathbb{N}$, $h \neq 0$, $\big(T^t(U_0), T^{t+h}(U_0)\big) \stackrel{d}{=} \big(U_0, T^h(U_0)\big)$ and $C_{T^t(U_0),T^{t+h}(U_0)} = C_{U_0,T^h(U_0)}$.*

Proof: The first result follows from the fact that, for $T \in \mathcal{T}^{\updownarrow}$ and $\varphi \in \mathcal{L}^1(\mu_T)$, the $\mathcal{T}^{\updownarrow}_{\varphi}$ process associated is stationary. In particular, the result holds for φ taken to be the identity map. The second result is a consequence of the process' stationarity and of Sklar's theorem. ■

As for the copulas related to $\mathcal{T}^{\updownarrow}_{\varphi}$ processes, for φ almost surely increasing, the following result holds.

Corollary 3.1. *Let $T \in \mathcal{T}^{\updownarrow}$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process. Then, for any $t, h \in \mathbb{N}$, $h \neq 0$,*

$$C_{X_t,X_{t+h}}(u,v) = C_{X_0,X_h}(u,v) = C_{U_0,T^h(U_0)}(u,v),$$

everywhere in I^2.

Proof: Direct consequence of Proposition 3.1 and Theorem 2.1. ■

Remark 3.3. Notice that the result in Corollary 3.1 actually holds in a much more general context when combined with Proposition 3.1, but for our purposes, we only need it as stated.

Now we turn our attention to determine the copula associated to any pair (X_p, X_q) of random variables, for any $p, q \in \mathbb{N}$, obtained from a $\mathcal{T}^{\updownarrow}_{\varphi}$ process with φ increasing almost everywhere. Let $T \in \mathcal{T}^{\updownarrow}_{\varphi}$ and assume that T has $s \geq 1$ nodes. Let μ_T be a T-invariant probability measure and F_0 be as before. In order to simplify the notation, let us define the functions $\mathscr{F}_{h,k} : I \to [F_0(a_{h,k-1}), F_0(a_{h,k})]$ by

$$\mathscr{F}_{h,k}(x) := F_0\Big(\mathcal{T}_{h,k}\big(F_0^{-1}(x)\big)\Big), \tag{3.5}$$

for $h > 0$ and $k \in \{1, \cdots, s^h\}$. For a given set S, we also define $\delta_S(u)$ as being 1, if $u \in S$, and 0 otherwise. We start with the case $T \in \mathcal{T}^{\updownarrow} \backslash (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$. We shall denote the copula related to a given $\mathcal{T}^{\updownarrow}_{\varphi}$ process $\{X_t\}_{t\in\mathbb{N}}$ by $C^{\uparrow}_{X_t,X_{t+h}}$, $C^{\downarrow}_{X_t,X_{t+h}}$ and $C^{\updownarrow}_{X_t,X_{t+h}}$ when T belongs to $\mathcal{T}^{\uparrow}$, $\mathcal{T}^{\downarrow}$ and $\mathcal{T}^{\updownarrow} \backslash (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, respectively.

Proposition 3.2. *Let $T \in \mathcal{T}^{\updownarrow} \backslash (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process. If we let $\{a_{h,k}\}_{k=0}^{s^h}$ be the net associated to the nodes of T^h, then*

$$\begin{aligned} C^{\updownarrow}_{X_t,X_{t+h}}(u,v) \;=\; & \sum_{k\in n_0^{\uparrow}} \big[\mathscr{F}_{h,k}(v) - F_0(a_{h,k-1})\big] + \big[\min\big\{u, \mathscr{F}_{h,n_0}(v)\big\} - F_0(a_{h,n_0-1})\big]\delta_{K_h^{\uparrow}}(n_0) + \\ & + \sum_{k\in n_0^{\downarrow}} \big[F_0(a_{h,k}) - \mathscr{F}_{h,k}(v)\big] + \max\big\{0, u - \mathscr{F}_{h,n_0}(v)\big\}\delta_{K_h^{\downarrow}}(n_0), \end{aligned} \tag{3.6}$$

where $n_0 := n_0(u;h) = \big\{k : u \in \big[F_0(a_{h,k-1}), F_0(a_{h,k})\big)\big\}$, and, with $K_h^{\uparrow}$ and $K_h^{\downarrow}$ as in (2.1),

$$n_0^{\uparrow} := \{1, \cdots, n_0 - 1\} \bigcap K_h^{\uparrow} \quad \text{and} \quad n_0^{\downarrow} := \{1, \cdots, n_0 - 1\} \bigcap K_h^{\downarrow}. \tag{3.7}$$

Proof: We start by determining the copula associated to the pair $\big(U_0, T^h(U_0)\big)$. Let $T \in \mathcal{T}^{\updownarrow}\backslash(\mathcal{T}^{\uparrow}\bigcup\mathcal{T}^{\downarrow})$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Also let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process and $H_{0,h}(\cdot,\cdot)$ denote the distribution function of the pair $\big(U_0, T^h(U_0)\big)$. Notice that, by Lemma 3.3, we have

$$\begin{aligned} H_{0,h}(x,y) &= \mathbb{P}(U_0 \le x, T^h(U_0) \le y) = \mathbb{P}\big(U_0 \le x, U_0 \in \textstyle\bigcup_{k=1}^{s^h} A_{h,k}(y)\big) \\ &= \mathbb{P}\big(U_0 \in [0,x] \textstyle\bigcap \bigcup_{k=1}^{s^h} A_{h,k}(y)\big). \\ &= \displaystyle\sum_{k=1}^{s^h}\Big[\mathbb{P}\big(U_0 \in A^{\uparrow}_{h,k}(y)\textstyle\bigcap[0,x]\big) + \mathbb{P}\big(U_0 \in A^{\downarrow}_{h,k}(y)\textstyle\bigcap[0,x]\big)\Big], \end{aligned}$$

where $A^{\uparrow}_{h,k}$ and $A^{\downarrow}_{h,k}$ are given in (3.3). Set $n_1 := n_1(x;h) = \big\{k : x \in [a_{h,k-1}, a_{h,k})\big\}$ and let

$$n_1^{\uparrow} := \{1,\cdots,n_1-1\}\textstyle\bigcap K_h^{\uparrow} \quad \text{and} \quad n_1^{\downarrow} := \{1,\cdots,n_1-1\}\textstyle\bigcap K_h^{\downarrow},$$

with $K_h^{\uparrow}$ and $K_h^{\downarrow}$ as in (2.1). Notice that $n_1^{\uparrow}\bigcap n_1^{\downarrow} = \emptyset$ and $n_1^{\uparrow}\bigcup n_1^{\downarrow} = \{1,\cdots,n_1-1\}$. It follows that

$$\begin{aligned} H_{0,h}(x,y) &= \sum_{k=1}^{s^h}\Big[\mathbb{P}\big(U_0\in A^{\uparrow}_{h,k}(y)\textstyle\bigcap[0,x]\big) + \mathbb{P}\big(U_0\in A^{\downarrow}_{h,k}(y)\textstyle\bigcap[0,x]\big)\Big] \\ &= \sum_{k\in n_1^{\uparrow}} \mathbb{P}\big(U_0\in A^{\uparrow}_{h,k}(y)\textstyle\bigcap[0,x]\big) + \mathbb{P}\big(U_0\in \underbrace{A^{\uparrow}_{h,n_1}(y)\textstyle\bigcap[a_{h,n_1},x]}_{:=Q^{\uparrow}}\big)\delta_{K_h^{\uparrow}}(n_1) + \\ &\quad + \sum_{k\in n_1^{\downarrow}} \mathbb{P}\big(U_0\in A^{\downarrow}_{h,k}(y)\textstyle\bigcap[0,x]\big) + \mathbb{P}\big(U_0\in \underbrace{A^{\downarrow}_{h,n_1}(y)\textstyle\bigcap[a_{h,n_1},x]}_{:=Q^{\downarrow}}\big)\delta_{K_h^{\downarrow}}(n_1) \\ &= \sum_{k\in n_1^{\uparrow}} \mu_T\big(\big[a_{h,k-1},\mathcal{T}_{h,k}(y)\big]\big) + \mu_T(Q^{\uparrow})\delta_{K_h^{\uparrow}}(n_1) + \\ &\qquad + \sum_{k\in n_1^{\downarrow}} \mu_T\big(\big[\mathcal{T}_{h,k}(y),a_{h,k-1}\big]\big) + \mu_T(Q^{\downarrow})\delta_{K_h^{\downarrow}}(n_1), \end{aligned}$$

where

$$Q^{\uparrow} = \big[a_{h,n_1-1}, \min\big\{x,\mathcal{T}_{h,n_1}(y)\big\}\big] \quad \text{and} \quad Q^{\downarrow} = \begin{cases} \emptyset, & \text{if} \quad x < \mathcal{T}_{h,n_1}(y), \\ \big[\mathcal{T}_{h,n_1}(y), x\big], & \text{if} \quad x \ge \mathcal{T}_{h,n_1}(y), \end{cases}$$

so that

$$\begin{aligned} H_{0,h}(x,y) = &\sum_{k\in n_1^{\uparrow}}\big[F_0\big(\mathcal{T}_{h,k}(y)\big) - F_0(a_{h,k-1})\big] + \sum_{k\in n_1^{\downarrow}}\big[F_0(a_{h,k}) - F_0\big(\mathcal{T}_{h,k}(y)\big)\big] + \\ &+\big[F_0\big(\min\big\{x,\mathcal{T}_{h,n_1}(y)\big\}\big) - F_0(a_{h,n_1-1})\big]\delta_{K_h^{\uparrow}}(n_1) + \max\big\{0, F_0(x) - F_0(\mathcal{T}_{h,n_1}(y))\big\}\delta_{K_h^{\downarrow}}(n_1). \end{aligned}$$

Upon noticing that $F_0\big(\min\big\{x,\mathcal{T}_{h,n_1}(y)\big\}\big) = \min\big\{F_0(x), F_0\big(\mathcal{T}_{h,n_1}(y)\big)\big\}$, by Sklar's Theorem, it follows that

$$\begin{aligned} C^{\updownarrow}_{U_0,T^h(U_0)}(u,v) &= H_{0,h}\big(F_0^{-1}(u), F_h^{-1}(v)\big) = H_{0,h}\big(F_0^{-1}(u), F_0^{-1}(v)\big) \\ &= \sum_{k\in n_0^{\uparrow}}\big[\mathscr{F}_{h,k}(v) - F_0(a_{h,k-1})\big] + \big[\min\big\{u,\mathscr{F}_{h,n_0}(v)\big\} - F_0(a_{h,n_0-1})\big]\delta_{K_h^{\uparrow}}(n_0) + \\ &\quad + \sum_{k\in n_0^{\downarrow}}\big[F_0(a_{h,k}) - \mathscr{F}_{h,k}(v)\big] + \max\big\{0, u - \mathscr{F}_{h,n_0}(v)\big\}\delta_{K_h^{\downarrow}}(n_0), \end{aligned}$$

where $n_0 := n_1\big(F_0^{-1}(u);h\big)$ and $n_0^{\uparrow}$ and $n_0^{\downarrow}$ are given by (3.7). Now by Proposition 3.1 and Corollary 3.1, $C^{\updownarrow}_{X_t,X_{t+h}} = C^{\updownarrow}_{U_0,T^h(U_0)}$ and the desired result follows. ■

Remark 3.4. Let $\{a_k\}_{k\in\mathbb{N}}$ be an arbitrary sequence of real numbers and let $D \subseteq \mathbb{N}$ be a set of indexes. As usual, if $D = \emptyset$, we shall set $\sum_{k\in D} a_k = 0$.

Next we shall derive the copula associated to any pair (X_p, X_q), for any $p, q \in \mathbb{N}$, of random variables obtained from a $\mathcal{T}_\varphi^\uparrow$ or $\mathcal{T}_\varphi^\downarrow$ process with φ increasing almost everywhere. The results follow from a specialization of Proposition 3.2.

Proposition 3.3. *Let $T \in \mathcal{T}^\uparrow \bigcup \mathcal{T}^\downarrow$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_\varphi^\uparrow$ or $\mathcal{T}_\varphi^\downarrow$ process. If we let $\{a_{h,k}\}_{k=0}^{s^h}$ be the net associated to the nodes of T^h and $n_0 := n_0(u;h) = \big\{k : u \in \big[F_0(a_{h,k-1}), F_0(a_{h,k})\big)\big\}$, then the following is true:*

(i). *If $T \in \mathcal{T}^\uparrow$, then, for any $t, h \in \mathbb{N}$, $h \neq 0$ and $(u, v) \in I^2$,*

$$C^\uparrow_{X_t,X_{t+h}}(u,v) = \left(\sum_{k=1}^{n_0-1} \mathscr{F}_{h,k}(v) - F_0(a_{h,k-1})\right)\delta_{\mathbb{N}^*}(n_0-1) + \\ + \min\big\{u, \mathscr{F}_{h,n_0}(v)\big\} - F_0(a_{h,n_0-1}), \tag{3.8}$$

where $\mathbb{N}^* := \mathbb{N}\backslash\{0\}$.

(ii). *If $T \in \mathcal{T}^\downarrow$ instead, then, for any $t, h \in \mathbb{N}$, $h \neq 0$ and $(u, v) \in I^2$,*

$$C^\downarrow_{X_t,X_{t+h}}(u,v) = \left(\sum_{k=1}^{n_0-1} F_0(a_{h,k}) - \mathscr{F}_{h,k}(v)\right)\delta_{\mathbb{N}^*}(n_0-1) + \max\big\{0, u - \mathscr{F}_{h,n_0}(v)\big\}. \tag{3.9}$$

Proof: If $T \in \mathcal{T}^\uparrow$, (i) follows from Proposition 3.2 by noticing that $K_h^\downarrow = \emptyset$ while if $T \in \mathcal{T}^\downarrow$, the opposite happens, namely, $K_h^\uparrow = \emptyset$ which implies (ii). ■

Remark 3.5. For computational purposes, it can be advantageous to write copulas (3.6), (3.8) and (3.9) explicitly as a function of μ_T. In this case, (3.6) becomes

$$\begin{aligned} C^\updownarrow_{X_t,X_{t+h}}(u,v) = &\sum_{k\in n_0^\uparrow} \mu_T\Big(\big[a_{h,k-1}, \mathcal{T}_{h,k}\big(F_0^{-1}(v)\big)\big]\Big) + \\ &+ \mu_T\Big(\big[a_{h,n_0-1}, \min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}\big]\Big)\delta_{K_h^\uparrow}(n_0) + \\ &+ \sum_{k\in n_0^\downarrow} \mu_T\Big(\big[\mathcal{T}_{h,k}\big(F_0^{-1}(v)\big), a_{h,k}\big]\Big) + \mu_T\Big(\big[\mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big), F_0^{-1}(u)\big]^+\Big)\delta_{K_h^\downarrow}(n_0), \end{aligned} \tag{3.10}$$

where $[a,b]^+$ equals $[a,b]$, if $b \geq a$, and $\emptyset$ otherwise. By their turn, copulas (3.8) and (3.9) become respectively

$$\begin{aligned} C^\uparrow_{X_t,X_{t+h}}(u,v) \;=\; &\sum_{k=1}^{n_0-1} \mu_T\Big(\big[a_{h,k-1}, \mathcal{T}_{h,k}\big(F_0^{-1}(v)\big)\big]\Big)\delta_{\mathbb{N}^*}(n_0-1) + \\ &+ \mu_T\Big(\big[a_{h,n_0-1}, \min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}\big]\Big), \end{aligned}$$

and

$$C^\downarrow_{X_t,X_{t+h}}(u,v) = \sum_{k=1}^{n_0-1} \mu_T\Big(\big[\mathcal{T}_{h,k}\big(F_0^{-1}(v)\big), a_{h,k}\big]\Big)\delta_{\mathbb{N}^*}(n_0-1) + \mu_T\Big(\big[\mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big), F_0^{-1}(u)\big]^+\Big).$$

In the next lemma we show that the relation $C_{X_t,X_{t+h}} = C_{X_0,X_h}$, valid when φ is increasing, still holds in the decreasing case.

Lemma 3.4. *Let $T \in \mathcal{T}^{\updownarrow}$ and let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process for $\varphi \in \mathcal{L}^1(\mu_T)$ an almost surely decreasing function. Then, the relation*

$$C_{X_t,X_{t+h}}(u,v) = u + v - 1 + C_{U_0,T^h(U_0)}(1-u,1-v) = C_{X_0,X_h}(u,v), \tag{3.11}$$

holds for all $(u,v) \in I^2$.

Proof: In Proposition 3.1, we have shown that $C_{T^t(U_0),T^{t+h}(U_0)} = C_{U_0,T^h(U_0)}$ (independently of φ). Now, since the inverse of an almost everywhere decreasing function is still decreasing almost everywhere, upon applying Theorem 2.1, it follows that

$$C_{X_t,\,X_{t+h}}(u,v) = C_{\varphi^{-1}(T^t(U_0)),\,\varphi^{-1}(T^{t+h}(U_0))}(u,v) = u + v - 1 + C_{T^t(U_0),\,T^{t+h}(U_0)}(1-u,1-v),$$

which proves the first equality. As for the second, it suffices to notice that

$$C_{X_0,X_h}(u,v) = C_{\varphi(U_0),\varphi(T^h(U_0))}(u,v) = u + v - 1 + C_{U_0,\,T^h(U_0)}(1-u,1-v),$$

everywhere in I^2, by Theorem 2.1, and the result follows by Proposition 3.1. ■

In the next propositions we present the copulas related to $\mathcal{T}^{\updownarrow}_{\varphi}$ processes in the case where $\varphi \in \mathcal{L}^1(\mu_t)$ is an almost everywhere decreasing function. In this case, we use the same notation for the copulas as before, but we add an asterisk in order to emphasize the difference on φ. We start by considering $T \in \mathcal{T}^{\updownarrow}\setminus(\mathcal{T}^{\uparrow}\bigcup\mathcal{T}^{\downarrow})$.

Proposition 3.4. *Let $T \in \mathcal{T}^{\updownarrow}\setminus(\mathcal{T}^{\uparrow}\bigcup\mathcal{T}^{\downarrow})$ be a transformation with $s > 1$ nodes, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere decreasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process. If we let $\{a_{h,k}\}_{k=0}^{s^h}$ be the net associated to the nodes of T^h, then*

$$\begin{aligned} C^{\updownarrow *}_{X_t,X_{t+h}}(u,v) \;=\; & u + v - 1 + \sum_{k\in n_0^{\uparrow *}} \big[\mathscr{F}_{h,k}(1-v) - F_0(a_{h,k-1})\big] + \sum_{k\in n_0^{\downarrow *}} \big[F_0(a_{h,k}) - \mathscr{F}_{h,k}(1-v)\big] + \\ & + \big[\min\big\{1-u, \mathscr{F}_{h,n_0^*}(1-v)\big\} - F_0(a_{h,n_0^*-1})\big]\delta_{K_h^{\uparrow}}(n_0^*) + \\ & + \max\big\{0, 1-u-\mathscr{F}_{h,n_0^*}(1-v)\big\}\delta_{K_h^{\downarrow}}(n_0^*), \end{aligned} \tag{3.12}$$

where $n_0^* := n_0^*(u;h) = \big\{k : u \in \big[1-F_0(a_{h,k}), 1-F_0(a_{h,k-1})\big)\big\}$,

$$n_0^{\uparrow *} := \{1,\cdots,n_0^*-1\}\bigcap K_h^{\uparrow} \quad \textit{and} \quad n_0^{\downarrow *} := \{1,\cdots,n_0^*-1\}\bigcap K_h^{\downarrow}.$$

Proof: By Lemma 3.4, we have

$$C^{\updownarrow *}_{X_t,X_{t+h}}(u,v) = u + v - 1 + C^{\updownarrow}_{U_0,T^h(U_0)}(1-u,1-v) \tag{3.13}$$

so that the result follows by Proposition 3.2 (with φ as the identity map) upon substituting (3.6) into (3.13). ■

Proposition 3.5. *Let $T \in \mathcal{T}^{\uparrow}\bigcup\mathcal{T}^{\downarrow}$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere decreasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}^{\uparrow}_{\varphi}$ or $\mathcal{T}^{\downarrow}_{\varphi}$ process. If we let $\{a_{h,k}\}_{k=0}^{s^h}$ be the net associated to the nodes of T^h and $n_0^* := n_0^*(u;h) = \big\{k : u \in \big[1-F_0(a_{h,k}), 1-F_0(a_{h,k-1})\big)\big\}$, then the following is true:*

(i). *If $T \in \mathcal{T}^{\uparrow}$, then, for any $t,h \in \mathbb{N}$, $h \neq 0$ and $(u,v) \in I^2$,*

$$\begin{aligned} C^{\uparrow *}_{X_t,X_{t+h}}(u,v) \;=\; & u + v - 1 + \left(\sum_{k=1}^{n_0^*-1} \mathscr{F}_{h,k}(1-v) - F_0(a_{h,k-1})\right)\delta_{\mathbb{N}^*}(n_0^*-1) + \\ & + \min\big\{1-u, \mathscr{F}_{h,n_0^*}(1-v)\big\} - F_0(a_{h,n_0^*-1}). \end{aligned} \tag{3.14}$$

(ii). *If $T \in \mathcal{T}^{\downarrow}$, then, for any $t, h \in \mathbb{N}$, $h \neq 0$ and $(u,v) \in I^2$,*

$$C^{\downarrow *}_{X_t, X_{t+h}}(u,v) = u + v - 1 + \left(\sum_{k=1}^{n_0^*-1} F_0(a_{h,k}) - \mathscr{F}_{h,k}(1-v) \right) \delta_{\mathbb{N}^*}(n_0^* - 1) + \\ + \max\{0, 1 - u - \mathscr{F}_{h,n_0^*}(1-v)\}. \tag{3.15}$$

Proof: Immediate from Proposition 3.4 and (3.13). ∎

All copulas derived in this section are singular in the sense that $\partial C_{X_t,X_{t+h}}(u,v)/\partial u \partial v = 0$ everywhere on I^2. A question that naturally arises is what are the support of these copulas? To answer that, the following notation will be useful. Let $T \in \mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow}$ and F_0 be the distribution associated to a T-invariant probability measure. Assume that T has $s > 1$ nodes and let h be a positive integer. We define, for all $k \in \{1, \cdots, s^h\}$, functions $\ell^{\uparrow}_{h,k}, \ell^{\downarrow}_{h,k} : \big[F_0(a_{h,k-1}), F_0(a_{h,k})\big] \to I$ by setting

$$\ell^{\uparrow}_{h,k}(x) := \frac{x - F_0(a_{h,k-1})}{F_0(a_{h,k}) - F_0(a_{h,k-1})} \quad \text{and} \quad \ell^{\downarrow}_{h,k}(x) := \frac{F_0(a_{h,k}) - x}{F_0(a_{h,k}) - F_0(a_{h,k-1})}$$

and $\ell^{\uparrow *}_{h,k}, \ell^{\downarrow *}_{h,k} : \big[1 - F_0(a_{h,k}), 1 - F_0(a_{h,k-1})\big] \to I$ by setting

$$\ell^{\uparrow *}_{h,k}(x) := \frac{x + F_0(a_{h,k-1}) - 1}{F_0(a_{h,k}) - F_0(a_{h,k-1})} \quad \text{and} \quad \ell^{\downarrow *}_{h,k}(x) := \frac{1 - F_0(a_{h,k}) - x}{F_0(a_{h,k}) - F_0(a_{h,k-1})}\,.$$

Observe that $\ell^{\uparrow}_{h,k}$ is just the linear function joining $(F_0(a_{h,k-1}), 0)$ and $(F_0(a_{h,k}), 1)$, while $\ell^{\downarrow}_{h,k}$ joins $(F_0(a_{h,k-1}), 1)$ and $(F_0(a_{h,k}), 0)$. In the next proposition we provide a characterization for the support of the copulas derived so far.

Proposition 3.6. *Let $T \in \mathcal{T}^{\updownarrow}$ and μ_T be a T-invariant probability measure. For $\varphi_1 \in \mathcal{L}^1(\mu_T)$ an almost everywhere increasing function and $\varphi_2 \in \mathcal{L}^1(\mu_T)$ an almost everywhere decreasing function, let $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$ denote respectively the associated $\mathcal{T}^{\updownarrow}_{\varphi_1}$ and $\mathcal{T}^{\updownarrow}_{\varphi_2}$ process. Also suppose that T has $s \geq 1$ nodes. Then, for any $t, h \in \mathbb{N}$ and $h > 0$,*

$$\operatorname{supp}\{C^{\updownarrow}_{X_t,X_{t+h}}\} = \Big(\bigcup_{k\in K^{\uparrow}} \big\{\big(u, \ell^{\uparrow}_{h,k}(u)\big) : u \in R_{h,k}\big\}\Big) \bigcup \Big(\bigcup_{k\in K^{\downarrow}} \big\{\big(u, \ell^{\downarrow}_{h,k}(u)\big) : u \in R_{h,k}\big\}\Big),$$

and

$$\operatorname{supp}\{C^{\updownarrow}_{Y_t,Y_{t+h}}\} = \Big(\bigcup_{k\in K^{\uparrow}} \big\{\big(u, \ell^{\downarrow *}_{h,k}(u)\big) : u \in R^*_{h,k}\big\}\Big) \bigcup \Big(\bigcup_{k\in K^{\downarrow}} \big\{\big(u, \ell^{\uparrow *}_{h,k}(u)\big) : u \in R^*_{h,k}\big\}\Big),$$

where $R_{h,k} := \big[F_0(a_{h,k-1}), F_0(a_{h,k})\big]$ and $R^*_{h,k} := \big[1 - F_0(a_{h,k}), 1 - F_0(a_{h,k-1})\big]$.

Proof: Let $R = [u_1, u_2] \times [v_1, v_2]$ be a rectangle in I^2. First assume that $T \in \mathcal{T}^{\uparrow}_{\varphi_1}$ and let $V_{C^{\uparrow}_{\boldsymbol{X}}}(R)$ denote the $C_{X_t,X_{t+h}}$-volume of R. Let $k \in \{1, \cdots, s^h\}$ be fixed and since for any copula C, the C-volume is a (doubly stochastic) measure, we can assume without loss of generality that $u_i \in R_{h,k}$, for $i = 1, 2$, so that $n_0 = k$ for all terms in the expression of $V_{C^{\uparrow}_{\boldsymbol{X}}}(R)$. Hence the summands and constants on the copula cancel out so that we have

$$\begin{aligned} V_{C^{\uparrow}_{\boldsymbol{X}}}(R) &= \min\big\{u_1, \mathscr{F}_{h,k}(v_1)\big\} + \min\big\{u_2, \mathscr{F}_{h,k}(v_2)\big\} - \min\big\{u_1, \mathscr{F}_{h,k}(v_2)\big\} - \min\big\{u_2, \mathscr{F}_{h,k}(v_1)\big\} \\ &= V_M\big([u_1, u_2] \times \big[\mathscr{F}_{h,k}(v_1), \mathscr{F}_{h,k}(v_2)\big]\big), \end{aligned}$$

where $M(u,v) = \min\{u, v\}$ is the Frechèt upper bound copula, whose support is the main diagonal in I^2. Since $[u_1, u_2] \times \big[\mathscr{F}_{h,k}(v_1), \mathscr{F}_{h,k}(v_2)\big] \subseteq R^2_{h,k}$, it follows that $V_{C^{\uparrow}_{\boldsymbol{X}}}(R) > 0$ if, and only if, $R \bigcap \big\{\big(u, \ell^{\uparrow}_{h,k}(u)\big) : u \in R_{h,k}\big\} \neq \emptyset$.

Now assume that $T \in \mathcal{T}^{\downarrow}_{\varphi_1}$ and, considering a rectangle R as before and $u_i \in R_{h,k}$, let $V_{C^{\downarrow}_{\boldsymbol{X}}}(R)$ denote its $C_{X_t,X_{t+h}}$-volume. Again the summands and constants on the copula cancel out and we have

$$\begin{aligned} V_{C^{\downarrow}_{\boldsymbol{X}}}(R) &= \max\left\{0, u_1 - \mathscr{F}_{h,k}(v_1)\right\} + \max\left\{0, u_2 - \mathscr{F}_{h,k}(v_2)\right\} - \max\left\{0, u_1 - \mathscr{F}_{h,k}(v_2)\right\} - \\ &\quad - \max\left\{0, u_2 - \mathscr{F}_{h,k}(v_1)\right\} \\ &= V_W\big([u_1, u_2] \times \left[1 - \mathscr{F}_{h,k}(v_1), 1 - \mathscr{F}_{h,k}(v_2)\right]\big), \end{aligned}$$

where $W(u,v) = \max\{0, u+v-1\}$ is the Frechèt lower bound copula, whose support is the secondary diagonal in I^2. Since $[u_1,u_2] \times \left[1 - \mathscr{F}_{h,k}(v_1), 1 - \mathscr{F}_{h,k}(v_2)\right] \subseteq R_{h,k} \times R^*_{h,k}$, it follows that $V_{C^{\downarrow}_{\boldsymbol{X}}}(R) > 0$ if, and only if, $R \bigcap \left\{\big(u, \ell^{\downarrow}_{h,k}(u)\big) : u \in R_{h,k}\right\} \neq \emptyset$.

Now suppose $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ and let $V_{C^{\updownarrow}_{\boldsymbol{X}}}(R)$ denote the $C_{X_t,X_{t+h}}$-volume of R. Let k be fixed as before and let $u_i \in R_{h,k}$, $i = 1, 2$. We can write

$$V_{C^{\updownarrow}_{\boldsymbol{X}}}(R) = V_{C^{\uparrow}_{\boldsymbol{X}}}(R)\delta_{K^{\uparrow}_h}(n_0) + V_{C^{\downarrow}_{\boldsymbol{X}}}(R)\delta_{K^{\downarrow}_h}(n_0),$$

so that the result follows from the previous cases and by observing that we can write $I = \bigcup_{k=1}^{s^h} R_{h,k}$.

As for $T \in \mathcal{T}^{\updownarrow}_{\varphi_2}$, the result follows similarly as the previous case, by noticing that for a rectangle R as before and $u_i \in R^*_{h,k}$, $i = 1, 2$,

$$V_{C^{\uparrow}_{\boldsymbol{Y}}}(R) = V_M\big([1-u_2, 1-u_1] \times \left[\mathscr{F}_{h,k}(1-v_2), \mathscr{F}_{h,k}(1-v_1)\right]\big)$$

and

$$V_{C^{\downarrow}_{\boldsymbol{Y}}}(R) = V_W\big([1-u_2, 1-u_1] \times \left[1 - \mathscr{F}_{h,k}(1-v_2), 1 - \mathscr{F}_{h,k}(1-v_1)\right]\big),$$

and by applying standard arguments. This completes the proof. ∎

We end this section with the following remark on the unicity of a T-invariant probability measure and the copulas related to the associated process.

Remark 3.6. A map $T \in \mathcal{T}^{\updownarrow}$ can have several different absolutely continuous T-invariant probability measures. This implies that, for a single given $T \in \mathcal{T}^{\updownarrow}$, the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process can have several different copulas associated to it depending on the choice of the T-invariant probability measure. This happens because U_0, which fundamentally defines the probability structure of the process, also depends completely on the choice of the T-invariant probability measure.

4 Multidimensional Case

In this section we shall extend our results from the previous bidimensional set up to a multidimensional one. That is, in this section we are interested in deriving the copulas related to n-dimensional vectors $(X_{t_1}, \cdots, X_{t_n})$ coming from a $\mathcal{T}^{\updownarrow}_{\varphi}$ process for φ an almost everywhere monotone function. It turns out that the bidimensional and the multidimensional case have more in common than one could expect and much of the work will be built over the results of last section.

First, let us establish some useful notation. Let $a, b \in \mathbb{N}$ with $a < b$. We shall write $a{:}b := \{a, \cdots, b\}$, $x_{a:b} := (x_a, \cdots, x_b)$ and, for a function f, we shall write $f(x_{a:b}) := \big(f(x_a), \cdots, f(x_b)\big)$. Again we shall denote the distribution of U_0 by F_0. The next proposition will be useful to simplify the proofs of the main results of this section and in establishing notation.

Proposition 4.1. *Let $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ and suppose T has $s \geq 1$ nodes. Denote by $\{a_{t,k}\}_{k=0}^{s^t}$ the net associated to the nodes of T^t. Let $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$, set $\boldsymbol{t} = (t, t + h_1, \cdots, t + h_n)$ and let $H_{\boldsymbol{t}}$ denote the distribution function of $\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_n}(U_0)\big)$. Then, for all $(x_0, \cdots, x_n) \in I^{n+1}$,*

$$H_{\boldsymbol{t}}(x_0, \cdots, x_n) = \sum_{k \in K^{\uparrow}_{h_n}} \mathbb{P}\big(U_0 \in \widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n}) \bigcap [0, x_0]\big) + \sum_{k \in K^{\downarrow}_{h_n}} \mathbb{P}\big(U_0 \in \widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n}) \bigcap [0, x_0]\big), \tag{4.1}$$

where

$$\widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n}) = \begin{cases} \big[a_{h_n,k-1}, b^{\uparrow}_{h_n,k}(x_{1:n})\big], & \text{if } k \in K^{\uparrow}_{h_n}, \\ \emptyset, & \text{otherwise,} \end{cases} \tag{4.2}$$

and

$$\widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n}) = \begin{cases} \big[b^{\downarrow}_{h_n,k}(x_{1:n}), a_{h_n,k}\big], & \text{if } k \in K^{\downarrow}_{h_n}, \\ \emptyset, & \text{otherwise,} \end{cases} \tag{4.3}$$

with,

$$b^{\uparrow}_{h_n,k}(x_{1:n}) = \min_{i=1,\cdots,n} \big\{c^{\uparrow}_i(x_i; h_n, k)\big\} \quad \text{and} \quad b^{\downarrow}_{h_n,k}(x_{1:n}) = \min_{i=1,\cdots,n} \big\{c^{\downarrow}_i(x_i; h_n, k)\big\} \tag{4.4}$$

where

$$c^{\uparrow}_i(x_i; h_n, k) = \begin{cases} a_{h_n,k-1}, & \text{if } B^{\uparrow}_i(x_i; h_n, k) = \emptyset, \\ B^{\uparrow}_i(x_i; h_n, k), & \text{otherwise,} \end{cases}$$

$$c^{\downarrow}_i(x_i; h_n, k) = \begin{cases} a_{h_n,k}, & \text{if } B^{\downarrow}_i(x_i; h_n, k) = \emptyset, \\ B^{\downarrow}_i(x_i; h_n, k), & \text{otherwise,} \end{cases}$$

$$B^{\uparrow}_i(x_i; h_n, k) = \min_{j=1,\cdots,s^{h_i}} \big\{\mathcal{T}_{h_i,j}(x_i) : \mathcal{T}_{h_i,j}(x_i) > a_{h_n,k-1} \text{ and } a_{h_i,j} < a_{h_n,k}\big\},$$

and

$$B^{\downarrow}_i(x_i; h_n, k) = \max_{j=1,\cdots,s^{h_i}} \big\{\mathcal{T}_{h_i,j}(x_i) : \mathcal{T}_{h_i,j}(x_i) < a_{h_n,k} \text{ and } a_{h_i,j} > a_{h_n,k-1}\big\}.$$

Proof: In view of Theorem 2.1, given $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$ it suffices to prove the result for the vector $\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_1}(U_0)\big)$. Let $H_{\boldsymbol{t}}$ denote the distribution function of $\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_1}(U_0)\big)$ and let $x_{0:n} \in (0,1)^{n+1}$. Also, for the sake of simplicity, let $Y_t := T^t(U_0)$. We have

$$\begin{aligned}
&H_{\boldsymbol{t}}(x_0, \cdots, x_n) = \mathbb{P}\big(T^t(U_0) \leq x_0, T^{t+h_1}(U_0) \leq x_1, \cdots, T^{t+h_n}(U_0) \leq x_n\big) \\
&= \mathbb{P}\big(Y_t \leq x_0, T^{h_1}(Y_t) \leq x_1, \cdots, T^{h_n}(Y_t)(U_0) \leq x_n\big) \\
&= \mathbb{P}\Big(Y_t \in [0, x_0], Y_t \in \bigcup_{k=1}^{s^{h_1}} \big(A^{\uparrow}_{h_1,k}(x_1) \bigcup A^{\downarrow}_{h_1,k}(x_1)\big), \cdots, Y_t \in \bigcup_{k=1}^{s^{h_n}} \big(A^{\uparrow}_{h_n,k}(x_n) \bigcup A^{\downarrow}_{h_n,k}(x_n)\big)\Big) \\
&= \mathbb{P}\Big(Y_t \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} \big(A^{\uparrow}_{h_i,k}(x_i) \bigcup A^{\downarrow}_{h_i,k}(x_i)\big)\Big) \\
&= \mathbb{P}\Big(U_0 \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} \big(A^{\uparrow}_{h_i,k}(x_i) \bigcup A^{\downarrow}_{h_i,k}(x_i)\big)\Big) \qquad (4.5) \\
&= \mathbb{P}\Big(U_0 \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} A^{\uparrow}_{h_i,k}(x_i)\Big) + \mathbb{P}\Big(U_0 \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} A^{\downarrow}_{h_i,k}(x_i)\Big) - \\
&\quad - \mathbb{P}\Big(U_0 \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} \big(A^{\uparrow}_{h_i,k}(x_i) \bigcap A^{\downarrow}_{h_i,k}(x_i)\big)\Big), \qquad (4.6)
\end{aligned}$$

where $A^{\uparrow}_{h_i,k}$ and $A^{\downarrow}_{h_i,k}$ are given in (3.3). The last term in (4.6) is equal to zero since the non-atomicity of the distribution of U_0 implies

$$\begin{aligned}
\mathbb{P}\Big(U_0 \in [0, x_0] \bigcap_{i=1}^{n} \bigcup_{k=1}^{s^{h_i}} \big(A^{\uparrow}_{h_i,k}(x_i) \bigcap A^{\downarrow}_{h_i,k}(x_i)\big)\Big) \leq \\
\leq \mathbb{P}\Big(U_0 \in [0, x_0] \bigcap \big[\big(\bigcup_{k=0}^{s^{h_n}} \{a_{h_n,k}\}\big) \bigcup \big(\bigcup_{k=1}^{s^{h_n}} \{\mathcal{T}_{h_n,k}(x)\}\big)\big]\Big) \\
\leq \mathbb{P}\Big(U_0 \in \big[\big(\bigcup_{k=0}^{s^{h_n}} \{a_{h_n,k}\}\big) \bigcup \big(\bigcup_{k=1}^{s^{h_n}} \{\mathcal{T}_{h_n,k}(x)\}\big)\big]\Big) = 0.
\end{aligned}$$

Considering $K^{\uparrow}_{h_n}$ and $K^{\downarrow}_{h_n}$ given by (2.1), let $b^{\uparrow}_{h_n,k}(x_{1:n})$, $b^{\downarrow}_{h_n,k}(x_{1:n})$, $\widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n})$ and $\widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n})$ be as in the enunciate. Notice that $\widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n})$ and $\widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n})$ are both proper closed subsets of $[a_{h_n,k-1}, a_{h_n,k}]$. Each $b^{\uparrow}_{h_n,k}(x_{1:n})$ actually is the smallest $\mathcal{T}_{h_i,j}(x_i)$ that lies on the k-th node of T^{h_n} (which determines the thinner partition among all T^{h_i}), so that $\widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n})$'s are just the intersection of all $A_{h_i,k}(x_i)$'s with the k-th node of T^{h_n}. In opposition, $b^{\downarrow}_{h_n,k}(x_{1:n})$ is the largest $\mathcal{T}_{h_i,j}(x_i)$ that lies on the k-th node of T^{h_n}. Also notice that the $\widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n})$'s are pairwise disjoint and so are the $\widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n})$'s. Now, with this machinery, we can rewrite (4.6) as

$$\begin{aligned} H_{\boldsymbol{t}}(x_0,\cdots,x_n) &= \mathbb{P}\Big(U_0 \in \bigcup\nolimits_{k\in K^{\uparrow}_{h_n}} \widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n}) \bigcap [0,x_0]\Big) + \mathbb{P}\Big(U_0 \in \bigcup\nolimits_{k\in K^{\downarrow}_{h_n}} \widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n}) \bigcap [0,x_0]\Big) \\ &= \sum_{k\in K^{\uparrow}_{h_n}} \mathbb{P}\big(U_0 \in \widetilde{A}^{\uparrow}_{h_n,k}(x_{1:n}) \bigcap [0,x_0]\big) + \sum_{k\in K^{\downarrow}_{h_n}} \mathbb{P}\big(U_0 \in \widetilde{A}^{\downarrow}_{h_n,k}(x_{1:n}) \bigcap [0,x_0]\big), \end{aligned}$$

which is the desired formula. ■

Proposition 4.2. *Let $T \in \mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow}$ and suppose T has $s \geq 1$ nodes and denote by $\{a_{t,k}\}_{k=0}^{s^t}$ the net associated to the nodes of T^t. Let $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$ and $H_{\boldsymbol{t}}$ denote the distribution function of $\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_n}(U_0)\big)$. Then, for all $(x_0,\cdots,x_n) \in I^{n+1}$,*

$$H_{\boldsymbol{t}}(x_0,\cdots,x_n) = \sum_{k=1}^{s^{h_n}} \mathbb{P}\big(U_0 \in \widetilde{A}_{h_n,k}(x_{1:n}) \bigcap [0,x_0]\big), \tag{4.7}$$

where

$$\widetilde{A}_{h_n,k}(x_{1:n}) = \begin{cases} \big[a_{h_n,k-1}, b^{\uparrow}_{h_n,k}(x_{1:n})\big], & \text{if } T \in \mathcal{T}^{\uparrow}, \\ \big[b^{\downarrow}_{h_n,k}(x_{1:n}), a_{h_n,k}\big], & \text{if } T \in \mathcal{T}^{\downarrow}, \end{cases} \tag{4.8}$$

with $b^{\uparrow}_{h_n,k}(x_{1:n})$ and $b^{\downarrow}_{h_n,k}(x_{1:n})$ given by (4.4).

Proof: With the notation of Proposition 4.1, the result follows by noticing that $\widetilde{A}_{h_n,k}$ is just a combination of (4.2) and (4.3) and that if $T \in \mathcal{T}^{\uparrow}$, then $K^{\downarrow}_{h_n} = \emptyset$, while if $T \in \mathcal{T}^{\uparrow}$, we then have $K^{\uparrow}_{h_n} = \emptyset$. ■

Proposition 4.3. *Let $T \in \mathcal{T}^{\updownarrow}$, μ_T be a T-invariant probability measure and let U_0 be distributed as μ_T. Then, for any $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$,*

$$\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_n}(U_0)\big) \overset{d}{=} \big(U_0, T^{h_1}(U_0), \cdots, T^{h_n}(U_0)\big). \tag{4.9}$$

Furthermore, $\big(T^t(U_0), T^{t+h_1}(U_0), \cdots, T^{t+h_n}(U_0)\big)$ and $\big(U_0, T^{h_1}(U_0), \cdots, T^{h_n}(U_0)\big)$ have the same copula.

Proof: Let $H_{\boldsymbol{h}}$ denote the distribution of $\big(U_0, T^{h_1}(U_0), \cdots, T^{h_n}(U_0)\big)$. Notice that it suffices to prove that, for any $(x_0,\cdots,x_n) \in I^{n+1}$, $H_{\boldsymbol{h}}$ is equal to (4.5). Indeed, we have

$$\begin{aligned} H_{\boldsymbol{h}}(x_0,\cdots,x_n) &= \mathbb{P}\big(U_0 \leq x_0, T^{h_1}(U_0) \leq x_1, \cdots, T^{h_n}(U_0) \leq x_n\big) \\ &= \mathbb{P}\Big(U_0 \in [0,x_0], U_0 \in \bigcup\nolimits_{k=1}^{s^{h_1}} \big(A^{\uparrow}_{h_1,k}(x_1) \bigcup A^{\downarrow}_{h_1,k}(x_1)\big), \cdots, U_0 \in \bigcup\nolimits_{k=1}^{s^{h_n}} \big(A^{\uparrow}_{h_n,k}(x_n) \bigcup A^{\downarrow}_{h_n,k}(x_n)\big)\Big) \\ &= \mathbb{P}\Big(U_0 \in [0,x_0] \bigcap\nolimits_{i=1}^{n} \bigcup\nolimits_{k=1}^{s^{h_i}} \big(A^{\uparrow}_{h_i,k}(x_i) \bigcup A^{\downarrow}_{h_i,k}(x_i)\big)\Big), \end{aligned}$$

which is precisely (4.5). The other assertion follows from Sklar's theorem in view of (4.9) and from the fact that $T^t(U_0) \overset{d}{=} U_0$, for all $t \in \mathbb{N}$. ■

Upon specializing the results seen so far, we can now determine the copulas associated to multidimensional vectors coming from a $\mathcal{T}_\varphi^\updownarrow$ process with φ increasing almost everywhere starting with the case $T\in\mathcal{T}^\updownarrow\setminus(\mathcal{T}^\uparrow\bigcup\mathcal{T}^\downarrow)$, which essentially embodies the other two cases.

Theorem 4.1. *Let $T\in\mathcal{T}^\updownarrow\setminus(\mathcal{T}^\uparrow\bigcup\mathcal{T}^\downarrow)$, μ_T be a T-invariant probability measure and $\varphi\in\mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_\varphi^\updownarrow$ process. If we let $\{a_{h_n,k}\}_{k=0}^{s^{h_n}}$ be the net associated to the nodes of T^{h_n}, and $C_{\boldsymbol{t}}$ be the copula associated to $(X_t,X_{t+h_1},\cdots,X_{t+h_n})$, for $t,h_1,\cdots,h_n\in\mathbb{N}$, $0<h_1<\cdots<h_n$, then*

$$\begin{aligned}C_{\boldsymbol{t}}(u_0,\cdots,u_n) &= \sum_{k\in n_0^\uparrow}\big[F_0\big(r^\uparrow_{h_n,k}(u_{1:n})\big)-F_0(a_{h_n,k-1})\big]+\sum_{k\in n_0^\downarrow}F_0(a_{h_n,k})-F_0\big(r^\downarrow_{h_n,k}(u_{1:n})\big)+\\ &\quad+\Big[\min\big\{u_0,F_0\big(r^\uparrow_{h_n,n_0}(u_{1:n})\big)\big\}-F_0(a_{h_n,n_0-1})\Big]\delta_{K^\uparrow_{h_n}}(n_0)+\\ &\quad+\max\big\{0,u_0-F_0\big(r^\downarrow_{h_n,n_0}(u_{1:n})\big)\big\}\,\delta_{K^\downarrow_{h_n}}(n_0),\end{aligned}$$

where $n_0=\big\{k:u_0\in\big(F_0(a_{h_n,k-1}),F_0(a_{h_n,k})\big]\big\}$, with $K^\uparrow_{h_n}$ and $K^\downarrow_{h_n}$ given in (2.1),

$$n_0^\uparrow=\{1,\cdots,n_0-1\}\bigcap K^\uparrow_{h_n},\quad n_0^\downarrow=\{1,\cdots,n_0-1\}\bigcap K^\downarrow_{h_n},$$

with $b^\uparrow_{h_n,k}$ and $b^\downarrow_{h_n,k}$ given in (4.4),

$$r^\uparrow_{h_n,k}(u_{1:n})=b^\uparrow_{h_n,k}\big(F_0^{-1}(u_{1:n})\big)\quad\textit{and}\quad r^\downarrow_{h_n,k}(u_{1:n})=b^\downarrow_{h_n,k}\big(F_0^{-1}(u_{1:n})\big).\tag{4.10}$$

Proof: In view of Proposition 4.3, we only need to show the result for the vector $\big(U_0,T^{h_1}(U_0),\cdots,T^{h_n}(U_0)\big)$, where $h_1,\cdots,h_n\in\mathbb{N}$, $0<h_1<\cdots<h_n$. Considering $K^\uparrow_{h_n}$ and $K^\downarrow_{h_n}$ as in (2.1), let $b^\uparrow_{h_n,k}(x_{1:n})$, $b^\downarrow_{h_n,k}(x_{1:n})$, $\widetilde{A}^\uparrow_{h_n,k}(x_{1:n})$ and $\widetilde{A}^\downarrow_{h_n,k}(x_{1:n})$ be respectively as in (4.4), (4.2) and (4.3). Set $n_1=\big\{k:x_0\in(a_{h_n,k-1},a_{h_n,k}]\big\}$ and let $n_1^\uparrow=\{1,\cdots,n_1-1\}\bigcap K^\uparrow_{h_n}$ and $n_1^\downarrow=\{1,\cdots,n_1-1\}\bigcap K^\downarrow_{h_n}$. Notice that $n_1^\uparrow\bigcap n_1^\downarrow=\emptyset$ and $n_1^\uparrow\bigcup n_1^\downarrow=\{1,\cdots,n_1-1\}$. We can now rewrite (4.1), in view of Proposition 4.3 and setting $\boldsymbol{h}=(0,h_1,\cdots,h_n)$, as

$$\begin{aligned}H_{\boldsymbol{h}}(x_0,\cdots,x_n) &= \sum_{k\in n_1^\uparrow}\mu_T\Big(\big[a_{h_n,k-1},b^\uparrow_{h_n,k}(x_{1:n})\big]\Big)+\sum_{k\in n_1^\downarrow}\mu_T\Big(\big[b^\downarrow_{h_n,k}(x_{1:n}),a_{h_n,k}\big]\Big)+\\ &\quad+\mu_T\Big(\big[a_{n_1-1},\min\big\{x_0,b^\uparrow_{h_n,n_1}(x_{1:n})\big\}\big]\Big)\delta_{K^\uparrow_{h_n}}(n_1)+\\ &\quad+\mu_T\Big(\big[b^\downarrow_{h_n,n_1}(x_{1:n}),x_0\big]\Big)\,\delta_{[b^\downarrow_{h_n,n_1}(x_{1:n}),1]}(x_0)\delta_{K^\downarrow_{h_n}}e(n_1)\\ &= \sum_{k\in n_1^\uparrow}\big[F_0\big(b^\uparrow_{h_n,k}(x_{1:n})\big)-F_0(a_{h_n,k-1})\big]+\sum_{k\in n_1^\downarrow}\big[F_0(a_{h_n,k})-F_0\big(b^\downarrow_{h_n,k}(x_{1:n})\big)\big]+\\ &\quad+\Big[F_0\big(\min\big\{x_0,b^\uparrow_{h_n,n_1}(x_{1:n})\big\}\big)-F_0(a_{h_n,n_1-1})\Big]\delta_{K^\uparrow_{h_n}}(n_1)+\\ &\quad+\max\big\{0,F_0(x_0)-F_0\big(b^\downarrow_{h_n,n_1}(x_{1:n})\big)\big\}\,\delta_{K^\downarrow_{h_n}}(n_1)\end{aligned}$$

Now, by Sklar's Theorem and (3.1), taking $n_0=\big\{k:u_0\in\big(F_0(a_{h_n,k-1}),F_0(a_{h_n,k})\big]\big\}$, it follows that

$$\begin{aligned}C_{\boldsymbol{h}}(u_0,\cdots,u_n) &= H_{\boldsymbol{h}}\big(F_0^{-1}(u_0),\cdots,F_0^{-1}(u_n)\big)\\ &= \sum_{k\in n_0^\uparrow}\big[F_0\big(r^\uparrow_{h_n,k}(u_{1:n})\big)-F_0(a_{h_n,k-1})\big]+\sum_{k\in n_0^\downarrow}\big[F_0(a_{h_n,k})-F_0\big(r^\downarrow_{h_n,k}(u_{1:n})\big)\big]+\\ &\quad+\Big[\min\big\{u_0,F_0\big(r^\uparrow_{h_n,n_0}(u_{1:n})\big)\big\}-F_0(a_{h_n,n_0-1})\Big]\delta_{K^\uparrow_{h_n}}(n_0)+\\ &\quad+\max\big\{0,u_0-F_0\big(r^\downarrow_{h_n,n_0}(u_{1:n})\big)\big\}\,\delta_{K^\downarrow_{h_n}}(n_0),\end{aligned}$$

where $r^\uparrow_{h_n,k}$ and $r^\downarrow_{h_n,k}$ are as in the enunciate. ∎

From Proposition 4.1 we have all we need to work with the case $T \in \mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow}$. This is the content of the next proposition.

Proposition 4.4. *Let $T \in \mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow}$, μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_{\varphi}^{\uparrow}$ or $\mathcal{T}_{\varphi}^{\downarrow}$ process. If we let $\{a_{h_n,k}\}_{k=0}^{s^{h_n}}$ be the net associated to the nodes of T^{h_n}, and $C_{\boldsymbol{t}}$ be the copula associated to $(X_t, X_{t+h_1}, \cdots, X_{t+h_n})$, for $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$, then*

(i). *if $T \in \mathcal{T}^{\uparrow}$,*

$$C_{\boldsymbol{t}}(u_0, \cdots, u_n) = \sum_{k=1}^{n_0-1} \left[F_0\big(r_{h_n,k}(u_{1:n})\big) - F_0(a_{h_n,k-1})\right] + \min\left\{u_0, F_0\big(r_{h_n,n_0}(u_{1:n})\big)\right\} - F_0(a_{h_n,n_0-1});$$

(ii). *if $T \in \mathcal{T}^{\downarrow}$,*

$$C_{\boldsymbol{t}}(u_0, \cdots, u_n) = \sum_{k=1}^{n_0-1} \left[F_0(a_{h_n,k}) - F_0\big(r_{h_n,k}(u_{1:n})\big)\right] + \max\left\{0, u_0 - F_0\big(r_{h_n,n_0}(u_{1:n})\big)\right\},$$

where $n_0 := \left\{k : u_0 \in \big(F_0(a_{h_n,k-1}), F_0(a_{h_n,k})\big]\right\}$, $b_{h_n,k}^{\uparrow}$ and $b_{h_n,k}^{\downarrow}$ given in (4.4), and

$$r_{h_n,k}(u_{1:n}) := \begin{cases} b_{h_n,k}^{\uparrow}\big(F_0^{-1}(u_{1:n})\big), & \text{if } T \in \mathcal{T}^{\uparrow}, \\ b_{h_n,k}^{\downarrow}\big(F_0^{-1}(u_{1:n})\big), & \text{if } T \in \mathcal{T}^{\downarrow}. \end{cases}$$

Proof: We observe that $r_{h_n,k}$ is just a combination of the two expressions in (4.10) and that if $T \in \mathcal{T}^{\uparrow}$, then $K_{h_n}^{\downarrow} = \emptyset$, so that $n_0^{\uparrow} = \{1, \cdots, n_0 - 1\}$ and $n_0^{\downarrow} = \emptyset$, while if $T \in \mathcal{T}^{\downarrow}$, then $K_{h_n}^{\uparrow} = \emptyset$ so that $n_0^{\downarrow} = \{1, \cdots, n_0 - 1\}$ and $n_0^{\uparrow} = \emptyset$. With this in mind, the results follow directly from Theorem 4.1. ■

The generalization to the n-dimensional case for φ an almost surely decreasing function leads to more complicated formulas in terms of the copulas in Theorem 4.1 and Proposition 4.4. Although the set up here is much more general than the one in Lopes and Pumi (2011), it is interesting to notice that the same result valid there can be applied here, so that we include it just for the sake of completeness.

Proposition 4.5. *Let $T \in \mathcal{T}^{\updownarrow}$, μ_T be a T-invariant probability measure and let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_{\varphi}^{\updownarrow}$ process for $\varphi \in \mathcal{L}^1(\mu_T)$ an almost surely decreasing function. Let $t, h_1, \cdots, h_n \in \mathbb{N}$, $0 < h_1 < \cdots < h_n$ and set $Y_k := T^{h_k}(U_0)$ and $Y_0 := U_0$. If we denote the copula associated to $(X_t, X_{t+h_1}, \cdots, X_{t+h_n})$ by $C_{\boldsymbol{t}}$, then the following relation holds*

$$\begin{aligned} C_{\boldsymbol{t}}(u_0, \cdots, u_n) = 1 - n + \sum_{i=0}^{n} u_i + \sum_{i=0}^{n} \sum_{j=i+1}^{n} C_{Y_i,Y_j}(1-u_i, 1-u_j) + \cdots + \\ + (-1)^{n-1} \sum_{k_1=0}^{n} \sum_{k_2=k_1+1}^{n} \cdots \sum_{k_{n-1}=k_{n-2}+1}^{n} C_{Y_{k_1},\cdots,Y_{k_{n-1}}}(1-u_{k_1}, \cdots, 1-u_{k_{n-1}}) + \\ + (-1)^n C_{U_0,Y_1,\cdots,Y_n}(1-u_0, \cdots, 1-u_n), \end{aligned} \tag{4.11}$$

everywhere in I^{n+1}.

Proof: See Proposition 4.1 in Lopes and Pumi (2011). ■

The copula in Proposition 4.5 can be explicitly calculated since (4.11) is written as sums of the copulas of vectors containing U_0 and $T^t(U_0)$ for different t's. Hence, Theorem 2.1 can be applied to derive the desired formulas in terms of the copulas in Theorem 4.1 and Proposition 4.4.

Remark 4.1. In principle, we could have assumed T belonging to the slightly broader family of $C^1_{1\text{-}1}$ transformations. However, for $T \in C^1_{1\text{-}1} \setminus C^{1+\alpha}_{1\text{-}1}(I)$, if an absolutely continuous T-invariant probability measure exists, it does not have a positive continuous Radon-Nikodym derivative. This implies that F_0^{-1} is not always well defined. In this case one can substitute F_0^{-1} by its pseudo-inverse, $F_0^{(-1)}$. Since $F_0 \circ F_0^{(-1)}$ and $F_0^{(-1)} \circ F_0$ are not the identity map, the expressions for the copulas become more complex. If $T \in C^1_{1\text{-}1} \setminus C^{1+\alpha}_{1\text{-}1}(I)$, the results in Sections 3 and 4 hold if F_0^{-1} is substituted by $F_0^{(-1)}$ and expressions of the form $F_0\big(F_0^{(-1)}(x)\big)$ and $F_0^{(-1)}\big(F_0(x)\big)$ will appear instead of the identities $F_0\big(F_0^{-1}(x)\big) = x$ and $F_0^{-1}\big(F_0(x)\big) = x$.

5 Numerical Approximations and Computational Issues

Let $T \in \mathcal{T}^{\updownarrow}$, $h > 0$, μ_T be a T-invariant probability measure and consider the associated $\mathcal{T}^{\updownarrow}_{\varphi}$ process for $\varphi \in \mathcal{L}^1(\mu_T)$ a monotone function. The computation and implementation of the copulas derived in Sections 3 and 4 depend on the knowledge of several elements. These elements are the invariant probability measure μ_T, the associated distribution function F_0 and its inverse F_0^{-1}, the inverse of T^h in each branch ($\{\mathcal{T}_{h,k}\}_{k=1}^{s^h}$) and the net $\{a_{h,k}\}_{k=0}^{s^h}$ associated to T^h, where $s > 1$ denotes, as usual, the number of branches of T.

As mentioned in Section 2, the general problem of determining the existence of a T-invariant absolutely continuous probability measure is usually a hard one. Therefore, one rarely finds an explicit formula for a T-invariant probability measure. Furthermore, although the computation of $T^h(y)$ is usually straightforward, the exact calculation of $\mathcal{T}_{h,k}(y)$ can be a highly complex task. Even for small values of h, the apparently simple calculation of the net associated to T^h can be troublesome. Nevertheless, one can still rely on approximations in order to implement, compute and study the copulas related to $\mathcal{T}^{\updownarrow}_{\varphi}$ processes.

With this in mind, our goal in this section is to present general results and conditions to construct simple approximations to the copulas derived in Sections 3 and 4 in such a way to guarantee its uniform convergence to the true copula.

5.1 Approximating the T-invariant probability measure and related functions

Perhaps the most appealing way to approximate a measure, assuming it is ergodic, is by using Birkhoff's ergodic theorem. Recall that a measure μ is called a *Sinai-Bowen-Ruelle measure* (SBR, for short) if the weak convergence

$$\frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}(A) \longrightarrow \mu(A) \tag{5.1}$$

holds for almost all $x \in I$ and all μ-continuity sets A. In this work, all the T-invariant probability measure discussed are assumed to be absolutely continuous with respect to the Lebesgue measure. However, if μ_T is ergodic and absolutely continuous with respect to the Lebesgue measure, it is also an SBR measure (see, for instance, Keller, 1998), so that we will only consider the latter case.

Given $T \in \mathcal{T}^{\updownarrow}$ and μ_T a T-invariant absolutely continuous SBR probability measure, one way to approximate μ_T is by truncating expression (5.1) for a reasonably large value of $n > 1$. That is, let $x_0 \in I$ be a point such that the weak convergence (5.1) holds. Set

$$\mu_n(A; T, x_0) := \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}(A), \tag{5.2}$$

for all μ_T-continuity sets A. This is not the only way to approximate an SBR measure, see for instance Lopes and Pumi (2011) and references therein. For simplicity and since no confusion will arise, we shall drop T and x_0 from the notation of μ_n.

From μ_n, an approximation for F_0 is obtained simply by setting $\hat{F}_n(x) = \mu_n\big([0,x];T,x_0\big)$. From a computational standpoint, fixed x_0 and truncation point $n > 0$, to calculate μ_n for different sets one needs to compute the iteration vector $\big(x_0, T(x_0), \cdots, T^{n-1}(x_0)\big)$ only once. Furthermore, by the choice of x_0, $T^i(x_0) \neq T^j(x_0)$ for all $i \neq j$, therefore, $\hat{F}_n$ will coincide with the empirical distribution function based on the iteration vector. To improve the performance pointwisely, one can use some interpolation method based on the iteration vector and $\hat{F}_n$. That is, we choose a sequence of interpolating functions[1] based on $\big(x_0, T(x_0), \cdots, T^{n-1}(x_0)\big)$ and its image by $\hat{F}_n$.

For approximations based on a vector of iterations, the limits taken are understood to be in terms of partitions in the following manner. Starting with a set of points $R_m = \{x_1, \cdots, x_m\}$, we consider refinements obtained by adding a single point to the set R_m. That is, we consider refinements of the form $R_{m+1} = R_m \bigcup \{x_{m+1}\}, \cdots, R_{m+k} = R_{m+k-1} \bigcup \{x_{m+k}\}$. Let $f_m(\cdot) := f(\,\cdot\,; R_m)$ be an approximation based on R_m. For a sequence of refinements $\{R_k\}_{k=m+1}^{\infty}$ we consider the sequence of functions $\{f_k\}_{k=m+1}^{\infty}$. If the sequence $\{f_k\}_{k=0}^{\infty}$ has a limit, we set $\lim\limits_{m\to\infty} f_m(\cdot) := \lim\limits_{m\to\infty} f(\,\cdot\,; R_m)$.

Proposition 5.1. *Let $S = \{x_1, \cdots, x_n\}$ be a given (ordered) sample of some continuous and monotone distribution F_0 and let $\hat{F}_n$ be the empirical distribution based on S. Let F_n be an approximation based on a sequence of interpolating functions defined from S and let $x_0 = 0$ and $x_{n+1} = 1$. Suppose that F_n satisfies, for each $x \in (x_i, x_i + 1)$, $F_n(x) \in \big(\hat{F}_n(x_i), \hat{F}_n(x_{i+1})\big)$, $i = 0, \cdots, n$ and $F_n(x_i) = \hat{F}_n(x_i)$, for all $i = 0, \cdots, n+1$. Then $F_n(x) \to F_0(x)$ uniformly over $x \in I$. If, in addition, F_n is continuous and monotone, then also $F_n^{-1}(x) \to F_0^{-1}(x)$ uniformly over $x \in I$.*

Proof: By the Glivenko-Cantelli theorem, $\widehat{F}_n(x) \to F_0(x)$ uniformly in $x \in [0,1]$, so that, given $\varepsilon > 0$, there exists $n_0 > 0$ depending on ε only such that $\big|\widehat{F}_n(x) - F_0(x)\big| < \varepsilon/2$, for all $x \in I$, whenever $n > n_0$. Now, for $x \in [0,1)$, there exists $k \in \{1, \cdots, n\}$ such that $x \in [x_k, x_{k+1})$ and since $F_n(x) \in \big[\hat{F}_n(x_i), \hat{F}_n(x_{i+1})\big)$, if $n > \max\{n_0, \lceil 2/\varepsilon \rceil\}$, it follows that

$$\begin{aligned}\big|F_n(x) - F_0(x)\big| &\leq \big|F_n(x) - \widehat{F}_n(x)\big| + \big|\widehat{F}_n(x) - F_0(x)\big| < \big|\widehat{F}_n(x_{k+1}) - \widehat{F}_n(x_k)\big| + \frac{\varepsilon}{2} \\ &\leq \sup_{i=1,\cdots,n-1} \big\{\big|\widehat{F}_n(x_{i+1}) - \widehat{F}_n(x_i)\big|\big\} + \frac{\varepsilon}{2} \;\leq\; \frac{1}{n} + \frac{\varepsilon}{2} \;<\; \varepsilon,\end{aligned}$$

uniformly in $x \in [0,1)$ and, by the continuity of F_0, it holds uniformly in I.

As for the inverse, notice that, by hypothesis, F_n^{-1} is (uniformly) continuous and monotone. Therefore, given $\varepsilon > 0$ and $y \in I$, there exists a $\delta > 0$ (depending on ε only) such that

$$|x - y| < \delta \implies |F_n^{-1}(x) - F_n^{-1}(y)| < \varepsilon.$$

By the Glivenko-Cantelli theorem, there exists $n_1 > 0$ (depending on δ only) such that $\big|F_n(x) - F_0(x)\big| < \delta$, whenever $n > n_1$, and the inequality holds for all $x \in [0,1]$. Furthermore, the monotonicity of F_0 implies the existence of $z_0 \in [0,1]$ such that $y = F_0(z_0)$. Now, taking $n > n_1$, it follows that

$$\big|F_n^{-1}(y) - F_0^{-1}(y)\big| = \big|F_n^{-1}\big(F_0(z_0)\big) - z_0\big| = \big|F_n^{-1}\big(F_0(z_0)\big) - F_n^{-1}\big(F_n(z_0)\big)\big| < \varepsilon,$$

and since n_1 does not depend on y, the convergence is uniform. ∎

[1] by interpolation function, we mean a function $f : [a,b] \to [c,d]$ such that, given a collection of pairs $P = \{(u_1, v_1), \cdots, (u_n, v_n)\}$, $u_i \in [a,b]$ and $v_i \in [c,d]$, $i = 1, \cdots, n$, and a point $x \in [a,b]$, f assigns a value $f(x;P)$ for each $x \in [a,b] \setminus \{u_1, \cdots, u_n\}$ and $f(u_i;P) = v_i$, for $i = 1, \cdots, n$. A simple linear interpolation or a spline interpolation are examples of such functions.

Suppose that T has $s > 1$ nodes. There is no general optimal way of approximating the net related to T^h, for a given $h > 1$, so the task has to be handled case by case. A general method, which may not be optimal in most cases, is as follows. Let $S := \{x_1, \cdots, x_m\}$ be an ordered set of points in I and, to avoid trivialities, suppose that $x_1 = 0$, $x_m = 1$ and $x_i \neq x_j$, for $i, j = 1, \cdots, m$, $i \neq j$. Let R denote the image of S by T^h. The next step is to identify possible intervals where the discontinuities of T^h lie on, by using the elements of R. In order to do that, S must reflect all the s^h discontinuities of T^h. Usually a relatively thin grid of equally spaced points solves the problem (obviously, $m > s^h$ points should be used).

If $T \in \mathcal{T}^{\uparrow}$, the identification can be done in the following manner. Note first that $a_{h,0} = 0$ and $a_{h,s^h} = 1$, for any h. Let $D^+ = \big\{i : T^h(x_i) > T^h(x_{i+1})\big\} \subset \{1, \cdots, m\}$ and let $\{d_j\}_{j=1}^{s^h-1}$ denote the ordered elements of D^+. The set D^+ contains the indexes $i \in \{1, \cdots, m\}$ for which the interval $[x_{d_k}, x_{d_{k+1}}]$ contains the k-th discontinuity of T^h. If $T \in \mathcal{T}^{\downarrow}$, we let $D^- = \big\{i : T^h(x_i) < T^h(x_{i+1})\big\}$. If $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, one has to define D as a mixture of D^+ and D^-. One obvious way to detect a poor choice of S is the number of elements in the respective set D^+, D^- or D, which should have exactly $s^h - 1$ distinct points.

Once the intervals containing the discontinuities of T^h are found, one can perform a simple bisection, a golden search or any other root-finder algorithm to find an approximation to $a_{h,k}$ up to a predetermined acceptable error. This gives an "instantaneous" estimate which, in terms of precision, should not improve as the number of elements in S increases. However, given the fact that, depending on the transformation, most of the times there are easier ways of calculating an approximation to $a_{h,k}$ based on S, our results cover this possibility as well. See also Remark 5.1.

As an example of this possibility, we mention the method presented in Lopes and Pumi (2011). Although presented in the context of Manneville-Pomeau transformations, it can be extended to a broader family of transformations as follows. Consider transformations of the type $T(x) = g(x) \pmod 1$ for a suitable increasing and differentiable function g, such that T has $s > 1$ full branches. Let S and D^+ be as before, and consider the function $T^*_{i,h} : [x_{d_i}, x_{d_i+1}] \to [0, 2]$ given by $T^*_{i,h}(x) := g\big(T^{h-1}(x)\big)$, for $i = 1, \cdots, s^h - 2$. Clearly $T^*_{i,h}(x_{d_i}) < 1$, $T^*_{i,h}(x_{d_i+1}) > 1$ and

$$T^h_s(x) = T^*_{i,h}(x) - \delta_{[1,2]}\big(T^*_{i,h}(x)\big),$$

for all $x \in [x_{d_i}, x_{d_i+1}]$. We denote the approximation to $a_{h,i}$, based on S, by $a^m_{h,i}$, and define it as the linear interpolation of 1 between the points $\big(x_{d_i}, T^*_{i,h}(x_{d_i})\big)$ and $\big(x_{d_i+1}, T^*_{i,h}(x_{d_i+1})\big)$. This approximation can be very crude if m is small, but the precision increases fast as m increases, especially for large h. The method can be easily adapted to cover the cases where g is decreasing and differentiable. In Section 6 we present a simpler example.

As for the approximation to $\mathcal{T}_{h,k}$, we shall use an interpolation argument based on a grid of values from T^h. Let $S = \{x_1, x_2, \cdots, x_m\}$, be an ordered set of points in I and assume that $x_1 = 0$, $x_m = 1$ and $x_i \neq x_j$, for $i, j = 1, \cdots, m$, $i \neq j$. Suppose the net $\{a_{h,k}\}_{k=0}^{s^h}$ related to T^h is known or estimates based on S, say $\{a^m_{h,k}\}_{k=0}^{s^h}$, are available and satisfy, for all $k = 0, \cdots, s^h$, $a^m_{h,k} \to a_{h,k}$, as m tends to infinity. Let D^+, D^- or D, depending on the case, be given as before. For $i = 1, \cdots, s^h$, let

$$S_{h,i} := \{x^{(1)}_{h,i}, \cdots, x^{(p_i)}_{h,i}\} := \{a^m_{h,i-1}, x_{d_i+1}, \cdots, x_{d_{i+1}}, a^m_{h,i}\}, \tag{5.3}$$

$$R_{h,i} := \{y^{(1)}_{h,i}, \cdots, y^{(p_i)}_{h,i}\} := \big\{0, T^h(x_{d_i+1}), \cdots, T^h(x_{d_{i+1}}), 1\big\}, \tag{5.4}$$

and $O_{h,i} := \big\{(y^{(k)}_{h,i}, x^{(k)}_{h,i})\big\}_{k=1}^{p_i}$, where $p_i := d_{i+1} - d_i + 2$. Notice that the elements on $S_{h,i}$ are just the values among $x_1, \cdots, x_m$ that lie on the i-th node of T^h and that $R_{h,i}$ is the image of $S_{h,i}$ by T^h. Suppose that a sequence of interpolation functions $h_{i,m} : I \to [a^m_{h,i-1}, a^m_{h,i}]$, $i = 1, \cdots, s^h$, is given. Given $y \in [0, 1] \setminus \{T^h(x_1), \cdots, T^h(x_m)\}$, there exists $y^{(k)}_{h,i} \in R_{h,i}$ such

that $y \in (y_{h,i}^{(k)}, y_{h,i}^{(k+1)})$. We define an approximation to $\mathcal{T}_{h,k}(y)$ based on S, which we denote by $\mathcal{T}_{h,i}^m(y)$, simply by

$$\mathcal{T}_{h,i}^m(y) := h_{i,m}(y; O_{h,i}), \tag{5.5}$$

for $i = 1, \cdots, s^h$.

Remark 5.1. The estimates $a_{h,k}^m$ and $\mathcal{T}_{h,k}^m$ are both assumed to be based on S. This is convenient because it allows the following updating scheme to determine both approximations: given a refinement of S, first we update the net, and secondly, the net new values are used to determine the approximation to $\mathcal{T}_{h,k}$. This type of scheme can be applied very often in practice. The proofs of the results in this section, however, assume solely that $a_{h,k}^m \to a_{h,k}$, as m goes to infinity. The inverse scheme (update $\mathcal{T}_{h,k}^m$ and then update $a_{h,k}^m$) also works, because $a_{h,k} = \mathcal{T}_{h,k}(1)$, for $k = 1, \cdots, s^h$, but it is computationally more delicate to implement than the former. This happens because in updating $\mathcal{T}_{h,k}^m$ to, say, $\mathcal{T}_{h,k}^{m+1}$, one uses $a_{h,k}^m$. Next, to obtain $a_{h,k}^{m+1}$, $T_{h,k}^{m+1}$ is applied which, by its turn, is based on the old value $a_{h,k}^m$. To account for this discrepancy, a posterior adjustment has to be made, but we shall not get into details here.

Proposition 5.2. *Let $T \in \mathcal{T}^{\updownarrow}$ and suppose T has $s > 1$ branches. Let $S = \{x_1, x_2, \cdots, x_m\}$, be an ordered set of points in I and assume that $x_1 = 0$, $x_m = 1$ and $x_i \neq x_j$, for $i, j = 1, \cdots, m$, $i \neq j$. For $h > 0$, let $\{a_{h,j}^m\}_{j=0}^{s^h}$ be approximations to $\{a_{h,j}\}_{j=0}^{s^h}$ based on S such that $a_{h,j}^m \to a_{h,j}$, for all $j = 0, \cdots, s^h$, as $m \to \infty$. Let $S_{h,i}$ and $R_{h,i}$ be as in (5.3) and (5.4), respectively. For a sequence of interpolation functions $\{h_{i,m}\}_{i=1}^{s^h}$ based on S, let $\mathcal{T}_{h,i}^m$ be as in (5.5). If, for all $y \in (y_{h,i}^{(k)}, y_{h,i}^{(k+1)})$, $\mathcal{T}_{h,i}^m(y) \in (x_{h,i}^{(k)}, x_{h,i}^{(k+1)})$, for $i = 1, \cdots, s^h$ and $k \in \{1, \cdots, p_i - 1\}$, then $\mathcal{T}_{h,i}^m \longrightarrow \mathcal{T}_{h,i}$ uniformly, as $m \to \infty$.*

Proof: The uniform continuity of $\mathcal{T}_{h,i}$ implies that, for any given $\varepsilon > 0$, there exists a $\delta > 0$ depending only on ε, such that

$$|x - y| < \delta \implies \left|\mathcal{T}_{h,i}(x) - \mathcal{T}_{h,i}(y)\right| < \varepsilon, \tag{5.6}$$

for all $x \in I$. Let $S := \{x_1, \cdots, x_{m_0}\}$ be an ordered set of points in I and assume that $x_1 = 0$, $x_{m_0} = 1$ and $x_i \neq x_j$, $i, j = 1, \cdots, m_0$, $i \neq j$. Without loss of generality, assume that $m_0 > 0$ is such that

$$\sup_{i=1,\cdots,m_0-1} \left\{|x_{i+1} - x_i|\right\} < \delta. \tag{5.7}$$

For $m > m_0$, let $S_m = \{x_1^*, \cdots, x_m^*\} \supset S$ be a size m refinement of S. By hypothesis, given $y \in (0,1)$, for each $i \in \{1, \cdots, s^h\}$, there exists a $k \in \{1, \cdots, p_i - 1\}$ such that $\mathcal{T}_{h,i}^m(y) \in (y_{h,i}^{(k)}, y_{h,i}^{(k+1)}) = \big(\mathcal{T}_{h,i}(x_{h,i}^{(k)}), \mathcal{T}_{h,i}(x_{h,i}^{(k+1)})\big)$. Also, since $T \in \mathcal{T}^{\updownarrow}$, $\mathcal{T}_{h,i}\big(x_{h,i}^{(k)}\big) \leq \mathcal{T}_{h,i}(y) < \mathcal{T}_{h,i}\big(x_{h,i}^{(k+1)}\big)$, so that, by (5.6) and (5.7),

$$\begin{aligned} \left|\mathcal{T}_{h,i}^m(y) - \mathcal{T}_{h,i}(y)\right| &\leq \left|\mathcal{T}_{h,i}\big(x_{h,i}^{(k+1)}\big) - \mathcal{T}_{h,i}\big(x_{h,i}^{(k)}\big)\right| \\ &\leq \sup_{j=1,\cdots,m-1} \left\{\left|\mathcal{T}_{h,i}(x_{j+1}) - \mathcal{T}_{h,i}(x_j)\right|\right\} < \varepsilon, \end{aligned}$$

independently of $y \in (0,1)$. For $y \in \{0,1\}$, by definition $\mathcal{T}_{h,i}^m(y) = \mathcal{T}_{h,i}(y)$, so that the result holds uniformly for all $y \in I$. ∎

5.2 Approximating the copulas

In this subsection we present a general approximation for the copulas in Section 3 and prove its uniform convergence to the true copula. As mentioned before, the formulas for the copulas presented in Remark 3.5 are the ones used for approximation purposes. Although we could

proceed by using the approximations developed in the last subsection, our convergence results hold in a more general setting. Let $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ and μ_T be a T-invariant probability measure. Let $\varphi \in \mathcal{L}^1(\mu_T)$ and $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_{\varphi}^{\updownarrow}$ process. Let μ_n be an approximation to μ_T and let F_n and F_n^{-1} be approximations to F_0 and F_0^{-1}, respectively. Suppose T has $s > 1$ nodes and, for $h > 0$, let $\{a_{h,i}^m\}_{i=0}^{s^h}$ be an approximation to the net related to T^h. Also let $\mathcal{T}_{h,i}^m$ be an approximation to $\mathcal{T}_{h,i}$, $i = 1, \cdots, s^h$. Define, for all $u, v \in I$,

$$\begin{aligned} C_{m,n}^{\updownarrow}(u,v;h) &= \sum_{k\in n_0^{\uparrow}} \mu_n\Big(\big[a_{h,k-1}^m, \mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big)\big]\Big) + \sum_{k\in n_0^{\downarrow}} \mu_n\Big(\big[\mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big), a_{h,k}^m\big]\Big) + \\ &\quad + \mu_n\Big(\big[a_{h,\widehat{n}_0-1}^m, \min\big\{F_n^{-1}(u), \mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big)\big\}\big]\Big)\delta_{K_h^{\uparrow}}(\widehat{n}_0) + \\ &\quad + \mu_n\Big(\big[\mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big), F_n^{-1}(u)\big]^+\Big)\delta_{K_h^{\downarrow}}(\widehat{n}_0), \end{aligned} \tag{5.8}$$

where $\widehat{n}_0 = n_0(m,n) = \big\{k : u \in \big[F_n(a_{h,k}^m), F_n(a_{h,k+1}^m)\big)\big\}$. Expression (5.8) can be used as an approximation to $C_{X_t,X_{t+h}}^{\updownarrow}$ in (3.6), for $t \geq 0$. In the next theorem, we show that, under certain simple conditions on the approximations in (5.8), $C_{m,n}^{\updownarrow}(u,v) \to C_{X_t,X_{t+h}}^{\updownarrow}(u,v)$ uniformly in $(u,v) \in I^2$ as m and n go to infinity.

Theorem 5.1. *Let $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ and μ_T be a T-invariant probability measure. Let $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function and let $\{X_t\}_{t\in\mathbb{N}}$ be the $\mathcal{T}_{\varphi}^{\updownarrow}$ associated process. Let μ_n be a sequence of measures converging weakly to μ_T and let F_n and F_n^{-1} be approximations to F_0 and F_0^{-1}, respectively, such that $F_n \to F_0$ and $F_n^{-1} \to F_0^{-1}$ uniformly. Suppose T has $s > 1$ nodes and, for $h > 1$, let $\{a_{h,i}^m\}_{i=0}^{s^h}$ be an approximation to the net related to T^h such that $a_{h,i}^m \to a_{h,i}$, for all $i = 0, \cdots, s^h$, as m goes to infinity. Also let $\{\mathcal{T}_{h,i}^m\}_{i=1}^{s^h}$ be an approximation to $\{\mathcal{T}_{h,i}\}_{i=1}^{s^h}$ and suppose $\mathcal{T}_{h,i}^m \to \mathcal{T}_{h,i}$ uniformly for all $i = 1, \cdots, s^h$. Let $C_{m,n}^{\updownarrow}$ be given by (5.8) with the approximations just defined. Then, for all $(u,v) \in I^2$,*

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\updownarrow}(u,v;h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\updownarrow}(u,v;h) = \lim_{m,n\to\infty} C_{m,n}^{\updownarrow}(u,v;h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\updownarrow}(u,v)$ given by (3.6), for all $t > 0$. Furthermore, the limits above are uniform in (u,v).

Proof: The hypothesis on the approximations in expression (5.8), allow us to apply Lemma 5.1 in Lopes and Pumi (2011) to conclude that, for any k,

$$\lim_{m,n\to\infty} \mu_n\Big(\big[a_{h,k-1}^m, \mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big)\big]\Big) = \mu_T\Big(\big[a_{h,k-1}, \mathcal{T}_{h,k}\big(F_0^{-1}(v)\big)\big]\Big),$$

and, by the same argument,

$$\lim_{m,n\to\infty} \mu_n\Big(\big[\mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big), a_{h,k}^m\big]\Big) = \mu_T\Big(\big[\mathcal{T}_{h,k}\big(F_0^{-1}(v)\big), a_{h,k}\big]\Big).$$

From the same lemma we also conclude that the iterated limits exist and are equal to the double ones above and all limits are uniform in v. By using the same argument as in the proof of Theorem 5.1 in Lopes and Pumi (2011), we conclude that

$$\lim_{m,n\to\infty} \mu_n\Big(\big[a_{h,\widehat{n}_0-1}^m, \min\big\{F_n^{-1}(u), \mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big)\big\}\big]\Big) = \mu_T\Big(\big[a_{h,n_0-1}, \min\big\{F_0^{-1}(u), \mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big)\big\}\big]\Big),$$

and that the iterated limits are equal to the double one above. Furthermore, the limits are all uniform in $u, v \in I$. So, it remains to show that

$$\lim_{m,n\to\infty} \mu_n\Big(\big[\mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big), F_n^{-1}(u)\big]^+\Big) = \mu_T\Big(\big[\mathcal{T}_{h,n_0}\big(F_0^{-1}(v)\big), F_0^{-1}(u)\big]^+\Big),$$

uniformly and that the iterated limits are equal to the double one. First, the existence and the uniform convergence of the iterated limits follow by the same argument in the proof of Lemma 5.1 in Lopes and Pumi (2011). As for the double limit, by the hypothesis on μ_T, the Radon-Nikodym theorem applies and implies the existence of a non-negative bounded (since we are restricted to I) continuous function h such that, for any $A \in \mathcal{B}(I)$,

$$\mu_T(A) = \int_A h(x)\mathrm{d}\lambda \leq M\lambda(A),$$

where $M = \sup_{x\in I}\{h(x)\} < \infty$. Now, the uniform convergence of F_n^{-1} to F_0^{-1} implies the existence of $n_0 > 0$ such that, if $n > n_0$, $F_n^{-1}(u) \in K(\varepsilon) := [F_0^{-1}(u) - \varepsilon/10M, F_0^{-1}(u) + \varepsilon/10M]$ and

$$\mu_T\big(K(\varepsilon)\big) \leq M\lambda\big(K(\varepsilon)\big) = \frac{\varepsilon}{5},$$

for all $u \in I$. The rest of the proof is carried out by mimicking the proof of Lemma 5.1 in Lopes and Pumi (2011) upon substituting $K_1(\varepsilon)$ there by $K(\varepsilon)$ just defined and noticing that n_0 here plays the role of m_1 there. This completes the proof of the theorem. ■

From Theorem 5.1, we derive approximations to the copulas in the case $T \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$ given in (3.8) and (3.9).

Corollary 5.1. *Let $T \in \mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow}$ and μ_T be a T-invariant probability measure. Let $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function and let $\{X_t\}_{t=0}^{\infty}$ be the $\mathcal{T}_{\varphi}^{\downarrow}$ or $\mathcal{T}_{\varphi}^{\uparrow}$ associated process. Let $\{\mu_n\}_{n=1}^{\infty}$ be a sequence of measures converging weakly to μ_T and let F_n and F_n^{-1} be approximations to F_0 and F_0^{-1}, respectively, such that $F_n \to F_0$ and $F_n^{-1} \to F_0^{-1}$ uniformly. Suppose T has $s > 1$ nodes and, for $h > 1$, let $\{a_{h,i}^m\}_{i=0}^{s^h}$ be an approximation to the net related to T^h such that $a_{h,i}^m \to a_{h,i}$, for all $i = 0, \cdots, s^h$, as m goes to infinity. Also let $\{\mathcal{T}_{h,i}^m\}_{i=1}^{s^h}$ be an approximation to $\{\mathcal{T}_{h,i}\}_{i=1}^{s^h}$ and suppose $\mathcal{T}_{h,i}^m \to \mathcal{T}_{h,i}$ uniformly for all $i = 1, \cdots, s^h$.*

(i). *If $T \in \mathcal{T}^{\uparrow}$, then, for all $(u, v) \in I^2$,*

$$\begin{aligned} C_{m,n}^{\uparrow}(u,v;h) \;=\; & \sum_{k=1}^{\widehat{n}_0-1} \mu_n\Big(\big[a_{h,k-1}^m, \mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big)\big]\Big)\delta_{\mathbb{N}^*}(\widehat{n}_0 - 1) + \\ & + \mu_n\Big(\big[a_{h,\widehat{n}_0-1}^m, \min\big\{F_n^{-1}(u), \mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big)\big\}\big]\Big) \end{aligned}$$

is an approximation to the copula (3.8)*,*

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\uparrow}(u,v;h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\uparrow}(u,v;h) = \lim_{m,n\to\infty} C_{m,n}^{\uparrow}(u,v;h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\uparrow}(u,v)$ given by (3.8)*.*

(ii). *If $T \in \mathcal{T}^{\downarrow}$, then, for all $(u, v) \in I^2$,*

$$C_{m,n}^{\downarrow}(u,v;h) = \sum_{k=1}^{\widehat{n}_0-1} \mu_n\Big(\big[\mathcal{T}_{h,k}^m\big(F_n^{-1}(v)\big), a_{h,k}^m\big]\Big)\delta_{\mathbb{N}^*}(\widehat{n}_0 - 1) + \mu_n\Big(\big[\mathcal{T}_{h,\widehat{n}_0}^m\big(F_n^{-1}(v)\big), F_n^{-1}(u)\big]^+\Big)$$

is an approximation to the copula (3.9)*,*

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\downarrow}(u,v;h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\downarrow}(u,v;h) = \lim_{m,n\to\infty} C_{m,n}^{\downarrow}(u,v;h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\downarrow}(u,v)$ given by (3.9)*. Furthermore, in all cases the limits are uniform in (u, v).*

Proof: If $T \in \mathcal{T}^{\uparrow}$, (i) follows from Theorem 5.1 by noticing that $K_h^{\downarrow} = \emptyset$, while if $T \in \mathcal{T}^{\downarrow}$, the opposite happens, namely, $K_h^{\uparrow} = \emptyset$, which implies (ii). ■

As an application of Theorem 5.1 and Lemma 3.4, we obtain the approximations for the copulas in the case where φ is an almost everywhere decreasing function.

Corollary 5.2. *Let $T \in \mathcal{T}^{\updownarrow}$ and μ_T be a T-invariant probability measure. Let $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere decreasing function and let $\{X_t\}_{t=0}^{\infty}$ be the $\mathcal{T}_{\varphi}^{\updownarrow}$ associated process. Let μ_n be a sequence of measures converging weakly to μ_T and let F_n and F_n^{-1} be approximations to F_0 and F_0^{-1}, respectively, such that $F_n \to F_0$ and $F_n^{-1} \to F_0^{-1}$ uniformly. Suppose T has $s > 1$ nodes and, for $h > 1$, let $\{a_{h,i}^m\}_{i=0}^{s^h}$ be an approximation to the net related to T^h such that $a_{h,i}^m \to a_{h,i}$, for all $i = 0, \cdots, s^h$, as m goes to infinity. Also let $\{\mathcal{T}_{h,i}^m\}_{i=1}^{s^h}$ be an approximation to $\{\mathcal{T}_{h,i}\}_{i=1}^{s^h}$ and suppose $\mathcal{T}_{h,i}^m \to \mathcal{T}_{h,i}$ uniformly for all $i = 1, \cdots, s^h$. For all $(u, v) \in I^2$, let*

$$C_{m,n}^{\updownarrow *}(u, v; h) = u + v - 1 + C_{m,n}^{\updownarrow}(1 - u, 1 - v; h),$$

$$C_{m,n}^{\uparrow *}(u, v; h) = u + v - 1 + C_{m,n}^{\uparrow}(1 - u, 1 - v; h),$$

and

$$C_{m,n}^{\downarrow *}(u, v; h) = u + v - 1 + C_{m,n}^{\downarrow}(1 - u, 1 - v; h).$$

Then

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\updownarrow *}(u, v; h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\updownarrow *}(u, v; h) = \lim_{m,n\to\infty} C_{m,n}^{\updownarrow *}(u, v; h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\updownarrow *}(u, v)$ given by (3.12),

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\uparrow *}(u, v; h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\uparrow *}(u, v; h) = \lim_{m,n\to\infty} C_{m,n}^{\uparrow *}(u, v; h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\uparrow *}(u, v)$ given by (3.14),

$$\lim_{n\to\infty}\lim_{m\to\infty} C_{m,n}^{\downarrow *}(u, v; h) = \lim_{m\to\infty}\lim_{n\to\infty} C_{m,n}^{\downarrow *}(u, v; h) = \lim_{m,n\to\infty} C_{m,n}^{\downarrow *}(u, v; h),$$

and the common limit is $C_{X_t,X_{t+h}}^{\downarrow *}(u, v)$ given by (3.15). *Furthermore, all the above limits are uniform in $(u, v) \in I^2$.*

Proof: Immediate application of Theorem 5.1 and Lemma 3.4. ■

Remark 5.2. The uniform convergence of the approximations is a crucial hypothesis for the results in this subsection and cannot be dropped. Nevertheless, if the uniform convergence of any of the approximations for F_0, F_0^{-1} or $\mathcal{T}_{h,i}$ is violated, but the pointwise convergence to its target is maintained, the results of Theorem 5.1 and its corollaries will hold pointwisely instead of uniformly. The proof of Theorem 5.1 can be easily adapted to cover this case and the details are left to the reader. We observe that the approximations developed in Subsection 5.1 satisfy the conditions of the theorems presented in this section.

5.3 Random Variate Generation

In view of Proposition 3.6, obtaining a random sample from the copulas derived in Section 3 is a trivial task. Given $T \in \mathcal{T}^{\updownarrow}$ a transformation with $s > 1$ nodes, let μ_T be a T-invariant probability measure and $\varphi \in \mathcal{L}^1(\mu_T)$ be an almost everywhere increasing function. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_{\varphi}^{\updownarrow}$ process and let $\{a_{h,k}\}_{k=0}^{s^h}$ be the net associated to the nodes of T^h, for $h > 0$. Let $C_{X_t,X_{t+h}}^{\updownarrow}$ denote the copula associated to $\{X_t\}_{t\in\mathbb{N}}$. To obtain a random pair from $C_{X_t,X_{t+h}}^{\updownarrow}$, the following simple algorithm can be used:

1. Generate an uniform $(0,1)$ variate u.
2. Let k_0 be the index for which $u \in \big[F_0(a_{h,k_0-1}), F_0(a_{h,k_0})\big)$. Set

$$v := \begin{cases} \ell^{\uparrow}_{h,k_0}(u), & \text{if } k_0 \in K_h^{\uparrow}; \\ \ell^{\downarrow}_{h,k_0}(u), & \text{if } k_0 \in K_h^{\downarrow}. \end{cases}$$

3. The desired pair is (u, v).

Also notice that given two transformations $T_1, T_2 \in \mathcal{T}^{\updownarrow}$ such that both have the same discontinuity points and, in each node, they are either both increasing or both decreasing, then a sample from the copula related to one of the respective $\mathcal{T}^{\uparrow}_{\varphi}$ processes cannot be distinguished from a sample to the other. This is so because from Proposition 3.6 both copulas have the same support, which, in this case, ultimately characterizes the sample.

Remark 5.3. By using a similar argument as in Proposition 3.6 associated with a cumbersome analysis of the possible cases, the support of the copulas derived in Section 5 can be obtained as well. Samples from those multidimensional copulas can be obtained in a similar fashion as in the bidimensional case.

6 Examples

In this section we present two examples of the copulas derived in the last sections. We concentrate ourselves on the bidimensional case, especially because there is no simple graphical representation of copulas in dimension higher than 2. The first presented example is the Manneville-Pomeau copula (MP copula for short) extracted from Lopes and Pumi (2011).

Example 6.1. For $s > 0$, consider the so-called Manneville Pomeau transformation $T_s : I \to I$, given by

$$T_s(x) = x + x^{1+s} (\mathrm{mod}\, 1).$$

Figure 1 shows the plot of the Manneville-Pomeau transformation for $s \in \{0.1, 0.5, 0.8\}$. If $s \in (0,1)$, there exists an absolutely continuous T_s-invariant probability measure, say μ_s, so that $T_s \in \mathcal{T}^{\uparrow}$. The MP copula is, therefore, given by (3.8), although a slightly different notation for n_0 is adopted in Lopes and Pumi (2011). No closed formula for μ_s is known, but it can be shown that μ_s is an SBR measure.

To approximate the MP copulas, the authors apply similar ideas to the ones presented here. To approximate μ_s, the authors use (5.2) and to approximate F_0^{-1}, a local linear interpolation function is applied. To approximate the net related to T_s^h and $\mathcal{T}_{h,k}$, the same scheme presented here, with a local linear interpolation function, is considered. For more details, see Lopes and Pumi (2011).

The next example is related to the well known Tent transformation (see Lopes and Lopes, 1998 and references therein).

Example 6.2. For $a \in (0,1)$, consider the piecewise linear transformation $T_a : I \to I$ given by

$$T_a(x) := \begin{cases} \dfrac{x}{a}, & \text{if } 0 \le x < a, \\ \dfrac{1-x}{1-a}, & \text{if } a \le x \le 1. \end{cases}$$

This is the so-called *Tent transformation*. Figure 6.1 presents the graphs of the Tent transformation for $a \in \{0.1, 0.5, 0.8\}$, where it can be seen that T_a has $s = 2$ full branches. Also, $T_a \in \mathcal{T}^{\updownarrow} \setminus (\mathcal{T}^{\uparrow} \bigcup \mathcal{T}^{\downarrow})$, since it can be shown that the Lebesgue measure in I^2 is a T_a-invariant

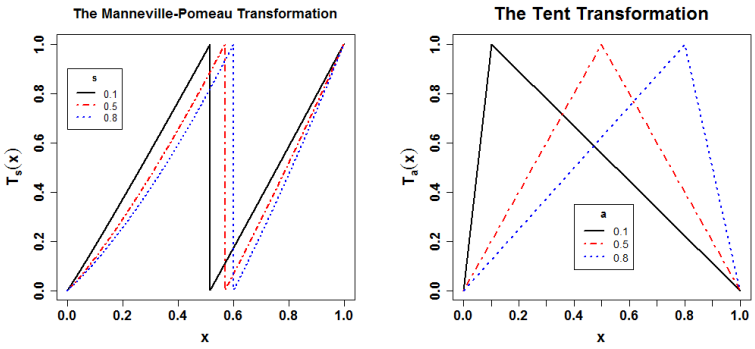


Figure 6.1: Plot of the Manneville-Pomeau transformation for $s \in \{0.1, 0.5, 0.8\}$ and the Tent transformation for $a \in \{0.1, 0.5, 0.8\}$.

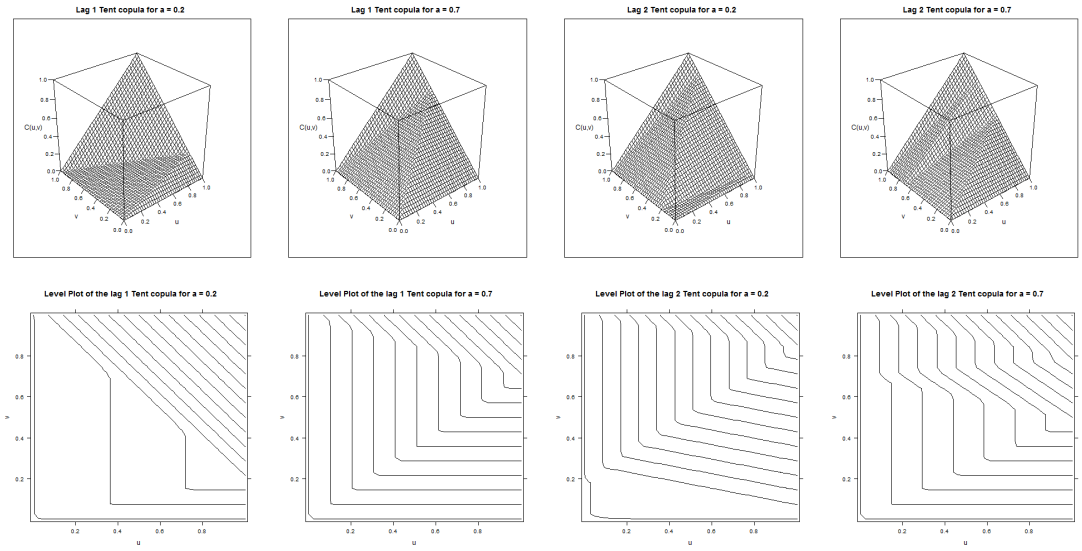


Figure 6.2: From left to right, three dimensional plots of the lag 1 and lag 2 Tent copula for $a \in \{0.2, 0.7\}$ (top panel) and respective level curves (bottom panel).

probability measure. This implies that $F_0(x) = x$ and $F_0^{-1}(y) = y$ everywhere. Let $\{X_t\}_{t\in\mathbb{N}}$ be the $\mathcal{T}_\varphi^\updownarrow$ process associated to T_a, with φ increasing almost everywhere. It is easy to see that, for any $h > 0$ and $t \geq 0$, the copula related to the pair (X_t, X_{t+h}), referred here as the *lag h Tent copula*, is given by

$$\begin{aligned} C_{X_t,X_{t+h}}(u,v) &= \sum_{k\in n_0^\uparrow} \left(\mathcal{T}_{h,k}(v) - a_{h,k-1}\right) + \left[\min\left\{u, \mathcal{T}_{h,n_0}(v)\right\} - a_{h,n_0-1}\right] \delta_{K_h^\uparrow}(n_0) + \\ &\quad + \sum_{k\in n_0^\downarrow} \left(a_{h,k} - \mathcal{T}_{h,k}(v)\right) + \max\left\{0, u - \mathcal{T}_{h,n_0}(v)\right\} \delta_{K_h^\downarrow}(n_0), \end{aligned}$$

where $K_h^\uparrow$ contains the odd numbers in $\{1, \cdots, 2^h\}$ and $K_h^\downarrow$ contains the even ones. Also notice that both, the net related to T^h and $\mathcal{T}_{h,k}$, can be determined exactly since, being each branch a linear function, any two points in the node suffice to give all information on the function to determine $\mathcal{T}_{h,k}$ and $a_{h,k}$. In this context, linear interpolation produces exact, instead of approximate, results.

Figure 6.2 shows the tree dimensional graphs of the lag 1 and 2 Tent copula for $a \in \{0.2, 0.7\}$ and its level curves. At the top panel of Figure 6.3, we present 500 sample points for the lag 1 and 2 Tent copula for $a \in \{0.2, 0.7\}$ and at the bottom panel, for the lag 7 Tent copula for $a \in \{0.2, 0.4, 0.5, 0.7\}$. Notice that, for small values of h, the sample resembles its support, but

as h increases, it becomes hard to guess whether the sample came from a singular copula or from a continuous one.

The case $h = 1$ is very simple and the Tent copula can be easily calculated. In this case, the copula coincides with the one presented in Example 3.3 in Nelsen (2006). In the aforementioned example, the copula is derived by using a purely geometrical argument based only on the support, which coincides with the support of the lag 1 Tent copula. More details on the Tent transformation can also be found in Lopes and Lopes (1998) and references therein.

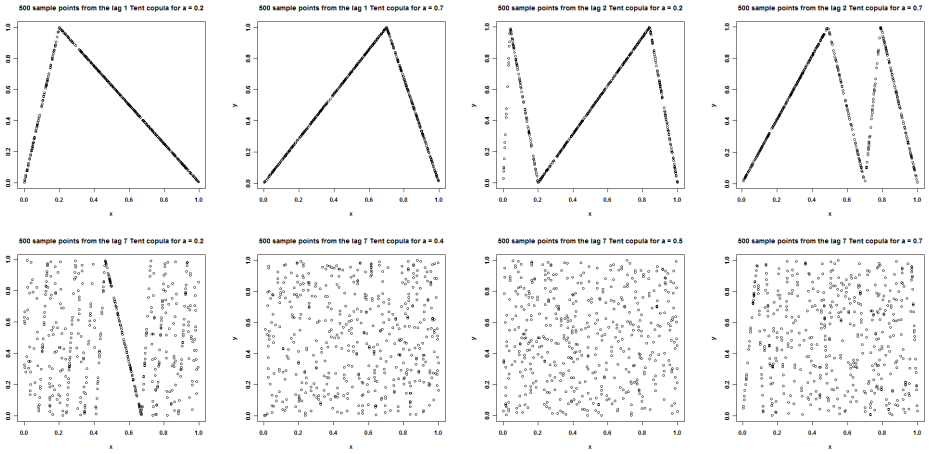


Figure 6.3: 500 sample points from the lag 1 and 2 Tent copula for $a \in \{0.2, 0.7\}$ (top panel) and the lag 7 Tent copula for $a \in \{0.2, 0.4, 0.5, 0.7\}$ (bottom panel).

7 Application

Let $T_{\boldsymbol{\theta}} \in \mathcal{T}^{\updownarrow}$, for $\boldsymbol{\theta} \in S \subseteq \mathbb{R}^p$, $p \geq 1$, with $s > 1$ nodes. In this section we apply the general theory of Section 3 to the problem of estimating the parameter $\boldsymbol{\theta}$ based on a sample path of the associated $\mathcal{T}_{\varphi}^{\updownarrow}$ process $\{X_t\}_{t\in\mathbb{N}}$, assuming that the parameter is identifiable through the knowledge of the discontinuity points of $T_{\boldsymbol{\theta}}$. The idea is to generalize the method described in Lopes and Pumi (2011), but the task will heavily depend on the transformation at hand. The general idea is as follows. According to Proposition 3.6, the support of the copula associated to a pair (X_t, X_{t+1}) from the $\mathcal{T}_{\varphi}^{\updownarrow}$ process is the graph of a piecewise linear function joining consecutive points of the net and its image by $T_{\boldsymbol{\theta}}$ (equal to either 0 or 1). Let $L_{\boldsymbol{\theta}}$ denote this function. This implies that all points in a sample from the lag 1 copula lie on the graph of $L_{\boldsymbol{\theta}}$. Suppose for the moment that φ is the identity map and let $x_1, \cdots, x_N$ be a sample from the process X_t. Let $u_k := \big(F_0(x_k), F_0(x_{k+1})\big)$, $k = 1, \cdots, N-1$, where F_0 denotes the distribution of X_0 (for now, assume it is known). By Sklar's theorem, the sequence $\{u_k\}_{k=0}^{N-1}$ can be regarded as correlated sample from the lag 1 copula associated to the $\mathcal{T}_{\varphi}^{\updownarrow}$ process. Often this simple situation makes an estimation possible.

Suppose that closed formulas for $F_0(\cdot;\boldsymbol{\theta})$ and $L_{\boldsymbol{\theta}}$, which may depend on $\boldsymbol{\theta}$, are available. Then the reasoning on the previous paragraph suggests the following optimization procedure to obtain an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Let $x_1, \cdots, x_N$ be a sample from the $\mathcal{T}_{\varphi}^{\updownarrow}$ process and assume that we have at least two points in each branch of $T_{\boldsymbol{\theta}}$. Let $\mathscr{D}(\cdot,\cdot) : \mathbb{R}^{N-1} \times \mathbb{R}^{N-1} \to \mathbb{R}$ be a given function measuring the distance between two vectors in $\mathbb{R}^{N-1}$. With the notation on the beginning of Section 4, we define the estimate of $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in S}{\operatorname{argmin}} \Big\{ \mathscr{D}\big(F_0(x_{2:N};\boldsymbol{\theta}), L_{\boldsymbol{\theta}}(x_{1:N-1})\big) \Big\}. \tag{7.9}$$

Usual choices of $\mathscr{D}$ are $\mathscr{D}(\boldsymbol{x},\boldsymbol{y}) = \sum_{k=1}^{N-1}(x_k - y_k)^2$ and $\mathscr{D}(\boldsymbol{x},\boldsymbol{y}) = \sum_{k=1}^{N-1}|x_k - y_k|$, where $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{N-1}$. The optimization procedure (7.9) is justified by noticing that for the correct $\boldsymbol{\theta}$, $\big(F_0(x_k;\boldsymbol{\theta}), F_0(x_{k+1};\boldsymbol{\theta})\big) = \big(F_0(x_k;\boldsymbol{\theta}), L_{\boldsymbol{\theta}}\big(F_0(x_{k+1};\boldsymbol{\theta})\big)\big)$, for all k, while for a misspecified $\boldsymbol{\theta}$, the equality does not hold. Most of the times, however, the particular form of F_0 is not known and often the net cannot be obtained analytically. This is the case of the Manneville-Pomeau transformation, for instance. A solution for this problem is to use approximations for F_0 and $\{a_k\}_{k=0}^s$ in the spirit of Subsection 5.1. However, an optimization procedure such as (7.9) performed by using the approximations of Subsection 5.1 may be computationally too expensive for all practical purposes, so we shall seek for another faster estimation method.

In practice many different situations may occur. For instance, if T_a is the Tent transformation in Example 6.2, then F_0 is the identity map, so that $u_k = (x_k, x_{k+1})$ in the notation of the previous paragraph. This means that any two points u_k in the same branch is sufficient to obtain the parameter a.

Another method can be devised upon noticing that the net of $T_{\boldsymbol{\theta}}$ and the net of $L_{\boldsymbol{\theta}}$ (whose graph is the support of the copula C associated to the $\mathcal{T}_\varphi^\updownarrow$ process) are the same. Let $v_k := (x_k, x_{k+1})$, $k = 1,\cdots,N-1$. By Sklar's theorem, the points $\{v_k\}_{k=1}^{N-1}$ lie on the graph of $T_{\boldsymbol{\theta}}$. At this point, it could be possible in some cases to obtain $\boldsymbol{\theta}$ by a more direct argument (see Example 7.1) or by applying an argument such as the preceding paragraphs. That is, for $\mathscr{D}$ a function as above, $\boldsymbol{\theta}$ can be obtained by performing the optimization procedure

$$\hat{\boldsymbol{\theta}} := \operatorname*{argmin}_{\boldsymbol{\theta}\in S}\Big\{\mathscr{D}\big(x_{2:N}, T_{\boldsymbol{\theta}}(x_{1:N-1})\big)\Big\}.$$

The methods above may not be applicable or may fail, especially for high values of s. In this case another computationally fast method is obtained by adapting the argument in Lopes and Pumi (2011). Let $T_{\boldsymbol{\theta}} \in \mathcal{T}^\updownarrow$, with $s > 1$ branches and $\boldsymbol{\theta} := (\theta_1,\cdots,\theta_p) \in S \subseteq \mathbb{R}^p$, for $1 \le p \le s-2$. Let $\{a_k\}_{k=0}^s$ denote the net associated to $T_{\boldsymbol{\theta}}$. Suppose that $\boldsymbol{\theta}$ can be uniquely determined by the knowledge of the net $\{a_k\}_{k=0}^s$ in the sense that, given $\{a_k\}_{k=0}^s$, $\boldsymbol{\theta}$ can be computed. That is, $\boldsymbol{\theta} := f(a_0,\cdots,a_s)$, where $f : I^{s+1} \to S$ is known. Notice that we allow for θ_i to be obtainable from two or more different a_k's as long as the value of θ_i agree in all cases. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_\varphi^\updownarrow$ process, assuming that φ is the identity map. The goal is to estimate the parameter $\boldsymbol{\theta}$ based on a sample $x_1,\cdots,x_N$ from the process X_t.

Given that $T_{\boldsymbol{\theta}} \in \mathcal{T}^\updownarrow$, F_0 is generally smooth so that, near $\{a_k\}_{k=0}^s$, $T_{\boldsymbol{\theta}}$ should locally behave like a linear function. With this in mind, the estimation of $\boldsymbol{\theta}$ can be performed as follows. Let $\{I_k\}_{k=1}^s$ be the nodes relative to $T_{\boldsymbol{\theta}}$. Let $P_k := \{x_{m+1} : x_m \in I_k, m = 1,\cdots,N-1\}$, $k = 1,\cdots,s$. Notice that, if $x_m \in P_k$, v_m lies on the graph of the k-th branch of $T_{\boldsymbol{\theta}}$. Set $x_k^+ := \max\{x : x \in P_k\}$ and $x_k^- := \min\{x : x \in P_k\}$. Let

$$\tilde{a}_k(y) := \frac{y-\beta_k}{\alpha_k}, \quad \text{where } \alpha_k := \frac{x_k^+ - x_k^-}{T_{\boldsymbol{\theta}}(x_k^+) - T_{\boldsymbol{\theta}}(x_k^-)} \quad \text{and} \quad \beta_k := T_{\boldsymbol{\theta}}(x_k^+) - \alpha_k x_k^+.$$

Notice that for each $y \in \mathbb{R}$, $\tilde{a}_k(y)$ is the inverse image of y by the linear function connecting $\big(x_k^-, T_{\boldsymbol{\theta}}(x_k^-)\big)$ and $\big(x_k^+, T_{\boldsymbol{\theta}}(x_k^+)\big)$. For $k \in \{1,\cdots,s-1\}$, we define the estimator $\hat{a}_k$ of a_k by

$$\hat{a}_k := \begin{cases} \tilde{a}_k(1), & \text{if } T_{\boldsymbol{\theta}} \text{ is increasing in } I_k, \\ \tilde{a}_k(0), & \text{if } T_{\boldsymbol{\theta}} \text{ is decreasing in } I_k. \end{cases} \tag{7.10}$$

Once we have the estimates $\{\hat{a}_k\}_{k=1}^{s-1}$ (recall that $a_0 = 0$ and $a_s = 1$), we obtain the estimate $\hat{\boldsymbol{\theta}} := f(\hat{a}_0,\cdots,\hat{a}_s)$.

Remark 7.1. Notice that by inverting the roles of $\tilde{a}_k(0)$ and $\tilde{a}_k(1)$ in (7.10) we obtain an estimate of a_{k-1}. That is, for $k \in \{2,\cdots,s\}$, define

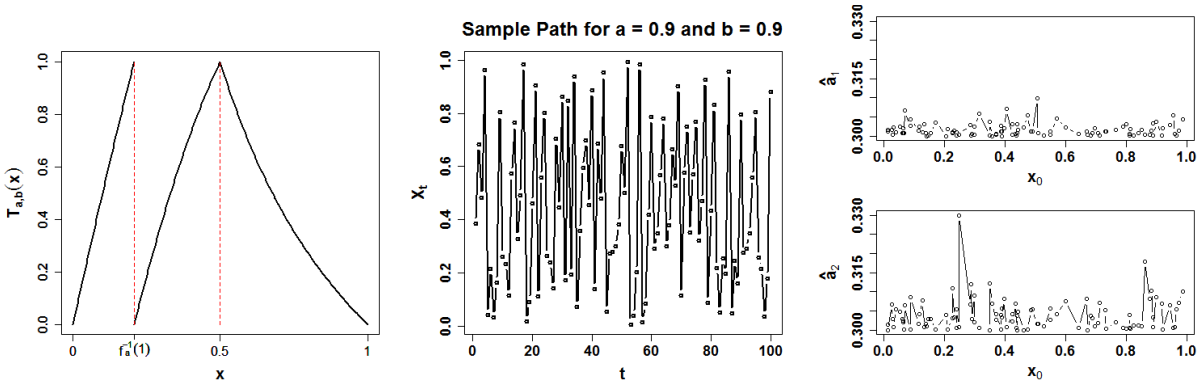
$$\hat{a}_{k-1} := \begin{cases} \tilde{a}_k(0), & \text{if } T_{\boldsymbol{\theta}} \text{ is increasing in } I_k, \\ \tilde{a}_k(1), & \text{if } T_{\boldsymbol{\theta}} \text{ is decreasing in } I_k. \end{cases} \tag{7.11}$$

Thus, for each $k = 2, \cdots, s-1$ we obtain an estimate of a_k from I_k by using (7.10) and another one from I_{k+1} by using (7.11). Of course, the net related to $T_{\boldsymbol{\theta}}$ is unique and we assume that $\boldsymbol{\theta}$ is uniquely determined by the knowledge of the net, but in practice, the estimates (7.10) and (7.11) usually do not agree. This implies that the estimated value of $\boldsymbol{\theta}$ depends on which estimator we apply. Hence, the question which one provides better results is a valid one. The answer is intuitive: the "best" one is usually the one obtained from the node which is closer to a straight line and for which the points v_k's are closer to $(a_k, 0)$ and $(a_{k+1}, 1)$. In this case the line connecting $\left(x_k^-, T_{\boldsymbol{\theta}}(x_k^-)\right)$ and $\left(x_k^+, T_{\boldsymbol{\theta}}(x_k^+)\right)$ is "closer" to the respective branch of the copula's support.

Example 7.1. Let $a \in [0,1]$ and $b \in (0,1)$ and consider the map

$$T_{a,b}(x) := \begin{cases} f_a(x)(\text{mod}1), & \text{if } 0 \leq x < \frac{1}{2}, \\ \frac{(2-b)(1-x)}{1-b+bx}, & \text{if } \frac{1}{2} \leq x \leq 1, \end{cases} \tag{7.12}$$

where $f_a(x) := 2[2 - 0.6a\sin(2.1a)]x + 0.6a\sin(4.2ax)$. The map $T_{a,b}$ has 3 full branches, two of them are increasing and one is decreasing. Figure 7.4(a) presents the typical graph of $T_{a,b}$. It is easy to see that $T_{a,b}$ is uniformly expanding and it can be shown that there exists an absolutely continuous $T_{a,b}$-invariant probability measure (cf. Theorem 1 and Remark 1 in Pianigiani, 1980) and so $T_{a,b} \in \mathcal{T}^{\updownarrow}$. Let $\{X_t\}_{t\in\mathbb{N}}$ be the associated $\mathcal{T}_\varphi^{\updownarrow}$ process. A sample of size 100 from the $\mathcal{T}_\varphi^{\updownarrow}$ process with $a = b = 0.9$, for φ the identity map is shown in Figure 7.4(b).



(a) (b) (c)

Figure 7.4: (a) A typical graph of $T_{a,b}$; (b) A sample of size 100 from the $\mathcal{T}_\varphi^{\updownarrow}$ process with $a = b = 0.9$ for φ the identity map and $X_0 = 0.385969$; (c) Behavior of the estimators $\hat{a}_1$ and $\hat{a}_2$ as a function of the initial point x_0. The parameters are $a = 0.3$ and $b = 0.5$ in all cases.

To exemplify the estimation procedure developed in this section, we perform the following Monte Carlo simulation study. For 100 randomly chosen initial points, we simulate a sample of size 100 from the $\mathcal{T}_\varphi^{\updownarrow}$ process $\{X_t\}_{t\in\mathbb{N}}$ for all combinations of parameters a and b ranging over the set $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. For simplicity, let $\{d_0, \cdots, d_3\}$ denote the net of $T_{a,b}$. In this case $d_0 = 0$, $d_2 = 0.5$ and $d_3 = 1$, so that d_1 is the only discontinuity to be estimated. This estimation is performed by using the two methods (7.10) based on the first branch and (7.11) based on the second branch, the estimators of d_1 being denoted by $\hat{d}_1^{(1)}$ and $\hat{d}_1^{(2)}$, respectively. From $\hat{d}_1^{(1)}$ and $\hat{d}_1^{(2)}$, we set $\hat{a}_i := f_a^{-1}(d_1^{(i)})$, $i = 1, 2$, where the inverse of f_a is obtained numerically. The parameter b needs no estimation since it can be exactly calculated by noticing that for a pair (x_k, x_{k+1}), with $0.5 < x_k < 1$, b is exactly obtained by the formula $b = \frac{2(1-x_k)-x_{k+1}}{(1-x_k)(1-x_{k+1})}$.

The simulation results are reported in Table 7.1, where the mean estimated value along with its mean square error (values in parenthesis $\times 10^{-4}$) are presented. The overall performance of $\hat{a}_1$ and $\hat{a}_2$ are very good with small bias and small variability. As expected (see Remark 7.1

Table 7.1: Simulation results based on samples of size 100 of the $\mathcal{T}_{\varphi}^{\updownarrow}$ process associated to (7.12) for 100 randomly chosen initial points and parameters a and b ranging over the set $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. Presented are the estimated values and the respective mean square error ($\times 10^{-4}$).

b	0.1		0.3		0.5		0.7		0.9	
a	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_2$
0.1	0.1005 (0.0045)	0.1011 (0.0251)	0.1007 (0.0089)	0.1012 (0.0303)	0.1005 (0.0055)	0.1011 (0.0224)	0.1005 (0.0054)	0.1011 (0.0248)	0.1005 (0.0039)	0.1011 (0.0241)
0.3	0.3018 (0.0589)	0.3040 (0.2491)	0.3019 (0.0690)	0.3045 (0.3625)	0.3017 (0.0585)	0.3039 (0.3394)	0.3016 (0.0486)	0.3033 (0.2056)	0.3015 (0.0439)	0.3035 (0.2462)
0.5	0.5033 (0.2031)	0.5064 (0.7755)	0.5027 (0.1332)	0.5072 (1.1065)	0.5024 (0.0924)	0.5054 (0.5878)	0.5027 (0.1215)	0.5054 (0.6567)	0.5030 (0.1528)	0.5064 (0.9126)
0.7	0.7046 (0.3845)	0.7130 (2.9859)	0.7039 (0.2515)	0.7097 (1.8811)	0.7042 (0.3369)	0.7099 (2.2650)	0.7040 (0.2767)	0.7087 (1.3976)	0.7044 (0.3638)	0.7090 (1.5546)
0.9	0.9051 (0.4326)	0.9185 (5.9339)	0.9043 (0.3095)	0.9172 (6.3317)	0.9049 (0.4276)	0.9174 (6.8437)	0.9050 (0.4598)	0.9182 (7.0888)	0.9049 (0.5110)	0.9162 (4.2946)

and Figure 7.4(a)), $\hat{a}_1$ outperforms $\hat{a}_2$ since the first branch of $T_{a,b}$ is "closer" to linear than the second one. Both estimators always overestimate the true parameter. The behavior of the estimators $\hat{a}_1$ and $\hat{a}_2$ as a function of the initial point x_0 for $a = 0.3$ and $b = 0.5$ are presented in Figure 7.4(c). Also notice that the particular value of the parameter b does not significantly affects the estimation of parameter a.

8 Conclusions

In this work we study the copulas related to pairs and vectors of random variables coming from a class of stochastic processes defined in terms of iterations of a certain smooth piecewise monotonic transformation of the interval to some initial random variable. More specifically, we study the copulas related to random variables coming from a stochastic process defined as $X_t = \varphi\big(T^t(U_0)\big)$ for T a piecewise monotonic transformation of the interval, U_0 an initial random variable distributed according to a T-invariant probability measure and a smooth function $\varphi : [0,1] \to \mathbb{R}$, taken to be monotonic. We show that the copulas depend only on the lag h between components and we derived formulas and properties for both, the joint distribution function and the copulas for pairs of random variables coming from the process $\{X_t\}_{t\in\mathbb{N}}$. As expected, the copulas heavily depend on the T-invariant probability measure. The multidimensional case is similar to the bidimensional one so we follow the same agenda as in the latter and we are able to show analogous results. We notice the similarities among the results of Sections 3 and 4 when φ is increasing. For φ decreasing, the multidimensional case poses some difficulties which ultimately lead to more complicated formulas compared to the bidimensional case. We also discuss random variate generation and approximations for the copulas derived on Section 3. The general theory is applied to the problem of parametric estimation in $\mathcal{T}_{\varphi}^{\updownarrow}$ processes. Examples and a simple Monte Carlo study are also provided.

Acknowledgements

G. Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil. S.R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1 and also by INCT *em Matemática*.

References

Bowen, R. (1979). "Invariant Measures for Markov Maps of the Interval". *Communications in Mathematical Physics*, **69**, 1-14.

Chen, X. and Fan, Y. (2006). "Estimation of Copula-Based Semiparametric Time Series Models". *Journal of Econometrics,* **130**, 307-335.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester: Wiley.

Frees, E.W. and Valdez, E.A. (1998). "Understanding Relationships Using Copulas". *North American Actuarial Journal*, **2**, 1-25.

Genest, C. and Favre, A.-C. (2007). "Everything you Always Wanted to Know About Copula Modeling but Were Afraid to Ask". *Journal of Hydrologic Engineering*, **12**, 347-368.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* New York: Chapman & Hall.

Keller, G. (1998). *Equilibrium States in Ergodic Theory*. Cambridge: Cambridge University Press.

Lasota, A. and Mackey, M.C. (1994). *Chaos, Fractals, and Noise. Stochastic Aspects of Dynamics.* New York: Springer-Verlag. 2nd. Edition.

Lasota, A. and Yorke, J.A. (1973). "On the Existence of Invariant Measures for Piecewise Monotonic Transformations". *Transactions of the American Mathematical Society*, **186**, 481-488.

Lopes, A. and Lopes, S.R.C. (1998). "Parametric Estimation and Spectral Analysis of Piecewise Linear Maps of the Interval". *Advances in Applied Probability*, **30**, 757-776.

Lopes, S.R.C. and Pumi, G. (2011). "Copulas Related to Manneville-Pomeau Processes". *Brazilian Journal of Probability and Statistics*. Forthcoming. Preprint: arXiv:1109.2628v1.

Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag. 2nd. Edition.

Pianigiani, G. (1980). "First Return Map and Invariant Measures". *Israel Journal of Mathematics*, **35**, 32-48.

Pianigiani, G. (1981). "Existence of Invariant Measures for Piecewise Continuous Transformations". *Annales Polonici Mathematici*, **40**, 39-45.

Rényi, A. (1957). "Representations for Real Numbers and their Ergodic Properties". *Acta Mathematica Academiae Scientiarum Hungaricae*, **8**, 477-493.

Schweizer, B. and Sklar, A. (2005). *Probabilistic Metric Spaces*. Mineola: Dover Publications.

Wang, D., Rachev, S.T. and Fabozzi, F.J. (2009). "Pricing of Credit Default Index Swap Tranches with One-Factor Heavy-Tailed Copula Models". *Journal of Empirical Finance*, **16**(2), 201-215.

Zebrowsky, J.J. (2001). "Intermittency in Human Heart Rate Variability". *Acta Physica Polonica B*, **32**, 1531-1540.

Anexo C

Artigo Pumi e Lopes (2011b)

Smoothed Semiparametric Estimation on Multivariate Long Memory Processes

Guilherme Pumi and Sílvia R.C. Lopes
Mathematics Institute
Federal University of Rio Grande do Sul

This version: 13/10/2011

Abstract

In this paper we propose and study a general class of Gaussian Semiparametric Estimators (GSE) of the fractional differencing parameter in the context of long-range dependent multivariate time series. We establish large sample properties of the estimator without assuming Gaussianity. The class of models considered here satisfies simple conditions on the spectral density function, restricted to a small neighborhood of the zero frequency and includes important class of VARFIMA processes. We also present a simulation study to assess the finite sample properties of the proposed estimator based on a smoothed version of the GSE which supports its competitiveness.

Keywords: Fractional integration; Long-range dependence; Semiparametric estimation; Smoothed periodogram; VARFIMA processes.

Mathematical Subject Classification (2010). Primary 62H12, 62F12, 62M10, 60G10, 62M15;

1 Introduction

Let $\boldsymbol{d} = (d_1, \cdots, d_q)' \in (-1/2, 1/2)^q$ and let $\mathcal{B}$ be the shift operator. Consider the q-dimensional weakly stationary process $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ obtained as a stationary solution of the difference equations

$$\operatorname*{diag}_{k\in\{1,\cdots,q\}} \big\{(1-\mathcal{B})^{d_k}\big\}\big(\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t)\big) = \boldsymbol{Y}_t, \tag{1.1}$$

where $\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$ is a q-dimensional weakly stationary process whose spectral density function $f_{\boldsymbol{Y}}$ is bounded and bounded away from zero. Each coordinate process in (1.1) exhibits long-range dependence whenever the respective parameter $d_i > 0$, in the sense that the spectral density function satisfies $f(\lambda) \sim K\lambda^{-2d_i}$, as $\lambda \to 0^+$, for some constant $K > 0$ and $i \in \{1, \cdots, q\}$.

Processes of the form (1.1) constitute the so-called fractionally integrated processes. As a particular case, consider the situation where the i-th coordinate process $\{Y_t^{(i)}\}_{t=0}^{\infty}$ follows an ARMA model. In this case, the associated coordinate process $\{X_t^{(i)}\}_{t=0}^{\infty}$ will be a classical ARFIMA process with the same AR and MA orders and differencing parameter d_i. If the process $\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$ is a vectorial ARMA process, then the resulting multivariate process will be the so-called VARFIMA process with differencing parameter $\boldsymbol{d} = (d_1, \cdots, d_q)'$. VARFIMA and, more generally, fractionally integrated processes, are widely used to model multivariate processes with long-range dependence. See, for instance, the recent work of Chiriac and Voev (2011) on modeling and forecasting high frequency data by using VARFIMA and fractionally integrated processes.

The parameter $\boldsymbol{d}$ in (1.1) determines the spectral density function behavior at the zero frequency as well as the long run autocovariance/autocorrelation structure. Hence, estimation becomes an important matter whenever the long run structure of the process is of interest.

Estimation of the parameter $\boldsymbol{d}$ in the multivariate case has seen a growing interest in the last years. A maximum likelihood approach was first considered in Sowell (1989), but the computational cost of the author's method is very high. A few years later, Luceño (1996) presented

a computationally cheaper alternative for the maximum likelihood approach based on rewriting and approximating the quadratic form of the Gaussian likelihood function. In a recent work, Tsay (2010) proposed an even faster approach to calculate the exact conditional likelihood based on the multivariate Durbin-Levinson algorithm. Although the maximum likelihood approach usually provides good results, it is still a computationally expensive method.

The works of Fox and Taqqu (1986), Giraitis and Surgailis (1990), among others, provided a rigorous asymptotic theory for (univariate) Gaussian parametric estimates which includes, for instance, $n^{1/2}$-consistency and asymptotic normality. One drawback is the crucial role played by the Gaussianity assumption in the theory, which also requires strong distributional and regularity conditions and is non-robust with respect to the parametric specification of the model, leading to inconsistent estimates under misspecification.

In the univariate case, Gaussian Semiparametric Estimation (GSE) was first introduced in Künsch (1987) and later rigorously developed by Robinson (1995b). It provides a more robust alternative compared to the parametric one, requiring less distributional assumptions and being more efficient. In the multivariate case, Robinson (1995a) was the first to study and develop a rigorous treatment of a semiparametric estimator. A two-step multivariate GSE has been studied in the work of Lobato (1999), which showed its asymptotic normality under mild conditions, but without relying on Gaussianity. A few years later, Shimotsu (2007) introduced a refinement of Lobato's two-step GSE, which is consistent and asymptotically normal under very mild conditions (Gaussianity is, again, nowhere assumed), but with smaller asymptotic variance than Lobato's estimator. The technique applied in Shimotsu (2007) was a multivariate extension of that in Robinson (1995b), powerful enough to show not only the consistency of the proposed estimator, but also the consistency of Lobato's two-step GSE. Recently, Nielsen (2011) extended the work of Shimotsu (2007) to include the non-stationary case by using the so-called extended periodogram.

The estimator introduced in Shimotsu (2007) is based on the specification of the spectral density function in a neighborhood of the zero frequency. Estimation of the differencing parameter $\boldsymbol{d}$ is obtained through minimization of an objective function, which is derived from the expression of the Gaussian log-likelihood function near the zero frequency. To obtain the objective function, the spectral density is estimated by the periodogram of the process. Although asymptotically unbiased, it is well known that the periodogram is not a consistent estimator of the spectral density and do not even converge to a random variable at all, cf. Grenander (1951). Some authors actually consider the periodogram "*an extremely poor (if not useless) estimate of the spectral density function*" (Priestley, 1981, page 420).

Our contribution to the theory of GSE is two-folded. First, being consistency a highly desirable property of an estimator, we study the consequences of substituting the periodogram in Shimotsu (2007)'s objective function by an arbitrary consistent estimator of the spectral density function. We prove the consistency of the proposed estimator under the same assumptions as in Shimotsu (2007) and no assumption on the spectral density estimator other than consistency. Second, considering Shimotsu (2007)'s objective function with the periodogram substituted by an arbitrary spectral density estimator, we derive necessary conditions under which GSE is consistent and satisfy a multivariate CLT. Gaussianity is nowhere assumed. In order to assess the finite sample properties of the estimators studied here and its competitiveness, we present a simulation study based on simulated VARFIMA process. We apply the so-called smoothed periodogram as an estimator of the spectral density function.

The paper is organized as follows. In the next section, we present some preliminaries concepts and results necessary for this work. In Section 3 we introduce the Smoothed GSE as well as a general class of estimators based on appropriate modifications of the GSE's objective function. Section 4 is devoted to derive the consistency of the proposed estimator while in Section 5, its asymptotic normality is studied. In Section 6 we present some Monte Carlo simulation results to

assess the finite sample performance of the proposed estimator. Conclusions and final remarks are reserved to Section 7. Proofs are presented in Appendix A.

2 Preliminaries

Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by (1.1) and assume that the spectral density matrix of $\boldsymbol{Y}_t$ satisfies $f_{\boldsymbol{Y}} \sim G$ for a real, symmetric, finite and positive definite matrix G. Let f be the spectral density matrix function of $\boldsymbol{X}_t$, so that

$$\mathbb{E}\big[\big(\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t)\big)\big(\boldsymbol{X}_{t+h} - \mathbb{E}(\boldsymbol{X}_t)\big)'\big] = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}h\lambda} f(\lambda) d\lambda,$$

for $h \in \mathbb{N}^* := \mathbb{N}\backslash\{0\}$. Following the reasoning in Shimotsu (2007), the spectral density matrix of $\boldsymbol{X}_t$ at the Fourier frequencies $\lambda_j = 2\pi j/n$, with $j = 1, \cdots, m$ and $m = o(n)$, can be written as

$$f(\lambda_j) \sim \Lambda_j(\boldsymbol{d}) G \overline{\Lambda_j(\boldsymbol{d})}', \quad \text{for} \quad \Lambda_j(\boldsymbol{d}) = \underset{k\in\{1,\cdots,q\}}{\operatorname{diag}} \{\Lambda_j^{(k)}(\boldsymbol{d})\} \quad \text{and} \quad \Lambda_j^{(k)}(\boldsymbol{d}) = \lambda_j^{-d_k} \mathrm{e}^{\mathrm{i}(\pi-\lambda_j)d_k/2}, \quad (2.1)$$

where, for a complex matrix A, $\overline{A}'$ denotes the conjugate transpose of A. Let

$$w_n(\lambda) := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{X}_t \mathrm{e}^{\mathrm{i}t\lambda} \quad \text{and} \quad I_n(\lambda) := w_n(\lambda)\overline{w_n(\lambda)}'$$

be the discrete Fourier transform and the periodogram of $\boldsymbol{X}_t$ at λ, respectively. In order to define the smoothed periodogram of $\boldsymbol{X}_t$, let $W_n(k) := \big(W_n^{ij}(k)\big)_{i,j=1}^{q}$ be an array of functions (called weights) and $\{\ell(k)\}_{k=0}^{\infty}$ be an increasing sequence of positive integers satisfying the following conditions

A1. $1/\ell(n) + \ell(n)/n \longrightarrow 0$, as n tends to infinity;

A2. $W_n^{ij}(k) = W_n^{ij}(-k)$ and $W_n^{ij}(k) \geq 0$, for all k;

A3. $\sum_{|k|\leq\ell(n)} W_n^{ij}(k) = 1$;

A4. $\sum_{|k|\leq\ell(n)} W_n^{ij}(k)^2 \longrightarrow 0$, as n tends to infinity.

Conditions **A**1 - **A**4 are standard in the literature of smoothed periodogram. See for instance Priestley (1981). For a Fourier frequency λ_j, let

$$\hat{f}_n(\lambda_j) := \sum_{|k|\leq\ell(n)} W_n(k) \odot w_n(\lambda_{j+k})\overline{w_n(\lambda_{j+k})}', \quad (2.2)$$

be an estimator of the spectral density matrix of $\boldsymbol{X}_t$, where $\odot$ denotes the Hadamard product. If, for some j and k, $\lambda_{j+k} \notin [-\pi, \pi]$, we take w_n as having period 2π. The function $\hat{f}_n$ can be extended to take values on $\mathbb{R}$. In order to do so, we need to introduce the following auxiliary function. For a fixed $n > 0$, given $\lambda \in [0, \pi]$, there exists $k_0 \in \mathbb{N}^*$ such that $\lambda_{k_0-1} \leq \lambda \leq \lambda_{k_0}$. Define

$$g(\lambda; n) := \begin{cases} \lambda_{k_0-1}, & \text{if} \quad \lambda - \lambda_{k_0-1} \leq \lambda_{k_0} - \lambda; \\ \lambda_{k_0}, & \text{if} \quad \lambda - \lambda_{k_0-1} > \lambda_{k_0} - \lambda. \end{cases}$$

If $\lambda \in [-\pi, 0)$, we define $g(\lambda; n) := g(-\lambda; n)$. Then, if λ is not a Fourier frequency, we extend (2.2) by setting

$$\hat{f}_n(\lambda) := \hat{f}_n\big(g(\lambda; n)\big).$$

Expression (2.2) extended in this way is a multivariate extension of the univariate smoothed periodogram function. In (2.2), we allow the use of different weight functions for different components (including cross spectrum components) of the spectral density matrix. In this manner

we accommodate the necessity, often observed in practice, to model different characteristics present in the spectral density matrix by using different weight functions. We refrain from presenting further discussion on the different types of weight functions in the literature, since they can be found in most text books on the subject. See, for example, Priestley (1981) and references therein.

In practice, at zero frequency, we use a slightly different estimative, namely,

$$\hat{f}_n(0) := \mathrm{Re}\Big[W_n(0) \odot w_n(\lambda_1)\overline{w_n(\lambda_1)}' + 2\sum_{k=1}^{\ell(n)} W_n(k) \odot w_n(\lambda_{k+1})\overline{w_n(\lambda_{k+1})}'\Big].$$

Considering (1.1), under conditions **A**1 to **A**4, the smoothed periodogram is $n^{1/2}$-consistent for the spectral density matrix whenever $\boldsymbol{d} \in [-1/2, 0]^q$. See, for instance, Grenander and Rosemblatt (1953) and Priestley (1981).

3 Smoothed Gaussian Semiparametric Estimation

From the local form of the spectral density at zero frequency (2.1) replaced in the frequency domain Gaussian log-likelihood localized at the origin, Shimotsu (2007) proposed a semiparametric estimator for the fractional differencing parameter $\boldsymbol{d}$ based on the objective function

$$R(\boldsymbol{d}) := \log\big(\det\{\tilde{G}(\boldsymbol{d})\}\big) - 2\sum_{k=1}^{q} d_k \frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j), \tag{3.1}$$

where

$$\tilde{G}(\boldsymbol{d}) := \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\big[\Lambda_j(\boldsymbol{d})^{-1} I_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\big]. \tag{3.2}$$

with $\Lambda_j(\boldsymbol{d})$ defined in (2.1). The estimator of $\boldsymbol{d}$ is then given by

$$\widetilde{\boldsymbol{d}} = \arg\min_{\boldsymbol{d}\in\Theta}\{R(\boldsymbol{d})\}, \tag{3.3}$$

where the space of admissible estimates is of the form $\Theta = [-1/2 + \epsilon_1, 1/2 - \epsilon_2]$, for arbitrarily small $\epsilon_i > 0$, $i = 1, 2$, henceforth fixed. In Section 6 we shall denote the estimator (3.3) by *Sh*. Shimotsu (2007) shows that the estimator based on the objective function (3.1) and (3.2) is consistent under mild conditions. The proof, however, is complicated and involves somewhat delicate results on the periodogram function I_n. A natural question is what happens if the periodogram I_n is replaced by a consistent estimator of the spectral density function? As we shall see later, replacing the periodogram I_n by a consistent estimator of the spectral density is sufficient to guarantee the consistency of the estimator (3.3) in the case of long-range dependence, $\boldsymbol{d} \in (0, 1/2)^q$. For the case $\boldsymbol{d} \in (-1/2, 0)^q$, surprisingly, consistency alone is not enough, but n^β-consistency, for some $\beta \in (0, 1)$, guarantees the consistency of the estimator for $\boldsymbol{d} \in (-\beta/2, 1/2)^q$. Under extra regularity conditions on f_n, the n^β-consistency can be relaxed (see Theorem 4.2). Proofs are significantly simpler than the original ones.

Let f_n be an arbitrary estimator for the spectral density function f. We are interested in estimators based on objective functions of the form

$$S(\boldsymbol{d}) := \log\big(\det\{\widehat{G}(\boldsymbol{d})\}\big) - 2\sum_{k=1}^{q} d_k \frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j), \tag{3.4}$$

with

$$\widehat{G}(\boldsymbol{d}) := \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\big]. \tag{3.5}$$

Notice that (3.4) is just (3.1) with the periodogram I_n in (3.2) replaced by f_n. The estimator of $\boldsymbol{d}$ is then defined analogously as

$$\widehat{\boldsymbol{d}} = \arg\min_{\boldsymbol{d}\in\Theta}\{S(\boldsymbol{d})\}. \tag{3.6}$$

The estimator (3.6) based on the smoothed periodogram shall be denoted in Section 6 by *SSh*.

Before proceeding with the results, we shall establish some notation. Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by (1.1) and let $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ such that $\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t) = \sum_{k=0}^{\infty} A_k\boldsymbol{\varepsilon}_{t-k}$. We define a function A by setting

$$A(\lambda) := \sum_{k=0}^{\infty} A_k \mathrm{e}^{\mathrm{i}k\lambda}. \tag{3.7}$$

The periodogram function associated to $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is denoted by $I_{\boldsymbol{\varepsilon}}$, that is,

$$I_{\boldsymbol{\varepsilon}}(\lambda) := w_{\boldsymbol{\varepsilon}}(\lambda)\overline{w_{\boldsymbol{\varepsilon}}(\lambda)}', \qquad \text{where} \qquad w_{\boldsymbol{\varepsilon}}(\lambda) := \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\mathrm{e}^{\mathrm{i}t\lambda}. \tag{3.8}$$

For a matrix M, we shall denote the r-th row and the s-th column of M by $(M)_{r\cdot}$ and $(M)_{\cdot s}$, respectively.

4 Consistency of the estimator

Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by (1.1) and f be its spectral density matrix. Suppose that the spectral density matrix of the weakly stationary process $\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$ in (1.1) satisfy $f_{\boldsymbol{Y}}(\lambda) \sim G_0$ for a real, symmetric and positive definite matrix $G_0 = (G_0^{rs})_{r,s=1}^{q}$. Let $\boldsymbol{d}_0 = (d_1^0, \cdots, d_q^0)'$ be the true fractional differencing vector parameter and assume that the following assumptions are satisfied:

B1. As $\lambda \to 0^+$,

$$f_{rs}(\lambda) = \mathrm{e}^{\mathrm{i}\pi(d_r^0 - d_s^0)/2}G_0^{rs}\lambda^{-d_r^0-d_s^0} + o(\lambda^{-d_r^0-d_s^0}), \quad \text{for all } r,s = 1,\cdots,q.$$

B2. Denoting the sup-norm by $\|\cdot\|_\infty$, assume that

$$\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t) = \sum_{k=0}^{\infty} A_k\boldsymbol{\varepsilon}_{t-k}, \quad \sum_{k=0}^{\infty}\|A_k\|_\infty^2 < \infty, \tag{4.1}$$

where $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is a process such that

$$\mathbb{E}(\boldsymbol{\varepsilon}_t|\mathscr{F}_{t-1}) = 0 \quad \text{and} \quad \mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathscr{F}_{t-1}) = \mathrm{I}_q, \quad \text{a.s.}$$

for all $t \in \mathbb{Z}$, where I_q is the $q\times q$ identity matrix and $\mathscr{F}_t$ denotes the σ-field generated by $\{\boldsymbol{\varepsilon}_s, s \le t\}$. Also assume that there exist a scalar random variable ξ and a constant $K > 0$ such that $\mathbb{E}(\xi^2) < \infty$ and $\mathbb{P}\big(\|\boldsymbol{\varepsilon}_t\|_\infty^2 > \eta\big) \le K\mathbb{P}(\xi^2 > \eta)$, for all $\eta > 0$.

B3. In a neighborhood $(0,\delta)$ of the origin, A given by (3.7) is differentiable and, as $\lambda \to 0^+$,

$$\frac{\partial}{\partial\lambda}\big(\overline{A(\lambda)}'\big)_{r\cdot} = O\big(\lambda^{-1}\big\|\big(\overline{A(\lambda)}'\big)_{r\cdot}\big\|_\infty\big).$$

B4. As $n \to \infty$,

$$\frac{1}{m} + \frac{m}{n} \longrightarrow 0.$$

Remark 4.1. Assumptions **B**1-**B**4 are the same as in Shimotsu (2007) and are multivariate extensions of the assumptions made in Robinson (1995b) and analogous to the ones used in Robinson (1995a) and Lobato (1999). Assumption **B**1 describes the true spectral density matrix behavior at the origin. Notice that, since $\lim_{\lambda\to 0^+} \mathrm{e}^{\mathrm{i}\lambda} - 1 = 0$, replacing $\mathrm{e}^{\mathrm{i}\pi(d_r^0-d_s^0)/2}$ by $\mathrm{e}^{\mathrm{i}(\pi-\lambda)(d_r^0-d_s^0)/2}$ makes no difference. Assumption **B**2 regards the causal representation of $\boldsymbol{X}_t$, and more specifically, the behavior of the innovation process which is assumed to be a not necessarily uncorrelated square integrable martingale difference uniformly dominated (in probability) by a scalar random variable with finite second moment. Assumption **B**3 is a regularity condition (also imposed in Fox and Taqqu, 1986 and Giraitis and Surgailis, 1990, among others, in the parametric case) and will be useful in proving Lemmas 4.1 and 4.2 below. Assumption **B**4 is minimal but necessary since m must go to infinity for consistency, but slower than n in view of Assumption **B**1, which specifies f only at a neighborhood of the zero frequency (recall that $\lambda_j = 2\pi j/n$, $j = 1, \cdots, m$).

Observe that assumptions **B**1-**B**4 are only concerned to the behavior of the spectral density matrix on a neighborhood of the origin and, apart from integrability (implied by the process' weakly stationarity property), no assumption whatsoever is made on the spectral density matrix behavior outside this neighborhood. For $\beta \in [0,1]$ and $q \in \mathbb{N}^*$, let

$$\Omega_\beta := \left[-\frac{\beta}{2}, \frac{1}{2}\right]^q \bigcap \left[-\frac{1}{2}+\epsilon_1, \frac{1}{2}-\epsilon_2\right]^q, \tag{4.2}$$

for $\epsilon_i > 0, i \in \{1,2\}$.

Lemma 4.1 establishes the consistency of $\widehat{G}(\boldsymbol{d}_0)$ given in (3.5) under the assumption of n^β-consistency of f_n and will be useful in proving Theorem 4.1. The proofs of all results in the paper, due to their lengths, are postponed to the Appendix A.

Lemma 4.1. *Let $\{\boldsymbol{X}_t\}_{t=0}^\infty$ be a q-dimensional process specified by* (1.1) *and f be its spectral density matrix. Let f_n be a n^β-consistent estimator for f. If $\boldsymbol{d}_0 \in \Omega_\beta$, then*

$$\widehat{G}(\boldsymbol{d}_0) = G_0 + o_{\mathbb{P}}(1).$$

Theorem 4.1 establishes the consistency of $\widehat{\boldsymbol{d}}$, given in (3.6), under assumptions **B**1-**B**4 and assuming n^β-consistency of the spectral density function estimator.

Theorem 4.1. *Let $\{\boldsymbol{X}_t\}_{t=0}^\infty$ be a q-dimensional process specified by* (1.1) *and f be its spectral density matrix. Let f_n be a n^β-consistent estimator of f, for $\beta \in [0,1]$, and let $\widehat{\boldsymbol{d}}$ be as in* (3.6). *Assume that assumptions **B**1-**B**4 hold and let $\boldsymbol{d}_0 \in \Omega_\beta$. Then, $\widehat{\boldsymbol{d}} \xrightarrow{\mathbb{P}} \boldsymbol{d}_0$, as $n \to \infty$.*

Lemma 4.2 will be useful in proving Theorem 4.2.

Lemma 4.2. *Let $\{\boldsymbol{X}_t\}_{t=0}^\infty$ be a q-dimensional process specified by* (1.1) *and f be its spectral density matrix. Let f_n be an estimator of f, and consider the estimator $\widehat{\boldsymbol{d}}$ based on f as in* (3.6). *Assume that assumptions **B**1-**B**4 hold. Suppose that f_n satisfies*

$$\mathbb{E}\Big(n^{-d_r^0-d_s^0}\Big|f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\Big|\Big) = O\big(j^{-d_r^0-d_s^0-\gamma}\big), \quad \text{as } n\to\infty, \tag{4.3}$$

for some $\gamma > 0$, for all $r, s \in \{1, \cdots, q\}$ and $\boldsymbol{d}_0 \in \Theta$, where A and $I_{\boldsymbol{\varepsilon}}$ are given by (3.7) *and* (3.8), *respectively. Then, for $1 \le u < v \le m$,*

$$\max_{r,s\in\{1,\cdots,q\}} \left\{ \sum_{j=u}^{v} \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2} \lambda_j^{d_r^0+d_s^0} f_n^{rs}(\lambda_j) - G_0^{rs} \right\} = \mathscr{A}_{uv} + \mathscr{B}_{uv},$$

where $\mathscr{A}_{uv}$ *and* $\mathscr{B}_{uv}$ *satisfy*

$$\mathbb{E}\big(|\mathscr{A}_{uv}|\big) = O(v^{1-\gamma}) \qquad \textit{and} \qquad \max_{1\leq u<v\leq m} \big\{\big|v^{-1}\mathscr{B}_{uv}\big|\big\} = o_{\mathbb{P}}(1).$$

Theorem 4.2, we derive a necessary condition for the consistency of $\widehat{\boldsymbol{d}}$ given in (3.6), when the consistency condition on f_n is relaxed.

Theorem 4.2. *Let* $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ *be a* q*-dimensional process specified by* (1.1) *and* f *be its spectral density matrix. Let* f_n *be an estimator of* f*, and consider the estimator* $\widehat{\boldsymbol{d}}$*, based on* f_n*, given in* (3.6)*. Assume that assumptions* ***B1-B4*** *hold. Suppose that* f_n *satisfies* (4.3)*, for some* $\gamma > 0$*, for all* $r, s \in \{1, \cdots, q\}$ *and* $\boldsymbol{d}_0 \in \Theta$*, where* A *and* $I_{\boldsymbol{\varepsilon}}$ *are given by* (3.7) *and* (3.8)*, respectively. Then,* $\widehat{\boldsymbol{d}} \xrightarrow{\mathbb{P}} \boldsymbol{d}_0$*, as* $n \to \infty$*.*

Corollary 4.1 establishes the consistency of the estimator (3.6) based on the smoothed periodogram under assumptions **A**1-**A**4. We shall also need an extra minimal condition involving the rate of convergence between m in (3.4) and $\ell(n)$.

A5. As $n \to \infty$, $\ell(n)/m \to 0$.

Corollary 4.1. *Let* $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ *be a* q*-dimensional process specified by* (1.1) *and* f *be its spectral density matrix. Assume that assumptions* ***B1-B4*** *hold. Let* $\{W_n^{ij}(k)\}_{i,j=1}^{q}$ *be a sequence of weights satisfying assumptions* ***A1-A5*** *and let* f_n *be the respective smoothed periodogram given in* (2.2)*. For* $\boldsymbol{d}_0 \in \Theta$*, consider the estimator* (3.6)*. Then,* $\widehat{\boldsymbol{d}} \xrightarrow{\mathbb{P}} \boldsymbol{d}_0$*, as* $n \to \infty$*.*

We observe that assumption **A**5 is not necessary when $\boldsymbol{d}_0 \in (0, 1/2)^q$, as it is clear from the proof of Corollary 4.1.

Remark 4.2. Recall that the smoothed periodogram is $n^{1/2}$-consistent (under **A**1-**A**4) for $\boldsymbol{d}_0 \in [-1/2, 0]^q$, so that Theorem 4.1 applies and we conclude that the estimator (3.6) is consistent for $\boldsymbol{d}_0 \in \Omega_{1/2} \cap (-1/2, 0] \subset [-1/4, 0]$. In contrast, as shown in Corollary 4.1, Theorem 4.2 also holds and establishes the consistency of (3.6) for all admissible parameter in Θ, under **A**1-**A**5. The latter is a far more interesting result, but to obtain it, we had to show that condition (4.3) holds (see the proof of Corollary 4.1), which required some cumbersome computations. The former result, in contrast, was automatically obtained from the properties of the spectral density estimator. Furthermore, Theorem 4.2 does not assume consistency of the underline spectral density estimator. In fact, the periodogram itself satisfies condition (4.3) (see lemma 1(a) in Shimotsu, 2007).

5 Asymptotic normality of the estimator

In this section we present a sufficient condition for the asymptotic normality of the GSE given by (3.6), under similar assumptions to Shimotsu (2007), with f_n an estimator of the spectral density function satisfying a single regularity condition. The asymptotic distribution of the estimator (3.6) will be the same as (3.3), established by Shimotsu (2007).

Again, let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by (1.1) and f be its spectral density matrix. Suppose that the spectral density matrix of the weakly stationary process $\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$ in (1.1) satisfies $f_{\boldsymbol{Y}}(\lambda) \sim G_0$ for a real, symmetric and positive definite matrix $G_0 = (G_0^{rs})_{r,s=1}^{q}$. Let $\boldsymbol{d}_0 = (d_1^0, \cdots, d_q^0)'$ be the true fractional differencing vector parameter. Assume that the following assumptions are satisfied

C1. For $\alpha \in (0, 2\,]$ and $r, s \in \{1, \cdots, q\}$,

$$f_{rs}(\lambda) = \mathrm{e}^{\mathrm{i}(\pi-\lambda)(d_r^0-d_s^0)/2}\lambda^{-d_r^0-d_s^0}G_0^{rs} + O(\lambda^{-d_r^0-d_s^0+\alpha}), \quad \text{as } \lambda \to 0^+.$$

C2. Assumption **B**2 holds and the process $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ has finite fourth moment.

C3. Assumption **B**3 holds.

C4. For any $\delta > 0$,

$$\frac{1}{m} + \frac{m^{1+2\alpha}\log(m)^2}{n^{2\alpha}} + \frac{\log(n)}{m^{\delta}} \longrightarrow 0, \quad \text{as } n \to \infty.$$

C5. There exists a finite real matrix M such that

$$\Lambda_j(\boldsymbol{d}_0)^{-1}A(\lambda_j) = M + o(1), \quad \text{as } \lambda_j \to 0.$$

Remark 5.1. Assumption **C**1 is a smoothness condition often imposed in spectral analysis. Compared to assumption 1 in Robinson (1995a), assumption **C**1 is slightly more restrictive. It is satisfied by certain VARFIMA processes. Assumption **C**2 imposes that the process $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}^*}$ is linear with finite fourth moment. This restriction in the innovation process is necessary in the proof of Theorem 5.1 in applying a CLT to a certain martingale difference derived from a quadratic form involving $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$, which must have finite variance. Assumption **C**4 is the same as assumption $4'$ in Shimotsu (2007) and is slightly stronger than the ones imposed in Robinson (1995b) and Lobato (1999) (see Shimotsu, 2007 p.283 for a discussion). It implies that $(m/n)^b = o\big(m^{-\frac{b}{2\alpha}}\ \log(m)^{-\frac{b}{\alpha}}\big)$, for $b \neq 0$. Assumption **C**5 is the same as assumption 5' in Shimotsu (2007) and is a mild regularity condition in the degree of approximation of $A(\lambda_j)$ by $\Lambda_j(\boldsymbol{d}_0)$. It is satisfied by general VARFIMA processes.

The next lemma will be useful in proving Theorem 5.1. The proofs of the results in this section are presented in Appendix A.

Lemma 5.1. *Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by* (1.1) *and f be its spectral density matrix. Let f_n be an estimator of f, and consider the estimator $\widehat{\boldsymbol{d}}$, based on f_n, given in* (3.6)*. Assume that assumptions **C**1-**C**5 hold. Suppose that f_n satisfies*

$$\sum_{j=1}^{v}\Big|f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot}I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}\,'\big)_{\cdot s}\Big| = o_{\mathbb{P}}\bigg(\frac{n^{d_r^0+d_s^0}}{\log(v)v^{d_r^0+d_s^0-1/2}}\bigg), \tag{5.1}$$

for all $r, s \in \{1, \cdots, q\}$, $1 \le v \le m$ and $\boldsymbol{d}_0 \in \Theta$, where $I_{\boldsymbol{\varepsilon}}$ and A are defined in (3.8) *and* (3.7)*, respectively. Then,*

(a) *uniformly in $1 \le v \le m$,*

$$\max_{r,s\in\{1,\cdots,q\}}\bigg\{\sum_{j=1}^{v}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big[f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot}I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}\,'\big)_{\cdot s}\Big]\bigg\} = o_{\mathbb{P}}\bigg(\frac{v^{1/2}}{\log(v)}\bigg); \tag{5.2}$$

(b) *uniformly in $1 \le v \le m$,*

$$\max_{r,s\in\{1,\cdots,q\}}\bigg\{\sum_{j=1}^{v}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}f_n^{rs}(\lambda_j) - G_0^{rs}\bigg\} = O_{\mathbb{P}}\bigg(\frac{v^{\alpha+1}}{n^{\alpha}} + v^{1/2}\log(v)\bigg). \tag{5.3}$$

The next theorem presents a necessary condition for the asymptotic normality of the GSE given in (3.6). We notice that the variance-covariance matrix of the limiting distribution is the same as the estimator in (3.3), as derived in Shimotsu (2007).

Theorem 5.1. *Let $\{\boldsymbol{X}_t\}_{t=0}^{\infty}$ be a q-dimensional process specified by* (1.1) *and f be its spectral density matrix. Let f_n be an estimator of f, and consider the estimator $\widehat{\boldsymbol{d}}$, based on f_n, given in* (3.6)*. Assume that assumptions **C1**-**C5** hold. Suppose that f_n satisfies* (5.1)*, for all $r, s \in \{1, \cdots, q\}$ and $\boldsymbol{d}_0 \in \Theta$. If $\widehat{\boldsymbol{d}} \xrightarrow{\mathbb{P}} \boldsymbol{d}_0$, for $\boldsymbol{d}_0 \in \Theta$, then*

$$m^{1/2}(\widehat{\boldsymbol{d}} - \boldsymbol{d}_0) \xrightarrow{d} N(\mathbf{0}, \Omega^{-1}),$$

as n tends to infinity, where

$$\Omega := 2\bigg[G_0 \odot G_0^{-1} + \mathrm{I}_q + \frac{\pi^2}{4}\big(G_0 \odot G_0^{-1} - \mathrm{I}_q\big)\bigg],$$

with I_q the $q \times q$ identity matrix.

6 Monte Carlo Simulation Study

In this section we present a Monte Carlo simulation study to assess the finite sample performance of the proposed estimator given in (3.6) (denoted *SSh*). For comparison purposes, we also calculate the estimator (3.3) (denoted *Sh*). All Monte Carlo simulations are based on time series of fixed sample size $n = 1,000$ obtained from bidimensional Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ processes for several different parameters $\boldsymbol{d}$ and correlations $\rho \in \{0, 0.3, 0.6, 0.8\}$. We perform 1,000 replications of each experiment. To generate the time series, we apply the traditional method of truncating the multidimensional infinite moving average representation of the process. The truncation point is fixed in 50,000 for all cases.

In all simulations, we apply the smoothed periodogram (2.2) with the same weights for all spectral density components, given by the so-called Bartlett's window, that is, we use

$$W_n^{ij}(k) := \frac{\sin^2(nk/2)}{2\pi n \sin^2(k/2)}, \quad \text{for all } i, j = 1, 2.$$

The truncation point of the smoothed periodogram function is of the form $\ell(n, \beta) := \lfloor n^{\beta} \rfloor$, for $\beta \in \{0.7, 0.9\}$, while the truncation point of the estimators *SSh* and *Sh* are both of the form $m := m(n) = \lfloor n^{\alpha} \rfloor$, for $\alpha \in \{0.65, 0.85\}$.

The routines are implemented in FORTRAN 95 language optimized by using OpenMP directives for parallel computing. All simulations were performed by using the computational resources from the (Brazilian) National Center of Super Computing (CESUP-UFRGS).

Table 6.1 reports the simulation results for $\boldsymbol{d} \in \big\{(0.1, 0.4), (0.4, 0.4), (0.2, 0.2), (0.2, 0.3)\big\}$. Presented are the estimated values (mean), their standard deviations (st.d.) and the mean square error of the estimates (mse).

We observe that in all cases both estimators perform well, but the *SSh* usually present a slightly better performance in terms of mse compared to the *Sh* estimator. In terms of mse, the best combination of (α, β) for the *SSh* estimator is $(0.85, 0.9)$ in all cases. For the *Sh* estimator, we found that $\alpha = 0.85$ produces better results than $\alpha = 0.65$, applied in the simulations presented in Shimotsu (2007). As the correlation in the innovation process increases, the estimated values slightly degrade.

In Table 6.1 we notice that the trade off between variance and bias, an usual feature in smoothing the periodogram, has little effect in the estimated values. In most cases, we observe that the higher the value of β, the higher the variance but smaller the bias of the estimate.

Table 6.1: Simulation results of the estimator (3.6) based on the smoothed periodogram with the Barttlet's spectral window (weights) and estimator (3.3) in VARFIMA$(0, \boldsymbol{d}, 0)$ processes. Presented are the estimated values (mean), its standard deviation (st.d) and the mean square error of the estimates (mse).

ρ	Method	β	$\hat{d}_i$	$\boldsymbol{d} = (0.1, 0.4)$						$\boldsymbol{d} = (0.4, 0.4)$					
				$\alpha = 0.65$			$\alpha = 0.85$			$\alpha = 0.65$			$\alpha = 0.85$		
				mean	st.d.	mse	mean	st.d.	mse	mean	st.d.	mse	mean	st.d.	mse
0	*SSh*	0.7	$\hat{d}_1$	0.1030	0.0544	0.0030	0.0955	0.0268	0.0007	0.4506	0.0709	0.0076	0.4119	0.0366	0.0015
			$\hat{d}_2$	0.4467	0.0789	0.0084	0.4140	0.0402	0.0018	0.4453	0.0781	0.0082	0.4138	0.0401	0.0018
		0.9	$\hat{d}_1$	0.1047	0.0447	0.0032	0.0953	0.0268	0.0007	0.4312	0.0601	0.0046	0.3944	0.0291	0.0009
			$\hat{d}_2$	0.4230	0.0539	0.0044	0.3949	0.0306	0.0010	0.4228	0.0624	0.0044	0.3948	0.0306	0.0010
	Sh	-	$\hat{d}_1$	0.1053	0.0581	0.0020	0.0955	0.0271	0.0005	0.4136	0.0590	0.0020	0.3851	0.0277	0.0008
			$\hat{d}_2$	0.3955	0.0606	0.0022	0.3822	0.0292	0.0006	0.3955	0.0606	0.0021	0.3822	0.0292	0.0007
0.3	*SSh*	0.7	$\hat{d}_1$	0.1100	0.0513	0.0027	0.0982	0.0257	0.0007	0.4497	0.0648	0.0067	0.4118	0.0346	0.0013
			$\hat{d}_2$	0.4468	0.0741	0.0077	0.4148	0.0384	0.0017	0.4478	0.0714	0.0074	0.4142	0.0375	0.0016
		0.9	$\hat{d}_1$	0.1076	0.0525	0.0028	0.0960	0.0256	0.0007	0.4299	0.0555	0.0040	0.3942	0.0277	0.0008
			$\hat{d}_2$	0.4243	0.0587	0.0040	0.3958	0.0288	0.0008	0.4253	0.0581	0.0040	0.3952	0.0286	0.0008
	Sh	-	$\hat{d}_1$	0.1038	0.0542	0.0030	0.0947	0.0259	0.0007	0.4109	0.0550	0.0031	0.3846	0.0265	0.0009
			$\hat{d}_2$	0.3985	0.0572	0.0033	0.3834	0.0273	0.0010	0.3991	0.0570	0.0032	0.3828	0.0273	0.0010
0.6	*SSh*	0.7	$\hat{d}_1$	0.1298	0.0483	0.0032	0.1073	0.0248	0.0007	0.4499	0.0566	0.0057	0.4125	0.0310	0.0011
			$\hat{d}_2$	0.4500	0.0665	0.0069	0.4182	0.0361	0.0016	0.4503	0.0607	0.0062	0.4143	0.0327	0.0013
		0.9	$\hat{d}_1$	0.1167	0.0467	0.0025	0.0990	0.0234	0.0006	0.4290	0.0485	0.0032	0.3943	0.0250	0.0007
			$\hat{d}_2$	0.4274	0.0525	0.0035	0.3981	0.0262	0.0007	0.4280	0.0505	0.0033	0.3953	0.0252	0.0007
	Sh	-	$\hat{d}_1$	0.1034	0.0478	0.0023	0.0939	0.0939	0.0006	0.4077	0.04077	0.0024	0.3840	0.0239	0.0008
			$\hat{d}_2$	0.4029	0.0501	0.0025	0.3858	0.0241	0.0008	0.4037	0.0498	0.0025	0.3836	0.0240	0.0008
0.8	*SSh*	0.7	$\hat{d}_1$	0.1570	0.0512	0.0059	0.1233	0.0278	0.0013	0.4506	0.0535	0.0054	0.4130	0.0290	0.0010
			$\hat{d}_2$	0.4669	0.0647	0.0087	0.4295	0.0374	0.0023	0.4512	0.0557	0.0057	0.4142	0.0298	0.0011
		0.9	$\hat{d}_1$	0.1307	0.0458	0.0030	0.1055	0.0231	0.0006	0.4288	0.0451	0.0029	0.3946	0.0231	0.0006
			$\hat{d}_2$	0.4372	0.0509	0.0040	0.4037	0.0258	0.0007	0.4290	0.0464	0.0030	0.3953	0.0230	0.0006
	Sh	-	$\hat{d}_1$	0.1060	0.0448	0.0020	0.0946	0.0218	0.0005	0.4066	0.0448	0.0020	0.3838	0.0221	0.0008
			$\hat{d}_2$	0.4071	0.0465	0.0022	0.3882	0.0221	0.0006	0.4056	0.0458	0.0021	0.3840	0.0218	0.0007
ρ	Method	β	$\hat{d}_i$	$\boldsymbol{d} = (0.2, 0.2)$						$\boldsymbol{d} = (0.2, 0.3)$					
				$\alpha = 0.65$			$\alpha = 0.85$			$\alpha = 0.65$			$\alpha = 0.85$		
				mean	st.d.	mse	mean	st.d.	mse	mean	st.d.	mse	mean	st.d.	mse
0	*SSh*	0.7	$\hat{d}_1$	0.2085	0.0570	0.0033	0.1946	0.0277	0.0008	0.2085	0.0569	0.0033	0.1946	0.0277	0.0008
			$\hat{d}_2$	0.1994	0.0572	0.0033	0.1950	0.0294	0.0009	0.3164	0.0643	0.0044	0.3000	0.0321	0.0010
		0.9	$\hat{d}_1$	0.2087	0.0565	0.0033	0.1924	0.0269	0.0008	0.2087	0.0565	0.0033	0.1924	0.0269	0.0008
			$\hat{d}_2$	0.1989	0.0567	0.0032	0.1926	0.0286	0.0009	0.3078	0.0580	0.0034	0.2919	0.0289	0.0009
	Sh	-	$\hat{d}_1$	0.2072	0.0581	0.0034	0.1915	0.0271	0.0008	0.2072	0.0581	0.0034	0.1915	0.0271	0.0008
			$\hat{d}_2$	0.1925	0.0601	0.0037	0.1902	0.0291	0.0009	0.2937	0.0603	0.0037	0.2860	0.0291	0.0010
0.3	*SSh*	0.7	$\hat{d}_1$	0.2072	0.0530	0.0029	0.1945	0.0264	0.0007	0.2097	0.0530	0.0029	0.1951	0.0264	0.0007
			$\hat{d}_2$	0.2014	0.0539	0.0029	0.1952	0.0275	0.0008	0.3171	0.0601	0.0039	0.3000	0.0301	0.0009
		0.9	$\hat{d}_1$	0.2073	0.0527	0.0028	0.1923	0.0257	0.0007	0.2085	0.0527	0.0028	0.1924	0.0257	0.0007
			$\hat{d}_2$	0.2009	0.0538	0.0029	0.1927	0.0268	0.0008	0.3093	0.0547	0.0031	0.2922	0.0270	0.0008
	Sh	-	$\hat{d}_1$	0.2050	0.0543	0.0030	0.1912	0.0260	0.0008	0.2051	0.0542	0.0030	0.1910	0.0259	0.0008
			$\hat{d}_2$	0.1955	0.0567	0.0032	0.1907	0.0272	0.0008	0.2967	0.0568	0.0032	0.2867	0.0272	0.0009
0.6	*SSh*	0.7	$\hat{d}_1$	0.2058	0.0464	0.0022	0.1946	0.0239	0.0006	0.2138	0.0470	0.0024	0.1970	0.0240	0.0006
			$\hat{d}_2$	0.2039	0.0473	0.0022	0.1952	0.0242	0.0006	0.3178	0.0525	0.0031	0.3001	0.0267	0.0007
		0.9	$\hat{d}_1$	0.2057	0.0462	0.0022	0.1923	0.0233	0.0006	0.2095	0.0461	0.0022	0.1929	0.0233	0.0006
			$\hat{d}_2$	0.2035	0.0473	0.0023	0.1928	0.0236	0.0006	0.3111	0.0481	0.0024	0.2926	0.0239	0.0006
	Sh	-	$\hat{d}_1$	0.2025	0.0478	0.0023	0.1909	0.0235	0.0006	0.2030	0.0477	0.0023	0.1904	0.0235	0.0006
			$\hat{d}_2$	0.1993	0.0496	0.0025	0.1911	0.0239	0.0007	0.3005	0.0496	0.0025	0.2877	0.0239	0.0007
0.8	*SSh*	0.7	$\hat{d}_1$	0.2054	0.0430	0.0019	0.1948	0.0222	0.0005	0.2192	0.0448	0.0024	0.1998	0.0228	0.0005
			$\hat{d}_2$	0.2049	0.0436	0.0019	0.1952	0.0221	0.0005	0.3208	0.0492	0.0028	0.3014	0.0249	0.0006
		0.9	$\hat{d}_1$	0.2051	0.0428	0.0019	0.1925	0.0216	0.0005	0.2117	0.0428	0.0020	0.1939	0.0216	0.0005
			$\hat{d}_2$	0.2046	0.0435	0.0019	0.1928	0.0215	0.0005	0.3131	0.0445	0.0021	0.2933	0.0219	0.0005
	Sh	-	$\hat{d}_1$	0.2016	0.0445	0.0020	0.1909	0.0218	0.0006	0.2026	0.0444	0.0020	0.1903	0.0218	0.0006
			$\hat{d}_2$	0.2008	0.0455	0.0021	0.1912	0.0217	0.0005	0.3024	0.0456	0.0021	0.2883	0.0217	0.0006

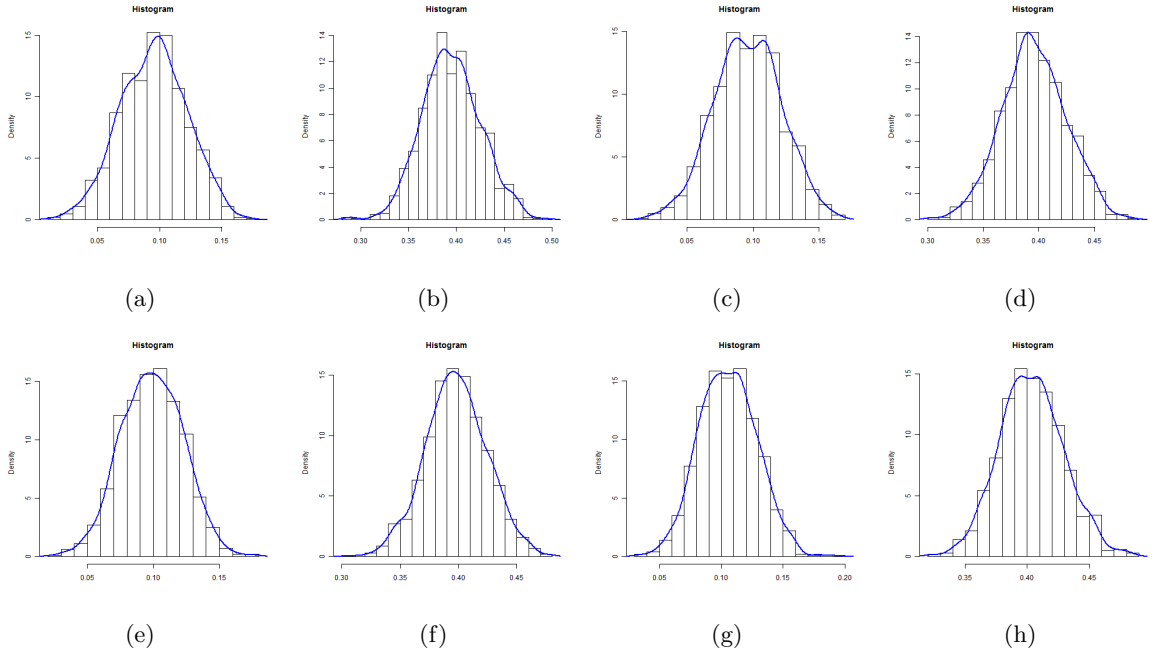


Figure 6.1: Histogram and kernel density of the *SSh* estimated values of $\boldsymbol{d}_0 = (0.1, 0.4)$ for (a) and (b), $\rho = 0$, (c) and (d), $\rho = 0.3$, (e) and (f), $\rho = 0.6$ and (g) and (h), $\rho = 0.8$.

The variance of the estimator responds stronger to changes in α than in β in the opposite direction, that is, the higher the α, the smaller the variance. The estimated value of $\boldsymbol{d}$ by the *SSh* seems to respond stronger to changes in α than in β.

Figure 6.1 presents the histogram and kernel density estimator of the *SSh* estimated values for $\boldsymbol{d}_0 = (0.1, 0.4)$ when $\alpha = 0.85$, $\beta = 0.9$. Figures 6.1(a) and 6.1(b) correspond to $\rho = 0$, Figures 6.1(c) and 6.1(d) to $\rho = 0.3$, Figures 6.1(e) and 6.1(f) to $\rho = 0.6$ and Figures 6.1(g) and 6.1(h) to $\rho = 0.8$. Notice that the resemblance of the histograms to the normal distribution is greater for $\boldsymbol{d}_0^{(2)} = 0.4$ than to 0.1. At this moment, we were not able to prove the asymptotic normality of the *SSh* estimator with the smoothed periodogram by direct verification of (5.1). However we conjecture that this is the case and Figure 6.1 supports this opinion.

7 Conclusions

In this work we propose and analyze a class of Gaussian semiparametric estimators of multivariate long-range dependent processes. The work is motivated by the semiparametric methodology presented in Shimotsu (2007). More specifically, we propose a class of estimators based on the method studied in Shimotsu (2007) by substituting the periodogram applied there for an arbitrary spectral density estimator. We analyze two frameworks. First we assume that the spectral density estimator is consistent for the spectral density estimator and we show that the proposed semiparametric estimator is also consistent under mild conditions. Second, we relax the consistency condition and derive necessary conditions for the consistency and asymptotic normality of the proposed estimator. We show that the variance-covariance matrix of the limiting distribution is the same as the one derived in Shimotsu (2007), under the same conditions imposed in the process.

In order to assess the finite sample performance and illustrate the usefulness of the estimator, we perform a Monte Carlo simulation based on VARFIMA$(0, \boldsymbol{d}, 0)$ processes. We applied the so-called smoothed periodogram (which is shown to satisfy the consistency conditions imposed in the theoretical results) with the Bartlett's weight function as the spectral density estimator.

For comparison we also compute the estimator proposed in Shimotsu (2007). Both estimators perform great but the one proposed here present generally better results.

The assumptions required in the asymptotic theory are mild ones and are commonly applied in the literature. The semiparametric methodology present several advantages compared to the parametric framework such as weaker distributional assumptions, robustness with respect to misspecification of the short run dynamics of the process and efficiency. The theory includes the fractionally integrated processes as well as the class of VARFIMA processes.

Acknowledgements

G. Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil. S.R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1 and also by INCT *em Matemática*. The authors are also grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

Appendix A: Proofs

In this section we present the proofs of the results in Sections 4 and 5. We establish lemmas and theorems in the same sequence as they appear in the text.

Proof of Lemma 4.1:

By hypothesis, $f_n(\lambda) = f(\lambda) + o_{\mathbb{P}}(n^{-\beta})$. Recalling the definition of Λ_j given in (2.1), we have

$$\begin{aligned}
\widehat{G}(\boldsymbol{d}_0) &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\big] \\
&= \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[\Lambda_j(\boldsymbol{d}_0)^{-1}\big(f(\lambda_j) + o_{\mathbb{P}}(n^{-\beta})\big)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\big] \\
&= G_0 + \sum_{k=1}^{q}\Big(\Big[\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d_k^0}\Big]o_{\mathbb{P}}(n^{-\beta})\Big) \;=\; G_0 + o_{\mathbb{P}}(1),
\end{aligned}$$

since

$$\begin{aligned}
\Big[\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d_k^0}\Big]o_{\mathbb{P}}(n^{-\beta}) &= \frac{1}{2d_k^0+1}\left(\frac{2\pi m}{n}\right)^{2d_k^0}\left[\frac{2d_k^0+1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2d_k^0}\right]o_{\mathbb{P}}(n^{-\beta}) \\
&= \frac{1}{2d_k^0+1}\left(\frac{2\pi m}{n}\right)^{2d_k^0}\big[O(m^{\beta-1})+1\big]o_{\mathbb{P}}(n^{-\beta}) \;=\; o_{\mathbb{P}}(1),
\end{aligned}$$

where the second equality follows from lemma 2 in Robinson (1995b), by taking $\gamma = 2d_k^0 + 1 > 0$. The last equality is justified as follows: recalling that $\beta \in (0,1)$, if $d_k^0 \geq 0$, $(m/n)^{2d_k^0} \to 0$, and the result is immediate; if $d_k^0 < 0$, $m^{2d_k^0} \to 0$ and the result follows from the assumption that $\boldsymbol{d}_0 \in \Omega_\beta$, which implies $2d_k^0 + \beta \geq 0$. This completes the proof. ■

Proof of Theorem 4.1:

Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)' := \boldsymbol{d} - \boldsymbol{d}_0$ and $L(\boldsymbol{d}) := S(\boldsymbol{d}) - S(\boldsymbol{d}_0)$. Let $0 < \delta < 1/2$ be fixed and let

$$N_\delta := \big\{\boldsymbol{d} : \|\boldsymbol{d} - \boldsymbol{d}_0\|_\infty > \delta\big\}.$$

Let $0 < \epsilon < 1/4$ and define $\Theta_1 := \big\{\boldsymbol{\theta} : \boldsymbol{\theta} \in [-1/2+\epsilon, 1/2]^q\big\}$ and $\Theta_2 = \Omega_\beta \setminus \Theta_1$ (possibly an empty set), where Ω_β is given by (4.2). Following Robinson (1995b) and Shimotsu (2007), we have

$$\begin{aligned}\mathbb{P}\big(\|\widehat{\boldsymbol{d}} - \boldsymbol{d}_0\|_\infty > \delta\big) &\leq \mathbb{P}\Big(\inf_{\overline{N_\delta}\cap\Omega_\beta}\big\{L(\boldsymbol{d})\big\} \leq 0\Big)\\ &\leq \mathbb{P}\Big(\inf_{\overline{N_\delta}\cap\Theta_1}\big\{L(\boldsymbol{d})\big\} \leq 0\Big) + \mathbb{P}\Big(\inf_{\Theta_2}\big\{L(\boldsymbol{d})\big\} \leq 0\Big) := P_1 + P_2 \qquad \text{(A.1)}\end{aligned}$$

where, for a given set $\mathcal{O}$, $\overline{\mathcal{O}}$ denotes the closure of $\mathcal{O}$. We shall first show that $P_1 \to 0$, as n tends to infinity. Rewrite $L(\boldsymbol{d})$ as

$$\begin{aligned}L(\boldsymbol{d}) &= \log\big(\det\{\widehat{G}(\boldsymbol{d})\}\big) - \log\big(\det\{\widehat{G}(\boldsymbol{d}_0)\}\big) - 2\sum_{k=1}^{q}\theta_k\frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j)\\ &= \log\big(\det\{\widehat{G}(\boldsymbol{d})\}\big) - \log\big(\det\{\widehat{G}(\boldsymbol{d}_0)\}\big) + \log\left(\frac{2\pi m}{n}\right)^{-2\sum_k\theta_k} -\\ &\qquad - 2\sum_{k=1}^{q}\theta_k\left(\frac{1}{m}\sum_{j=1}^{m}\log(j) - \log(m)\right) - \sum_{k=1}^{q}\log(2\theta_k+1)\\ &= \log\big(\mathcal{A}(\boldsymbol{d})\big) - \log\big(\mathcal{B}(\boldsymbol{d})\big) - \log\big(\mathcal{A}(\boldsymbol{d}_0)\big) + \log\big(\mathcal{B}(\boldsymbol{d}_0)\big) + \mathcal{R}(\boldsymbol{d})\\ &= Q_1(\boldsymbol{d}) - Q_2(\boldsymbol{d}) + \mathcal{R}(\boldsymbol{d}), \qquad \text{(A.2)}\end{aligned}$$

where

$$\begin{aligned}Q_1(\boldsymbol{d}) &:= \log\big(\mathcal{A}(\boldsymbol{d})\big) - \log\big(\mathcal{B}(\boldsymbol{d})\big), \qquad Q_2(\boldsymbol{d}) := \log\big(\mathcal{A}(\boldsymbol{d}_0)\big) + \log\big(\mathcal{B}(\boldsymbol{d}_0)\big),\\ \mathcal{A}(\boldsymbol{d}) &:= \left(\frac{2\pi m}{n}\right)^{-2\sum_k\theta_k}\det\{\widehat{G}(\boldsymbol{d})\}, \qquad \mathcal{B}(\boldsymbol{d}) := \det\{G_0\}\prod_{k=1}^{q}\frac{1}{2\theta_k+1},\\ \text{and} \quad \mathcal{R}(\boldsymbol{d}) &:= 2\sum_{k=1}^{q}\theta_k\left(\log(m) - \frac{1}{m}\sum_{j=1}^{m}\log(j)\right) - \sum_{k=1}^{q}\log(2\theta_k+1).\end{aligned}$$

By lemma 2 in Robinson (1995b), $\log(m) - m^{-1}\sum_{j=1}^{m}\log(j) = 1 + O(m^{-1}\log(m))$, so that

$$\mathcal{R}(\boldsymbol{d}) = \sum_{k=1}^{q}\theta_k\Big[2\theta_k - \log(2\theta_k+1)\Big] + O\left(\frac{\log(m)}{m}\right).$$

Since $x - \log(x+1)$ has a unique global minimum in $(-1,\infty)$ at $x=0$ and $x - \log(x+1) \geq x^2/4$, for $|x| \leq 1$, it follows that

$$\inf_{\overline{N_\delta}\cap\Theta_1}\big\{\mathcal{R}(\boldsymbol{d})\big\} \geq \frac{1}{4}\Big(2\max_k\{\theta_k\}\Big)^2 \geq \delta^2 > 0\,.$$

As for $Q_1(\boldsymbol{d})$ and $Q_2(\boldsymbol{d})$ in (A.2), it suffices to show the existence of a function $h(\boldsymbol{d}) > 0$ satisfying

$$\text{(i)}\ \sup_{\Theta_1}\big\{\big|\mathcal{A}(\boldsymbol{d}) - h(\boldsymbol{d})\big|\big\} = o_{\mathbb{P}}(1); \quad \text{(ii)}\ h(\boldsymbol{d}) \geq \mathcal{B}(\boldsymbol{d}); \quad \text{(iii)}\ h(\boldsymbol{d}_0) = \mathcal{B}(\boldsymbol{d}_0),$$

as n goes to infinity, because (ii) implies $\inf_{\Theta_1}\big\{h(\boldsymbol{d})\big\} \geq \inf_{\Theta_1}\big\{\mathcal{B}(\boldsymbol{d})\big\} > 0$, so that, uniformly in Θ_1,

$$Q_1(\boldsymbol{d}) \geq \log\big(\mathcal{A}(\boldsymbol{d})\big) - \log\big(h(\boldsymbol{d})\big) = \log\big(h(\boldsymbol{d}) + o_{\mathbb{P}}(1)\big) - \log\big(h(\boldsymbol{d})\big) = o_{\mathbb{P}}(1),$$

and (iii) implies $Q_2(\boldsymbol{d}) = \log\big(h(\boldsymbol{d}_0) + o_{\mathbb{P}}(1)\big) - \log\big(h(\boldsymbol{d}_0)\big) = o_{\mathbb{P}}(1)$, from which $P_1 \longrightarrow 0$ follows. To show (i), recall that

$$\begin{aligned}\Lambda_j(\boldsymbol{d})^{-1} &= \operatorname*{diag}_{k\in\{1,\cdots,q\}}\{\lambda_j^{d_k}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)d_k/2}\} = \operatorname*{diag}_{k\in\{1,\cdots,q\}}\{\lambda_j^{(d_k-d_k^0)}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_k-d_k^0)/2} \times \lambda_j^{d_k^0}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)d_k^0/2}\}\\ &= \Lambda_j(\boldsymbol{d}-\boldsymbol{d}_0)^{-1}\Lambda_j(\boldsymbol{d}_0)^{-1} = \Lambda_j(\boldsymbol{\theta})^{-1}\Lambda_j(\boldsymbol{d}_0)^{-1},\end{aligned}$$

so that we can write

$$\widehat{G}(\boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[\Lambda_j(\boldsymbol{\theta})^{-1}\widehat{G}(\boldsymbol{d}_0)\overline{\Lambda_j(\boldsymbol{\theta})^{-1}}'\big] \tag{A.3}$$

and

$$\begin{aligned}\mathcal{A}(\boldsymbol{d}) &= \left(\frac{2\pi m}{n}\right)^{-2\sum_k \theta_k} \times \det\left\{\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[\Lambda_j(\boldsymbol{\theta})^{-1}\widehat{G}(\boldsymbol{d}_0)\overline{\Lambda_j(\boldsymbol{\theta})^{-1}}'\big]\right\} \\ &= \det\left\{\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta})\widehat{G}(\boldsymbol{d}_0)\overline{M_j(\boldsymbol{\theta})}'\big]\right\},\end{aligned} \tag{A.4}$$

where $M_j(\boldsymbol{\theta}) := \mathrm{diag}_{k\in\{1,\cdots,q\}}\left\{\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)\theta_k/2}(j/m)^{\theta_k}\right\}$. To determine the function h in (i), we first show that

$$\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta})\widehat{G}(\boldsymbol{d}_0)\overline{M_j(\boldsymbol{\theta})}'\big] = \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta})G_0\overline{M_j(\boldsymbol{\theta})}'\big] + o_{\mathbb{P}}(1), \tag{A.5}$$

uniformly in Θ_1. By Lemma 4.1, it follows that

$$\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\Big[M_j(\boldsymbol{\theta})\Big(\widehat{G}(\boldsymbol{d}_0) - G_0\Big)\overline{M_j(\boldsymbol{\theta})}'\Big] = \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta})o_{\mathbb{P}}(1)\overline{M_j(\boldsymbol{\theta})}'\big]. \tag{A.6}$$

The (r,s)-th element in (A.6) is given by

$$\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} o_{\mathbb{P}}(1)\right] = \left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s}\right]o_{\mathbb{P}}(1) \;=\; o_{\mathbb{P}}(1),$$

where the last equality follows from lemma 2 in Robinson (1995b) by taking $\gamma = \theta_r + \theta_s + 1 \in [2\epsilon, 1]$. Therefore, (A.5) follows uniformly in Θ_1. The next step is to derive an approximation to the RHS of (A.5). First, since $\mathrm{e}^{\mathrm{i}(\lambda-\pi)(\theta_r-\theta_s)/2} = \mathrm{e}^{\mathrm{i}\pi(\theta_r-\theta_s)/2} + O(\lambda)$, the (r,s)-th element of the RHS of (A.5), omitting the Re[$\cdot$] operator, can be written as

$$\begin{aligned}&\frac{1}{m}\sum_{j=1}^{m}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} G_0^{rs} = \\ &\qquad = \mathrm{e}^{-\mathrm{i}\pi(\theta_r-\theta_s)/2}\left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} G_0^{rs}\right] + \left[\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} G_0^{rs}\right]O\Big(\frac{m}{n}\Big).\end{aligned} \tag{A.7}$$

Now, from lemma 2 in Robinson (1995b), it follows that

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} &= \frac{1}{1+\theta_r+\theta_s}\left(\frac{2\pi m}{n}\right)^{\theta_r+\theta_s}\left[\frac{1+\theta_r+\theta_s}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s}\right] \\ &= \frac{1}{1+\theta_r+\theta_s}\left(\frac{2\pi m}{n}\right)^{\theta_r+\theta_s}\Big(1+O(m^{-2\epsilon})\Big).\end{aligned} \tag{A.8}$$

Substituting (A.8) on (A.7), it follows that

$$\frac{1}{m}\sum_{j=1}^{m}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{m}\right)^{\theta_r+\theta_s} o_{\mathbb{P}}(1) = \frac{\mathrm{e}^{-\mathrm{i}\pi(\theta_r-\theta_s)/2}}{1+\theta_r+\theta_s}G_0^{rs} + O\left(\frac{m}{n}\right) + O(m^{-2\epsilon}). \tag{A.9}$$

Define the matrices

$$\mathcal{E}(\boldsymbol{\theta}) := \big(\mathrm{e}^{-\mathrm{i}\pi(\theta_r-\theta_s)/2}\big)_{r,s=1}^{q} \quad \text{and} \quad \mathcal{M}(\boldsymbol{\theta}) := \left(\frac{1}{1+\theta_r+\theta_s}\right)_{r,s=1}^{q}$$

and notice that, from (A.9), we have

$$\frac{1}{m}\sum_{j=1}^{m} M_j(\boldsymbol{\theta})G_0\overline{M_j(\boldsymbol{\theta})}' = \mathcal{E}(\boldsymbol{\theta})\odot\mathcal{M}(\boldsymbol{\theta})\odot G_0 + O\left(\frac{m}{n}\right) + O(m^{-2\epsilon}). \tag{A.10}$$

Upon defining,

$$h(\boldsymbol{d}) := \det\Big\{\mathrm{Re}\big[\mathcal{E}(\boldsymbol{\theta})\big] \odot \mathcal{M}(\boldsymbol{\theta}) \odot G_0\Big\},$$

the continuity of the determinant, the finiteness of $\mathcal{E}(\boldsymbol{\theta}), \mathcal{M}(\boldsymbol{\theta})$ and G_0, for $\boldsymbol{\theta} \in \Theta_1$, and (A.10), imply (i). Inequality (ii) follows upon observing that $V(\boldsymbol{\theta}) := \mathrm{Re}\big[\mathcal{E}(\boldsymbol{\theta})\big] \odot \mathcal{M}(\boldsymbol{\theta})$ is positive semidefinite (cf. Shimotsu, 2007, p.293) and, from Oppenheim's inequality,

$$h(\boldsymbol{d}) \geq \prod_{k=1}^{q} V_{kk}(\boldsymbol{\theta}) \det(G_0) = \prod_{k=1}^{q} \mathcal{M}_{kk}(\boldsymbol{\theta}) \det(G_0) = \mathcal{B}(\boldsymbol{d}).$$

By definition, $\mathcal{E}_{rs}(\mathbf{0}) = 1$, for all $r, s = 1, \cdots, q$, so that

$$h(\boldsymbol{d}_0) = \det\big\{\mathcal{M}(\mathbf{0}) \odot G_0\big\} = \mathcal{B}(\boldsymbol{d}_0),$$

and (iii) follows.

Now we move to bound P_2 in (A.1). Expression (A.3) can be used to rewrite $L(\boldsymbol{d})$ as

$$\begin{aligned} L(\boldsymbol{d}) &= \log\big(\det\{\widehat{G}(\boldsymbol{d})\}\big) - \log\big(\det\{\widehat{G}(\boldsymbol{d}_0)\}\big) - 2\sum_{k=1}^{q}\theta_k \frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j) \\ &= \log\Big(\det\big\{\widehat{\mathcal{D}}(\boldsymbol{d})\big\}\Big) - \log\Big(\det\big\{\widehat{\mathcal{D}}(\boldsymbol{d}_0)\big\}\Big), \end{aligned} \tag{A.11}$$

where

$$\widehat{\mathcal{D}}(\boldsymbol{d}) := \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\big[\mathcal{P}_j(\boldsymbol{\theta})^{-1}\widehat{G}(\boldsymbol{d}_0)\overline{\mathcal{P}_j(\boldsymbol{\theta})^{-1}}'\big],$$

with

$$\mathcal{P}_j(\boldsymbol{\theta}) := \underset{k\in\{1,\cdots,q\}}{\mathrm{diag}}\left\{\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)\theta_k/2}\left(\frac{j}{p}\right)^{\theta_k}\right\} \quad \text{and} \quad p := \exp\left(\frac{1}{m}\sum_{j=1}^{m}\log(j)\right),$$

and, as m tends to infinity, $p \sim m/\mathrm{e}$. Observe that $\widehat{\mathcal{D}}(\boldsymbol{d})$ is positive semidefinite since each summand of $\widehat{\mathcal{D}}$ is. For $\kappa \in (0,1)$, define

$$\widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) := \frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{Re}\big[\mathcal{P}_j(\boldsymbol{\theta})^{-1}\widehat{G}(\boldsymbol{d}_0)\overline{\mathcal{P}_j(\boldsymbol{\theta})^{-1}}'\big] \text{ and } \mathcal{Q}_\kappa(\boldsymbol{d}) := \frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{Re}\big[\mathcal{P}_j(\boldsymbol{\theta})G_0\overline{\mathcal{P}_j(\boldsymbol{\theta})}'\big],$$

where $[x]$ denotes the integer part of x. By Lemma 4.1,

$$\widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) - \mathcal{Q}_\kappa(\boldsymbol{d}) = \frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{Re}\big[\mathcal{P}_j(\boldsymbol{\theta})\big(\widehat{G}(\boldsymbol{d}_0) - G_0\big)\overline{\mathcal{P}_j(\boldsymbol{\theta})}'\big] = \frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{Re}\big[\mathcal{P}_j(\boldsymbol{\theta})o_{\mathbb{P}}(1)\overline{\mathcal{P}_j(\boldsymbol{\theta})}'\big].$$

The (r,s)-th element of $\widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) - \mathcal{Q}_\kappa(\boldsymbol{d})$ is then

$$\begin{aligned} \Big(\widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) - \mathcal{Q}_\kappa(\boldsymbol{d})\Big)_{rs} &= \mathrm{Re}\left[\frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{p}\right)^{\theta_r+\theta_s}o_{\mathbb{P}}(1)\right] \\ &= \left(\frac{m}{p}\right)^{\theta_r+\theta_s}\frac{1}{m}\sum_{j=[m\kappa]}^{m}\left(\frac{j}{m}\right)^{\theta_r+\theta_s}o_{\mathbb{P}}(1) \\ &= \left(\frac{m}{p}\right)^{\theta_r+\theta_s}O(1)o_{\mathbb{P}}(1) \;=\; o_{\mathbb{P}}(1), \end{aligned}$$

uniformly in $\boldsymbol{\theta} \in \Theta_2$, where the penultimate equality follows from lemma 5.4 in Shimotsu and Philips (2005). Therefore, as $n \to \infty$, for any $\kappa \in (0,1)$,

$$\sup_{\Theta_2}\big\{\big|\det\{\widehat{\mathcal{D}}_\kappa(\boldsymbol{d})\} - \det\{\mathcal{Q}_\kappa(\boldsymbol{d})\}\big|\big\} = o_{\mathbb{P}}(1).$$

Now, for all $\boldsymbol{d} \in \Theta_2$, one can obtain $\Delta \in (0, 0.1)$ so that, for κ sufficiently small and m sufficiently large (cf. Shimotsu, 2007, p.295),

$$\inf_{\Theta_2} \big\{ \det \big\{ \widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) \big\} \big\} = \inf_{\Theta_2} \big\{ \det \big\{ \mathcal{Q}_\kappa(\boldsymbol{d}) \big\} \big\} + o_{\mathbb{P}}(1) \geq \det\{G_0\}(1+\Delta) + o_{\mathbb{P}}(1).$$

Now, since $\det\big\{\widehat{\mathcal{D}}(\boldsymbol{d}_0)\big\} = \det\big\{\widehat{G}(\boldsymbol{d}_0)\big\} \xrightarrow{\mathbb{P}} \det\{G_0\}$, as $n \to \infty$, by Lemma 4.1, it follows that

$$\mathbb{P}\Big(\inf_{\Theta_2} \big\{ \det \big\{ \widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) \big\} \big\} - \det \big\{ \widehat{\mathcal{D}}(\boldsymbol{d}_0) \big\} \leq 0 \Big) \longrightarrow 0, \quad \text{as } n \to \infty.$$

Also, since $\det\big\{\widehat{\mathcal{D}}(\boldsymbol{d})\big\} \geq \det\big\{\widehat{\mathcal{D}}_\kappa(\boldsymbol{d})\big\}$ we conclude that

$$\mathbb{P}\Big(\inf_{\Theta_2} \big\{ \det \big\{ \widehat{\mathcal{D}}(\boldsymbol{d}) \big\} \big\} - \det \big\{ \widehat{\mathcal{D}}(\boldsymbol{d}_0) \big\} \leq 0 \Big) \longrightarrow 0, \quad \text{as } n \to \infty. \tag{A.12}$$

Since the logarithm is a monotone increasing function of its argument and in view of (A.11), (A.12) implies $P_2 \underset{n\to\infty}{\longrightarrow} 0$ and the proof is complete. ■

Proof of Lemma 4.2:

For fixed $r, s \in \{1, \cdots, q\}$, let $\mathscr{A}_{uv} := \sum_{j=u}^{v} \mathscr{A}_j$ and $\mathscr{B}_{uv} := \sum_{j=u}^{v} \mathscr{B}_j$, where

$$\mathscr{A}_j := \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(d_r^0 - d_s^0)/2} \lambda_j^{d_r^0 + d_s^0} \big[f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j) \big(\overline{A(\lambda_j)}\,'\big)_{\cdot s} \big], \tag{A.13}$$

and

$$\mathscr{B}_j := \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(d_r^0 - d_s^0)/2} \lambda_j^{d_r^0 + d_s^0} \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j) \big(\overline{A(\lambda_j)}\,'\big)_{\cdot s} - G_0^{rs}. \tag{A.14}$$

Hence, for each j, $\mathscr{A}_j + \mathscr{B}_j = \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(d_r^0 - d_s^0)/2} \lambda_j^{d_r^0 + d_s^0} f_n^{rs}(\lambda_j) - G_0^{rs}$. For fixed $u \leq j \leq v$, we have

$$\begin{aligned} \mathbb{E}\big(|\mathscr{A}_j|\big) &= (2\pi j)^{d_r^0 + d_s^0} \mathbb{E}\Big(n^{-d_r^0 - d_s^0} \Big| f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j) \big(\overline{A(\lambda_j)}\,'\big)_{\cdot s} \Big| \Big) \\ &= (2\pi j)^{d_r^0 + d_s^0} O\left(\frac{1}{j^{d_r^0 + d_s^0 + \gamma}} \right) \;=\; O(j^{-\gamma}). \end{aligned}$$

Therefore, $\max_{r,s} \big\{ \mathbb{E}\big(\big| \sum_{j=u}^{v} \mathscr{A}_j \big| \big) \big\} = O(v^{1-\gamma})$ and the result on $\mathscr{A}_{uv}$ follows. As for $\mathscr{B}_j$, from the proof of lemma 1(a) in Shimotsu (2007) (notice that $\mathscr{B}_j$ does not depend on f_n) it follows that $\sum_{j=u}^{v} \mathscr{B}_j = o_{\mathbb{P}}(v)$ uniformly in u and v and hence the desired result on $\mathscr{B}_{uv}$ follows. ■

Proof of Theorem 4.2:

From a careful inspection of the proof of Theorem 4.1, we observe that it suffices to show (with the same notation as in that proof)

$$\frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta}) \widehat{G}(\boldsymbol{d}_0) \overline{M_j(\boldsymbol{\theta})}\,' \big] = \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\big[M_j(\boldsymbol{\theta}) G_0 \overline{M_j(\boldsymbol{\theta})}\,' \big] + o_{\mathbb{P}}(1), \tag{A.15}$$

uniformly in Θ_1 and that $\widehat{\mathcal{D}}_\kappa(\boldsymbol{d}) - \mathcal{Q}_\kappa(\boldsymbol{d}) = o_{\mathbb{P}}(1)$, uniformly in Θ_2. To show (A.15), notice that the (r,s)-th component of the LHS in (A.15) is given by

$$\frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\left[\mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(\theta_r - \theta_s)/2} \left(\frac{j}{m} \right)^{\theta_r + \theta_s} f_n^{rs}(\lambda_j) \Big(\Lambda_j^{(r)}(\boldsymbol{d}_0) \overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}\,' \Big)^{-1} \right].$$

Summation by parts (see Zygmund, 2002, p.3) yields

$$\sup_{\Theta_1} \left\{ \left| \frac{1}{m} \sum_{j=1}^{m} \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(\theta_r - \theta_s)/2} \left(\frac{j}{m} \right)^{\theta_r + \theta_s} \Big[f_n^{rs}(\lambda_j) \Big(\Lambda_j^{(r)}(\boldsymbol{d}_0) \overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}\,' \Big)^{-1} - G_0^{rs} \Big] \right| \right\} \leq$$

$$
\begin{aligned}
&\leq \frac{1}{m}\sum_{k=1}^{m-1}\sup_{\Theta_1}\left\{\left|\mathrm{e}^{\mathrm{i}(\lambda_k-\pi)(\theta_r-\theta_s)/2}\left(\frac{k}{m}\right)^{\theta_r+\theta_s}-\mathrm{e}^{\mathrm{i}(\lambda_{k+1}-\pi)(\theta_r-\theta_s)/2}\left(\frac{k+1}{m}\right)^{\theta_r+\theta_s}\right|\right\}\times\\
&\quad\times\left|\sum_{j=1}^{k}\left[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\right]\right|+\left|\frac{1}{m}\sum_{j=1}^{m}\left[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\right]\right|\\
&\leq \mathcal{C}\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)^{2\epsilon}\frac{1}{k^2}\left|\sum_{j=1}^{k}\left[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\right]\right|+\\
&\quad+\left|\frac{1}{m}\sum_{j=1}^{m}\left[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\right]\right|, \qquad \text{(A.16)}
\end{aligned}
$$

where $0<\mathcal{C}<\infty$ is a constant. Now, from Lemma 4.2,

$$
\begin{aligned}
\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)^{2\epsilon}\frac{1}{k^2}\left|\sum_{j=1}^{k}\left[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\right]\right| &\leq\\
&\leq\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)^{2\epsilon}\frac{1}{k^2}\Big(\big|\mathscr{A}_{1k}\big|+\big|\mathscr{B}_{1k}\big|\Big)\\
&=\frac{1}{m^{2\epsilon}}\sum_{k=1}^{m-1}k^{2(\epsilon-1)}\big|\mathscr{A}_{1k}\big|+\frac{1}{m^{2\epsilon}}\sum_{k=1}^{m-1}k^{2(\epsilon-1)}o_{\mathbb{P}}(k)\\
&=o_{\mathbb{P}}(1)+\frac{m(m-1)^{2\epsilon-1}}{m^{2\epsilon}}\,o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1),
\end{aligned}
$$

uniformly in (r,s), where the penultimate equality follows from Lemma 4.2 which implies

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{m^{2\epsilon}}\sum_{k=1}^{m-1}k^{2(\epsilon-1)}\big|\mathscr{A}_{1k}\big|\right)&=\frac{1}{m^{2\epsilon}}\sum_{k=1}^{m-1}k^{2(\epsilon-1)}\mathbb{E}\Big(\big|\mathscr{A}_{1k}\big|\Big)\\
&\leq\frac{(m-1)^{2\epsilon-1}}{m^{2\epsilon}}\,O\big((m-1)^{1-\gamma}\big)=o(1),
\end{aligned}
$$

uniformly in (r,s). The other term in (A.16) is also $o_{\mathbb{P}}(1)$, uniformly in (r,s), by the same argument and, hence, (A.15) follows. On the other hand, the (r,s)-th element of $\widehat{\mathcal{D}}_\kappa(\boldsymbol{d})-\mathcal{Q}_\kappa(\boldsymbol{d})$ is given by

$$
\begin{aligned}
&\frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{Re}\left[\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{p}\right)^{\theta_r+\theta_s}\Big[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\Big]\right]=\\
&\quad=\left(\frac{m}{p}\right)^{\theta_r+\theta_s}\mathrm{Re}\left[\frac{1}{m}\sum_{j=[m\kappa]}^{m}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\left(\frac{j}{m}\right)^{\theta_r+\theta_s}\Big[f_n^{rs}(\lambda_j)\Big(\Lambda_j^{(r)}(\boldsymbol{d}_0)\overline{\Lambda_j^{(s)}(\boldsymbol{d}_0)}'\Big)^{-1}-G_0^{rs}\Big]\right]\\
&\quad=o_{\mathbb{P}}(1),
\end{aligned}
$$

uniformly in $\boldsymbol{\theta}\in\Theta_2$, where the last equality is derived similarly to (A.16) from summation by parts and lemma 5.4 in Shimotsu and Phillips (2005). This completes the proof. ■

Proof of Corollary 4.1:

From the proof of Lemma 4.2, in order to show (4.3) it suffices to show that, for $\mathscr{A}_j$ as in (A.13) and f_n as in the enunciate, $\mathbb{E}(|\mathscr{A}_j|)=O(j^{-\gamma})$, for some $\gamma>0$. From the proof of lemma 1(a) in Shimotsu (2007) (see also theorem 2 in Robinson, 1995a), we have

$$
\mathbb{E}\big(I_n(\lambda_j)\big)=f(\lambda_j)\left(1+O\left(\frac{\log(j+1)}{j}\right)\right); \qquad \text{(A.17a)}
$$

$$
\mathbb{E}\big(I_{\boldsymbol{\varepsilon}}(\lambda_j)\big)=\frac{\mathrm{I}_q}{2\pi}+O\left(\frac{\log(j+1)}{j}\right); \qquad \text{(A.17b)}
$$

$$\mathbb{E}\big(w_n^r(\lambda_j)\overline{w_{\boldsymbol{\varepsilon}}^s(\lambda_j)}'\big) = \frac{\big(A(\lambda_j)\big)_{r\cdot}}{2\pi} + O\left(\frac{\log(j+1)}{j\lambda_j^{d_r^0}}\right), \quad \text{for } j = 1, \cdots, m. \tag{A.17c}$$

Assumption **A**3 implies

$$\begin{aligned}\mathbb{E}(|\mathscr{A}_j|) &= \mathbb{E}\Bigg(\Bigg|\sum_{|k|\le\ell(n)} W_n^{rs}(k)\left(\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\big[I_n^{rs}(\lambda_{j+k}) - \big(A(\lambda_j)\big)_{r\cdot}I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\big]\right)\Bigg|\Bigg) \\ &\le \sum_{|k|\le\ell(n)} W_n^{rs}(k)\mathbb{E}\Big(\Big|\,\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big[I_n^{rs}(\lambda_{j+k}) - \big(A(\lambda_j)\big)_{r\cdot}I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\Big]\Big|\Big). \end{aligned} \tag{A.18}$$

Rewrite the expression inside the squared brackets on the RHS of (A.18) as

$$\Big[\big(w_n^r(\lambda_{j+k}) - \big(A(\lambda_j)\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)\big)\Big]\overline{w_n^s(\lambda_{j+k})}' + \big(A(\lambda_j)\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)\Big[\overline{w_n^s(\lambda_{j+k})}' - \overline{\big(A(\lambda_j)\big)_{\cdot s}w_{\varepsilon}(\lambda_j)}'\Big].$$

Now, observe that

$$\begin{aligned} w_n^r(\lambda_{j+k})\overline{w_{\boldsymbol{\varepsilon}}^s(\lambda_j)}' &= \frac{1}{2\pi n}\left(\sum_{t=1}^n \boldsymbol{X}_t^{(r)}\mathrm{e}^{\mathrm{i}t\lambda_{j+k}}\right)\left(\sum_{t=1}^n \boldsymbol{\varepsilon}_t^{(s)}\mathrm{e}^{-\mathrm{i}t\lambda_j}\right)' = \frac{1}{2\pi n}\sum_{a=1}^n\sum_{b=1}^n \boldsymbol{X}_a^{(r)}\boldsymbol{\varepsilon}_b^{(s)\prime}\mathrm{e}^{\mathrm{i}(a-b)\lambda_j}\mathrm{e}^{\mathrm{i}a\lambda_k} \\ &= \frac{1}{2\pi n}\sum_{a=1}^n\sum_{b=1}^n\left(\sum_{t=0}^\infty A_t^r\boldsymbol{\varepsilon}_{a-t}^{(r)}\right)\boldsymbol{\varepsilon}_b^{(s)\prime}\mathrm{e}^{\mathrm{i}(a-b)\lambda_j}\mathrm{e}^{\mathrm{i}a\lambda_k} = \frac{1}{2\pi n}\sum_{t=0}^\infty A_t^r\mathrm{e}^{\mathrm{i}t\lambda_j}I_{\boldsymbol{\varepsilon}}^{rs}(\lambda_j)O(1), \end{aligned}$$

so that, by using (A.17b), we conclude that

$$\mathbb{E}\Big(w_n^r(\lambda_{j+k})\overline{w_{\boldsymbol{\varepsilon}}^s(\lambda_j)}'\Big) = \frac{\big(A(\lambda_j)\big)_{r\cdot}}{2\pi}O(1) + O\left(\frac{\log(j+1)}{j\lambda_j^{d_r^0}}\right). \tag{A.19}$$

From (A.17a), (A.17b), (A.17c), $\big(A(\lambda_j)\big)_{r\cdot}\big(\overline{A(\lambda_j)}'\big)_{r\cdot}/2\pi = f^{rr}(\lambda_j)$, $f^{rr}(\lambda_j)^{2d_r^0} \sim G_0^{rr}$, assumption **A**5 and because the function $\log(|x|+1)/|x|$ is increasing in $(-\infty, 0)$ and decreasing in $(0, \infty)$, it follows that

$$\begin{aligned}\mathbb{E}\Big(\Big|\big(w_n^r(\lambda_{j+k}) - \big(A(\lambda_j)\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)\big)\Big|^2\Big) &= \mathbb{E}\Big(I_n^{rr}(\lambda_{j+k}) - w_n^r(\lambda_{j+k})\overline{w_{\boldsymbol{\varepsilon}}(\lambda_j)}'\big(\overline{A(\lambda_j)}'\big)_{r\cdot} + \\ &\qquad + \big(A(\lambda_j)\big)_{r\cdot}I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s} - \big(\overline{A(\lambda_j)}'\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)w_n^r(\lambda_{j+k})\Big) \\ &= O\left(\frac{\log\big(\min\{|j+k|, j\}+1\big)}{\min\{|j+k|, j\}\lambda_{j+k}^{2d_r^0}}\right) = O\big(\psi(j,k,d_r^0)\big),\end{aligned}$$

where $\psi(j,k,d_r^0) := \frac{\log(\min\{|j+k|,j\}+1)}{\min\{|j+k|,j\}\lambda_{j+k}^{2d_r^0}}$, and similarly for $\mathbb{E}\big(\big|\overline{w_n^s(\lambda_j)}' - \overline{\big(A(\lambda_j)\big)_{\cdot s}w_{\varepsilon}(\lambda_j)}'\big|^2\big)$. Also, $\mathbb{E}\big(I_n^{ss}(\lambda_{j+k})\big) = O(\lambda_{j+k}^{-2d_s^0})$. By the Cauchy-Schwartz inequality,

$$\begin{aligned}\mathbb{E}(|\mathscr{A}_j|) &\le \sum_{|k|\le\ell(n)} W_n^{rs}(k)\big|\lambda_j^{d_r^0+d_s^0}\big|\Big[\mathbb{E}\Big(\Big|\big(w_n^r(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)\big)\Big|^2\Big)^{\frac12}\mathbb{E}\big(\big|w_n^r(\lambda_{j+k})\big|^2\big)^{\frac12} + \\ &\qquad + \mathbb{E}\big(\big|\big(A(\lambda_j)\big)_{r\cdot}w_{\boldsymbol{\varepsilon}}(\lambda_j)\big|^2\big)^{\frac12}\mathbb{E}\Big(\Big|\overline{w_n^s(\lambda_j)}' - \overline{\big(A(\lambda_j)\big)_{\cdot s}w_{\varepsilon}(\lambda_j)}'\Big|^2\Big)^{\frac12}\Big] \\ &= \sum_{|k|\le\ell(n)} W_n^{rs}(k)\big|\lambda_j\big|^{d_r^0+d_s^0}O(\sqrt{\psi(j,k,d_r^0)})O(|\lambda_{j+k}|^{-d_s^0}) \\ &= \sum_{|k|\le\ell(n)} W_n^{rs}(k)O\left(\frac{\log\big(\min\{|j+k|,j\}+1\big)^{1/2}}{\min\{|j+k|,j\}^{1/2}}\right)O\left(\left|\frac{\lambda_j}{\lambda_{j+k}}\right|^{d_r^0+d_s^0}\right) = O(j^{-\gamma_0}),\end{aligned}$$

for $0 < \gamma_0 < 1/2$, where the last equality follows from assumption **A**5. This concludes the proof.■

Proof of Lemma 5.1:

(a) For $r, s \in \{1, \cdots, q\}$ fixed, ignoring the maximum in expression (5.2) for a while, its argument can be written as

$$\begin{aligned}\left|\sum_{j=1}^{v} \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big[f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\Big]\right| \leq \\ \leq \sum_{j=1}^{v}\left(\frac{2\pi j}{n}\right)^{d_r^0+d_s^0}\left|f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\right| \\ \leq O\left(\frac{v^{d_r^0+d_s^0}}{n^{d_r^0+d_s^0}}\right) o_{\mathbb{P}}\left(\frac{n^{d_r^0+d_s^0}}{\log(v) v^{d_r^0+d_s^0-1/2}}\right) \;=\; o_{\mathbb{P}}\left(\frac{v^{1/2}}{\log(v)}\right),\end{aligned}$$

uniformly in $1 \leq v \leq m$ and the result follows.

(b) Rewrite the argument of the summation in (5.3) as $\mathscr{A}_j + \mathscr{B}_j + \mathscr{C}_j$, where

$$\begin{aligned}\mathscr{A}_j &:= \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big[f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\Big], \\ \mathscr{B}_j &:= \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big[\big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s} - f_{rs}(\lambda_j)\Big], \\ \mathscr{C}_j &:= \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0} f_{rs}(\lambda_j) - G_0^{rs}.\end{aligned}$$

Part (a) yields $\max_{r,s}\left\{\sum_{j=1}^{v}\left|\mathscr{A}_j\right|\right\} = o_{\mathbb{P}}\Big(v^{1/2}\big(\log(v)\big)^{-1}\Big)$, while, from the proof of lemma 1(b2) in Shimotsu (2007), we obtain $\max_{r,s}\left\{\sum_{j=1}^{v}\left|\mathscr{B}_j\right|\right\} = O_{\mathbb{P}}\big(v^{1/2}\log(v)\big)$. Assumption **C**1 implies $\max_{r,s}\left\{\sum_{j=1}^{v}\left|\mathscr{C}_j\right|\right\} = O\big(v^{\alpha+1}n^{-\alpha}\big)$. The result now follows by noticing that $v^{1/2}\big(\log(v)\big)^{-1} = O\big(v^{1/2}\log(v)\big)$. ■

Proof of Theorem 5.1:

The idea of the proof is similar to that of Lobato (1999) with similar adaptations as in Shimotsu (2007). By hypothesis,

$$\mathbf{0} = \frac{\partial S(\boldsymbol{d})}{\partial \boldsymbol{d}}\bigg|_{\widehat{\boldsymbol{d}}} = \frac{\partial S(\boldsymbol{d})}{\partial \boldsymbol{d}}\bigg|_{\boldsymbol{d}_0} + \left(\frac{\partial^2 S(\boldsymbol{d})}{\partial \boldsymbol{d}\partial \boldsymbol{d}'}\bigg|_{\overline{\mathbf{d}}}\right)(\widehat{\boldsymbol{d}} - \boldsymbol{d}_0),$$

with probability tending to 1, as n tends to infinity, for some $\overline{\mathbf{d}}$ such that $\|\overline{\mathbf{d}} - \boldsymbol{d}_0\|_\infty \leq \|\widehat{\boldsymbol{d}} - \boldsymbol{d}_0\|_\infty$. We observe that $\widehat{\boldsymbol{d}}$ has the stated limiting distribution if

$$\frac{\partial S(\boldsymbol{d})}{\partial \boldsymbol{d}}\bigg|_{\boldsymbol{d}_0} \xrightarrow{\;d\;} N(0, \Omega) \tag{A.20}$$

and

$$\frac{\partial^2 S(\boldsymbol{d})}{\partial \boldsymbol{d}\partial \boldsymbol{d}'}\bigg|_{\overline{\mathbf{d}}} \xrightarrow{\;\mathbb{P}\;} \Omega. \tag{A.21}$$

We shall prove (A.20) first. In order to do that, we use a Crámer-Wold device. Let η be an arbitrary vector in $\mathbb{R}^q$. Observe that, for $r \in \{1, \cdots, q\}$,

$$\sqrt{m}\,\frac{\partial S(\boldsymbol{d})}{\partial d_r} = -\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log(\lambda_j) + \mathrm{tr}\left[\widehat{G}(\boldsymbol{d})^{-1}\sqrt{m}\,\frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_r}\right].$$

Let $\mathrm{I}_{(r)}$ denote a $q \times q$ matrix whose (r,r)-th element is 1 and all other elements are zero. Define a function $\varphi : (0, \infty) \to \mathbb{C}$ by

$$\varphi(x) := \log(x) + \mathrm{i}\left(\frac{x - \pi}{2}\right). \tag{A.22}$$

Since $\Lambda_j(\boldsymbol{d})^{-1} = \underset{k\in\{1,\cdots,q\}}{\operatorname{diag}} \big\{\lambda_j^{d_k} \mathrm{e}^{\mathrm{i}(\lambda_j-\pi)d_k/2}\big\}$ and $\mathrm{Re}\big[(a+\mathrm{i}b)(c+\mathrm{i}d)\big] = ac - bd$, we can write

$$\begin{aligned}
\sqrt{m}\,\frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_r}\Big|_{\boldsymbol{d}_0} &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \mathrm{Re}\Big[\varphi(\lambda_j)\Lambda_j(\boldsymbol{d}_0)^{-1}\mathrm{I}_{(r)} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big] + \\
&\quad + \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \mathrm{Re}\Big[\overline{\varphi(\lambda_j)}\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j)\mathrm{I}_{(r)}\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big] \\
&= \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \log(\lambda_j)\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1}\big(\mathrm{I}_{(r)} f_n(\lambda_j) + f_n(\lambda_j)\mathrm{I}_{(r)}\big)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big] + \\
&\quad + \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \Big[\frac{\lambda_j-\pi}{2}\Big]\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1}\big(-\mathrm{I}_{(r)} f_n(\lambda_j) + f_n(\lambda_j)\mathrm{I}_{(r)}\big)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big], \\
&:= \mathscr{H}_1(r) + \mathscr{H}_2(r). \qquad \text{(A.23)}
\end{aligned}$$

Therefore, from (A.23), we obtain

$$\begin{aligned}
\eta'\sqrt{m}\,\frac{\partial S(\boldsymbol{d})}{\partial \boldsymbol{d}}\Big|_{\boldsymbol{d}_0} &= \sum_{k=1}^{q}\eta_k\sqrt{m}\,\frac{\partial S(\boldsymbol{d})}{\partial d_k}\Big|_{\boldsymbol{d}_0} = \\
&= \sum_{k=1}^{q}\eta_k\Big[-\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log(\lambda_j) + \mathrm{tr}\,\big[\widehat{G}(\boldsymbol{d}_0)^{-1}\mathscr{H}_1(k)\big]\Big] + \sum_{k=1}^{q}\eta_k\,\mathrm{tr}\,\big[\widehat{G}(\boldsymbol{d}_0)^{-1}\mathscr{H}_2(k)\big], \\
&:= \mathscr{R}_1 + \mathscr{R}_2.
\end{aligned}$$

We analyze $\mathscr{R}_1$ first. By letting

$$a_j := \log(\lambda_j) - \frac{1}{m}\sum_{k=1}^{m}\log(\lambda_k) = \log(j) - \frac{1}{m}\sum_{k=1}^{m}\log(k) = O\big(\log(m)\big),$$

we can write

$$\begin{aligned}
-\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log(\lambda_j) + \mathrm{tr}\,\big[\widehat{G}(\boldsymbol{d}_0)^{-1}\mathscr{H}_1(k)\big] &= \mathrm{tr}\,\Big[\widehat{G}(\boldsymbol{d}_0)^{-1}\Big(\mathscr{H}_1(k) - \frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log(\lambda_j)\widehat{G}(\boldsymbol{d}_0)\mathrm{I}_{(k)}\Big)\Big] \\
&= \mathrm{tr}\,\Big[\widehat{G}(\boldsymbol{d}_0)^{-1}\frac{2}{\sqrt{m}}\sum_{j=1}^{m} a_j \mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big]\mathrm{I}_{(k)}\Big]. \qquad \text{(A.24)}
\end{aligned}$$

By Lemma 5.1(b), (A.24) can be written as

$$\Big[(G_0^{-1})_{k\cdot} + o_{\mathbb{P}}(1)\Big]\frac{2}{\sqrt{m}}\sum_{j=1}^{m} a_j\Big(\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big]\Big)_{\cdot k}. \qquad \text{(A.25)}$$

Now, by Lemma 5.1(a),

$$\begin{aligned}
&\Big|\Big(\sum_{j=1}^{m} a_j\Lambda_j(\boldsymbol{d}_0)^{-1}\Big(f_n(\lambda_j) - A(\lambda_j)I_{\boldsymbol{\varepsilon}}(\lambda_j)\overline{A(\lambda_j)}'\Big)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big)_{rs}\Big| \le \\
&\quad \le O\big(\log(m)\big)\max_{v=1,\cdots,m}\Big\{\Big|\sum_{j=1}^{v}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}\Big(f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j)\big(\overline{A(\lambda_j)}'\big)_{\cdot s}\Big)\Big|\Big\} \\
&\quad = O\big(\log(m)\big)o_{\mathbb{P}}\Big(\frac{\sqrt{m}}{\log(m)}\Big) \;=\; o_{\mathbb{P}}(\sqrt{m}),
\end{aligned}$$

uniformly in $r, s \in \{1,\cdots,q\}$. Therefore,

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m} a_j\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' =$$

$$= \frac{1}{\sqrt{m}} \sum_{j=1}^{m} a_j \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' + \frac{1}{\sqrt{m}} o_{\mathbb{P}}(\sqrt{m})$$

$$= \frac{1}{\sqrt{m}} \sum_{j=1}^{m} a_j \Big[\Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' - G_0 \Big] + o_{\mathbb{P}}(1), \tag{A.26}$$

where the last equality follows from $\sum_{j=1}^{m} a_j = 0$. For the moment, ignore the $o_{\mathbb{P}}(1)$ term in (A.25), which will be dealt with later. By doing so and applying (A.26), it follows that

$$\mathscr{R}_1 = \frac{2}{\sqrt{m}} \sum_{k=1}^{q} \eta_k \sum_{j=1}^{m} a_j \Big[\big(G_0^{-1}\big)_{k\cdot} \Big(\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \Big] \Big)_{\cdot k} - 1 \Big] + o_{\mathbb{P}}(1).$$

Writing

$$I_{\boldsymbol{\varepsilon}}(\lambda_j) = \frac{1}{2\pi n} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' + \frac{1}{2\pi n} \sum_{\substack{u,v=1 \\ u \neq v}}^{n} \boldsymbol{\varepsilon}_u \boldsymbol{\varepsilon}_v' \mathrm{e}^{\mathrm{i}(u-v)\lambda_j} := J_1(n) + J_2(j;n), \tag{A.27}$$

we can rewrite $\mathscr{R}_1$ as

$$\mathscr{R}_1 = \frac{2}{\sqrt{m}} \sum_{k=1}^{q} \eta_k \sum_{j=1}^{m} a_j \Big[\big(G_0^{-1}\big)_{k\cdot} \Big(\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) J_1(n) \overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \Big] \Big)_{\cdot k} - 1 \Big] +$$

$$+ \frac{2}{\sqrt{m}} \sum_{k=1}^{q} \eta_k \sum_{j=1}^{m} a_j \Big[\big(G_0^{-1}\big)_{k\cdot} \Big(\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) J_2(j;n) \overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \Big] \Big)_{\cdot k} \Big] + o_{\mathbb{P}}(1). \tag{A.28}$$

The first part of the RHS in (A.28) is $o_{\mathbb{P}}(1)$, which follows (cf. Shimotsu, 2007, p.297) from applying the proof of lemma 2 in appendix D of Lobato (1999), so that we can write

$$\mathscr{R}_1 = \sum_{u=2}^{n} \boldsymbol{\varepsilon}_u' \sum_{v=1}^{t-1} \zeta_{u-v} \boldsymbol{\varepsilon}_v + o_{\mathbb{P}}(1) \quad \text{with} \quad \zeta_t := \frac{1}{\pi n \sqrt{m}} \sum_{j=1}^{m} a_j \mathrm{Re}\Big[\Omega_j \mathrm{e}^{-\mathrm{i}t\lambda_j} + \Omega_j' \mathrm{e}^{\mathrm{i}t\lambda_j} \Big],$$

where

$$\Omega_j := \sum_{k=1}^{q} \eta_k \big(\overline{A(\lambda_j)}' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \big)_{\cdot k} \big(G_0^{-1}\big)_{k\cdot} \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j).$$

By writing

$$\xi_t := \frac{1}{\pi n \sqrt{m}} \sum_{j=1}^{m} a_j \mathrm{Re}\big[\Omega_j + \Omega_j' \big] \cos(t\lambda_j),$$

$\mathscr{R}_1$ can be further simplified, by using the argument in the proof of theorem 2, p.297, in Shimotsu (2007), as

$$\mathscr{R}_1 = \sum_{u=2}^{n} \boldsymbol{\varepsilon}_u' \sum_{v=1}^{t-1} \xi_{u-v} \boldsymbol{\varepsilon}_v + o_{\mathbb{P}}(1). \tag{A.29}$$

As for $\mathscr{R}_2$, proceeding analogously to $\mathscr{R}_1$, we obtain

$$\mathrm{tr}\Big[\widehat{G}(\boldsymbol{d}_0)^{-1} \mathscr{H}_2(k) \Big] = \Big[\big(G_0^{-1}\big)_{k\cdot} + o_{\mathbb{P}}(1) \Big] \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\lambda_j - \pi) \Big(\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \Big] \Big)_{\cdot k}. \tag{A.30}$$

On one hand,

$$\Bigg(\sum_{j=1}^{m} (\lambda_j - \pi) \Lambda_j(\boldsymbol{d}_0)^{-1} \Big(f_n(\lambda_j) - A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}' \Big) \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}' \Bigg)_{rs} =$$

$$= \sum_{j=1}^{m} (\lambda_j - \pi) \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(d_r^0 - d_s^0)/2} \lambda_j^{d_r^0 + d_s^0} \Big(f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j) \big(\overline{A(\lambda_j)}' \big)_{\cdot s} \Big)$$

$$\leq \pi \Big(\frac{m}{n} - 1 \Big) \sum_{j=1}^{m} \mathrm{e}^{\mathrm{i}(\lambda_j - \pi)(d_r^0 - d_s^0)/2} \lambda_j^{d_r^0 + d_s^0} \Big(f_n^{rs}(\lambda_j) - \big(A(\lambda_j)\big)_{r\cdot} I_{\boldsymbol{\varepsilon}}(\lambda_j) \big(\overline{A(\lambda_j)}' \big)_{\cdot s} \Big)$$

$$= O(1)\, o_{\mathbb{P}}\left(\frac{\sqrt{m}}{\log(m)}\right) \;=\; o_{\mathbb{P}}\left(\frac{\sqrt{m}}{\log(m)}\right). \tag{A.31}$$

On the other hand, by using (A.27),

$$\begin{aligned}
\sum_{j=1}^{m} \lambda_j \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' \leq \\
&\leq \frac{2\pi m}{n} \sum_{j=1}^{m} \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' \\
&= \frac{m}{n} \sum_{j=1}^{m} \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) \left(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \right) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' + \\
&\quad + \frac{m}{n} \sum_{j=1}^{m} \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) \left(\frac{1}{n} \sum_{\substack{u,v=1 \\ u \neq v}}^{n} \boldsymbol{\varepsilon}_u \boldsymbol{\varepsilon}_v' \mathrm{e}^{\mathrm{i}(u-v)\lambda_j} \right) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' \\
&:= \mathscr{S}_1 + \mathscr{S}_2.
\end{aligned}$$

From theorem 1 in Heyde and Senata (1972) and assumption **C**2, $\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' = \mathrm{I}_q + O_{\mathbb{P}}(n^{-1/2})$, so that assumption **C**5 implies, for n sufficiently large, $\mathscr{S}_1 = o_{\mathbb{P}}(\sqrt{m})$. As for $\mathscr{S}_2$, by the proof of lemma 1(b2) of Shimotsu (2007),

$$\mathscr{S}_2 = \frac{m}{n}\, O_{\mathbb{P}}(m^{1/2} \log m) = o_{\mathbb{P}}(\sqrt{m}),$$

where the last equality follows by assumption **C**4. Therefore,

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{m} \lambda_j \Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' = o_{\mathbb{P}}(1). \tag{A.32}$$

Combining (A.31) and (A.32), we obtain

$$\begin{aligned}
\frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\lambda_j - \pi) \operatorname{Im}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' \Big] = \\
&= -\frac{\pi}{\sqrt{m}} \sum_{j=1}^{m} \operatorname{Im}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} A(\lambda_j) I_{\boldsymbol{\varepsilon}}(\lambda_j) \overline{A(\lambda_j)}\,' \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}\,' \Big] + o_{\mathbb{P}}(1).
\end{aligned}$$

Again, by ignoring the $o_{\mathbb{P}}(1)$ term in (A.30) and proceeding as in p.298 of Shimotsu (2007) $\mathscr{R}_2$ can be approximated as

$$\mathscr{R}_2 = \sum_{u=2}^{n} \boldsymbol{\varepsilon}_u' \sum_{v=1}^{u-1} \tilde{\xi}_{u-v} \boldsymbol{\varepsilon}_v + o_{\mathbb{P}}(1), \quad \text{with} \quad \tilde{\xi}_t := \frac{1}{2n\sqrt{m}} \sum_{j=1}^{m} \operatorname{Re}\big[\Omega_j - \Omega_j'\big] \sin(t\lambda_j). \tag{A.33}$$

Let

$$Z_1 := 0 \quad \text{and} \quad Z_t := \boldsymbol{\varepsilon}_t' \sum_{u=1}^{t-1} \big(\xi_{t-u} + \tilde{\xi}_{t-u}\big) \boldsymbol{\varepsilon}_u, \quad \text{for } t = 2, 3, \cdots, n.$$

Combining (A.29) and (A.33), we conclude that

$$\sum_{k=1}^{q} \eta_k \sqrt{m}\, \frac{\partial S(\boldsymbol{d})}{\partial d_k}\bigg|_{\boldsymbol{d}_0} = \sum_{t=1}^{n} Z_t + o_{\mathbb{P}}(1).$$

Observing that $\{Z_t\}_{t=1}^{\infty}$ is a zero mean martingale difference, by a standard martingale difference CLT, (A.20) follows if (cf. Billingsley, 1995, theorem 35.12)

$$\sum_{t=1}^{n} \mathbb{E}(Z_t^2 | \mathscr{F}_{t-1}) - \sum_{k=1}^{q} \sum_{l=1}^{q} \eta_k \eta_l \Omega_{kl} \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \to \infty, \tag{A.34}$$

and

$$\sum_{t=1}^{n} \mathbb{E}\big(Z_t^2 I(|Z_t| > \delta) \xrightarrow{\mathbb{P}} 0, \qquad \text{for all } \delta > 0, \text{ as } n \to \infty. \tag{A.35}$$

Both results follow from the arguments in the proof of theorem 2, p.298, in Shimotsu (2007), since (A.34) and (A.35) are (29) and (30), respectively, in the aforementioned paper. Also observe that the $o_{\mathbb{P}}(1)$ terms in (A.25) and (A.30) do not affect the results, since, from our arguments,

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{m} a_j \mathrm{Re}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big] \quad \text{and} \quad \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \mathrm{Im}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'\Big]$$

are both $O_{\mathbb{P}}(1)$. This completes the proof of (A.20).

We now move to show (A.21). For fixed $\delta > 0$, let $\boldsymbol{\theta} := \boldsymbol{d} - \boldsymbol{d}_0$ and define

$$\mathcal{M} := \Big\{\boldsymbol{d} : \log(n)^4 \|\boldsymbol{d} - \boldsymbol{d}_0\|_\infty < \delta\Big\} = \Big\{\boldsymbol{\theta} : \log(n)^4 \|\boldsymbol{\theta}\|_\infty < \delta\Big\}.$$

First we show that $\mathbb{P}\big(\overline{\mathbf{d}} \in \mathcal{M}\big) \to 1$, as $n \to \infty$. Assuming the same notation as in the proof of Theorem 4.2, recall the decomposition of $L(\boldsymbol{d}) = S(\boldsymbol{d}) - S(\boldsymbol{d}_0)$ given in expression (A.2). By applying the same argument as in the proof of Theorem 4.2, we first obtain

$$\inf_{\Theta_1 \setminus \mathcal{M}} \big\{\mathcal{R}(\boldsymbol{d})\big\} \geq \delta^2 \log(n)^8,$$

and upon applying Lemma 5.1, we obtain

$$\sup_{\Theta_1} \Big\{\big|\mathcal{A}(\boldsymbol{d}) - h(\boldsymbol{d})\big|\Big\} = O_{\mathbb{P}}\bigg(\frac{m^\alpha}{n^\alpha} + \frac{\log(m)}{m^{2\epsilon}} + \frac{m}{n}\bigg).$$

It follows, uniformly in Θ_1 (cf. Shimotsu, 2007, p.300), that

$$\begin{aligned}
\log\big(\mathcal{A}(\boldsymbol{d})\big) - \log\big(\mathcal{B}(\boldsymbol{d})\big) &\geq \log\big(h(\boldsymbol{d}) + o_{\mathbb{P}}\big(\log(n)^{-8}\big)\big) - \log\big(h(\boldsymbol{d})\big) = o_{\mathbb{P}}\big(\log(n)^{-8}\big) \\
\log\big(\mathcal{A}(\boldsymbol{d}_0)\big) - \log\big(\mathcal{B}(\boldsymbol{d}_0)\big) &= \log\big(h(\boldsymbol{d}_0) + o_{\mathbb{P}}\big(\log(n)^{-8}\big)\big) - \log\big(h(\boldsymbol{d}_0)\big) = o_{\mathbb{P}}\big(\log(n)^{-8}\big).
\end{aligned}$$

Hence, $\mathbb{P}\big(\inf_{\Theta_1 \setminus \mathcal{M}} L(\boldsymbol{d}) \leq 0\big) \to 0$ and $\mathbb{P}\big(\overline{\mathbf{d}} \in \mathcal{M}\big) \to 1$, as $n \to \infty$, follows.

Now, observe that

$$\frac{\partial^2 S(\boldsymbol{d})}{\partial d_r \partial d_s} = \mathrm{tr}\bigg[-\widehat{G}(\boldsymbol{d})^{-1} \frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_r} \widehat{G}(\boldsymbol{d})^{-1} \frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_s} + \widehat{G}(\boldsymbol{d})^{-1} \frac{\partial^2 \widehat{G}(\boldsymbol{d})}{\partial d_r \partial d_s}\bigg].$$

The derivatives of $\widehat{G}(\boldsymbol{d})$ are given by

$$\frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_r} = \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\varphi(\lambda_j) \mathrm{I}_{(r)} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] + \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\overline{\varphi(\lambda_j)} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \mathrm{I}_{(r)}\Big], \tag{A.36}$$

and

$$\begin{aligned}
\frac{\partial^2 \widehat{G}(\boldsymbol{d})}{\partial d_r \partial d_s} = &\frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\varphi(\lambda_j)^2 \mathrm{I}_{(r)} \mathrm{I}_{(s)} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] + \\
&+ \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\big|\overline{\varphi(\lambda_j)}\big|^2 \mathrm{I}_{(r)} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \mathrm{I}_{(s)}\Big] + \\
&+ \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\big|\overline{\varphi(\lambda_j)}\big|^2 \mathrm{I}_{(s)} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \mathrm{I}_{(r)}\Big] + \\
&+ \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}\Big[\overline{\varphi(\lambda_j)}^2 \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \mathrm{I}_{(r)} \mathrm{I}_{(s)}\Big],
\end{aligned}$$

where φ is given by (A.22). For $k = 0, 1, 2$, let

$$\mathcal{R}_k(\boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m} \log(\lambda_j)^k \mathrm{Re}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big]$$
$$\mathcal{I}_k(\boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m} \log(\lambda_j)^k \mathrm{Im}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big],$$

so that, we can write

$$\frac{\partial \widehat{G}(\boldsymbol{d})}{\partial d_r} = \mathrm{I}_{(r)}\mathcal{R}_1(\boldsymbol{d}) + \mathcal{R}_1(\boldsymbol{d})\mathrm{I}_{(r)} + \frac{\pi}{2}\Big(\mathrm{I}_{(r)}\mathcal{I}_0(\boldsymbol{d}) - \mathcal{I}_0(\boldsymbol{d})\mathrm{I}_{(r)}\Big) + o_{\mathbb{P}}\left(\frac{1}{\log(n)}\right), \tag{A.37}$$

and

$$\begin{aligned}\frac{\partial^2 \widehat{G}(\boldsymbol{d})}{\partial d_r \partial d_s} = \frac{\pi^2}{4}\Big[&\mathrm{I}_{(r)}\mathrm{I}_{(s)}\mathcal{R}_0(\boldsymbol{d}) + \mathrm{I}_{(r)}\mathcal{R}_0(\boldsymbol{d})\mathrm{I}_{(s)} + \mathrm{I}_{(s)}\mathcal{R}_0(\boldsymbol{d})\mathrm{I}_{(r)} + \mathcal{R}_0(\boldsymbol{d})\mathrm{I}_{(s)}\mathrm{I}_{(r)}\Big] + \\ &+ \pi\Big[\mathrm{I}_{(r)}\mathrm{I}_{(s)}\mathcal{I}_1(\boldsymbol{d}) + \mathcal{I}_1(\boldsymbol{d})\mathrm{I}_{(r)}\mathrm{I}_{(s)}\Big] + \mathrm{I}_{(r)}\mathrm{I}_{(s)}\mathcal{R}_2(\boldsymbol{d}) + \mathrm{I}_{(r)}\mathcal{R}_2(\boldsymbol{d})\mathrm{I}_{(s)} + \\ &+ \mathrm{I}_{(s)}\mathcal{R}_2(\boldsymbol{d})\mathrm{I}_{(r)} + \mathcal{R}_2(\boldsymbol{d})\mathrm{I}_{(s)}\mathrm{I}_{(r)} + o_{\mathbb{P}}(1).\end{aligned} \tag{A.38}$$

The order of the remainder term in (A.37) is obtained as follows. Rewrite the first term on the RHS of (A.36) as

$$\begin{aligned}&\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\Big[\varphi(\lambda_j)\mathrm{I}_{(r)}\Big(\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] + \mathrm{i}\,\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big]\Big)\Big] = \\ &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\Big[\log(\lambda_j)\mathrm{I}_{(r)}\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] + \mathrm{i}\log(\lambda_j)\mathrm{I}_{(r)}\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] + \\ &\quad + \mathrm{i}\left(\frac{\lambda_j - \pi}{2}\right)\mathrm{I}_{(r)}\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] - \left(\frac{\lambda_j - \pi}{2}\right)\mathrm{I}_{(r)}\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big]\Big] \\ &= \frac{1}{m}\sum_{j=1}^{m} \Big[\log(\lambda_j)\mathrm{I}_{(r)}\mathrm{Re}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] - \left(\frac{\lambda_j - \pi}{2}\right)\mathrm{I}_{(r)}\mathrm{Im}\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big]\Big].\end{aligned}$$

By summation by parts, Lemma 5.1 and assumption **C**4

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m} \lambda_j\Big[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\Big] &\le \frac{1}{m}\sum_{j=1}^{m-1} |\lambda_j - \lambda_{j+1}|\left\|\sum_{k=1}^{j} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\right\|_\infty + \\ &\quad + \frac{\lambda_m}{m}\left\|\sum_{k=1}^{j} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}'\right\|_\infty \\ &\le \frac{1}{m}\frac{m-1}{n}\left[O_{\mathbb{P}}\left((m-1)^{1/2}\log(m-1) + \frac{(m-1)^{\alpha+1}}{n^\alpha}\right) + O\left(\frac{1}{m}\right)\right] + \\ &\quad + O\left(\frac{1}{n}\right)\left[O_{\mathbb{P}}\left(m^{1/2}\log(m) + \frac{m^{\alpha+1}}{n^\alpha}\right) + O\left(\frac{1}{m}\right)\right] \\ &= O_{\mathbb{P}}\left(\frac{m^{1/2}\log(m)}{n} + \frac{m^\alpha}{n^\alpha}\right) + O\left(\frac{1}{mn}\right) \;=\; o_{\mathbb{P}}\left(\frac{1}{\log(n)}\right),\end{aligned}$$

where the last equality follows from

$$\frac{m^{1/2}\log(m)}{n}\,\log(n) = \frac{m^{1/2}}{n^{1/2}}\,\frac{\log(m)}{n^{1/4}}\,\frac{\log(n)}{n^{1/4}} \longrightarrow 0,$$

as n tends to infinity, by assumption **C**4. The other term in (A.36) is dealt analogously. The remainder term in (A.38) involves

$$\frac{1}{m}\sum_{j=1}^{m} \rho(\lambda_j)\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d})^{-1}}',$$

for $\rho(\lambda_j)$ proportional to λ_j, λ_j^2 and $\lambda_j \log(\lambda_j)$. The order of the terms proportional to λ_j has already been obtained, while the one proportional to λ_j^2 is dealt analogously since $\lambda_j^2 = O(\lambda_j)$. The term proportional to $\lambda_j \log(\lambda_j)$ is $o_{\mathbb{P}}(1)$ since, by summation by parts, Lemma 5.1 and assumption **C**4,

$$\begin{aligned}
\left\| \frac{1}{m} \sum_{j=1}^{m} \lambda_j \log(\lambda_j) \left[\Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \right] \right\|_\infty \leq \\
\leq \frac{1}{m} \sum_{j=1}^{m-1} \left| \lambda_j \log(\lambda_j) - \lambda_{j+1} \log(\lambda_{j+1}) \right| \left\| \sum_{k=1}^{j} \Lambda_k(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_k(\boldsymbol{d})^{-1}}' \right\|_\infty + \\
+ \frac{\lambda_m \log(\lambda_m)}{m} \left\| \sum_{j=1}^{m} \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' \right\|_\infty \\
= \frac{1}{m} O_{\mathbb{P}} \left((m-1)^{1/2} \log(m-1) + \frac{(m-1)^{\alpha+1}}{n^\alpha} \right) o(1) + o_{\mathbb{P}}(1) \;=\; o_{\mathbb{P}}(1).
\end{aligned}$$

In order to finish the proof, it suffices to show that,

$$\mathcal{R}_k(\boldsymbol{d}) = G_0 \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j)^k + o_{\mathbb{P}}\big(\log(n)^{k-2} \big) \tag{A.39}$$

and

$$\mathcal{I}_k(\boldsymbol{d}) = o_{\mathbb{P}}\big(\log(n)^{k-2} \big), \tag{A.40}$$

uniformly in $\boldsymbol{d} \in \mathcal{M}$. Indeed, if (A.39) and (A.40) hold, upon defining for a matrix M,

$$T_1(M, r) := \mathrm{I}_{(r)} M + M \mathrm{I}_{(r)}, \qquad T_2(M, r, s) := \mathrm{I}_{(r)} \mathrm{I}_{(s)} M + \mathrm{I}_{(r)} M \mathrm{I}_{(s)} + \mathrm{I}_{(s)} M \mathrm{I}_{(r)} + M \mathrm{I}_{(r)} \mathrm{I}_{(s)}$$

and

$$T_3(M, r, s) := -\mathrm{I}_{(r)} \mathrm{I}_{(s)} M + \mathrm{I}_{(r)} M \mathrm{I}_{(s)} + \mathrm{I}_{(s)} M \mathrm{I}_{(r)} - M \mathrm{I}_{(r)} \mathrm{I}_{(s)},$$

it follows that (cf. Shimotsu, 2007, p.301)

$$\widehat{G}(\overline{\mathbf{d}})^{-1} = G_0^{-1} + o_{\mathbb{P}}\big(\log(n)^{-2} \big), \qquad \frac{\partial \widehat{G}(\overline{\mathbf{d}})}{\partial d_r} = \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j) T_1(G_0, r) + o_{\mathbb{P}}\big(\log(n)^{-1} \big),$$

$$\text{and} \quad \frac{\partial^2 \widehat{G}(\overline{\mathbf{d}})}{\partial d_r \partial d_s} = \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j)^2 T_2(G_0, r, s) + \frac{\pi^2}{4} T_3(G_0, r, s) + o_{\mathbb{P}}(1).$$

Since $\mathrm{tr}\left[G_0^{-1} T_1(G_0, r) G_0^{-1} T_1(G_0, s) \right] = \mathrm{tr}\left[G_0^{-1} T_2(G_0, r, s) \right]$ and

$$\frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j)^2 - \left(\frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j) \right)^2 \longrightarrow 1,$$

we obtain

$$\frac{\partial^2 S(\boldsymbol{d})}{\partial d_r \partial d_s} = \mathrm{tr}\left[G_0^{-1} T_2(G_0, r, s) + \frac{\pi^2}{4} G_0^{-1} T_3(G_0, r, s) \right] + o_{\mathbb{P}}(1),$$

from where (A.21) follows. We proceed to show (A.39) and (A.40). For $k = 0, 1, 2$, let

$$\mathcal{F}_k(\boldsymbol{\theta}) := \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j)^k \Lambda_j(\boldsymbol{\theta})^{-1} G_0 \overline{\Lambda_j(\boldsymbol{\theta})^{-1}}'.$$

Then, (A.39) and (A.40) follow if

$$\sup_{\boldsymbol{d} \in \mathcal{M}} \left\{ \left\| \frac{1}{m} \sum_{j=1}^{m} \log(\lambda_j)^k \Lambda_j(\boldsymbol{d})^{-1} f_n(\lambda_j) \overline{\Lambda_j(\boldsymbol{d})^{-1}}' - \mathcal{F}_k(\boldsymbol{\theta}) \right\|_\infty \right\} = o_{\mathbb{P}}\big(\log(n)^{k-2} \big), \tag{A.41}$$

and

$$\sup_{\boldsymbol{d}\in\mathcal{M}}\left\{\left\|\mathcal{F}_k(\boldsymbol{\theta})-G_0\frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j)^k\right\|_\infty\right\}=o\big(\log(n)^{k-2}\big). \tag{A.42}$$

Following Shimotsu (2007), p.302, notice that, by applying (A.3), we can rewrite (A.41) as

$$\sup_{\boldsymbol{d}\in\mathcal{M}}\left\{\left\|\frac{1}{m}\sum_{j=1}^{m}\log(\lambda_j)^k\Lambda_j(\boldsymbol{\theta})^{-1}\Big[\Lambda_j(\boldsymbol{d}_0)^{-1}f_n(\lambda_j)\overline{\Lambda_j(\boldsymbol{d}_0)^{-1}}'-G_0\Big]\overline{\Lambda_j(\boldsymbol{\theta})^{-1}}'\right\|_\infty\right\}.$$

Define $b_j(\boldsymbol{\theta};k):=\log(\lambda_j)^k\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(\theta_r-\theta_s)/2}\lambda_j^{\theta_r+\theta_s}$, for $k=0,1,2$. Then, by omitting the supremum, the (r,s)-th element of (A.41) is equal to

$$\begin{aligned}\left|\frac{1}{m}\sum_{j=1}^{m}b_j(\boldsymbol{\theta};k)\Big[\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}f_n^{rs}(\lambda_j)-G_0^{rs}\Big]\right|\leq\\ \leq\frac{1}{m}\sum_{j=1}^{m-1}\Big|b_j(\boldsymbol{\theta};k)-b_{j+1}(\boldsymbol{\theta};k)\Big|\left|\sum_{l=1}^{j}\mathrm{e}^{\mathrm{i}(\lambda_l-\pi)(d_r^0-d_s^0)/2}\lambda_l^{d_r^0+d_s^0}f_n^{rs}(\lambda_l)-G_0^{rs}\right|\\ +\frac{b_m(\boldsymbol{\theta};k)}{m}\left|\sum_{j=1}^{m}\mathrm{e}^{\mathrm{i}(\lambda_j-\pi)(d_r^0-d_s^0)/2}\lambda_j^{d_r^0+d_s^0}f_n^{rs}(\lambda_j)-G_0^{rs}\right|,\end{aligned} \tag{A.43}$$

where the inequality follows from summation by parts. Now, since

$$b_j(\boldsymbol{\theta};k)-b_{j+1}(\boldsymbol{\theta};k)=O\left(\frac{\log(n)^k}{j}\right)\quad\text{and}\quad b_m(\boldsymbol{d};k)=O\big(\log(n)^k\big),$$

uniformly in $\boldsymbol{\theta}\in\mathcal{M}$, for any $k=0,1,2$, it follows by Lemma 5.1 and Remark 5.1 that (A.43) can be rewritten as

$$\begin{aligned}O\left(\frac{\log(n)^k}{m}\right)\frac{1}{m}O_{\mathbb{P}}\left(\frac{m^{\alpha+1}}{n^\alpha}+m^{1/2}\log(m)\right)+\\ +O\big(\log(n)^k\big)\frac{1}{m}O_{\mathbb{P}}\left(\frac{m^{\alpha+1}}{n^\alpha}+m^{1/2}\log(m)\right)=o_{\mathbb{P}}\big(\log(n)^{k-2}\big),\end{aligned}$$

where the last equality follows from assumption **C**4 (see also Remark 5.1), because

$$\begin{aligned}\log(n)^2\frac{1}{m}O_{\mathbb{P}}\left(\frac{m^{\alpha+1}}{n^\alpha}+m^{1/2}\log(m)\right)&=\left[\frac{\log(n)^2}{m^{1/2}\log(m)}+\frac{\log(m)}{m^{1/2}}\frac{\log(n)^2}{m^{1/2}}\right]O_{\mathbb{P}}(1)\\ &=o(1)O_{\mathbb{P}}(1)=o_{\mathbb{P}}(1).\end{aligned}$$

The other term is dealt analogously, so that (A.41) follows. As for (A.42), the result follows from the proof of theorem 2, p.302, in Shimotsu (2007) (notice that it does not depend on f_n). This completes the proof. ■

References

[1] Billingsley, P. (1995). *Probability and Measure*. 3rd ed., New York: John Wiley.

[2] Chiriac, R. and Voev, V. (2011). "Modeling and Forecasting Multivariate Realized Volatility". *Journal of Applied Econometrics*, **26**(6), 922-947.

[3] Fox, R. and Taqqu, M.S. (1986). "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, **14**(2), 517-532.

[4] Giraitis, L. and Surgailis, D. (1990)."A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotic Normality of Whittle's Estimate". *Probability Theory and Related Fields*, **86**, 87-104.

[5] Grenander, U. (1951). "On Empirical Spectral Analysis of Stochastic Process". *Arkiv för Matematik*, **1**, 197-277.

[6] Grenander, U. and Rosenblatt, M. (1953). "Statistical Spectral Analysis of Time Series Arising From Stationary Stochastic Processes". *Annals of Mathematical Statistics*, **24**, 537-558.

[7] Heyde, C.C. and Senata, E. (1972). "Estimation Theory for Growth and Immigration Rates in a Multiplicative Process". *Journal of Applied Probability*, **9**, 235-256.

[8] Künsch, H. (1987). "Statistical Aspects of Self-Similar Processes". In Prokhorov, Yu. and Sazanov, V.V. (Eds.), *Proceedings of the First World Congress of the Bernoulli Society*. Utrecht: VNU Science Press, 67-74.

[9] Lobato, I.N. (1999). "A Semiparametric Two-Step Estimator in a Multivariate Long Memory Model". *Journal of Econometrics*, **90**, 129-153.

[10] Luceño, A. (1996). "A Fast Likelihood Approximation for Vector General Linear Processes with Long Series: Application to Fractional Differencing". *Biometrika*, **83**(3), 603-614.

[11] Nielsen, F.S. (2011). "Local Whittle Estimation of Multi-Variate Fractionally Integrated Processes". *Journal of Time Series Analysis*, **32**(3), 317-335.

[12] Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

[13] Robinson, P.M. (1995a). "Log-Periodogram Regression of Time Series with Long Range Dependence". *Annals of Statistics*, **23**(3), 1048-1072.

[14] Robinson, P.M. (1995b). "Gaussian Semiparametric Estimation of Long Range Dependence". *Annals of Statistics*, **23**(5), 1630-1661.

[15] Shimotsu, K. (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes". *Journal of Econometrics*, **137**, 277-310.

[16] Shimotsu, K. and Phillips, P.C.B. (2005). "Exact Local Whittle Estimation of Fractional Integration". *Annals of Statistics*, **33**(4), 1890-1933.

[17] Sowell, F. (1989). "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models". Working Paper, Carnegie-Mellon University.

[18] Tsay, W-J. (2010). "Maximum Likelihood Estimation of Stationary Multivariate ARFIMA Processes". *Journal of Statistical Computation and Simulation*, **80**(7-8), 729-745.

[19] Zygmund, A. (2002). *Trigonometric Series. Vol. I, II*. Cambridge: Cambridge University Press.

Anexo D

Artigo Lopes et al. (2011) e Addendum

Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes

Sílvia R.C. Lopes[‡], Guilherme Pumi and Karine Zaniol
Mathematics Institute
Federal University of Rio Grande do Sul

September 14, 2011

Abstract

In this work we present an extensive simulation study on Mallows distance in the context of Gaussian and non-Gaussian VARFIMA processes. Our main goal is to analyze the dependence among the components of VARFIMA processes through the Mallows distance point of view. A possible relationship between the Mallows distance and the fractional differencing parameter $\boldsymbol{d}$, the type and level of dependence in the innovation process as well as its marginal behavior is investigated. We study the behavior of the Kendall's τ dependence coefficient under the same framework for comparison purposes. For the Mallows distance, we consider an estimator based on the empirical marginal distribution function. Based on our simulation results, we propose both a semiparametric estimator for the fractional differencing parameter and a testing procedure to assess the presence of strong long range dependence in the components of VARFIMA processes of any (finite) dimension.

Keywords: Mallows Distance; VARFIMA Processes; Copulas; Empirical Estimation; Long Range Dependence, Kendall's τ Coefficient, Semiparametric Estimation, Hypothesis Test.

Mathematical Subject Classification (2010). Primary 62H20, 60G10, 62M10, 62E10;

1 Introduction

The Mallows distance was introduced by Mallows (1972) as a tool to prove the asymptotic normality for sums of independent random variables. After this work, several other applications for the Mallows distance were found, especially in proving convergence of random variables and certain central limit theorem-type results. In Bickel and Freedman (1981) the Mallows distance is used as a tool to provide asymptotic results for the bootstrap technique. An account of the theory and history of the Mallows distance can also be found there and references therein.

By its turn, the class of VARFIMA processes was introduced by Sowell (1989). It can be seen as a natural multidimensional extension of the classical ARFIMA processes, on which each component follows an ARFIMA process (see, for instance, Lopes, 2008 and Sena Junior, Reisen and Lopes, 2006), but the components in the innovation process can be correlated to each other. VARFIMA processes have been applied in a variety of fields such as hydrology, econometrics, statistics among others. Applications include modeling of stock prices' volatility, modeling and forecasting high frequency data, among others. See, for instance, Chiriac and Voev (2011), Diongue (2010) and references therein.

Our goal in this work is to investigate through extensive Monte Carlo simulations the Mallows distance behavior among the components of VARFIMA processes in several different settings. More specifically, we are interested in a possible relationship between the Mallows distance and the fractional differencing parameter $\boldsymbol{d}$, the type and level of dependence induced in the innovation process as well as its marginal behavior.

[‡]Corresponding author. E-mail: silvia.lopes@ufrgs.br

The study is based on Monte Carlo simulations of VARFIMA$(0, \boldsymbol{d}, 0)$ processes with Gaussian innovations as well as parametric copula-type innovations. Different dependence parameters and different types of marginal behavior are also considered. In the Gaussian innovation case, the degree of dependence is measured by the correlation coefficient. In the copula innovations case, the degree of dependence is measured by its parameter. In the simulations we also consider the Kendall's τ dependence coefficient, which is used as a benchmark to compare with the results obtained by using the Mallows distance.

The Mallows distance estimator considered in the simulations is introduced in Section 3. It is based on the empirical quantile function of the process' marginals. Time series are generated by using the infinite moving average representation of the individual components in the process, truncated at a certain cut-off point. In the simulations, several different features are studied in the Mallows distance point of view. These features are usually introduced directly into the innovation process. Some copula tools are also explored in order to separate the joint dependence in the process from its marginals. The advantage of splitting the process' joint behavior from its marginal structure is the possibility to compare time series with the exactly same joint behavior but completely different marginal configurations.

Our simulation results are applied to construct a semiparametric estimator for the fractional differencing parameter of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any (finite) dimension, including the univariate case. The main advantage of the estimator is to be fast while providing good estimates. The estimator is especially useful in very high dimensional estimation problems. In those cases, most estimators of the fractional differencing parameter depend on maximization procedures which require an initial guess for the value of $\boldsymbol{d}$. Usually a poor choice of the initial point will lead to slow convergence of the procedure. Given that the estimator we propose is easy to implement, fast to calculate and usually provide reasonable estimates, it is a good candidate for the initial guess necessary in the mentioned maximization procedures. In high dimension, this simple procedure may provide a significant reduction in computational time. We assess the estimator's performance through simple Monte Carlo experiments. The obtained simulations results are further applied to construct an approximate test for the presence of strong long range dependence in the components of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any (finite) dimension, including the univariate case. The performance of the test is also assessed by simple Monte Carlo experiments.

The paper is organized as follows. In the next section we present some preliminary concepts and results necessary for this work. In Section 3 we introduce the Mallows distance estimator considered in the simulation studies. In Section 4 we present the simulation results on the Mallows distance among the components of VARFIMA processes in several different settings. In Section 5 we present the simulation results, in the same configuration as in Section 4, by considering the Kendall's τ coefficient as dependence measure, instead of the Mallows distance. We also compare the results with the ones obtained in Section 4. As an application of the estimation results, we present in Section 6 the construction of a semiparametric estimator for the fractional differencing parameter as well as a simple test to assess the presence of strong long range dependence in the components of VARFIMA processes. Conclusions and final remarks are reserved to Section 7.

2 Preliminaries Concepts and Results

In this section we present some basic definitions and results necessary for this work. For $\alpha > 0$, let $\mathscr{F}_\alpha$ denote the space of all distribution functions satisfying $\int_{\mathbb{R}} |x|^\alpha \mathrm{d}F < \infty$.

Definition 2.1. (Mallows α-distance). Let $\alpha > 0$ and let F and G be two distribution functions in $\mathscr{F}_\alpha$. The *Mallows α-distance* of F and G is given by

$$\mathscr{D}_\alpha(F,G) := \inf_{A(F,G)} \Big\{ \mathbb{E}\big(|X-Y|^\alpha\big)^{\frac{1}{\alpha}} \Big\}, \tag{2.1}$$

where $A(F,G) := \big\{(X,Y) : X \sim F, Y \sim G\big\}$, that is, $A(F,G)$ is the set of all pairs (X,Y) of random variables with marginals given by F and G, respectively.

It can be shown that, for $\alpha \geq 1$, $\mathscr{D}_\alpha(\cdot,\cdot)$ is a metric in $\mathscr{F}_\alpha$, while for $\alpha < 1$, $\mathscr{D}_\alpha^\alpha(\cdot,\cdot)$ is a metric in $\mathscr{F}_\alpha$ (see Bickel and Freedman, 1981). If $\alpha \geq 1$, $\mathscr{D}_\alpha(\cdot,\cdot)$ given by (2.1) can be shown to be equivalent to a much simpler expression (cf. Major, 1978) as follows: let $U \sim U(0,1)$ be a uniformly distributed random variable, let $F, G \in \mathscr{F}_\alpha$ and set $X^* := F^{-1}(U)$ and $Y^* := G^{-1}(U)$. Then, it can be shown that

$$\mathscr{D}_\alpha^\alpha(F,G) = \mathbb{E}\big(|X^* - Y^*|^\alpha\big). \tag{2.2}$$

Alternative expressions for (2.2) are the following (see Rachev and Rüschendorf, 1998 and Major, 1978):

$$\mathscr{D}_\alpha^\alpha(F,G) = \iint_{\mathbb{R}^2} |x-y|^\alpha \mathrm{d}\mu = \int_0^1 \big|F^{-1}(u) - G^{-1}(u)\big|^\alpha \mathrm{d}u, \tag{2.3}$$

where μ is the probability measure in $\mathbb{R}^2$ with joint distribution function given by $H(x,y) = \min\{F(x), G(y)\}$, that is, μ is defined in the semi-ring of rectangles $R = [x_1, x_2] \times [y_1, y_2] \subseteq \mathbb{R}^2$ by

$$\begin{aligned} \mu(R) \;=\; & \min\{F(x_1), G(y_1)\} + \min\{F(x_2), G(y_2)\} - \\ & - \min\{F(x_1), G(y_2)\} - \min\{F(x_2), G(y_1)\}. \end{aligned}$$

The last expression in (2.3) will be useful in defining the estimator introduced on Section 3.

Now let X and Y be two continuous random variables defined in a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let (X_1, Y_1) and (X_2, Y_2) be two independent copies of (X,Y). Given $\omega \in \Omega$, the pairs $\big(X_1(\omega), Y_1(\omega)\big)$ and $\big(X_2(\omega), Y_2(\omega)\big)$ are called *concordant* if $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) > 0$ and *discordant* if $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) < 0$ (equality happens with probability 0).

Definition 2.2. The *Kendall's τ coefficient* between X and Y, denoted by $\tau_{X,Y}$ (or simply by τ if no confusion is possible), is defined as the probability of concordance minus the probability of discordance, that is,

$$\tau = \tau_{X,Y} := \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) > 0\big) - \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) < 0\big).$$

Next we define the class of the so-called VARFIMA$(p, \boldsymbol{d}, q)$ processes.

Definition 2.3. An m-dimensional process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ with mean $\boldsymbol{\mu}$ is called a VARFIMA$(p, \boldsymbol{d}, q)$ process if it is a stationary solution of the difference equations

$$\boldsymbol{\Phi}(\mathcal{B})\,\mathrm{diag}\big\{(1-\mathcal{B})^{\boldsymbol{d}}\big\}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(\mathcal{B})\boldsymbol{\varepsilon}_t, \tag{2.4}$$

where $\mathcal{B}$ is the backward shift operator, $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is an m-dimensional stationary process (the innovation process), $\boldsymbol{\Phi}(\mathcal{B})$ and $\boldsymbol{\Theta}(\mathcal{B})$ are $m \times m$ matrices in $\mathcal{B}$, given by the equations

$$\boldsymbol{\Phi}(\mathcal{B}) = \sum_{\ell=0}^{p} \boldsymbol{\phi}_\ell \mathcal{B}^\ell \quad \text{and} \quad \boldsymbol{\Theta}(\mathcal{B}) = \sum_{\ell=0}^{q} \boldsymbol{\theta}_\ell \mathcal{B}^\ell,$$

assumed to have no common roots, where $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_q$ real $m \times m$ matrices and $\boldsymbol{\phi}_0 = \boldsymbol{\theta}_0 = \boldsymbol{I}_{m\times m}$, the $m \times m$ identity matrix.

Notice that Definition 2.3 is more general than the classical definition of VARFIMA processes, introduced by Sowell (1989), in which $\boldsymbol{\varepsilon}_t$ is assumed to be Gaussian. In this work we analyze only the simpler case $p = q = 0$, for which equation (2.4) simplifies to

$$\operatorname{diag}\big\{(1-\mathcal{B})^{\boldsymbol{d}}\big\}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t, \quad \text{for all } t \in \mathbb{Z}.$$

It can be shown that, if $\boldsymbol{\varepsilon}_t$ has finite variance for all $t \in \mathbb{Z}$, a necessary and sufficient condition for the existence of stationary solutions for (2.4) is that $\boldsymbol{d} \in (-\infty, 0.5)^m$. On the other hand, to guarantee that the stationary solution is also causal and invertible, it can be shown that we must have $\det\big(\boldsymbol{\Theta}(z)\big) \neq 0$ and $\det\big(\boldsymbol{\Phi}(z)\big) \neq 0$, for all $z \in \mathbb{C}$ such that $|z| \leq 1$, and also $\boldsymbol{d} \in (-0.5, 0.5)^m$, so that $(-0.5, 0.5)^m$ will be the range we shall assume for the fractional differencing parameter $\boldsymbol{d}$.

VARFIMA processes can be seen as a natural extension of the classical ARFIMA processes. For the ARFIMA case, the so-called *long range dependence* occurs whenever the fractional differencing parameter $d \in (0, 0.5)$ while if d is in the $(-0.5, 0)$ zone, some authors refer to it as the *intermediate dependence* (see, for instance Lopes, 2008 and Sena Junior, Reisen and Lopes, 2006). In the case of VARFIMA process, we shall say that the k-th coordinate process $\{X_t^{(k)}\}_{t\in\mathbb{Z}}$ present *long range dependence* (*intermediate dependence*) whenever $d_k \in (0, 0.5)$ ($d_k \in (-0.5, 0)$), for $k \in \{1, \cdots, m\}$. In this work, the case where $d > 0.3$ is given special attention.

Definition 2.4. Let $\{X_t\}_{t\in\mathbb{N}}$ be an ARFIMA(p, d, q) process. We say that X_t is *strong long range dependent*, or present *strong long range dependence* (SLRD for short), if $d > 0.3$.

Several applications and extensions for VARFIMA processes (and general fractionally differentiated multivariate models) have been studied in recent years, see for instance, Chiriac and Voev (2011), Diongue (2010) and references therein. For estimation in VARFIMA processes, see Lobato (1999), Shimotsu (2007), Tsay (2010) and references therein. More details can also be found in Sowell (1989) and Luceño (1996).

A few results on copulas will also be necessary. The literature on the subject has grown very rapidly in the last decade especially in finance, statistics and econometrics where copulas have been widely used as tools for analyzing and modeling financial time series. An m-dimensional copula is a distribution function whose marginals are uniformly distributed on $[0, 1]$ and whose support is the $[0, 1]^m$ hypercube. The main theorem in the theory is the celebrated Sklar's theorem, which elucidates the usefulness of copulas.

Theorem 2.1 (Sklar's Theorem)**.** *Let $X_1, \cdots, X_n$ be random variables with marginals $F_1, \cdots, F_n$, respectively, and joint distribution function H. Then, there exists a copula C such that,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all } (x_1, \cdots, x_n) \in \mathbb{R}^n. \tag{2.5}$$

If F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\operatorname{Ran}(F_1) \times \cdots \times \operatorname{Ran}(F_n)$*, where, for a function f,* $\operatorname{Ran}(f)$ *denotes the range of f. The converse also holds. Furthermore,*

$$C(u_1, \cdots, u_n) = H\big(F_1^{(-1)}(u_1), \cdots, F_n^{(-1)}(u_n)\big), \quad \textit{for all } (u_1, \cdots, u_n) \in I^n,$$

where for a function F, $F^{(-1)}$ denotes its pseudo-inverse given by $F^{(-1)}(x) = \inf\big\{u \in \operatorname{Ran}(F) : F(u) \geq x\big\}$.

A proof of Sklar's Theorem in the bidimensional case can be found in Nelsen (2006).

Remark 2.1. Among many applications of Sklar's Theorem, one will be particularly useful in the simulations. Suppose we have an m-dimensional continuous random vector $\boldsymbol{X}$, with marginal distributions $F_1, \cdots, F_m$, for $m > 1$. Suppose that we want to estimate some quantity and investigate what happens when the marginal behavior of $\boldsymbol{X}$ is changed to, say, $G_1, \cdots, G_m$, but the joint dependence structure is kept as intact as possible. Letting $C_{\boldsymbol{X}}$ denote the copula associated to $\boldsymbol{X}$ (that is, a function satisfying (2.5)), if the copula $C_{\boldsymbol{X}}$ is known, this problem can be easily solved. Let $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ be a sample from the copula $C_{\boldsymbol{X}}$, where $\boldsymbol{u}_k = (u_k^{(1)}, \cdots, u_k^{(m)})$, $k = 1, \cdots, n$. Consider the following samples based on $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$:

a) $\{\boldsymbol{x}_k\}_{k=1}^n$, where $x_k^{(j)} = F_j^{-1}(u_k^{(j)})$, for $k = 1, \cdots, n$ and $j = 1, \cdots, m$.

b) $\{\boldsymbol{y}_k\}_{k=1}^n$, where $y_k^{(j)} = G_j^{-1}(u_k^{(j)})$, for $k = 1, \cdots, n$ and $j = 1, \cdots, m$.

By Sklar's Theorem, $\{\boldsymbol{x}_k\}_{k=1}^n$ and $\{\boldsymbol{y}_k\}_{k=1}^n$ are samples with the same joint dependence as $\boldsymbol{X}$, but the former has marginals $F_1, \cdots, F_m$, while the latter has marginals $G_1, \cdots, G_m$. One can now calculate and compare the quantity of interest by using $\{\boldsymbol{x}_k\}_{k=1}^n$ and $\{\boldsymbol{y}_k\}_{k=1}^n$. This method allows one to study how the marginal behavior affects some quantity of interest by keeping the joint behavior of the sample (determined by $C_{\boldsymbol{X}}$) fixed and introducing the features of interest directly into the marginals.

3 An Estimator for the Mallows Distance

We proceed to define the Mallows distance estimator considered in the simulations, prove its strong consistency and derive its limiting distribution. Given two i.i.d. samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ from distributions F and G, respectively, let $\widehat{F}_n$ and $\widehat{G}_m$ denote the empirical distribution functions based on these samples. For any $\alpha \geq 1$, the Mallows α-distance estimator is given by

$$\widehat{\mathscr{D}}_\alpha(F, G) := \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \left(\int_0^1 \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_m^{(-1)}(u) \right|^\alpha \mathrm{d}u \right)^{1/\alpha}. \tag{3.1}$$

If F and G are absolutely continuous distributions and $m = n$, the expression (3.1) takes a very simple form. Let $X_{(1)}, \cdots, X_{(n)}$ and $Y_{(1)}, \cdots, Y_{(n)}$ be the ordered samples. Since repetition in the sample occurs with probability 0, $\widehat{F}_n^{-1}(t) = X_{(i)}$, for all $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right)$ and similarly for $\widehat{G}_n^{-1}(t)$. Therefore, in this case,

$$\widehat{\mathscr{D}}_\alpha(F, G) = \left(\frac{1}{n} \sum_{i=1}^n |X_{(i)} - Y_{(i)}|^\alpha \right)^{1/\alpha}. \tag{3.2}$$

In the case where F and/or G is not absolutely continuous or the samples do not have the same length, expression (3.1) will become a summation depending on the jump points of both functions. In this case the estimated Mallows α-distance can be obtained by calculating (3.1) directly upon implementation of the functions $\widehat{F}_n^{(-1)}$ and $\widehat{G}_n^{(-1)}$ to derive the step function $s(u) = \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_n^{(-1)}(u) \right|^\alpha$. We then calculate the area under $s(u)$ in $[0, 1]$ and take it to the $1/\alpha$ power to obtain the estimate. Most statistical softwares already have a built in empirical quantile function (such as the function `quantile` on the software R), which facilitates the implementation of the estimator.

4 Simulation Results: Mallows Distance

In this section we present the Monte Carlo simulation results regarding the Mallows distance between the components of a bidimensional VARFIMA$(0, \boldsymbol{d}, 0)$. In the simulations, the frac-

tional differencing parameter $\boldsymbol{d} := (d_1, d_2)$ is taken to range over all combinations of $d_i \in \{-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4\}$, $i = 1, 2$. In this work, we always calculate the Mallows α-distance for $\alpha = 2$, and refer to it simply by Mallows distance. The estimator used is the one presented in (3.1).

All Monte Carlo simulations are based on time series of fixed sample size 2,000 obtained from bidimensional VARFIMA$(0, \boldsymbol{d}, 0)$ processes. We perform 1,000 replications of each experiment. To generate the time series, we apply the traditional method of truncating the multidimensional infinite moving average representation of the process. The truncation point is fixed in 50,000 for all $\boldsymbol{d}$.

All simulations are performed using the computational resources from the (Brazilian) National Center of Super Computing (CESUP-UFRGS). The routines are all implemented in FORTRAN 95 language optimized by using OpenMP directives for parallel computing.

4.1 Gaussian innovations with equal variances

Figure A.1 shows the graph of d_2 by Mallows distance for different correlations. The results are based on Gaussian innovations with fixed mean $\boldsymbol{\mu} = (0, 0)$ and variance $\boldsymbol{\sigma}^2 = (1, 1)$, for correlations $\rho \in \{0, 0.5, 0.95\}$.

An interesting feature shown in Figure A.1[1] is that, for small values of $\boldsymbol{d}$ (both coordinates smaller than 0.1), the Mallows distance behaves homogeneously across different correlation values. This behavior suggests that the Mallows distance is not significatively sensitive to the correlation for small values of d_i, $i = 1, 2$.

There is, however, difference when both coordinates start to increase. A clear differentiation across the correlation appears when the parameters d_1 and d_2 are both greater or equal than 0.2. From the graphs on Figure A.1, we can infer that the greater the long range dependence in each coordinate is (or, equivalently, the greater the values of d_1 and d_2 are) the greater the difference among the estimated Mallows distance values for different correlations. Also notice that, when at least one coordinate of $\boldsymbol{d}$ is small (less or equal than -0.1), the Mallows distance values behave like an increasing function of the other coordinate whenever the latter is greater or equal than -0.2. We shall explore this fact in Subsection 6.1 to construct a simple and fast semiparametric estimator for the fractional differencing parameter $\boldsymbol{d}$.

From the graphs, it is clear that the correlation starts to influence and differentiate the Mallows distance values only when both coordinates are greater than 0.1. Furthermore, it appears that the magnitude of the values in the fractional differencing parameter $\boldsymbol{d}$ has more influence in differentiating the Mallows distance than the magnitude of the correlation itself. However, as expected, the Mallows distance values decrease as the correlation increases for almost all cases.

We can summarize our findings as follows:

1. The Mallows distance generally decreases as the correlation increases when one coordinate of $\boldsymbol{d}$ is greater or equal than 0.1 and behaves close to an exponential when the coordinates of $\boldsymbol{d}$ assume values over 0.1;

2. The Mallows distance appears not be affected by the correlation when at least one coordinate in $\boldsymbol{d}$ is smaller than 0.1.

[1] Tables containing the results from which the graphs are draw from are not presented here due to the restriction on the number of pages. They can be found, along with additional graphs and information, as an addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm.

3. The higher the parameter $\boldsymbol{d}$, the greater the difference among the Mallows distance estimates across different correlations. This suggests that the Mallows distance is less sensitive to the correlation than to the parameter $\boldsymbol{d}$.

4.2 Gaussian innovations with unequal variances

The results from the previous subsection bring some light into the dynamics between the components of Gaussian VARFIMA processes from the Mallows distance point of view. All simulations involving Gaussian innovations are performed by using a variance-covariance matrix whose values in the main diagonal are identical (equal to one, to be precise).

In this section we investigate the following question: how (if at all) the Mallows distance behavior change when the innovation's second moments are altered? In other words, we study what happens if the innovation process' marginals have different variances. A simple approach is to compare the case where the marginals have equal variances to the case where they are different in the spirit of Remark 2.1.

Figure A.2 shows the graphs of d_2 by Mallows distance for fixed d_1 and correlation $\rho \in \{0, 0.5, 0.95\}$. The variance was taken to be $\boldsymbol{\sigma}^2 = (1, 2)$. From the graphs, it is clear that increasing the variance of one innovation component also increases the Mallows distance in comparison to the equal variances case (Figure A.1). Increasing the correlation in the unequal variances case produces little to no difference in the global behavior of the Mallows distance, except when both innovation components present strong long range dependence, particularly when $\boldsymbol{d} = (0.4, 0.4)$. Notice that the magnitude of the Mallows distance is larger in the unequal variances case, as can be seen by the scale on the graphs. Figures A.1 and A.2 show that the equal variances case present a more erratic Mallows distance behavior across correlation when compared to the smooth curves on the unequal variances one. As a function of d_2, high values of d_1 (> 0.2) produce a more erratic behavior for the Mallows distance, which becomes sensitive to the correlation for $d_2 > 0.2$. We also observe that the Mallows distance behaviors in Figure A.2(a)-(f) are all very similar to each other. Surprisingly, the results suggest that the different variances in the marginals stabilize the Mallows distance behavior, weakening the influence of the fractional differencing parameter $\boldsymbol{d}$.

We can summarize our findings as follows:

1. Compared to the equal variances case, the unequal variances situation is more stable with respect to the Mallows distance and to the parameter $\boldsymbol{d}$, so that for $d_i \leq 0.2$, there is little to no difference in the Mallows distance values within d_2.
2. The magnitude of the Mallows distance is generally larger than the equal variances case.
3. The difference in the marginal variances appear to reduce the influence of the fractional differencing parameter $\boldsymbol{d}$ causing the Mallows distance to be less sensitive to the correlation in the innovation.

4.3 More on the unequal variances

After studying the differences when the innovations process have unequal variances ($\boldsymbol{\sigma}^2 = (1, 1)$ and $\boldsymbol{\sigma}^2 = (1, 2)$ cases), two questions naturally arise:

1. How (if at all) does the magnitude of the components of $\boldsymbol{\sigma}^2$ influence the Mallows distance? That is, compared to the case $\boldsymbol{\sigma}^2 = (1, 2)$, will the Mallows distance significantly change if we take $\boldsymbol{\sigma}^2 = (2, 3)$?

2. How (if at all) does the magnitude of the difference between the components of $\boldsymbol{\sigma}^2$ influence the Mallows distance? That is, compared to the case $\boldsymbol{\sigma}^2 = (1, 2)$, will the Mallows distance change much if we take $\boldsymbol{\sigma}^2 = (1, 3)$?

In order to answer the questions above, we apply the ideas explained in Remark 2.1 to simulate Gaussian VARFIMA processes with innovation marginal variances equal to $\boldsymbol{\sigma}^2 = (1, 3)$ and $\boldsymbol{\sigma}^2 = (2, 3)$. We compare the results with the case $\boldsymbol{\sigma}^2 = (1, 2)$ from last subsection (the simulations were actually performed all together using the ideas in Remark 2.1). We apply the Gaussian copula with parameter ρ and the marginals are taken to be normally distributed with zero mean and the desired marginal variances. Also, since the marginals are normally distributed, the parameter ρ still represents the correlation between the components.

Figure A.3 and A.4 show the graphs of d_2 by Mallows distance for fixed d_1 for variances $\boldsymbol{\sigma}^2 = (1, 3)$ and $\boldsymbol{\sigma}^2 = (2, 3)$, respectively. These are the analogous of Figure A.1 (equal variances case) and A.2 ($\boldsymbol{\sigma}^2 = (1, 2)$). Upon analyzing the graphs, one notice that cases $\boldsymbol{\sigma}^2 = (1, 2)$ and $\boldsymbol{\sigma}^2 = (2, 3)$ present very similar magnitudes for the Mallows distance values, while the case $\boldsymbol{\sigma}^2 = (1, 3)$ present considerably larger values. The equal variances case present the smallest values among all. This indicates that the difference between the components in $\boldsymbol{\sigma}^2$ has a strong influence on the magnitude of the Mallows distance, stronger than the magnitude of the components in $\boldsymbol{\sigma}^2$.

For $d_1 < 0.2$ (Figures A.2, A.3 and A.4 (a)-(e)), we notice that the overall behavior of the Mallows distance is similar among the unequal variances case and basically no differentiation across correlation can be observed in any of the graphs. For $d_1 > 0.2$, we observe that the cases $\boldsymbol{\sigma}^2 = (1, 2)$ and $\boldsymbol{\sigma}^2 = (2, 3)$ are closer to each other than to any other cases. For $d_1 = 0.3$, in the unequal variances case, a small differentiation across correlation start to surface for $d_2 > 0$, which become stronger when $d_1 = 0.4$. We notice that the differentiation for $d_1 > 0.2$ is stronger when $\boldsymbol{\sigma}^2 = (2, 3)$ which may indicate that the magnitude of the $\boldsymbol{\sigma}^2$ components somehow affects the Mallows distance differentiation across correlation. Also the Mallows distance behavior for $d_1 = 0.4$ is similar for $\boldsymbol{\sigma}^2 \in \{(1, 1), (1, 2), (2, 3)\}$, but clearly different for $\boldsymbol{\sigma}^2 = (1, 3)$. In comparison to the equal variances case, it seems that the components of $\boldsymbol{\sigma}^2$ strongly influences the Mallows distance behavior, making it more stable across the correlation. Notice the resemblance of the graphs for $d_1 = 0.3$ (frame (g) in the respective figures) in the unequal variance case to the one for $d_1 = 0.2$ (in Figure A.1(f)) in the case of equal variances. This also happens for the case $d_1 = 0.4$ which is close to the case $d_1 = 0.3$ in the equal variances case.

We conclude that the magnitude of the Mallows distance responds positively to both, the magnitude and the difference between the components of $\boldsymbol{\sigma}^2$, but clearly the response is stronger to the latter. This is no surprise, since the higher the difference between the variances of two normally distributed random variables with same mean, the more distant the values assumed by a sample of each are, which is directly reflected into the Mallows distance values.

We can summarize our findings as follows:

1. The Mallows distance respond positively to the difference between the components of $\boldsymbol{\sigma}^2$. The higher the difference, the higher the magnitude of the Mallows distance.

2. The magnitude of $\boldsymbol{\sigma}^2$ seems to influence positively the Mallows distance sensitivity regarding the correlation, especially for high $\boldsymbol{d}$. That is, the higher the magnitude of $\boldsymbol{\sigma}^2$, the more sensitive the Mallows distance become with respect to the correlation in the presence of strong long range dependence.

3. The overall Mallows distance behavior seem to be unaffected by the unequal variances in the innovation process, except for the magnitude of the values.

4. For $d_1 < 0.2$, the correlation has no affect in the Mallows distance in the unequal variances case. Also, it appears that the influence of the fractional differencing parameter $\boldsymbol{d}$ is attenuated in the unequal variance case.

4.4 Non-Gaussian innovation and heavy-tailed marginals

So far, all results presented are based on the bivariate Gaussian distribution. A question that naturally arises is does the Mallows distance behave in the same way if the innovations generating the VARFIMA process have a distribution other than the bivariate Gaussian? In order to partially answer this question, we simulate innovations from the Frank copula (see Nelsen, 2006 for its definition and properties) with different parameters and added to it, via Sklar's theorem, standard normal marginals. In this way we obtain a sample with dependence following a Frank copula, but standard normal marginals, from which we generate the VARFIMA$(0, \boldsymbol{d}, 0)$ process and estimate the Mallows distance between the components of the process. Our aim is to compare the results obtained this way with the bivariate Gaussian with standard normal marginals ones presented in Subsection 4.1.

In order to make a fair comparison, we would like to somehow match the dependence strength in the Frank innovation case with the Gaussian innovation case. We use the Kendall's τ theoretical values as a measure of the dependence strength in the innovation. The Kendall's τ for the Gaussian case with correlation ρ and for the Frank copula with parameter θ are given, respectively, by

$$\tau_\rho = \frac{2}{\pi}\arcsin(\rho) \quad \text{and} \quad \tau_\theta = 1 - \frac{4}{\theta^2}\left(\theta - \int_0^\theta \frac{t}{e^t - 1}\,\mathrm{d}t\right).$$

To match the Kendall's τ in the Gaussian case for $\rho \in \{0, 0.5, 0.95\}$, a good approximation is $\theta \in \{0, 3.3, 18\}$. These are the values chosen for the simulations.

The results for the Mallows distance between the components of VARFIMA$(0, \boldsymbol{d}, 0)$ process with Frank innovations and standard normal marginals is shown in Figure A.5. This is the analogous of Figure A.1 in the Gaussian case. From the figures, we observe in both the same shape and almost the same magnitude of the Mallows distance. We observe that when both components of $\boldsymbol{d}$ have high values (≥ 0.2), there is a clear differentiation across the parameters, especially between $\rho = 0$ and $\rho \in \{3.3, 18\}$. The higher the dependence in the innovation, the higher the Mallows distance sensitivity to this dependence in both cases. For high values of d_1, we observe that the Mallows distance values are slightly higher in the Frank-Normal case.

Another question not discussed so far is whether or not the marginals distribution tails influence the Mallows distance in any way. In other words, is there any difference in the Mallows distance behavior if we only change the type of the marginal from, say, Gaussian marginals to heavy-tailed ones?

To answer this question we apply the ideas of Remark 2.1 to simulate Frank copula innovations with parameter $\theta \in \{0, 3.3, 18\}$ coupled with standard normal, t_3 and t_7 marginals, where, as usual, t_ν stands for the Student's t distribution with ν degrees of freedom. We recall that our previous experiments indicate that the innovation variances interfere in the Mallows distance. Since t_3, t_7 and normal distributions have different variances, we use a standardized version of the t distribution with unitary variance to avoid any differences in the Mallows distance from sources other than the marginals' tail behavior. For simplicity, whenever t_3 and t_7 marginal cases are mentioned, we mean the respective standardized version.

Figures A.5 shows the simulation results for the Frank-Normal couple. Figure A.6, the Frank-t_3 couple results are shown. In Figure A.7, the case Frank-t_7 is presented. Comparing

Figures A.5 and A.7, we notice that they all look very similar. In fact, the absolute difference on the estimated Mallows distance between the case t_7 and standard normal marginals ranges on $[0.003, 0.051]$, for all $\boldsymbol{d}$. Also in these cases (Frank-Normal and Frank-t_7), the Mallows distance behavior follows a similar pattern to the Gaussian case with equal variances.

In Figure A.6, we observe that the Mallows distance values in the Frank-t_3 case are higher than the respective ones in the other cases, but the overall curve pattern is basically the same. A differentiation across the parameter occurs only when at least one coordinate of $\boldsymbol{d}$ is greater or equal than 0.2. Otherwise, little differentiation appears.

The similarities between the cases t_7 and standard normal marginals are not really surprising because the difference in the tail of these distribution is small. However, the t_3 marginals case shows that the Mallows distance is sensitive to tail fatness in the innovation, which is reflected mainly in the magnitude of the Mallows distance. The pattern followed by the Mallows distance however, does not seem to be significatively affected neither by the innovation's non-Gaussianity, nor by the marginals' tails. We can summarize our findings as follows:

1. The Mallows distance behavior does not seem to be affected by the type of innovations, within the same dependence strength (here measured by the innovation theoretical Kendall's τ).

2. The marginals' tail seem to have little influence in the Mallows distance. Nevertheless, when present, this effect seem basically to be reflected in the magnitude of the Mallows distance values, which are slightly higher in the presence of heavy-tailed marginals.

Remark 4.1. The tables from which the graphs presented in this section were drawn from, along with the estimates' standard deviation and extra graphs, can be found as an addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm. The standard deviation for the estimated Mallows distance may look, at first glance, high compared to the magnitude of the respective estimate. This is because the sample distribution of the estimated Mallows distance (which is always non-negative) are generally skewed to the right, but concentrated at the mean.

5 Simulation Results: Kendall's τ comparison

In the previous section we investigated the behavior of the Mallows distance between the components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes in several contexts. So, as a measure of how close are two processes, are our findings natural or surprising? Are they shared for all types of dependence measure or are they unique to the Mallows distance? In order to provide a comparison, we perform the same experiment presented in Section 4 applying the Kendall's τ as a dependence measure instead of the Mallows distance. To insure fidelity, we performed the calculations using the same methodology as in the previous section. We start by presenting the Gaussian noise with equal variance case.

5.1 Gaussian innovations with equal variances

Figure A.8[2] presents the plots of d_2 by Kendall's τ for fixed d_1 and $\rho \in \{0, 0.5, 0.95\}$. This is the analogous of Figure A.1 in the Mallows distance case. Notice that the Kendall's τ is

[2]Tables containing the results from which the graphs are draw can also be found in the addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm. The tables also contain the standard deviation for the estimated Kendall's τ presented in this section. They are generally small, as one could expect given its asymptotic distribution.

much more sensitive to the correlation in the innovation than the Mallows distance. As the difference between the parameters d_1 and d_2 increases, the difference on the Kendall's τ between the components become higher and the range of the Kendall's τ values also increases as $|d_1|$ increases.

Although it is difficult to see in Figure A.8 given the scale, for $\rho = 0$, a simple hypothesis test show that in most cases (58 out of 64), the components can be regarded as statistically independent[3]. The exceptions are exactly the values of d_i for which $d_1 + d_2 \geq 0.6$, that is, when there is SLRD. This result is interesting since it reinforces the findings of Tsay and Chung (2000), where the authors study the presence of spurious regression in stationary and ergodic processes with long range dependence. In the aforementioned paper, the authors detected spurious effects in long range dependent time series whenever $d_1 + d_2 > 0.5$. In the present case, we observed that whenever $d_1 + d_2 > 0.6$, even though the components of the process are independent of each other by construction, the null hypothesis of independence is rejected, which indicates the presence of spurious effects. For more details on the matter, see Tsay and Chung (2000).

5.2 Gaussian innovations with unequal variances

In this subsection we present simulation results analogous to those in Subsection 4.3 in the context of the Kendall's τ. Figures A.8 to A.11 present the graphs of d_2 by the Kendall's τ for fixed d_1 for the cases $\boldsymbol{\sigma}^2$ equal to $(1,1), (1,2), (1,3)$ and $(2,3)$, respectively. From all graphs, it is clear that the variance strongly influences the Kendall's τ between the components. We notice the differences in the scale between unequal and equal variances cases. The much smaller values in the unequal variances case indicates that the different variances in the innovation somehow balances the number of concordant and discordant pairs, so that, at the Kendall's τ point of view, the components are more distinct compared to the equal variances case. For $\rho \neq 0$, even though the magnitude of the Kendall's τ is small, the components cannot be considered statistically independent regardless $\boldsymbol{d}$. For $\rho = 0$ and $\boldsymbol{\sigma}^2 = (1,2)$, the independence hypothesis is always rejected when $d_i > 0$, for $i = 1, 2$ and for 3 (out of 48) combinations the hypothesis of independence is rejected when at least one parameter in $\boldsymbol{d}$ is negative. This suggests that differences in the process' variances amplify the spurious effect present in the equal variances case and give evidence of spurious regression in a broader context than reported in Tsay and Chung (2000). Similar results hold for $\boldsymbol{\sigma}^2 \in \{(1,3), (2,3)\}$.

The Kendall's τ behavior as the correlation increases is erratic, especially compared to the smooth behavior in the equal variances case. In all cases, a more or less similar pattern is followed and, arguably, the cases $\boldsymbol{\sigma}^2 = (1,2)$ and $\boldsymbol{\sigma}^2 = (2,3)$ are more alike then any other combination. Based on these results, it is clear that the variance heavily influences the Kendall's τ behavior. However, it is not clear what influences the most the value of τ, if the magnitude of the difference between the components on the variance, or the magnitude of the components themselves. Also notice that there are no overlaps/crossings within the correlation, which means that in all cases, the Kendall's τ is sensitive to the correlation in the innovation process.

[3]At 95% confidence level, the critical point of the two sided test H_0 : *the components are independent* is 0.0009356. The test is obtained by the normal approximation to the Kendall's τ, which rejects H_0 if

$$|\hat{\tau}| > u_{\alpha/2}\sqrt{\frac{2(2n+5)}{9n(n-1)}},$$

where u_k stands for the $100 \times k$-percentile of the standard normal distribution. In this work, all hypothesis tests are performed at 95% significance level and $n = 2,000$.

5.3 Non-Gaussian innovation and heavy-tailed marginals

We also repeated the experiment of Section 4.4 and the simulation results are shown in Figures A.12 to A.14, which are the analogous of Figures A.5 to A.7, respectively. As expected, given the Kendall's τ nature, there is no relevant difference among the estimated values of the Kendall's τ between the three cases studied (see Section 4.4 for details on the experiment).

6 Applications

In this section we shall present two simple applications of the results obtained in the previous sections. In the first one, we use the empirical results obtained in Section 4 to propose a semiparametric estimator for the fractional differencing parameter $\boldsymbol{d}$ for VARFIMA processes of any finite dimension. We present a simple Monte Carlo experiment to assess the estimator's performance. In the second application we propose a Mallows distance based test to assess the presence of SLRD components in VARFIMA process of any finite dimension. We also perform a Monte Carlo experiment to analyze its performance.

6.1 Estimation in Gaussian ARFIMA$(0, \boldsymbol{d}, 0)$ processes

In this section we shall present a simple application of the results obtained in Section 4. The idea is based on the observation that the results of Subsection 4.1 (see Figure A.1) show that in the context of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process with independent components, when one coordinate of the fractional differencing vector $\boldsymbol{d} = (d_1, d_2)$ is negative, say $d_1 < 0$, the Mallows distance behaves like an increasing function of d_2, whenever $d_2 > 0$. We shall explore this result to develop a simple and fast way to estimate the fractional differencing parameter in Gaussian (univariate) ARFIMA$(0, d, 0)$ processes.

Let us illustrate the idea by considering Figure A.1. Suppose that $\{X_t\}_{t=1}^n$ is a Gaussian ARFIMA$(0, d, 0)$ process with $d \in (0, 1/2)$. Our goal is to estimate the fractional differencing parameter d. The method starts by choosing d_1, among the possibilities in Figure A.1, and fitting some one-to-one curve, say $h(\cdot)$, to the portion of the graph for which $d_2 > 0$ and $\rho = 0$. For example, consider $d_1 = -0.3$ corresponding to Figure A.1(b). In that graph, the Mallows distance behaves closely to an exponential function for $d_2 > 0$ and $\rho = 0$. To illustrate the procedure we fit three different curves to the points in the positive portion of Figure A.1(c), for $\rho = 0$, as shown in Figure 6.1. The parametric form of the curves are the following

$$h_1(t; \boldsymbol{a}) := a_1 t \exp\big(a_2 t^2 + a_3 t + a_4\big), \quad h_2(t, \boldsymbol{b}) := b_1(t - 0.1)^3 + b_2(t - 0.1) + b_3,$$

for $\boldsymbol{a} := (a_1, a_2, a_3, a_4)'$ and $\boldsymbol{b} := (b_1, b_2, b_3)'$. The third curve h_3 is defined as the piecewise linear function joining the points in the graphs and extrapolated below and above by the nearest fitted line. The parameters were chosen to minimize the squared error between the observed and fitted values. In Figure 6.1 we present the fitted curves for the data from Figure A.1(b), where the parameters are

$$\boldsymbol{a} = (0.0862, 22.1222, -9.0890, 2.8109)' \quad \text{and} \quad \boldsymbol{b} = (15.8494, 0.0417, 0.0789)'.$$

The next step is to simulate an auxiliary Gaussian ARFIMA$(0, d_1, 0)$ process, say $\{Y_t^{(1)}\}_{t=1}^n$, independent of X_t, and calculate the estimated Mallows distance between the two processes, say $\widehat{\mathscr{D}}_{(1)}$. Notice that it is advantageous to simulate the auxiliary process with the same length as

the one fixed in order to apply the fast computable expression (3.2). The process is repeated m times to obtain a sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{m}$ from which we define the estimator $\widehat{d}$ of d by

$$\widehat{d} := h^{-1}\big(\overline{\mathscr{D}}\big), \quad \text{where} \quad \overline{\mathscr{D}} := \frac{1}{m}\sum_{k=1}^{m}\widehat{\mathscr{D}}_{(k)}. \tag{6.1}$$

To measure the performance of the method, we carry out a simple experiment whose results are

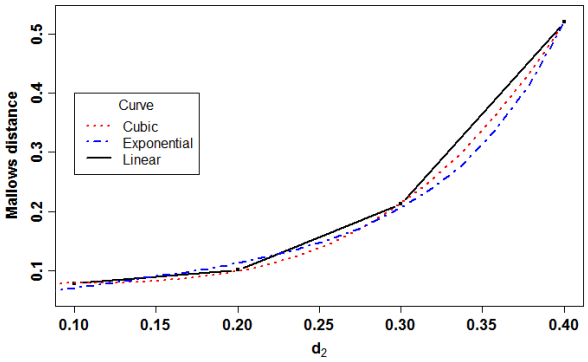


Figure 6.1: Fitted exponential, cubic and piecewise linear functions to the positive part of Figure A.1(b) for $\rho = 0$.

presented in Table 6.1. For each $d_2 \in \{0.10, 0.15, \cdots, 0.45\}$ in Table 6.1, we simulate a single Gaussian ARFIMA$(0, d, 0)$, say $\{X_t\}_{t=1}^{n}$, for $n = 2{,}000$, with the objective of estimating d. Next, we simulate $m = 1{,}000$ auxiliary Gaussian ARFIMA$(0, -0.3, 0)$ processes with the same length n as X_t. We then calculate the Mallows distance between X_t and each auxiliary series to obtain the sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$. From $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$ the estimator is calculated by (6.1) with h_1, h_2 and h_3 in the role of h. The estimated values of d, along with its bias (in parenthesis) and the standard deviation of the sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$ (at the bottom), are presented in Table 6.1. Notice that the inverse of $h_1(\cdot)$ has no closed form, so the value $h_1^{-1}(x)$ has to be numerically calculated. The inverse of h_2 can be explicitly derived after some tedious calculations, and is given by

$$h_2^{-1}(y) := \frac{\ell_0(y)^{1/3}}{6a_1} - \frac{2a_2}{\big(a_1^2\ell_0(y)\big)^{1/3}} + 0.1,$$

where

$$\ell_0(y) := 108(y - a_3) + 12\left(\frac{12a_2^3}{a_1} + \big[9(a_3 - y)\big]^2\right)^{\frac{1}{2}}.$$

Given the rather small sample to which the curves were fitted, the estimated values presented in Table 6.1 are relatively good, although some high bias can be seen for $d_2 = 0.1$, especially for the cubic fit (due to a root very close to 0). The estimators tend to perform better for d_2 corresponding to values for which data is available, as expected. In extrapolating to $d_2 = 0.45$, the estimates obtained from h_1 and h_2 are close to the true value, while for h_3, as one could expect given the exponential trend on the values, the estimative is off. Overall, the exponential curve (h_1) yield better results, even though, visually, it appears that h_2 provides a better fit for the data. At first sight, the standard deviation seems high compared to the respective estimated value, especially when d is high. However, since the estimates are skewed to the right, higher values of standard deviation are expected, especially compared to normally distributed estimators.

Table 6.1: Estimation results for $\widehat{d}$, based on the approximations shown in Figure 6.1. Presented are the estimated values along with the bias (in parenthesis).

Method	d_2							
	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
Exponential	0.0762 (-0.0238)	0.1453 (-0.0047)	0.1765 (-0.0235)	0.2793 (0.0293)	0.3198 (0.0198)	0.3512 (0.0012)	0.4020 (0.0020)	0.4486 (-0.014)
Cubic	0.0067 (-0.0933)	0.1749 (0.0249)	0.2044 (0.0044)	0.2792 (0.0292)	0.3128 (0.0128)	0.3430 (-0.0070)	0.4030 (0.0030)	0.4756 (0.0256)
Linear	0.0655 (-0.0345)	0.1353 (-0.0147)	0.1684 (-0.0316)	0.2518 (0.0018)	0.3011 (0.0011)	0.3256 (-0.0244)	0.3953 (-0.0047)	0.5252 (0.0752)
Std. Dev.	0.0234	0.0398	0.0598	0.1124	0.1638	0.2158	0.5755	0.6471

On one hand, for m small the method provide reasonable approximation and is extremely fast. Basically, the time spent for computing the estimator is dominated by the time spent in simulating the auxiliary processes. On the other hand, the procedure yield more precise estimative for m large. The limitations of the method, as seen in Table 6.1, are basically due to three sources: firstly we rely on approximated curves from few empirically calculated points to obtain the estimates, which may not be close enough to the true values to allow a good fitting; secondly, the results are based on independent pairs of process, while in the procedure, one of the coordinates is keep fixed and the other varies. Therefore, in doing so, some level of bias is expected, which is reflected on the estimated Mallows distance values and ultimately repassed to the final estimative; thirdly, the shape of the data near 0.1 is almost flat, causing high bias, but small variability in the estimates observed in Table 6.1. In contrast, near 0.4 the data increase rapidly, causing high variability but smaller bias on the estimates for large d. Generally, the method would certainly benefit from more empirically calculated values for the Mallows distance.

The method can be useful to calculate a first step for a more complicated estimation method, or to determine a "good" initial guess for estimators based on maximization procedures, especially if the objective function is non-smooth with several local maximums. The method is even more useful in the multivariate case, especially for high dimensional process. Estimators like the one proposed by Shimotsu (2007), are based on multidimensional optimization problems dependent on an initial guess for the parameter of differentiation. In that case, one could apply the method componentwise to obtain a reasonable initial guess for the optimization procedure. Furthermore, each auxiliary process can be used to calculate the estimated Mallows distance with respect to all components. For high dimensional problems, this could considerably reduce the usually time consuming optimization procedure, especially if no initial guess is available and a random one has to be used.

6.2 Hypothesis test

In this section we shall develop a test based on the results of Section 4.1 to test the presence of a certain level of dependence in multidimensional Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ processes. We start with the bidimensional case $\boldsymbol{d} = (d_1, d_2)$. The goal is to construct a test to detect the presence of a SLRD component in bivariate Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process.

The idea of the test is as follows. Let $\boldsymbol{\theta} := (d_1, d_2, \rho)$ and let $\mathscr{D}_2(\boldsymbol{\theta})$ denote the theoretical Mallows distance between the components of a Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process whose in-

novations are independent and identically distributed $\mathcal{N}(\mathbf{0}, \Omega)$ with variance-covariance matrix given by $\Omega := \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Let VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ denote such a process. Observe that in calculating each estimate presented on Figure A.1, we obtained, as a by-product of the replications, independent sequences of estimates from the theoretical unknown Mallows distance associated to each given Gaussian VARFIMA$(0, \boldsymbol{d}, 0; \rho)$. Each of these sequences, denoted by $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta})\right\}_{k=1}^n$, can be regarded as an independent and identically distributed approximated sample from $\mathscr{D}_2(\boldsymbol{\theta})$. We shall explore this fact to construct an approximated test of level α for assessing the presence of a SLRD component in the process.

Suppose we have a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process $\{\boldsymbol{X}_t\}_{t=0}^n$, with unknown $d_1 \neq d_2$ and $\rho \geq 0$, and we would like to test the presence of a SLRD component in the process. From $\boldsymbol{X}_t$ we can calculate an estimate for the Mallows distance between its components, say $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$, by using either (3.1) or (3.2). Our goal is to obtain a suitable test statistic and critical values for the test

$$H_0\text{: } \boldsymbol{X}_t \text{ has no SLRD component} \quad \text{vs} \quad H_1 : \text{ at least one component of } \boldsymbol{X}_t \text{ is SLRD.} \tag{6.2}$$

The problem of testing (6.2) is important in contexts where estimation near the non-stationarity zone $d \geq 0.5$ cannot be performed with precision, such as in spectral density based estimators (see Lopes, 2008). We start by analyzing the following related test

$$\begin{aligned} &H_0' : \mathscr{D}_2(\boldsymbol{\theta}) < \mathscr{D}_2(\boldsymbol{\theta}_0) \quad \text{vs} \quad H_1' : \mathscr{D}_2(\boldsymbol{\theta}) > \mathscr{D}_2(\boldsymbol{\theta}_0). \\ &\text{Rule: reject } H_0' \text{ for large values of } \widehat{\mathscr{D}}_2(\boldsymbol{\theta}). \end{aligned} \tag{6.3}$$

Testing (6.2) is the same as simultaneously test (6.3) for all combinations of $\boldsymbol{\theta}_0 := (d_1^0, d_2^0, \rho^0)$ for which at least one $d_i^0 > 0.3$. To avoid multiple testing procedures, we observe that from Figure A.1, the estimated Mallows distance is strictly increasing and consistently assumes high values whenever $\boldsymbol{d}$ contains at least one component greater or equal than 0.3 (except when $d_1 = d_2$). Consider the particular case of (6.3), obtained by taking

$$\mathscr{D}_2(\boldsymbol{\theta}_0) = \mathscr{D}_2(\boldsymbol{\theta}^*),$$

where $\boldsymbol{\theta}^*$ is the triple obtained from the combinations of $\boldsymbol{d}$ and ρ presented in Figure A.1 (and from which the sequences $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta})\right\}_{k=1}^n$ are available) such that at least one $d_i^* > 0.3$ and the value $\mathscr{D}_2(\boldsymbol{\theta}^*)$ is minimum among all possible candidates in Figure A.1. The reasoning behind such a choice is as follows: if there is no sufficient statistical evidence to reject H_0' with this choice, this means that $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$ is too small to be considered a member of the alternative class, relatively to $\mathscr{D}_2(\boldsymbol{\theta}^*)$. But if that is the case, clearly it won't be large enough for any other possible candidate in Figure A.1, hence $\mathscr{D}_2(\boldsymbol{\theta}^*)$ minimizes the type I error uniformly among all possible choices, provided $d_1 \neq d_2$. Of course, this protectiveness against type I error will eventually lead to an increase on the type II error. In the present case, considering all possible choices in Figure A.1, the best one in the sense just explained is $\boldsymbol{\theta}^* = (0.3, 0.2, 0.95)$, corresponding to Figure A.1(g).

Returning to the initial problem of testing (6.2), we notice that the uniformity of the choice $\mathscr{D}_2(\boldsymbol{\theta}^*)$ allows one to consider the test statistic $T(\boldsymbol{X}) := \widehat{\mathscr{D}}_2(\boldsymbol{\theta})$ and the rule

$$\text{Reject } H_0 \text{ if } \widehat{\mathscr{D}}_2(\boldsymbol{\theta}) > \widehat{q}_\alpha^{\,0}(\boldsymbol{\theta}^*), \tag{6.4}$$

where $\widehat{q}_\alpha^{\,0}(\boldsymbol{\theta}^*)$ is the $100 \times \alpha\%$-percentile obtained from the empirical distribution function of $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta}^*)\right\}_{k=1}^{1000}$. Taking $\alpha = 0.05$ in (6.4), it is routine to show that the test based on $\widehat{q}_{0.05}^{\,0}(\boldsymbol{\theta}^*) = \widehat{\mathscr{D}}_2^{(951)}(\boldsymbol{\theta}^*)$, where $\widehat{\mathscr{D}}_2^{(951)}(\boldsymbol{\theta}^*)$ is the 951-th order statistic of $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta}^*)\right\}_{k=1}^{1000}$, is of level 0.05 (see, for instance, Bickel and Doksum, 2007, p.271). In our situation[4], $\widehat{q}_{0.05}^{\,0}(\boldsymbol{\theta}^*) \approx 0.2751$.

[4]the simulated data used throughout this paper is available upon request.

To assess the performance of the test (6.2) with rule (6.4) for $\alpha = 0.05$, we generate 500 replications of a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process of sample size 2,000 for eight combinations of (d_1, d_2) and $\rho \in \{0, 0.5, 0.95\}$. In Table 6.2 we present the rejection rate (in percentage) of the test (6.2) with rule (6.4). The first four columns of Table 6.2 correspond to values of $\boldsymbol{d}$ for which H_0 holds, so that we expect the rejection rate to be small, while in the last four columns the null hypothesis does not hold and a large rejection rate is expected. Half of the values of $\boldsymbol{d}$ in the table are borderline (identified by *) and were included to assess the test performance under very weak evidence of either hypothesis, which usually induce type I and type II errors.

Table 6.2: Hypothesis testing results based on 500 replications of a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process. Presented is the rejection ratio in testing (6.2) by using rule (6.4) with $\boldsymbol{\theta}^* = (0.3, 0.2, 0.95)$.

ρ	$(-0.2, 0.2)$	$(-0.1, 0.2)$	$(-0.4, 0.25)^*$	$(-0.3, 0.25)^*$	$(0.1, 0.45)$	$(0.4, -0.3)$	$(0.2, 0.35)^*$	$(0.3, -0.1)^*$
0	0.4%	0.8%	7.2%	6.4%	99.8%	80.0%	53.4%	28.2%
0.5	1.0%	0.8%	9.8%	9.2%	99.6%	82.2%	47.4%	32.0%
0.95	0.4%	0.8%	5.8%	8.2%	100.0%	81.4%	41.4%	28.2%

* Borderline cases

From the results in Table 6.2, we notice that the test's overall performance is good when the values of $\boldsymbol{d}$ are not the borderline ones. The test performs well in terms of type I error even in borderline cases such as $(-0.4, 0.25)$ and $(-0.3, 0.25)$. As expected, the performance of the test in terms of type II error deteriorate in the borderline case $(0.2, 0.35)$ and is especially poor in the extreme case $\boldsymbol{d} = (0.3, -0.1)$. The test results clearly reflect the particular choice of $\mathscr{D}_2(\boldsymbol{\theta}_0)$, which was chosen to be conservative in terms of type I error, but permissive in terms of type II error.

Now, suppose we have $\{X_t\}_{t=1}^n$ a Gaussian ARFIMA$(0, d, 0)$ process for which we would like to test $H_0' : d < 0.3$ versus $H_1' : d > 0.3$. In order to do that, we reduce the problem to the previous case by considering the following construction. Choose $d_0 \in (-0.5, 0.1)$ and simulate a Gaussian ARFIMA$(0, d_0, 0)$ process $\{Y_t\}_{t=1}^n$ of same length as X_t. By considering $\boldsymbol{d} = (d, d_0)$, $\{(\boldsymbol{X}_t, \boldsymbol{Y}_t)\}_{t=1}^n$ is a VARFIMA$(0, \boldsymbol{d}, 0; 0)$, for which the Mallows distance estimate $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (d, d_0, 0)$, can be calculated. Upon applying the test (6.2) with rule (6.4), we obtain the conclusion. Obviously, if H_0 in (6.2) is refuted, we conclude $H_1' : d > 0.3$. One can replicate Y_t to obtain more information on the rejection rate and improve the conclusion. Also, as long as the sample size is not too small, even for $n \neq 1,000$ the test should yield valid conclusions. The same reasoning covers the case of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ processes of dimension higher than two by pairwise testing or by applying the previous idea componentwise.

We end up this section by observing that the results in this section were based on Subsection 4.1, but, naturally, the ideas could be easily adapted to any case presented in Section 4.

7 Conclusions and Final Remarks

In this work we present an extensive empirical analysis on the dependence among the components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes through the Mallows distance point of view. We examine several cases, including Gaussian and non-Gaussian innovation processes, heavy-tailed marginals, equal and unequal marginal variances, among others. These marginal features are introduced in the process at the innovation's level, mostly by using copula tools. The goal is to investigate a possible relationship between the Mallows distance, the fractional differencing parameter $\boldsymbol{d}$, the type and dependence in the innovation process as well as its marginal behavior.

To estimate the Mallows distance, in Section 3 we propose an estimator based on the

marginals' empirical quantiles. Section 4 is dedicated to present the simulation results. It is divided in 4 subsections. In Subsection 4.1 we present the results for the case where the innovation process is Gaussian, for several combinations of $\boldsymbol{d}$ and ρ. In this standard case, our findings suggest that the Mallows distance is not generally sensitive to the correlation on the innovation, except in the presence of strong long range dependence. As expected, the Mallows distance decreases as the correlation increases.

In Subsections 4.2 and 4.3 we study the case where the innovation process is still Gaussian, but the marginal variances are different. We conclude that the different variances do not affect the Mallows distance behavior, but do affect its magnitude. We discover that the higher the difference between the variance components, the higher the magnitude of the Mallows distance. We find that equal variances produce smaller Mallows distance compared to unequal variances. The Mallows distance also responds to the magnitude of the variance components. The higher the magnitude, the more sensitive the Mallows distance becomes with respect to correlation in the innovation process in the presence of long range dependence. The overall Mallows distance behavior seems to be indifferent with respect to the marginal variances, except, as mentioned, for its magnitude. We also conclude that the fractional differencing parameter influence is attenuated by differences in the variances.

In Subsection 4.4 we attack the problem of non-Gaussianity in the innovation process by considering innovations with Frank copula distribution, but Gaussian marginals. We compared innovations with the same dependence strength, measured here by the Kendall's τ. We find that, as long as the dependence strength is kept at the same level, non-Gaussian innovations produce no change in the Mallows distance. We also investigate whether heavy-tailed marginals influence the Mallows distance at all. We discover that, except for a small change in the magnitude of the estimates, the Mallows distance behavior do not change under heavy-tailed innovations.

Are the results unique to the Mallows distance or are they shared by other dependence measures? To partially answer this question, we repeat all the simulations presented in Section 4 calculating the Kendall's τ instead of the Mallows distance. We find that the Kendall's τ, in clear opposition to the Mallows distance, is highly sensitive to the innovations dependence in all experiments, but indifferent for most marginal changes we have applied. A bold exception is the unequal variances case, for which the Kendall's τ changes from its usual smooth behavior to an erratic one. Also worth noticing is the evidence of spurious regression when the process' components present long range dependence. This finding reinforces (and even extend) the conclusions on the paper by Tsay and Chung (2000), where the authors study the existence of spurious effects in stationary ergodic processes presenting strong long range dependence.

In Section 6 we present some applications of our simulation results. We propose a Mallows distance based semiparametric estimator for the fractional differencing parameter on Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any finite dimension. Its usefulness lies especially in its simplicity and fast computability, which makes it an attractive first step estimator for high dimensional problems. Its performance is assessed by a simple Monte Carlo experiment. We also propose a test to identify the presence of strong long range dependence in the components of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any finite dimension based solely on our simulation results. Its properties are assessed by simple Monte Carlo experiments which shows that the test performance is reasonably good, especially in terms of type I error.

Overall we conclude that the Mallows distance is usually indifferent to changes in the innovation dependence, except in the presence of strong long range dependence and that the fractional differencing parameter plays an important role in determining the dependence structure of VARFIMA$(0, \boldsymbol{d}, 0)$ processes whatever the marginal or joint dependence considered is.

Acknowledgements

S.R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by INCT *em Matemática* and also by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1. G. Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil. K. Zaniol was partially supported by CNPq-Brazil and by UFRGS. The authors are also grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources, to the editor and an anonymous referee whose comments lead to considerable improvement of this paper.

References

[1] Bickel, P.J. and Doksum, K.A. (2007). *Mathematical Statistics: Basic Ideas and Selected Topics*. New Jersey: Prentice Hall. 2nd edition.

[2] Bickel, P.J. and Freedman, D.A. (1981). "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics*, **9**, 1196-1217.

[3] Chiriac, R. and Voev, V. (2011). "Modeling and Forecasting Multivariate Realized Volatility". *Journal of Applied Econometrics*, **26**(6), 922-947.

[4] Dionge, A.K. (2010). "A Multivariate Generalized Long Memory Model". *Comptes Rendus de l'Académie des Sciences - Série I.* **348**, 327-330.

[5] Lobato, I.N. (1999). "A Semiparametric two Steps Estimator in a Multivariate Long Memory Model". *Journal of Econometrics*, **90**, 129-153.

[6] Lopes, S.R.C. (2008). "Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting". *Progress in Probability*, **60**, 497-525.

[7] Luceño, A. (1996). "A Fast Likelihood Approximation for Vector General Linear Processes with Long Series: Application to Fractional Differencing". *Biometrika*, **83**, 603-14.

[8] Major, P. (1978). "On the Invariance Principle for Sums of Independent Identically Distributed Random Variables". *Journal of Multivariate Analysis*. **8**(4), 487-517

[9] Mallows, C.L. (1972). "A Note on Asymptotic Joint Normality". *The Annals of Mathematical Statistics*, **43**, 508-515.

[10] Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer Verlag. 2nd edition.

[11] Rachev, S.T. and Rüschendorf, L. (1998). *Mass transportation problems. Vol. I. Theory*. New York: Springer-Verlag.

[12] Sena Jr, M.R.; Reisen, V.A. and Lopes, S.R.C. (2006). "Correlated Errors in the Parameters Estimation of ARFIMA Model: A Simulated Study". *Communications in Statistics. Simulation and Computation*, **35**(2), 789-802.

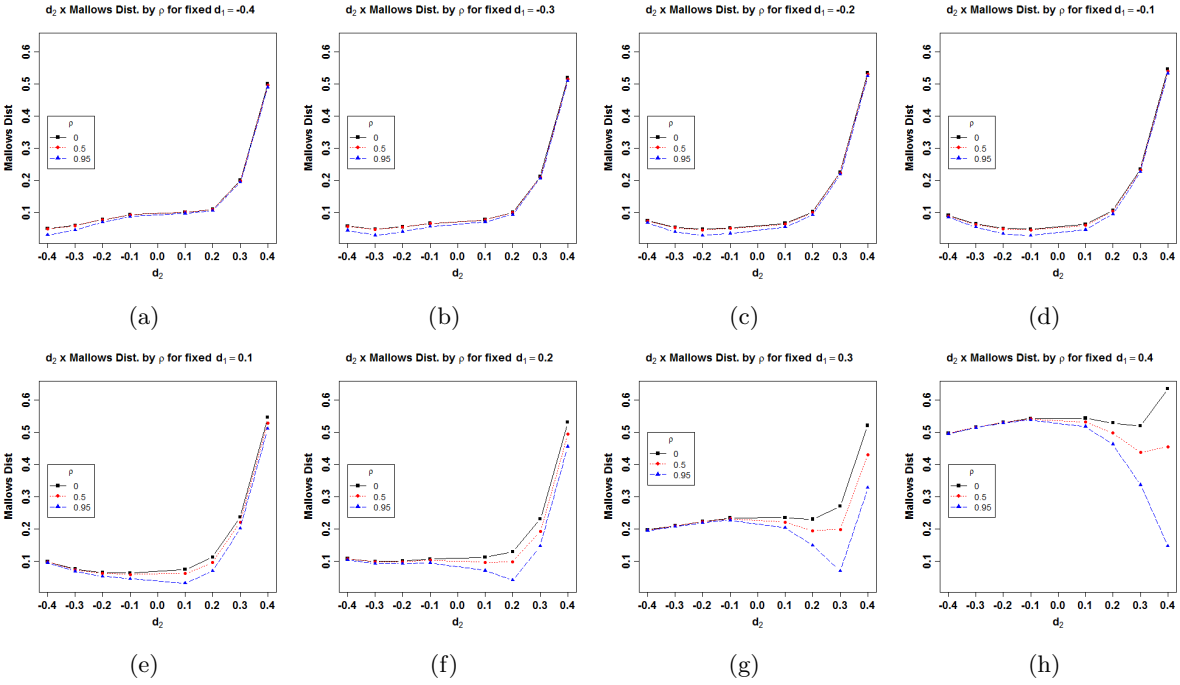
[13] Shimotsu, K. (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes". *Journal of Econometrics*, **137**, 277-310.

[14] Sowell, F. (1989). "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models", Working Paper, Carnegie-Mellon University.

[15] Tsay, W.-J. (2010). "Maximum Likelihood Estimation of Stationary Multivariate ARFIMA Processes", *Journal of Statistical Computation and Simulation*, **80**, 729 - 745.

[16] Tsay, W.-J. and Chung, C.-F. (2000). "The spurious regression of fractionally integrated processes". *Journal of Econometrics*, **96**(1), 155-182.

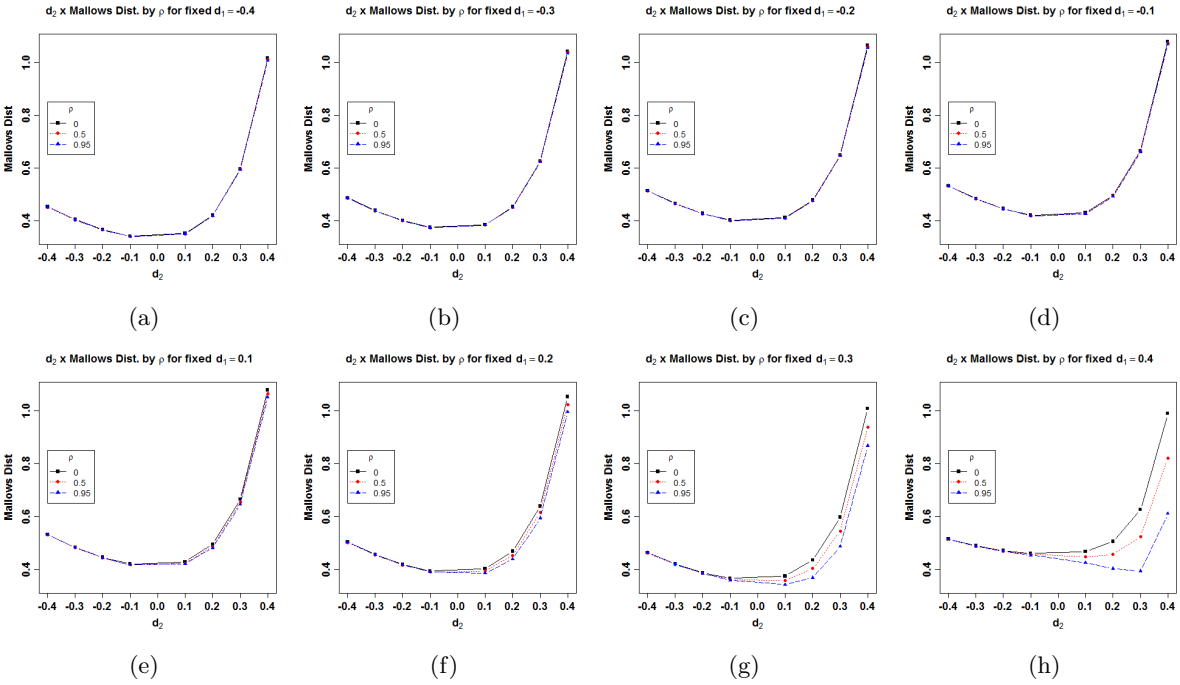
Appendix: Figures



(a) (b) (c) (d)

(e) (f) (g) (h)

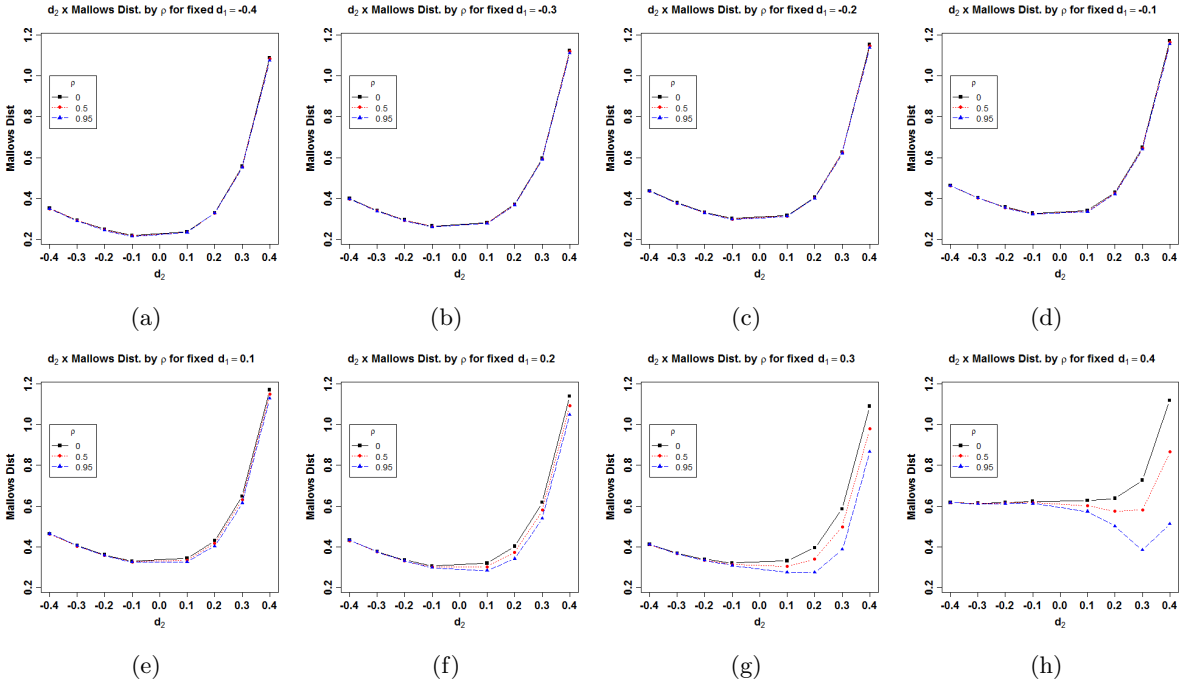
Figure A.1: Plots of d_2 by Mallows distance for fixed d_1.



(a) (b) (c) (d)

(e) (f) (g) (h)

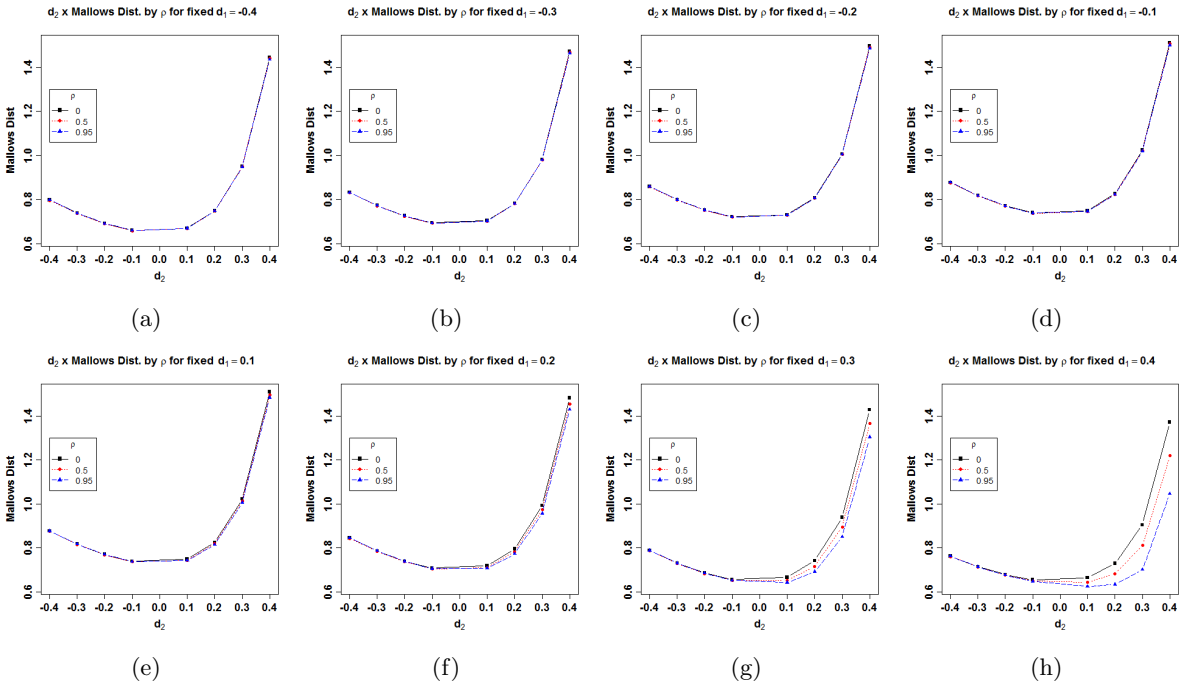
Figure A.2: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure A.3: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure A.4: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

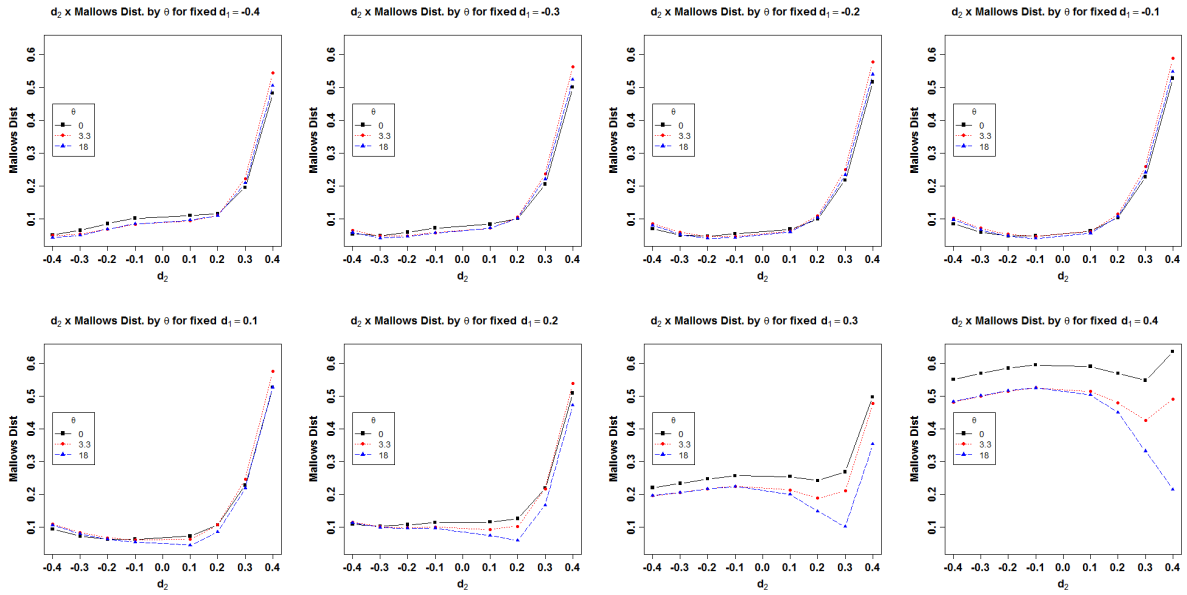


Figure A.5: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

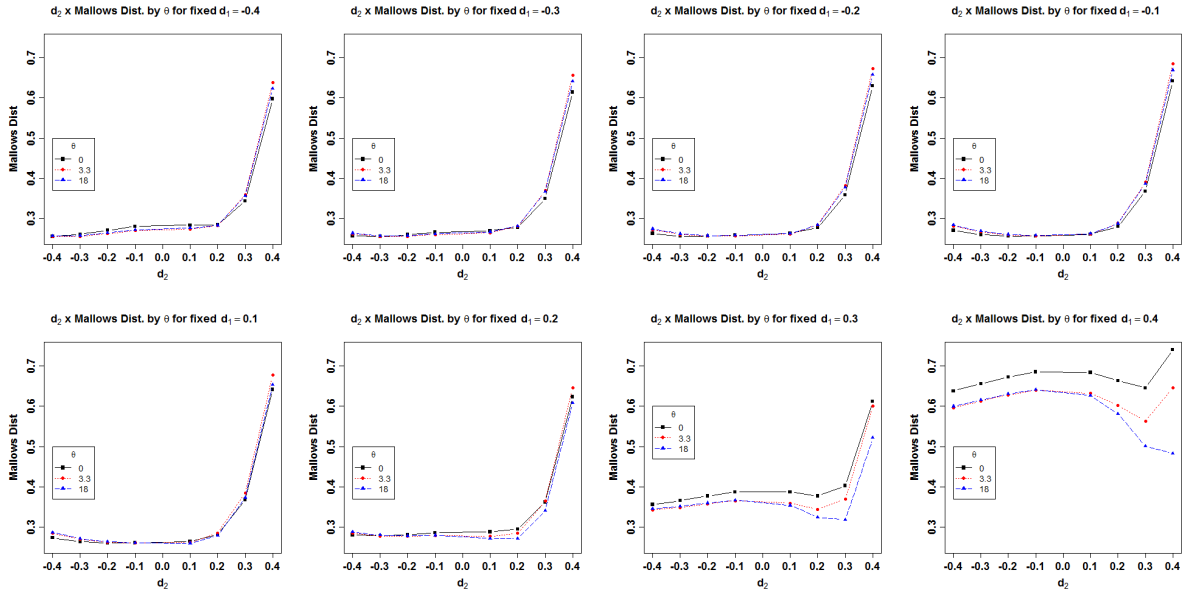


Figure A.6: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

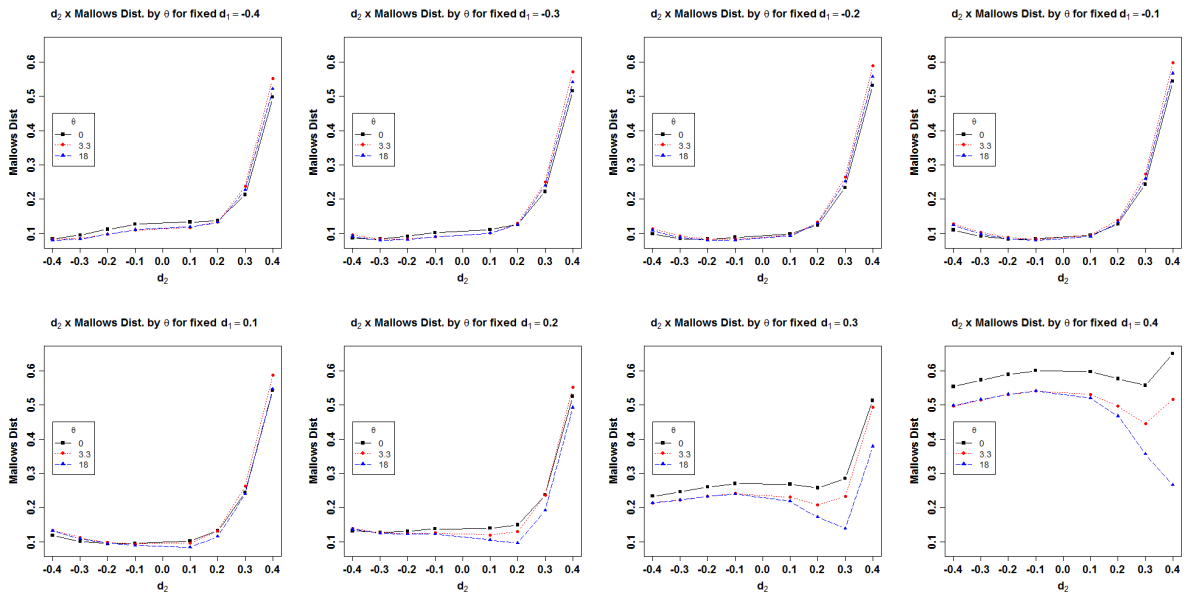


Figure A.7: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.

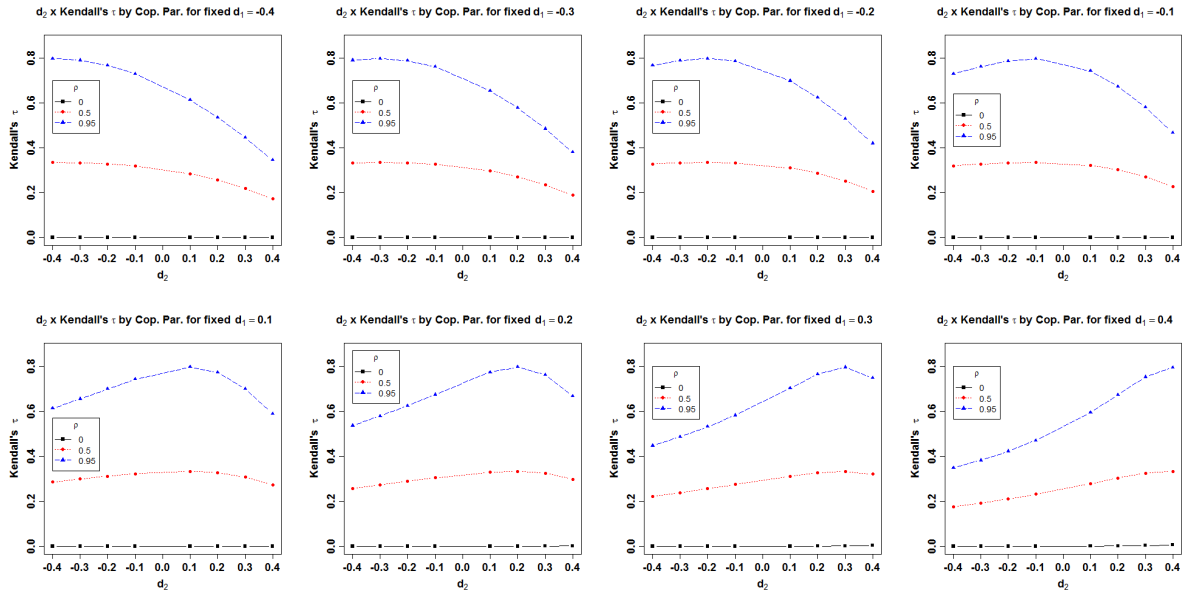


Figure A.8: Plots of d_2 by Kendall's τ for fixed d_1.

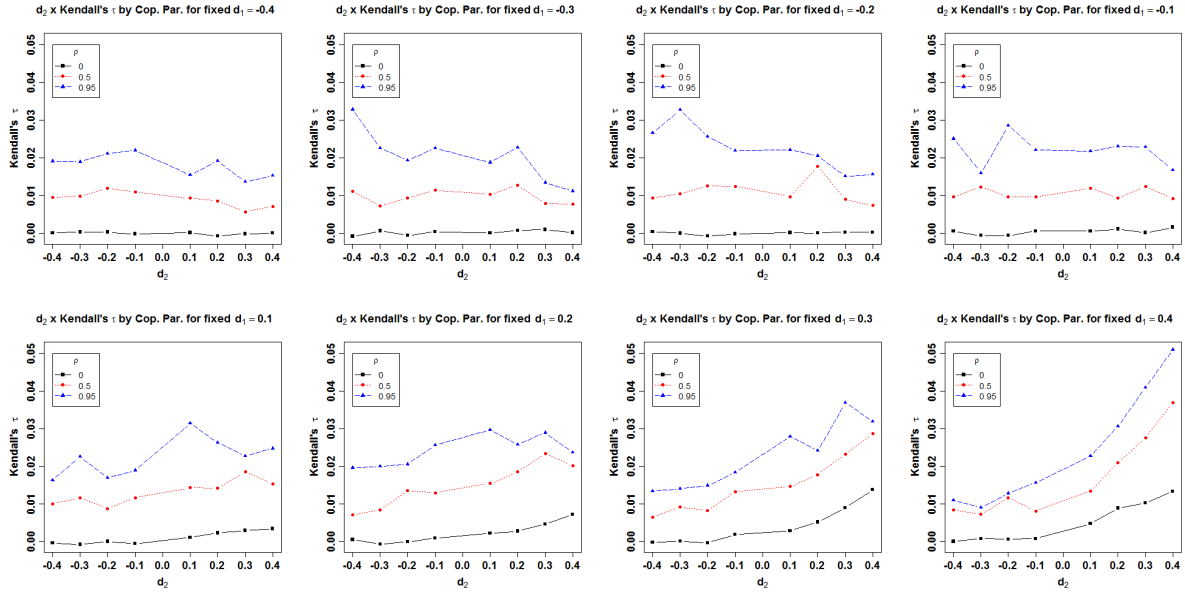


Figure A.9: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.

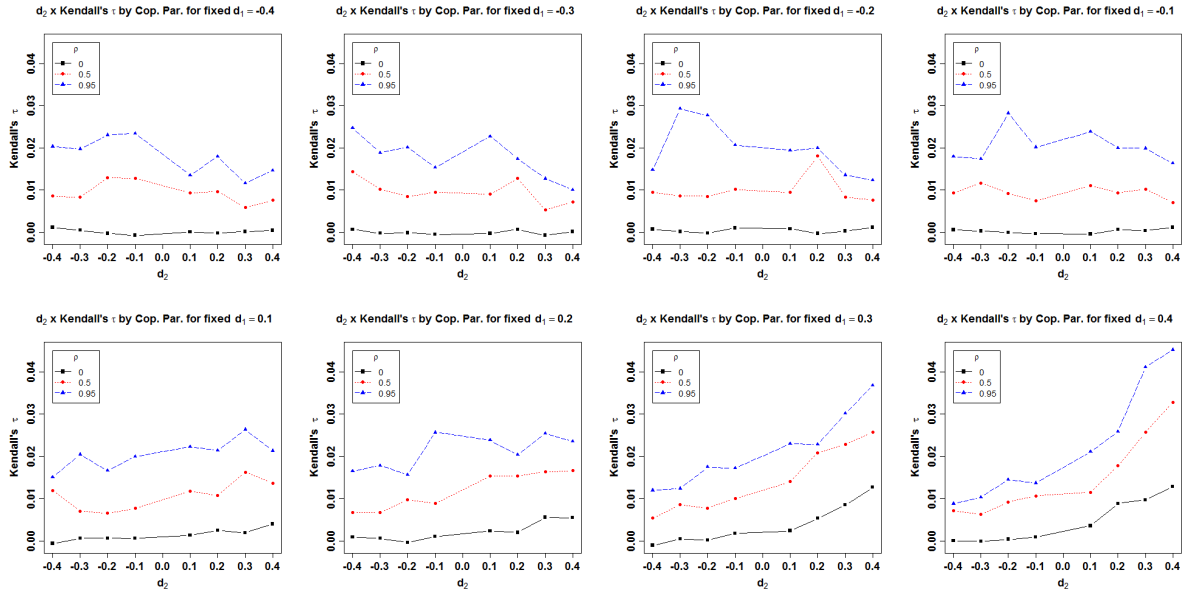


Figure A.10: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

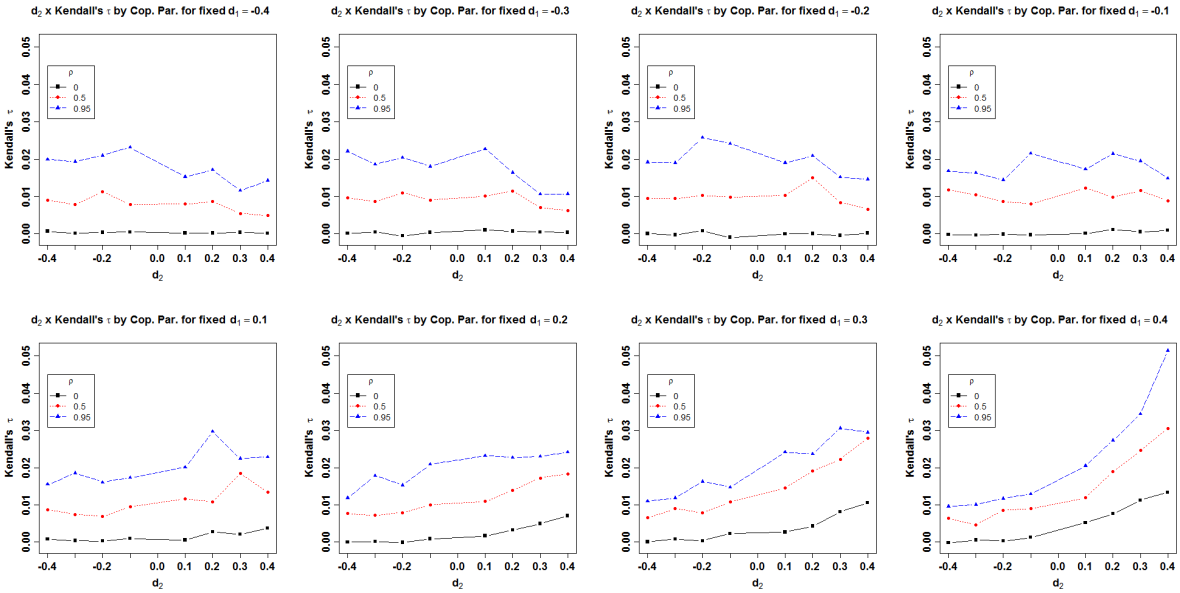


Figure A.11: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.

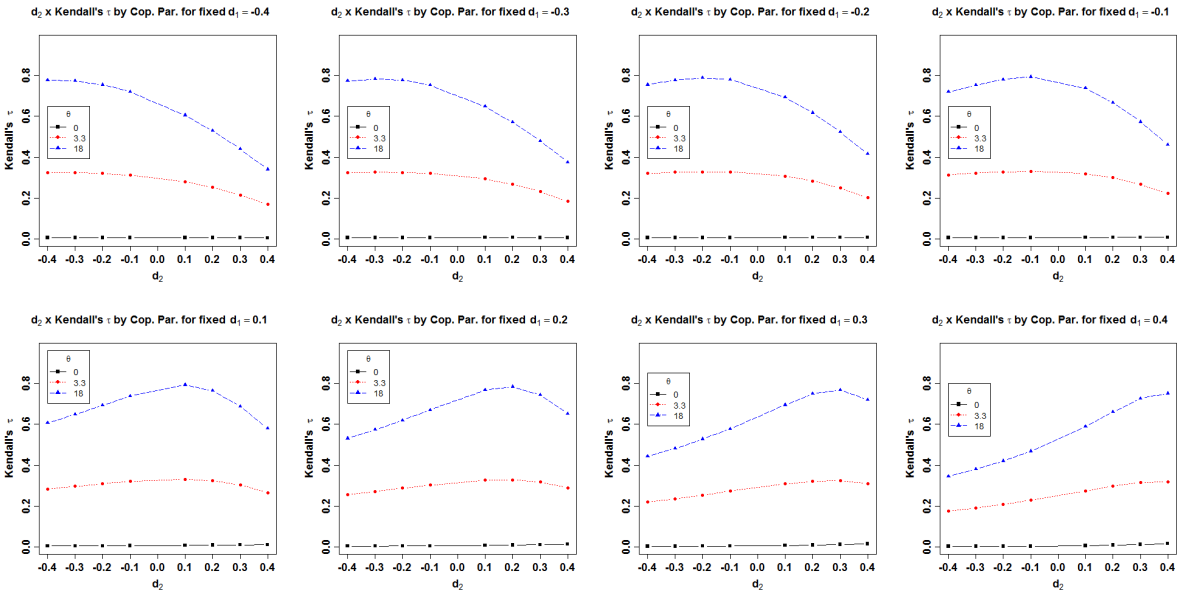


Figure A.12: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

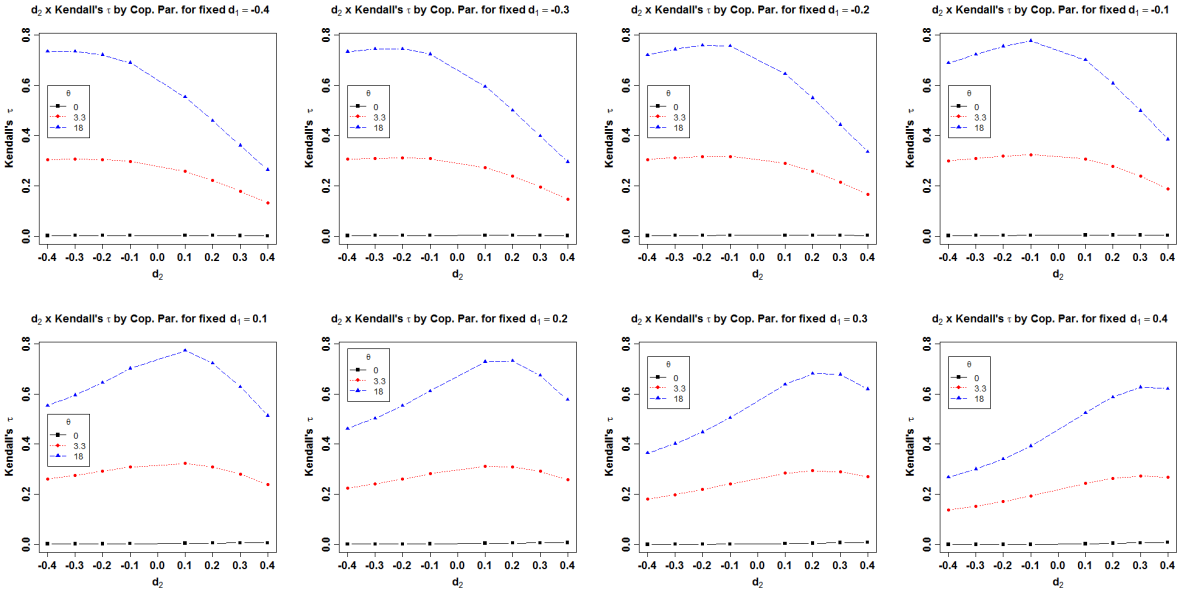


Figure A.13: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

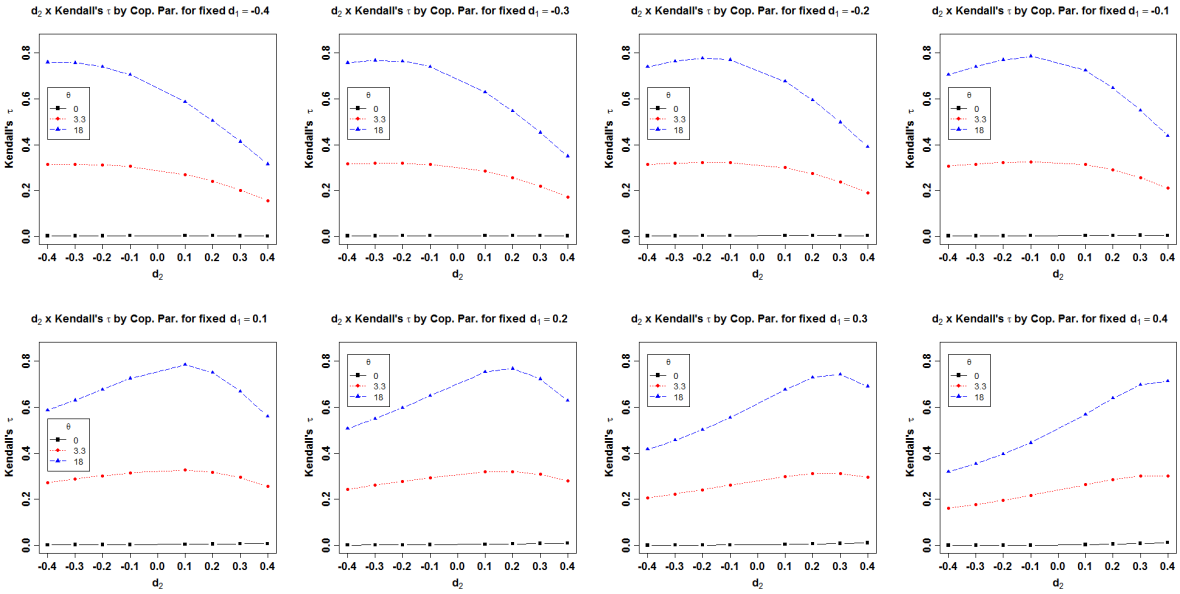


Figure A.14: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.

Addendum to "Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes"

Sílvia R.C. Lopes, Guilherme Pumi and Karine Zaniol
Federal University of Rio Grande do Sul
Mathematics Institute

Introduction and Goals

This report is intended to present an addendum to the article by Lopes, Pumi and Zaniol (2011), "Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes", in press at the *Communications in Statistics - Simulation and Computation*. In the article, the authors apply the Mallows distance as a tool to measure and interpret the dependence among components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes. To provide grounds for comparison, the well known Kendall's τ dependence measure is applied as well.

The article is based on extensive Monte Carlo simulations which yield a large amount of information. Due to restrictions in the number of pages, in the article this information is presented only in graphical format. In this addendum, we present the complete simulation results from which the graphs in the article are derived. We provide the results not only in tables, but also in graphical format. The extra graphs presented here contain the same information as the ones presented in the article, but from a different perspective: instead of considering the graphs as a function of the parameter $\boldsymbol{d}$, we consider them as a function of the correlation or the copula parameter, whichever applies.

It is not our intention to provide a full account of the results in the article here, but solely to provide more material to facilitate its understanding. Therefore, we refer the reader to the original article for information on the mathematical set up, general and specific definitions, simulation method, conclusions and other information pertinent to the data presented here. We hope that the tables and graphs we compile here will help the reader to better understand the dynamics of the Mallows distance in VARFIMA$(0, \boldsymbol{d}, 0)$ processes.

The tables and figures presented in this report are organized to match the ones on Lopes, Pumi and Zaniol (2011) as follows:

- Tables and Figures 1 to 4 are related respectively to Figures A.1 to A.4 in Lopes, Pumi and Zaniol (2011). They present the estimation results for the Mallows distance ($\hat{\mathscr{D}}_2$) and the estimates' standard deviation (sd) in the Gaussian case in both, equal and unequal variances cases. See Subsections 4.1, 4.2 and 4.3 in the article for more details.

- Tables and Figures 5 to 7 are related respectively to Figures A.5 to A.7 in Lopes, Pumi and Zaniol (2011). They present the estimation results for the Mallows distance ($\hat{\mathscr{D}}_2$) and the estimates' standard deviation (sd) in the non-Gaussian case and in the case of heavy tailed marginals. See Subsection 4.4 in the article for more details.

- Tables and Figures 8 to 11 are related respectively to Figures A.8 to A.11 in Lopes, Pumi and Zaniol (2011). They present the estimation results for the Kendall's τ ($\hat{\tau}$)

and the estimates' standard deviation (sd) in the Gaussian case in both, equal and unequal variances cases. See Subsections 5.1 and 5.2 in the article for more details.

- Tables and Figures 12 to 14 are related respectively to Figures A.12 to A.14 in Lopes, Pumi and Zaniol (2011). They present the estimation results for the Kendall's τ ($\hat{\tau}$) and the estimates' standard deviation (sd) in the non-Gaussian case and in the case of heavy tailed marginals. See Subsections 5.3 in the article for more details.

Reference

Lopes, S.R.C.; Pumi, G. and Zaniol, K. (2011). "Mallows Distance in VARFIMA $(0, \boldsymbol{d}, 0)$ Processes". *Communications in Statistics - Simulation and Computation.* In press.

*This version: July, 24th 2011.

Table 1: Simulation results for the Mallows distance in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 1)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1, 1)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.0509	0.0116	0.0495	0.0105	0.0309	0.0041	0.1	-0.4	0.0995	0.0236	0.0985	0.0225	0.0963	0.0174
	-0.3	0.0604	0.0163	0.0599	0.0153	0.0467	0.0068		-0.3	0.0766	0.0224	0.0750	0.0213	0.0699	0.0187
	-0.2	0.0783	0.0204	0.0784	0.0188	0.0701	0.0080		-0.2	0.0658	0.0214	0.0633	0.0201	0.0539	0.0191
	-0.1	0.0937	0.0216	0.0941	0.0198	0.0882	0.0089		-0.1	0.0644	0.0218	0.0603	0.0191	0.0467	0.0168
	0.1	0.1015	0.0241	0.1022	0.0229	0.0982	0.0181		0.1	0.0752	0.0320	0.0626	0.0210	0.0323	0.0056
	0.2	0.1103	0.0495	0.1101	0.0485	0.1058	0.0500		0.2	0.1133	0.0566	0.0968	0.0420	0.0708	0.0264
	0.3	0.2005	0.1240	0.1992	0.1188	0.1952	0.1229		0.3	0.2373	0.1162	0.2214	0.0976	0.2031	0.0863
	0.4	0.5008	0.2832	0.4966	0.2633	0.4903	0.2750		0.4	0.5468	0.2692	0.5299	0.2380	0.5123	0.2377
-0.3	-0.4	0.0587	0.0162	0.0573	0.0147	0.0451	0.0068	0.2	-0.4	0.1086	0.0502	0.1077	0.0501	0.1050	0.0501
	-0.3	0.0493	0.0110	0.0479	0.0099	0.0299	0.0039		-0.3	0.0998	0.0520	0.0985	0.0518	0.0938	0.0527
	-0.2	0.0558	0.0143	0.0550	0.0134	0.0411	0.0061		-0.2	0.1017	0.0520	0.0998	0.0510	0.0936	0.0511
	-0.1	0.0668	0.0173	0.0664	0.0162	0.0568	0.0074		-0.1	0.1076	0.0514	0.1041	0.0487	0.0960	0.0463
	0.1	0.0781	0.0225	0.0780	0.0213	0.0714	0.0192		0.1	0.1131	0.0576	0.0976	0.0438	0.0716	0.0277
	0.2	0.1011	0.0511	0.1003	0.0496	0.0943	0.0521		0.2	0.1305	0.0774	0.0997	0.0522	0.0420	0.0138
	0.3	0.2123	0.1199	0.2110	0.1141	0.2071	0.1174		0.3	0.2317	0.1329	0.1941	0.0996	0.1477	0.0632
	0.4	0.5197	0.2763	0.5155	0.2562	0.5094	0.2676		0.4	0.5314	0.2826	0.4944	0.2352	0.4565	0.2149
-0.2	-0.4	0.0587	0.0162	0.0573	0.0147	0.0451	0.0068	0.3	-0.4	0.1980	0.1254	0.1979	0.1253	0.1958	0.1260
	-0.3	0.0493	0.0110	0.0479	0.0099	0.0299	0.0039		-0.3	0.2103	0.1213	0.2100	0.1208	0.2077	0.1207
	-0.2	0.0558	0.0143	0.0550	0.0134	0.0411	0.0061		-0.2	0.2238	0.1170	0.2230	0.1158	0.2201	0.1148
	-0.1	0.0668	0.0173	0.0664	0.0162	0.0568	0.0074		-0.1	0.2345	0.1140	0.2322	0.1113	0.2278	0.1086
	0.1	0.0781	0.0225	0.0780	0.0213	0.0714	0.0192		0.1	0.2362	0.1180	0.2222	0.1042	0.2048	0.0898
	0.2	0.1011	0.0511	0.1003	0.0496	0.0943	0.0521		0.2	0.2309	0.1343	0.1958	0.1039	0.1500	0.0671
	0.3	0.2123	0.1199	0.2110	0.1141	0.2071	0.1174		0.3	0.2717	0.1849	0.1985	0.1273	0.0703	0.0367
	0.4	0.5197	0.2763	0.5155	0.2562	0.5094	0.2676		0.4	0.5217	0.3229	0.4310	0.2435	0.3287	0.1609
-0.1	-0.4	0.0911	0.0214	0.0901	0.0196	0.0862	0.0087	0.4	-0.4	0.4966	0.2823	0.4970	0.2820	0.4961	0.2819
	-0.3	0.0647	0.0173	0.0632	0.0157	0.0550	0.0073		-0.3	0.5157	0.2756	0.5159	0.2751	0.5148	0.2747
	-0.2	0.0513	0.0122	0.0492	0.0108	0.0347	0.0051		-0.2	0.5318	0.2703	0.5316	0.2692	0.5299	0.2683
	-0.1	0.0492	0.0104	0.0467	0.0091	0.0288	0.0038		-0.1	0.5433	0.2668	0.5419	0.2644	0.5390	0.2623
	0.1	0.0645	0.0217	0.0610	0.0182	0.0472	0.0168		0.1	0.5439	0.2694	0.5319	0.2570	0.5183	0.2449
	0.2	0.1078	0.0504	0.1042	0.0461	0.0958	0.0453		0.2	0.5295	0.2826	0.4980	0.2533	0.4636	0.2226
	0.3	0.2359	0.1125	0.2324	0.1045	0.2269	0.1053		0.3	0.5201	0.3241	0.4379	0.2557	0.3375	0.1712
	0.4	0.5470	0.2670	0.5411	0.2455	0.5338	0.2552		0.4	0.6357	0.4473	0.4558	0.3126	0.1478	0.0950

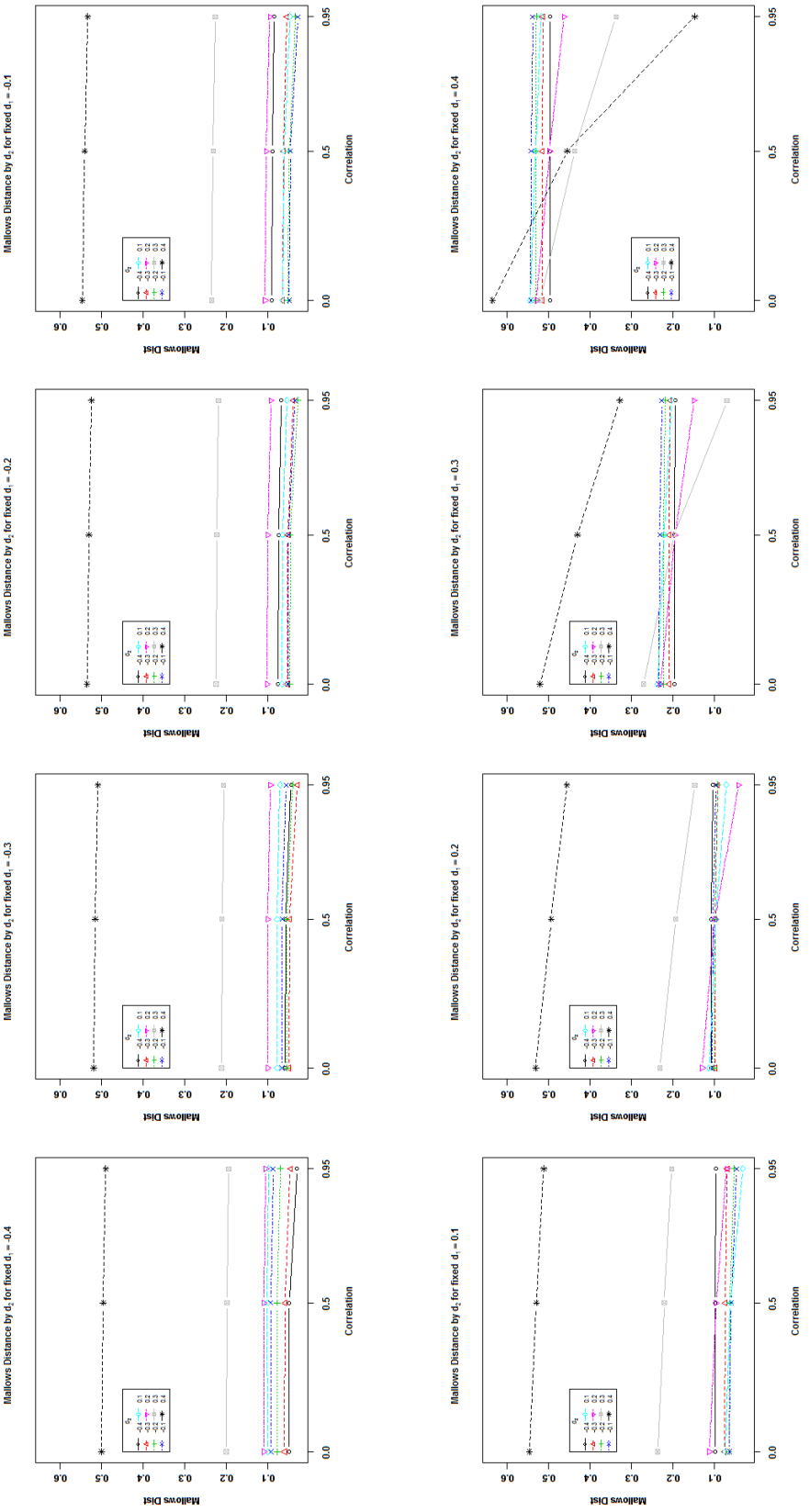


Figure 1: Plots of Mallows distance by correlation for fixed d_1. Gaussian case with equal variances.

Table 2: Simulation results for the Mallows distance in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 2)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1, 2)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
-0.4	-0.4	0.4529	0.0318	0.4515	0.0290	0.4519	0.0126	0.1	-0.4	0.5317	0.0302	0.5303	0.0285	0.5317	0.0180
	-0.3	0.4048	0.0306	0.4033	0.0279	0.4036	0.0116		-0.3	0.4836	0.0290	0.4821	0.0272	0.4832	0.0163
	-0.2	0.3668	0.0296	0.3653	0.0270	0.3655	0.0111		-0.2	0.4458	0.0280	0.4439	0.0260	0.4445	0.0145
	-0.1	0.3415	0.0289	0.3398	0.0263	0.3400	0.0113		-0.1	0.4203	0.0273	0.4179	0.0252	0.4177	0.0128
	0.1	0.3518	0.0301	0.3502	0.0280	0.3505	0.0172		0.1	0.4294	0.0293	0.4242	0.0258	0.4207	0.0111
	0.2	0.4206	0.0443	0.4194	0.0421	0.4193	0.0373		0.2	0.4955	0.0429	0.4881	0.0355	0.4816	0.0211
	0.3	0.5956	0.1100	0.5948	0.1036	0.5933	0.1060		0.3	0.6653	0.1042	0.6556	0.0889	0.6458	0.0796
	0.4	1.0167	0.3156	1.0129	0.2905	1.0077	0.3055		0.4	1.0790	0.3046	1.0651	0.2701	1.0510	0.2735
-0.3	-0.4	0.4869	0.0312	0.4856	0.0286	0.4861	0.0131	0.2	-0.4	0.5037	0.0338	0.5023	0.0327	0.5038	0.0259
	-0.3	0.4387	0.0300	0.4373	0.0274	0.4376	0.0119		-0.3	0.4565	0.0333	0.4549	0.0320	0.4560	0.0251
	-0.2	0.4007	0.0290	0.3992	0.0265	0.3994	0.0111		-0.2	0.4194	0.0330	0.4172	0.0313	0.4177	0.0239
	-0.1	0.3753	0.0283	0.3735	0.0258	0.3737	0.0109		-0.1	0.3946	0.0329	0.3913	0.0305	0.3906	0.0218
	0.1	0.3851	0.0294	0.3833	0.0273	0.3835	0.0161		0.1	0.4038	0.0369	0.3944	0.0295	0.3867	0.0131
	0.2	0.4528	0.0428	0.4514	0.0404	0.4512	0.0351		0.2	0.4693	0.0525	0.4539	0.0376	0.4390	0.0141
	0.3	0.6257	0.1065	0.6247	0.1001	0.6231	0.1020		0.3	0.6390	0.1144	0.6169	0.0867	0.5943	0.0603
	0.4	1.0438	0.3101	1.0399	0.2852	1.0348	0.2998		0.4	1.0540	0.3147	1.0244	0.2653	0.9956	0.2509
-0.2	-0.4	0.5139	0.0307	0.5125	0.0283	0.5133	0.0137	0.3	-0.4	0.4639	0.0760	0.4626	0.0758	0.4639	0.0740
	-0.3	0.4657	0.0295	0.4642	0.0271	0.4647	0.0123		-0.3	0.4209	0.0800	0.4193	0.0795	0.4201	0.0776
	-0.2	0.4276	0.0285	0.4261	0.0261	0.4263	0.0113		-0.2	0.3878	0.0837	0.3853	0.0825	0.3852	0.0800
	-0.1	0.4021	0.0278	0.4003	0.0254	0.4005	0.0108		-0.1	0.3661	0.0866	0.3618	0.0837	0.3598	0.0795
	0.1	0.4115	0.0289	0.4096	0.0267	0.4094	0.0150		0.1	0.3750	0.0909	0.3584	0.0751	0.3429	0.0564
	0.2	0.4785	0.0416	0.4768	0.0391	0.4762	0.0330		0.2	0.4355	0.1019	0.4036	0.0693	0.3692	0.0275
	0.3	0.6498	0.1038	0.6484	0.0971	0.6465	0.0983		0.3	0.5982	0.1552	0.5460	0.0999	0.4869	0.0288
	0.4	1.0654	0.3058	1.0612	0.2808	1.0559	0.2948		0.4	1.0089	0.3475	0.9391	0.2635	0.8680	0.1970
-0.1	-0.4	0.5323	0.0304	0.5310	0.0281	0.5320	0.0146	0.4	-0.4	0.5154	0.2727	0.5145	0.2729	0.5153	0.2724
	-0.3	0.4841	0.0292	0.4826	0.0269	0.4833	0.0130		-0.3	0.4895	0.2817	0.4883	0.2817	0.4885	0.2810
	-0.2	0.4460	0.0282	0.4444	0.0258	0.4448	0.0117		-0.2	0.4717	0.2886	0.4697	0.2878	0.4687	0.2866
	-0.1	0.4204	0.0274	0.4186	0.0251	0.4187	0.0108		-0.1	0.4614	0.2931	0.4574	0.2906	0.4541	0.2879
	0.1	0.4296	0.0285	0.4272	0.0262	0.4266	0.0138		0.1	0.4684	0.2953	0.4474	0.2785	0.4258	0.2623
	0.2	0.4961	0.0410	0.4937	0.0378	0.4923	0.0305		0.2	0.5058	0.2969	0.4575	0.2564	0.4036	0.2129
	0.3	0.6663	0.1022	0.6640	0.0945	0.6612	0.0945		0.3	0.6266	0.3209	0.5249	0.2297	0.3942	0.1068
	0.4	1.0802	0.3030	1.0751	0.2771	1.0689	0.2899		0.4	0.9897	0.4654	0.8217	0.3109	0.6115	0.0887

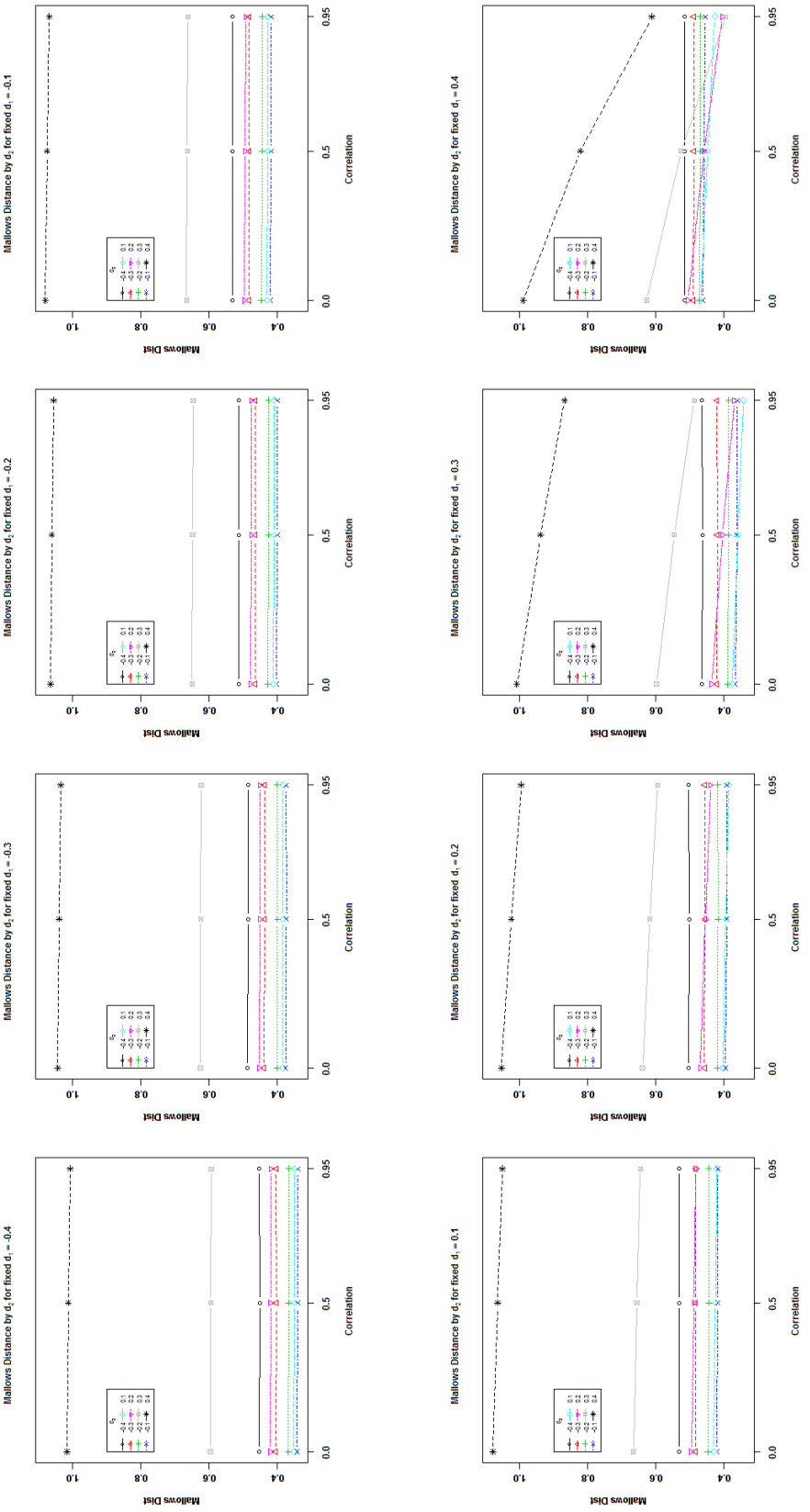


Figure 2: Plots of Mallows distance by correlation for fixed d_1 for $\boldsymbol{\sigma}^2 = (1, 2)$.

Table 3: Simulation results for the Mallows distance in the Gaussian case, $\boldsymbol{\sigma}^2 = (2, 3)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (2,3)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.3513	0.0407	0.3496	0.0369	0.3482	0.0144	0.1	-0.4	0.4634	0.0386	0.4617	0.0361	0.4628	0.0222
	-0.3	0.2933	0.0391	0.2914	0.0354	0.2894	0.0136		-0.3	0.4054	0.0372	0.4034	0.0346	0.4038	0.0203
	-0.2	0.2480	0.0376	0.2460	0.0340	0.2436	0.0133		-0.2	0.3598	0.0361	0.3572	0.0333	0.3566	0.0183
	-0.1	0.2184	0.0363	0.2161	0.0329	0.2136	0.0138		-0.1	0.3296	0.0354	0.3257	0.0322	0.3234	0.0159
	0.1	0.2368	0.0395	0.2349	0.0367	0.2335	0.0254		0.1	0.3434	0.0402	0.3338	0.0335	0.3247	0.0128
	0.2	0.3291	0.0667	0.3277	0.0636	0.3264	0.0601		0.2	0.4294	0.0631	0.4160	0.0493	0.4025	0.0277
	0.3	0.5564	0.1604	0.5552	0.1513	0.5523	0.1553		0.3	0.6482	0.1488	0.6319	0.1251	0.6149	0.1108
	0.4	1.0890	0.4222	1.0834	0.3897	1.0757	0.4089		0.4	1.1712	0.4045	1.1500	0.3578	1.1287	0.3602
-0.3	-0.4	0.3987	0.0399	0.3971	0.0363	0.3963	0.0148	0.2	-0.4	0.4314	0.0493	0.4297	0.0479	0.4309	0.0409
	-0.3	0.3403	0.0384	0.3385	0.0349	0.3372	0.0136		-0.3	0.3762	0.0504	0.3740	0.0487	0.3744	0.0420
	-0.2	0.2946	0.0371	0.2926	0.0336	0.2907	0.0130		-0.2	0.3335	0.0519	0.3303	0.0495	0.3293	0.0423
	-0.1	0.2643	0.0360	0.2620	0.0326	0.2600	0.0131		-0.1	0.3056	0.0534	0.3002	0.0493	0.2965	0.0402
	0.1	0.2807	0.0385	0.2785	0.0356	0.2771	0.0226		0.1	0.3190	0.0613	0.3012	0.0455	0.2827	0.0199
	0.2	0.3709	0.0628	0.3692	0.0594	0.3677	0.0547		0.2	0.4015	0.0845	0.3727	0.0562	0.3409	0.0164
	0.3	0.5953	0.1535	0.5938	0.1443	0.5908	0.1477		0.3	0.6172	0.1674	0.5793	0.1238	0.5392	0.0809
	0.4	1.1244	0.4133	1.1187	0.3810	1.1112	0.3997		0.4	1.1397	0.4207	1.0942	0.3521	1.0495	0.3278
-0.2	-0.4	0.4365	0.0393	0.4349	0.0359	0.4347	0.0156	0.3	-0.4	0.4113	0.1405	0.4099	0.1402	0.4106	0.1387
	-0.3	0.3779	0.0378	0.3761	0.0344	0.3753	0.0140		-0.3	0.3686	0.1497	0.3666	0.1491	0.3663	0.1475
	-0.2	0.3319	0.0365	0.3299	0.0332	0.3284	0.0129		-0.2	0.3380	0.1576	0.3347	0.1560	0.3326	0.1538
	-0.1	0.3013	0.0355	0.2990	0.0322	0.2971	0.0126		-0.1	0.3198	0.1632	0.3134	0.1592	0.3077	0.1549
	0.1	0.3164	0.0378	0.3137	0.0346	0.3119	0.0202		0.1	0.3309	0.1681	0.3035	0.1452	0.2732	0.1208
	0.2	0.4049	0.0601	0.4026	0.0561	0.4005	0.0499		0.2	0.3948	0.1784	0.3402	0.1291	0.2730	0.0649
	0.3	0.6269	0.1484	0.6247	0.1387	0.6212	0.1410		0.3	0.5852	0.2392	0.4974	0.1553	0.3861	0.0358
	0.4	1.1529	0.4065	1.1467	0.3738	1.1389	0.3917		0.4	1.0899	0.4729	0.9809	0.3550	0.8669	0.2506
-0.1	-0.4	0.4624	0.0387	0.4608	0.0356	0.4612	0.0167	0.4	-0.4	0.6169	0.4312	0.6165	0.4312	0.6158	0.4313
	-0.3	0.4038	0.0373	0.4020	0.0340	0.4017	0.0149		-0.3	0.6132	0.4337	0.6125	0.4333	0.6110	0.4333
	-0.2	0.3576	0.0360	0.3555	0.0328	0.3545	0.0134		-0.2	0.6164	0.4329	0.6148	0.4316	0.6120	0.4309
	-0.1	0.3269	0.0350	0.3243	0.0318	0.3225	0.0124		-0.1	0.6219	0.4311	0.6179	0.4276	0.6123	0.4250
	0.1	0.3412	0.0373	0.3376	0.0338	0.3348	0.0176		0.1	0.6267	0.4352	0.6016	0.4140	0.5730	0.3952
	0.2	0.4286	0.0585	0.4248	0.0534	0.4214	0.0449		0.2	0.6373	0.4494	0.5743	0.3996	0.5009	0.3499
	0.3	0.6487	0.1451	0.6449	0.1339	0.6400	0.1343		0.3	0.7270	0.4891	0.5811	0.3743	0.3831	0.2274
	0.4	1.1725	0.4020	1.1649	0.3680	1.1558	0.3842		0.4	1.1190	0.6557	0.8665	0.4450	0.5122	0.1110

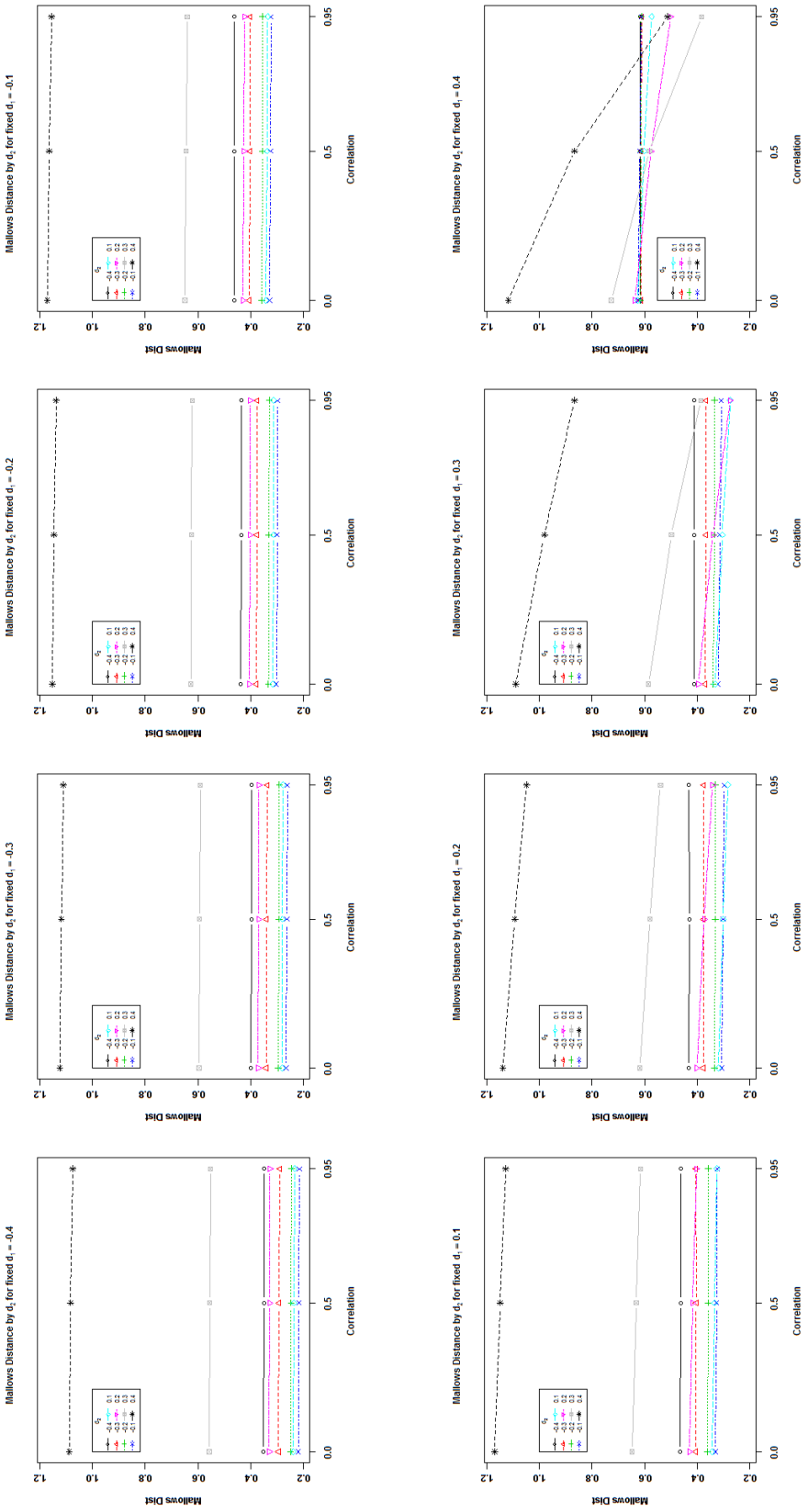


Figure 3: Plots of Mallows distance by correlation for fixed d_1 for the case $\boldsymbol{\sigma}^2 = (2, 3)$.

Table 4: Simulation results for the Mallows distance in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 3)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1,3)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.7982	0.0367	0.7966	0.0339	0.7979	0.0174	0.1	-0.4	0.8765	0.0352	0.8748	0.0336	0.8769	0.0228
	-0.3	0.7390	0.0352	0.7373	0.0325	0.7386	0.0159		-0.3	0.8174	0.0337	0.8156	0.0320	0.8174	0.0206
	-0.2	0.6923	0.0340	0.6905	0.0313	0.6917	0.0149		-0.2	0.7706	0.0324	0.7686	0.0305	0.7700	0.0186
	-0.1	0.6609	0.0330	0.6589	0.0304	0.6601	0.0147		-0.1	0.7392	0.0315	0.7366	0.0294	0.7375	0.0167
	0.1	0.6709	0.0341	0.6689	0.0320	0.6700	0.0198		0.1	0.7487	0.0330	0.7441	0.0301	0.7424	0.0154
	0.2	0.7493	0.0466	0.7478	0.0445	0.7482	0.0378		0.2	0.8255	0.0455	0.8194	0.0399	0.8151	0.0259
	0.3	0.9503	0.1137	0.9497	0.1073	0.9484	0.1090		0.3	1.0229	0.1098	1.0149	0.0959	1.0068	0.0873
	0.4	1.4445	0.3502	1.4408	0.3219	1.4356	0.3391		0.4	1.5101	0.3409	1.4974	0.3038	1.4841	0.3098
-0.3	-0.4	0.8324	0.0362	0.8307	0.0336	0.8322	0.0180	0.2	-0.4	0.8452	0.0371	0.8435	0.0360	0.8457	0.0279
	-0.3	0.7732	0.0347	0.7715	0.0321	0.7728	0.0164		-0.3	0.7864	0.0358	0.7845	0.0345	0.7864	0.0260
	-0.2	0.7264	0.0334	0.7246	0.0309	0.7258	0.0152		-0.2	0.7400	0.0347	0.7377	0.0332	0.7391	0.0240
	-0.1	0.6950	0.0325	0.6929	0.0300	0.6941	0.0147		-0.1	0.7088	0.0340	0.7057	0.0320	0.7062	0.0217
	0.1	0.7048	0.0335	0.7027	0.0314	0.7036	0.0192		0.1	0.7184	0.0367	0.7109	0.0318	0.7065	0.0160
	0.2	0.7825	0.0456	0.7810	0.0435	0.7812	0.0365		0.2	0.7954	0.0512	0.7833	0.0405	0.7731	0.0194
	0.3	0.9820	0.1113	0.9812	0.1049	0.9799	0.1062		0.3	0.9934	0.1172	0.9753	0.0928	0.9572	0.0700
	0.4	1.4732	0.3455	1.4695	0.3174	1.4643	0.3342		0.4	1.4827	0.3494	1.4559	0.2983	1.4296	0.2877
-0.2	-0.4	0.8594	0.0357	0.8578	0.0334	0.8594	0.0188	0.3	-0.4	0.7881	0.0600	0.7864	0.0597	0.7885	0.0567
	-0.3	0.8002	0.0342	0.7985	0.0318	0.7999	0.0170		-0.3	0.7312	0.0612	0.7293	0.0607	0.7310	0.0575
	-0.2	0.7535	0.0329	0.7516	0.0305	0.7529	0.0156		-0.2	0.6866	0.0624	0.6840	0.0613	0.6850	0.0576
	-0.1	0.7220	0.0320	0.7199	0.0296	0.7210	0.0148		-0.1	0.6568	0.0634	0.6526	0.0610	0.6523	0.0557
	0.1	0.7316	0.0330	0.7294	0.0309	0.7301	0.0185		0.1	0.6666	0.0678	0.6531	0.0551	0.6426	0.0373
	0.2	0.8089	0.0449	0.8071	0.0427	0.8071	0.0351		0.2	0.7422	0.0826	0.7169	0.0566	0.6925	0.0221
	0.3	1.0072	0.1094	1.0061	0.1028	1.0044	0.1035		0.3	0.9386	0.1471	0.8959	0.0984	0.8511	0.0394
	0.4	1.4961	0.3419	1.4921	0.3137	1.4867	0.3299		0.4	1.4288	0.3772	1.3664	0.2926	1.3042	0.2352
-0.1	-0.4	0.8779	0.0354	0.8763	0.0333	0.8780	0.0198	0.4	-0.4	0.7623	0.2138	0.7609	0.2141	0.7626	0.2133
	-0.3	0.8187	0.0339	0.8170	0.0317	0.8185	0.0178		-0.3	0.7150	0.2224	0.7132	0.2224	0.7145	0.2213
	-0.2	0.7719	0.0326	0.7700	0.0303	0.7713	0.0161		-0.2	0.6788	0.2296	0.6761	0.2288	0.6764	0.2270
	-0.1	0.7404	0.0316	0.7383	0.0293	0.7393	0.0150		-0.1	0.6552	0.2348	0.6503	0.2320	0.6485	0.2284
	0.1	0.7499	0.0327	0.7474	0.0305	0.7477	0.0176		0.1	0.6648	0.2377	0.6435	0.2186	0.6242	0.1987
	0.2	0.8269	0.0445	0.8246	0.0418	0.8239	0.0333		0.2	0.7293	0.2414	0.6829	0.1961	0.6344	0.1456
	0.3	1.0244	0.1083	1.0225	0.1009	1.0201	0.1004		0.3	0.9051	0.2808	0.8118	0.1856	0.7012	0.0599
	0.4	1.5117	0.3395	1.5068	0.3105	1.5006	0.3255		0.4	1.3719	0.4776	1.2194	0.3199	1.0464	0.1225

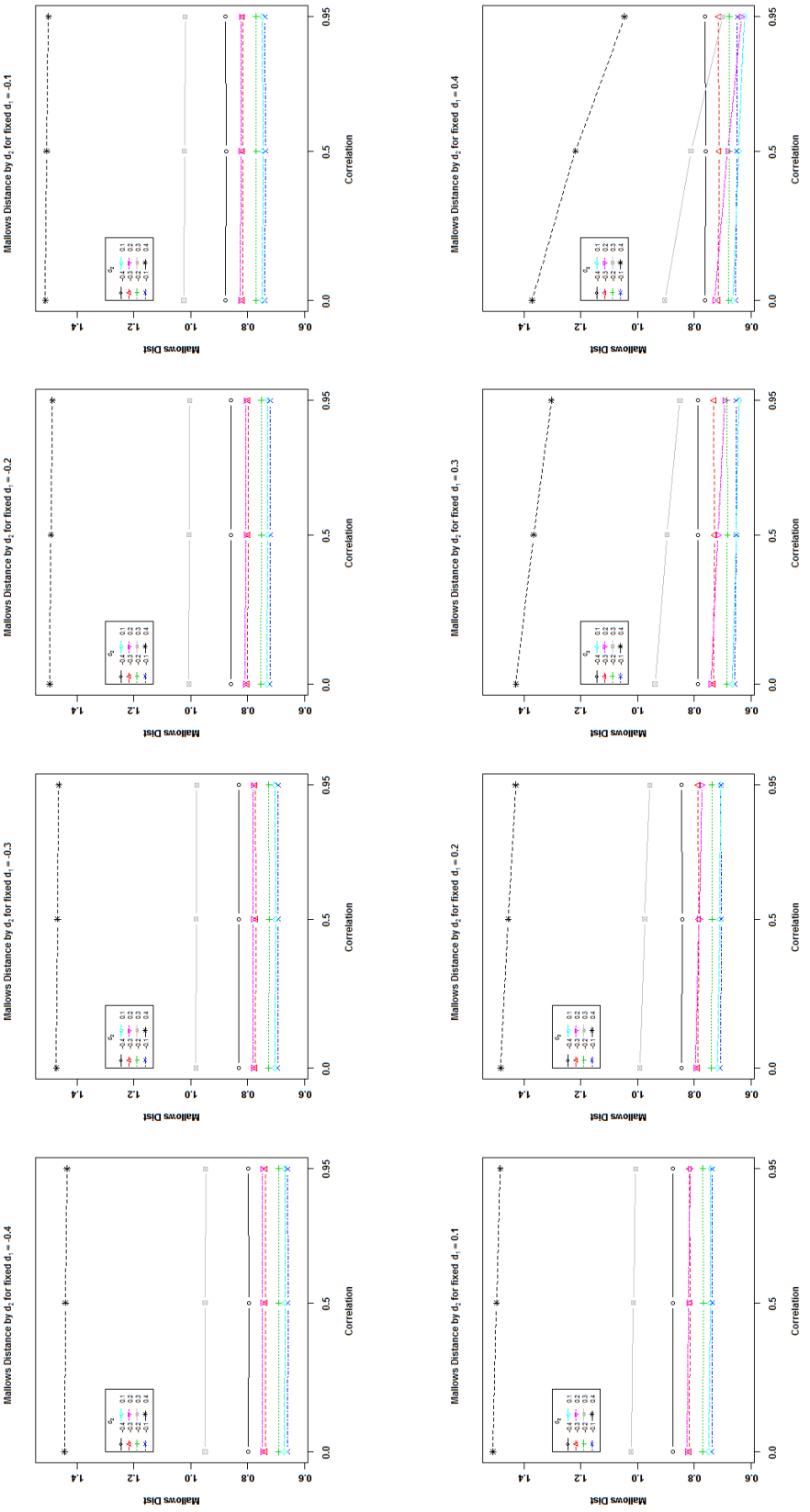


Figure 4: Plots of Mallows distance by correlation for fixed d_1 for the case $\boldsymbol{\sigma}^2 = (1, 3)$.

Table 5: Simulation results for the Mallows distance in the Frank-$\mathcal{N}(0,1)$ case.

$\boldsymbol{\varepsilon_t} \sim$ Frank-$\mathcal{N}(0,1)$															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.0510	0.0115	0.0502	0.0113	0.0430	0.0085	0.1	-0.4	0.0940	0.0224	0.1103	0.0220	0.1061	0.0196
	-0.3	0.0655	0.0168	0.0531	0.0127	0.0501	0.0100		-0.3	0.0720	0.0212	0.0834	0.0213	0.0793	0.0196
	-0.2	0.0858	0.0191	0.0689	0.0168	0.0690	0.0124		-0.2	0.0631	0.0202	0.0674	0.0205	0.0628	0.0189
	-0.1	0.1022	0.0196	0.0840	0.0183	0.0853	0.0135		-0.1	0.0633	0.0203	0.0608	0.0193	0.0546	0.0165
	0.1	0.1100	0.0222	0.0939	0.0223	0.0960	0.0206		0.1	0.0724	0.0298	0.0631	0.0207	0.0452	0.0099
	0.2	0.1154	0.0493	0.1111	0.0539	0.1101	0.0541		0.2	0.1069	0.0557	0.1068	0.0430	0.0860	0.0311
	0.3	0.1960	0.1248	0.2232	0.1310	0.2091	0.1299		0.3	0.2278	0.1147	0.2468	0.1063	0.2193	0.0936
	0.4	0.4823	0.2782	0.5440	0.2957	0.5058	0.2799		0.4	0.5271	0.2624	0.5754	0.2682	0.5268	0.2423
-0.3	-0.4	0.0542	0.0131	0.0653	0.0164	0.0586	0.0122	0.2	-0.4	0.1093	0.0514	0.1155	0.0489	0.1129	0.0487
	-0.3	0.0490	0.0108	0.0483	0.0107	0.0420	0.0083		-0.3	0.1034	0.0523	0.1015	0.0515	0.0994	0.0511
	-0.2	0.0595	0.0148	0.0493	0.0112	0.0460	0.0090		-0.2	0.1075	0.0509	0.0986	0.0514	0.0966	0.0501
	-0.1	0.0725	0.0167	0.0582	0.0140	0.0574	0.0109		-0.1	0.1142	0.0490	0.1006	0.0492	0.0973	0.0461
	0.1	0.0836	0.0218	0.0721	0.0214	0.0724	0.0209		0.1	0.1159	0.0529	0.0930	0.0437	0.0745	0.0283
	0.2	0.1018	0.0521	0.1057	0.0544	0.1021	0.0555		0.2	0.1262	0.0727	0.1029	0.0521	0.0594	0.0198
	0.3	0.2056	0.1210	0.2365	0.1256	0.2217	0.1244		0.3	0.2194	0.1292	0.2177	0.1054	0.1674	0.0729
	0.4	0.5007	0.2712	0.5628	0.2885	0.5245	0.2725		0.4	0.5094	0.2735	0.5395	0.2628	0.4723	0.2209
-0.2	-0.4	0.0698	0.0174	0.0860	0.0186	0.0803	0.0138	0.3	-0.4	0.2203	0.1295	0.1957	0.1247	0.1963	0.1242
	-0.3	0.0502	0.0116	0.0593	0.0145	0.0531	0.0113		-0.3	0.2336	0.1242	0.2050	0.1208	0.2059	0.1198
	-0.2	0.0476	0.0103	0.0469	0.0100	0.0411	0.0082		-0.2	0.2471	0.1191	0.2165	0.1158	0.2169	0.1141
	-0.1	0.0544	0.0126	0.0467	0.0097	0.0427	0.0082		-0.1	0.2572	0.1153	0.2250	0.1112	0.2239	0.1081
	0.1	0.0685	0.0216	0.0629	0.0202	0.0605	0.0204		0.1	0.2544	0.1157	0.2139	0.1034	0.2004	0.0895
	0.2	0.0999	0.0526	0.1097	0.0523	0.1041	0.0533		0.2	0.2422	0.1285	0.1883	0.1038	0.1487	0.0693
	0.3	0.2178	0.1166	0.2497	0.1202	0.2342	0.1184		0.3	0.2685	0.1753	0.2113	0.1301	0.1018	0.0517
	0.4	0.5165	0.2655	0.5782	0.2825	0.5396	0.2660		0.4	0.4970	0.3123	0.4779	0.2647	0.3533	0.1751
-0.1	-0.4	0.0849	0.0190	0.1024	0.0191	0.0973	0.0144	0.4	-0.4	0.5506	0.3005	0.4819	0.2784	0.4839	0.2777
	-0.3	0.0592	0.0146	0.0725	0.0164	0.0671	0.0126		-0.3	0.5693	0.2933	0.5001	0.2712	0.5018	0.2703
	-0.2	0.0478	0.0102	0.0538	0.0125	0.0479	0.0101		-0.2	0.5848	0.2874	0.5154	0.2650	0.5165	0.2637
	-0.1	0.0481	0.0100	0.0468	0.0097	0.0407	0.0082		-0.1	0.5954	0.2831	0.5255	0.2601	0.5252	0.2576
	0.1	0.0635	0.0218	0.0621	0.0191	0.0567	0.0181		0.1	0.5907	0.2821	0.5147	0.2522	0.5038	0.2404
	0.2	0.1038	0.0517	0.1152	0.0489	0.1076	0.0487		0.2	0.5698	0.2908	0.4801	0.2491	0.4499	0.2199
	0.3	0.2280	0.1132	0.2587	0.1153	0.2420	0.1122		0.3	0.5482	0.3222	0.4259	0.2557	0.3326	0.1771
	0.4	0.5279	0.2617	0.5880	0.2774	0.5485	0.2599		0.4	0.6362	0.4309	0.4909	0.3252	0.2145	0.1331

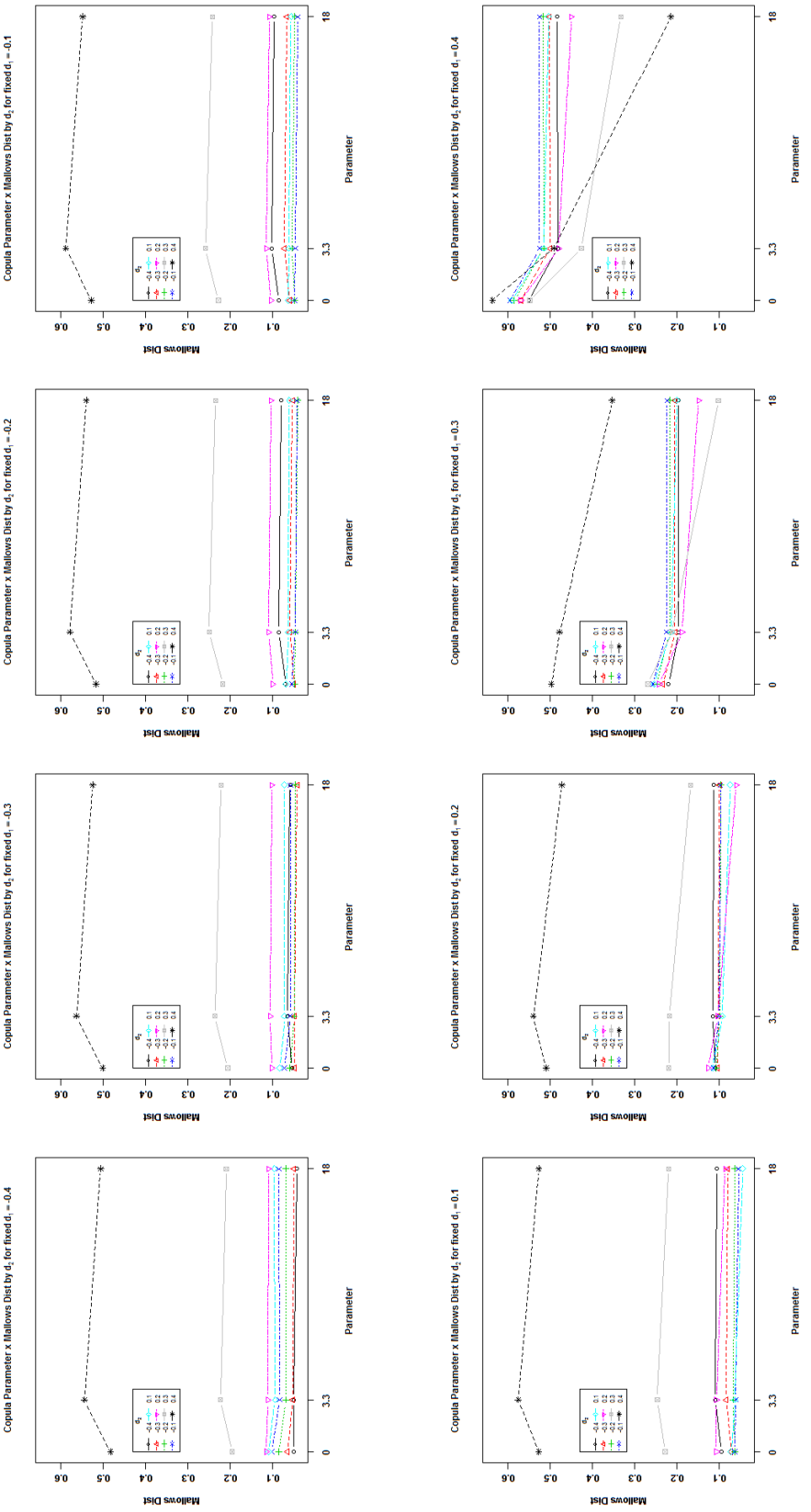


Figure 5: Plots of Mallows distance by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be standard normal.

Table 6: Simulation results for the Mallows distance in the Frank-t_3 case

$\boldsymbol{\varepsilon_t} \sim$ Frank-t_3															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.2552	0.2165	0.2550	0.2036	0.2567	0.1980	0.1	-0.4	0.2739	0.2056	0.2847	0.1953	0.2870	0.1890
	-0.3	0.2609	0.2141	0.2558	0.2012	0.2573	0.1956		-0.3	0.2641	0.2070	0.2707	0.1967	0.2722	0.1909
	-0.2	0.2707	0.2116	0.2622	0.1983	0.2639	0.1929		-0.2	0.2607	0.2079	0.2633	0.1976	0.2643	0.1925
	-0.1	0.2803	0.2097	0.2697	0.1957	0.2719	0.1906		-0.1	0.2612	0.2081	0.2607	0.1979	0.2613	0.1935
	0.1	0.2835	0.2090	0.2737	0.1949	0.2760	0.1897		0.1	0.2651	0.2078	0.2628	0.1979	0.2603	0.1942
	0.2	0.2847	0.2111	0.2819	0.1965	0.2832	0.1916		0.2	0.2827	0.2070	0.2854	0.1952	0.2795	0.1912
	0.3	0.3433	0.2232	0.3594	0.2101	0.3562	0.2050		0.3	0.3683	0.2141	0.3849	0.2000	0.3734	0.1925
	0.4	0.5977	0.3182	0.6383	0.3209	0.6229	0.3054		0.4	0.6416	0.3050	0.6776	0.3020	0.6539	0.2808
-0.3	-0.4	0.2563	0.2127	0.2615	0.2009	0.2635	0.1952	0.2	-0.4	0.2811	0.2077	0.2855	0.1978	0.2881	0.1917
	-0.3	0.2552	0.2121	0.2554	0.2004	0.2569	0.1950		-0.3	0.2788	0.2074	0.2785	0.1976	0.2803	0.1919
	-0.2	0.2597	0.2108	0.2560	0.1991	0.2576	0.1940		-0.2	0.2817	0.2068	0.2774	0.1970	0.2787	0.1917
	-0.1	0.2660	0.2095	0.2597	0.1975	0.2617	0.1927		-0.1	0.2864	0.2060	0.2791	0.1960	0.2799	0.1913
	0.1	0.2698	0.2087	0.2645	0.1964	0.2665	0.1916		0.1	0.2884	0.2065	0.2772	0.1970	0.2722	0.1931
	0.2	0.2781	0.2095	0.2804	0.1961	0.2813	0.1914		0.2	0.2948	0.2094	0.2850	0.1981	0.2713	0.1952
	0.3	0.3496	0.2189	0.3699	0.2067	0.3665	0.2014		0.3	0.3624	0.2223	0.3651	0.2042	0.3407	0.1953
	0.4	0.6140	0.3119	0.6565	0.3155	0.6413	0.2995		0.4	0.6230	0.3162	0.6461	0.3024	0.6079	0.2738
-0.2	-0.4	0.2628	0.2090	0.2718	0.1981	0.2739	0.1920	0.3	-0.4	0.3564	0.2224	0.3430	0.2110	0.3461	0.2047
	-0.3	0.2559	0.2098	0.2604	0.1990	0.2619	0.1933		-0.3	0.3661	0.2194	0.3491	0.2077	0.3516	0.2014
	-0.2	0.2556	0.2099	0.2560	0.1991	0.2574	0.1940		-0.2	0.3776	0.2167	0.3580	0.2046	0.3600	0.1982
	-0.1	0.2585	0.2094	0.2562	0.1985	0.2578	0.1939		-0.1	0.3875	0.2147	0.3659	0.2020	0.3672	0.1956
	0.1	0.2629	0.2085	0.2614	0.1972	0.2628	0.1926		0.1	0.3872	0.2153	0.3600	0.2019	0.3539	0.1936
	0.2	0.2775	0.2078	0.2837	0.1951	0.2840	0.1905		0.2	0.3777	0.2221	0.3444	0.2072	0.3241	0.1971
	0.3	0.3589	0.2152	0.3816	0.2036	0.3778	0.1980		0.3	0.4028	0.2465	0.3705	0.2207	0.3185	0.2056
	0.4	0.6297	0.3067	0.6731	0.3107	0.6577	0.2943		0.4	0.6123	0.3490	0.6004	0.3122	0.5221	0.2647
-0.1	-0.4	0.2703	0.2063	0.2816	0.1960	0.2837	0.1896	0.4	-0.4	0.6380	0.3373	0.5964	0.3102	0.5997	0.3037
	-0.3	0.2598	0.2079	0.2669	0.1974	0.2684	0.1916		-0.3	0.6559	0.3319	0.6127	0.3044	0.6156	0.2979
	-0.2	0.2558	0.2089	0.2592	0.1985	0.2603	0.1934		-0.2	0.6724	0.3272	0.6281	0.2993	0.6306	0.2926
	-0.1	0.2561	0.2092	0.2565	0.1987	0.2577	0.1942		-0.1	0.6852	0.3237	0.6398	0.2952	0.6415	0.2880
	0.1	0.2608	0.2083	0.2617	0.1975	0.2623	0.1931		0.1	0.6834	0.3232	0.6323	0.2923	0.6262	0.2788
	0.2	0.2800	0.2065	0.2881	0.1942	0.2875	0.1896		0.2	0.6632	0.3308	0.6017	0.2958	0.5813	0.2718
	0.3	0.3676	0.2127	0.3910	0.2012	0.3864	0.1952		0.3	0.6456	0.3570	0.5632	0.3109	0.5010	0.2637
	0.4	0.6421	0.3032	0.6854	0.3067	0.6692	0.2897		0.4	0.7403	0.4545	0.6457	0.3765	0.4834	0.2866

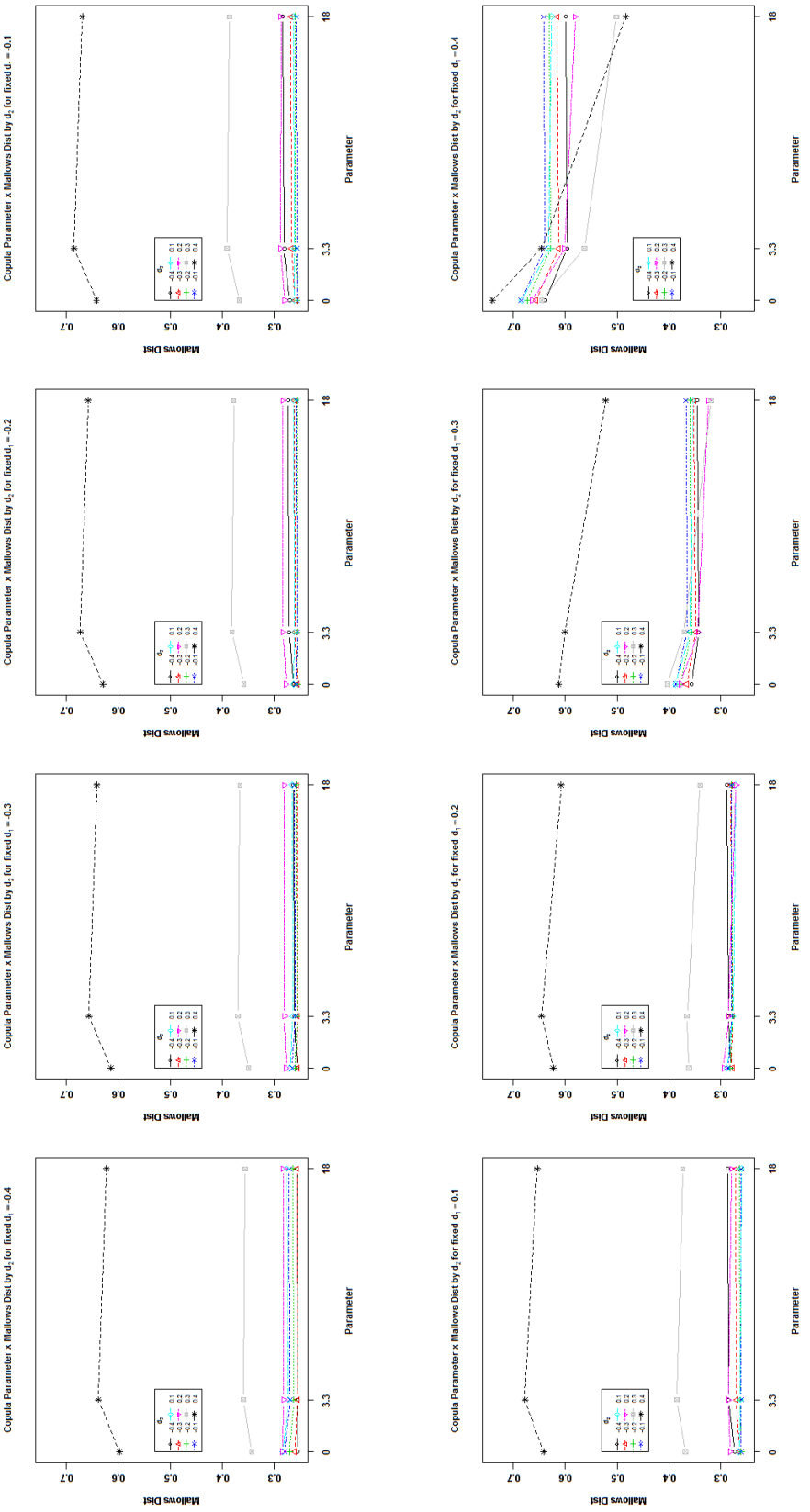


Figure 6: Plots of Mallows distance by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be (standardized) t_3.

Table 7: Simulation results for the Mallows distance in the Frank-t_7 case

$\boldsymbol{\varepsilon_t} \sim$ Frank-t_7															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	d_1	d_2	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd	$\hat{\mathscr{D}}_2$	sd
-0.4	-0.4	0.0841	0.0282	0.0839	0.0281	0.0796	0.0274	0.1	-0.4	0.1180	0.0299	0.1340	0.0288	0.1308	0.0269
	-0.3	0.0954	0.0289	0.0857	0.0283	0.0836	0.0268		-0.3	0.1006	0.0298	0.1111	0.0291	0.1078	0.0280
	-0.2	0.1122	0.0288	0.0973	0.0285	0.0974	0.0256		-0.2	0.0943	0.0297	0.0983	0.0295	0.0948	0.0288
	-0.1	0.1266	0.0283	0.1096	0.0283	0.1108	0.0247		-0.1	0.0951	0.0297	0.0932	0.0296	0.0889	0.0287
	0.1	0.1331	0.0291	0.1178	0.0296	0.1195	0.0274		0.1	0.1026	0.0346	0.0957	0.0305	0.0836	0.0284
	0.2	0.1375	0.0488	0.1331	0.0522	0.1326	0.0526		0.2	0.1314	0.0544	0.1323	0.0440	0.1153	0.0361
	0.3	0.2132	0.1208	0.2376	0.1267	0.2268	0.1265		0.3	0.2438	0.1135	0.2630	0.1048	0.2392	0.0938
	0.4	0.4974	0.2781	0.5527	0.2956	0.5221	0.2808		0.4	0.5430	0.2636	0.5875	0.2689	0.5460	0.2448
-0.3	-0.4	0.0864	0.0283	0.0959	0.0285	0.0914	0.0270	0.2	-0.4	0.1314	0.0498	0.1380	0.0483	0.1360	0.0477
	-0.3	0.0835	0.0283	0.0835	0.0282	0.0797	0.0276		-0.3	0.1271	0.0502	0.1261	0.0500	0.1244	0.0493
	-0.2	0.0918	0.0284	0.0840	0.0282	0.0821	0.0272		-0.2	0.1310	0.0492	0.1236	0.0500	0.1219	0.0486
	-0.1	0.1022	0.0280	0.0904	0.0281	0.0900	0.0262		-0.1	0.1372	0.0479	0.1253	0.0485	0.1225	0.0457
	0.1	0.1107	0.0293	0.1008	0.0297	0.1011	0.0287		0.1	0.1396	0.0516	0.1197	0.0451	0.1047	0.0345
	0.2	0.1262	0.0506	0.1295	0.0522	0.1270	0.0534		0.2	0.1494	0.0699	0.1298	0.0523	0.0959	0.0321
	0.3	0.2220	0.1178	0.2507	0.1222	0.2392	0.1217		0.3	0.2368	0.1272	0.2361	0.1042	0.1914	0.0757
	0.4	0.5158	0.2714	0.5721	0.2886	0.5413	0.2736		0.4	0.5252	0.2756	0.5525	0.2645	0.4924	0.2255
-0.2	-0.4	0.0982	0.0288	0.1130	0.0284	0.1089	0.0259	0.3	-0.4	0.2332	0.1249	0.2129	0.1208	0.2135	0.1199
	-0.3	0.0845	0.0283	0.0921	0.0282	0.0880	0.0272		-0.3	0.2460	0.1205	0.2212	0.1177	0.2220	0.1164
	-0.2	0.0833	0.0282	0.0832	0.0282	0.0799	0.0280		-0.2	0.2596	0.1160	0.2321	0.1135	0.2326	0.1115
	-0.1	0.0885	0.0280	0.0830	0.0282	0.0808	0.0276		-0.1	0.2699	0.1126	0.2406	0.1093	0.2397	0.1060
	0.1	0.0988	0.0299	0.0946	0.0296	0.0929	0.0294		0.1	0.2682	0.1132	0.2307	0.1027	0.2183	0.0889
	0.2	0.1246	0.0511	0.1334	0.0506	0.1293	0.0517		0.2	0.2572	0.1257	0.2080	0.1036	0.1718	0.0712
	0.3	0.2336	0.1142	0.2640	0.1175	0.2518	0.1163		0.3	0.2851	0.1725	0.2334	0.1297	0.1390	0.0612
	0.4	0.5319	0.2659	0.5883	0.2826	0.5571	0.2672		0.4	0.5131	0.3150	0.4941	0.2681	0.3786	0.1852
-0.1	-0.4	0.1105	0.0289	0.1275	0.0279	0.1237	0.0251	0.4	-0.4	0.5536	0.3010	0.4966	0.2784	0.4983	0.2775
	-0.3	0.0910	0.0284	0.1027	0.0279	0.0988	0.0264		-0.3	0.5729	0.2938	0.5148	0.2714	0.5163	0.2704
	-0.2	0.0835	0.0282	0.0885	0.0280	0.0847	0.0277		-0.2	0.5893	0.2879	0.5304	0.2654	0.5315	0.2639
	-0.1	0.0841	0.0281	0.0835	0.0281	0.0800	0.0282		-0.1	0.6006	0.2836	0.5410	0.2606	0.5408	0.2579
	0.1	0.0951	0.0304	0.0945	0.0293	0.0907	0.0290		0.1	0.5970	0.2828	0.5309	0.2537	0.5204	0.2416
	0.2	0.1278	0.0507	0.1387	0.0482	0.1328	0.0482		0.2	0.5766	0.2916	0.4968	0.2525	0.4674	0.2230
	0.3	0.2435	0.1114	0.2734	0.1130	0.2600	0.1107		0.3	0.5572	0.3225	0.4455	0.2619	0.3564	0.1853
	0.4	0.5436	0.2621	0.5989	0.2776	0.5666	0.2612		0.4	0.6505	0.4333	0.5163	0.3326	0.2659	0.1580

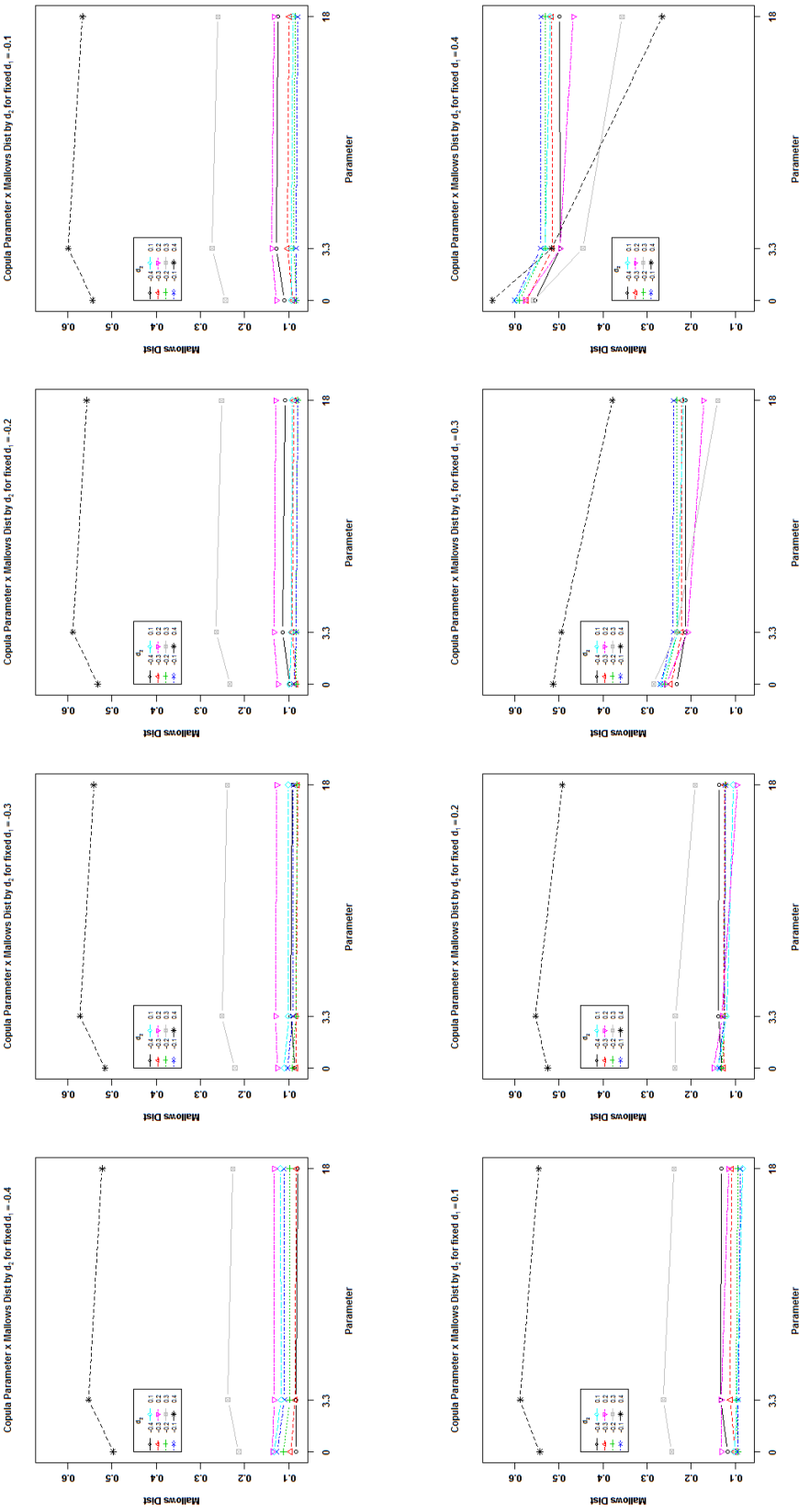


Figure 7: Plots of Mallows distance by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be (standardized) t_7.

Table 8: Simulation results for the Kendall's τ in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 1)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1,1)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0000	0.0165	0.3347	0.0139	0.7987	0.0048	0.1	-0.4	0.0001	0.0147	0.2856	0.0125	0.6137	0.0071
	-0.3	0.0000	0.0162	0.3332	0.0138	0.7911	0.0048		-0.3	0.0000	0.0147	0.2991	0.0125	0.6559	0.0067
	-0.2	-0.0001	0.0159	0.3282	0.0136	0.7678	0.0050		-0.2	0.0000	0.0147	0.3117	0.0125	0.6999	0.0061
	-0.1	-0.0001	0.0156	0.3192	0.0134	0.7296	0.0054		-0.1	-0.0001	0.0148	0.3226	0.0125	0.7436	0.0054
	0.1	-0.0003	0.0147	0.2844	0.0128	0.6131	0.0072		0.1	-0.0001	0.0154	0.3335	0.0130	0.7977	0.0045
	0.2	-0.0004	0.0139	0.2561	0.0124	0.5359	0.0092		0.2	-0.0001	0.0160	0.3279	0.0136	0.7736	0.0053
	0.3	-0.0006	0.0127	0.2187	0.0124	0.4459	0.0129		0.3	-0.0001	0.0169	0.3092	0.0149	0.7010	0.0110
	0.4	-0.0007	0.0112	0.1726	0.0133	0.3455	0.0182		0.4	0.0001	0.0182	0.2730	0.0183	0.5903	0.0217
-0.3	-0.4	0.0000	0.0162	0.3333	0.0137	0.7913	0.0047	0.2	-0.4	0.0001	0.0139	0.2580	0.0122	0.5367	0.0093
	-0.3	0.0000	0.0160	0.3346	0.0136	0.7986	0.0047		-0.3	0.0000	0.0140	0.2735	0.0122	0.5789	0.0092
	-0.2	0.0000	0.0158	0.3328	0.0134	0.7897	0.0047		-0.2	0.0000	0.0142	0.2892	0.0123	0.6251	0.0089
	-0.1	-0.0001	0.0155	0.3269	0.0132	0.7622	0.0050		-0.1	-0.0001	0.0145	0.3045	0.0124	0.6750	0.0083
	0.1	-0.0003	0.0147	0.2979	0.0127	0.6553	0.0068		0.1	0.0000	0.0159	0.3289	0.0135	0.7745	0.0054
	0.2	-0.0004	0.0140	0.2716	0.0124	0.5780	0.0091		0.2	0.0002	0.0175	0.3329	0.0148	0.7972	0.0051
	0.3	-0.0006	0.0130	0.2352	0.0126	0.4855	0.0134		0.3	0.0006	0.0200	0.3248	0.0172	0.7630	0.0079
	0.4	-0.0007	0.0115	0.1884	0.0140	0.3804	0.0196		0.4	0.0013	0.0237	0.2986	0.0217	0.6680	0.0185
-0.2	-0.4	0.0000	0.0159	0.3285	0.0135	0.7681	0.0049	0.3	-0.4	0.0000	0.0128	0.2215	0.0121	0.4472	0.0131
	-0.3	0.0000	0.0158	0.3330	0.0133	0.7899	0.0046		-0.3	0.0000	0.0130	0.2382	0.0123	0.4870	0.0137
	-0.2	0.0000	0.0156	0.3345	0.0132	0.7985	0.0045		-0.2	-0.0001	0.0134	0.2559	0.0125	0.5318	0.0141
	-0.1	-0.0001	0.0154	0.3322	0.0131	0.7877	0.0046		-0.1	-0.0001	0.0140	0.2746	0.0129	0.5826	0.0142
	0.1	-0.0003	0.0148	0.3106	0.0127	0.6993	0.0061		0.1	0.0002	0.0168	0.3120	0.0147	0.7033	0.0112
	0.2	-0.0004	0.0142	0.2873	0.0125	0.6241	0.0088		0.2	0.0007	0.0199	0.3267	0.0171	0.7651	0.0078
	0.3	-0.0006	0.0134	0.2528	0.0128	0.5302	0.0138		0.3	0.0017	0.0250	0.3325	0.0215	0.7963	0.0077
	0.4	-0.0007	0.0122	0.2061	0.0147	0.4207	0.0211		0.4	0.0033	0.0325	0.3213	0.0283	0.7489	0.0140
-0.1	-0.4	0.0001	0.0156	0.3197	0.0132	0.7300	0.0052	0.4	-0.4	-0.0001	0.0113	0.1767	0.0129	0.3478	0.0183
	-0.3	0.0000	0.0155	0.3272	0.0131	0.7626	0.0048		-0.3	-0.0001	0.0117	0.1929	0.0134	0.3829	0.0198
	-0.2	0.0000	0.0154	0.3325	0.0130	0.7879	0.0045		-0.2	-0.0001	0.0123	0.2111	0.0140	0.4235	0.0212
	-0.1	-0.0001	0.0152	0.3343	0.0129	0.7983	0.0044		-0.1	-0.0001	0.0133	0.2314	0.0147	0.4711	0.0224
	0.1	-0.0003	0.0149	0.3217	0.0127	0.7430	0.0054		0.1	0.0004	0.0182	0.2786	0.0175	0.5946	0.0216
	0.2	-0.0004	0.0146	0.3027	0.0126	0.6740	0.0081		0.2	0.0013	0.0237	0.3036	0.0212	0.6729	0.0180
	0.3	-0.0005	0.0140	0.2713	0.0131	0.5808	0.0139		0.3	0.0032	0.0327	0.3248	0.0283	0.7536	0.0130
	0.4	-0.0007	0.0132	0.2261	0.0155	0.4679	0.0222		0.4	0.0062	0.0454	0.3332	0.0390	0.7950	0.0146

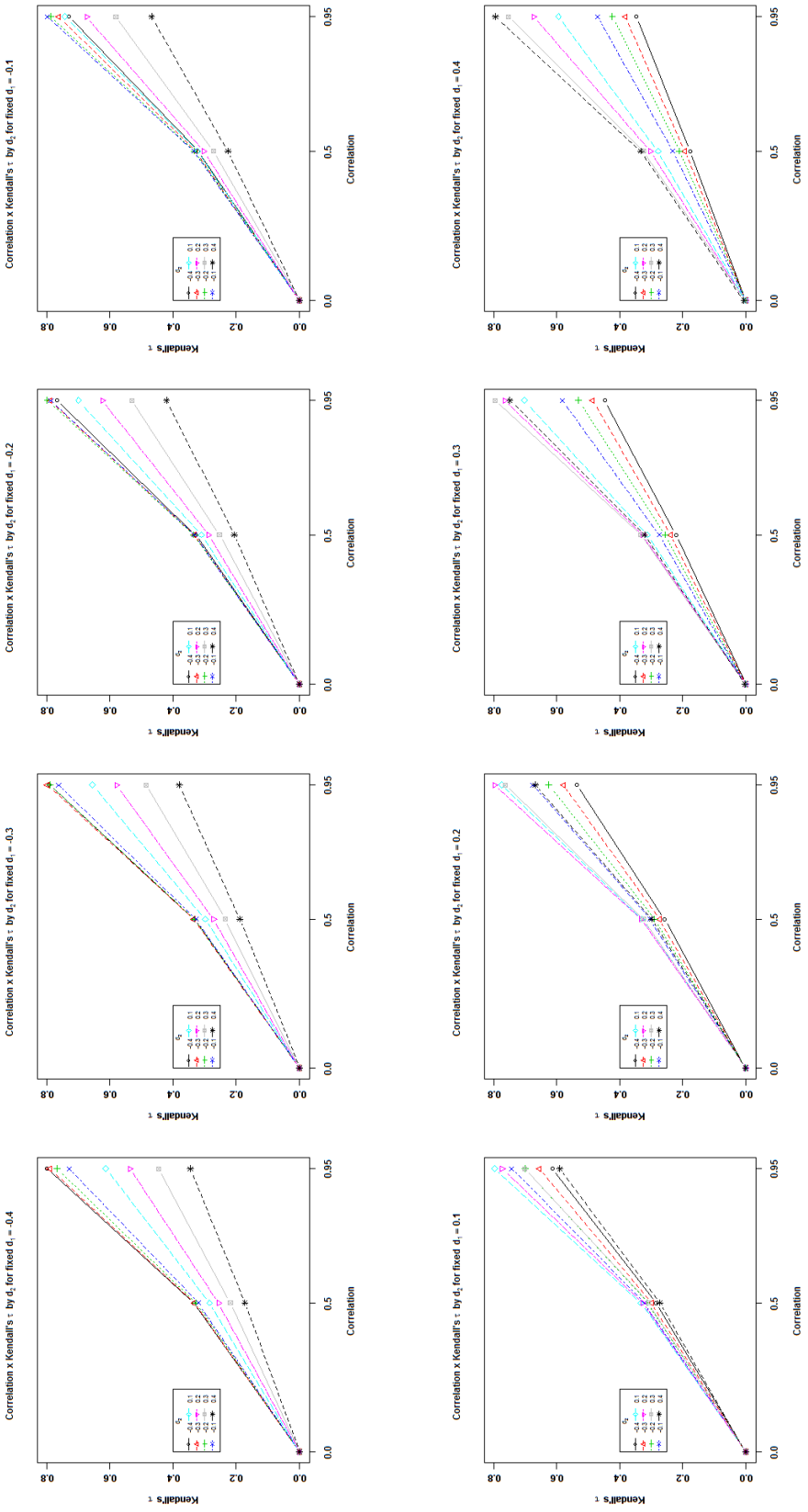


Figure 8: Plots of Kendall's τ by correlation for fixed d_1. Gaussian case with equal variances.

Table 9: Simulation results for the Kendall's τ in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 2)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1,2)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0000	0.0152	0.0095	0.0269	0.0191	0.0569	0.1	-0.4	-0.0004	0.0144	0.0100	0.0262	0.0163	0.0422
	-0.3	0.0003	0.0157	0.0099	0.0290	0.0190	0.0515		-0.3	-0.0008	0.0138	0.0116	0.0330	0.0226	0.0547
	-0.2	0.0003	0.0157	0.0119	0.0328	0.0211	0.0544		-0.2	0.0000	0.0148	0.0087	0.0281	0.0170	0.0467
	-0.1	-0.0002	0.0153	0.0109	0.0330	0.0220	0.0599		-0.1	-0.0006	0.0150	0.0117	0.0298	0.0189	0.0444
	0.1	0.0001	0.0146	0.0093	0.0265	0.0155	0.0392		0.1	0.0010	0.0154	0.0143	0.0380	0.0315	0.0807
	0.2	-0.0007	0.0137	0.0085	0.0244	0.0192	0.0424		0.2	0.0022	0.0156	0.0143	0.0295	0.0263	0.0626
	0.3	-0.0002	0.0128	0.0057	0.0210	0.0137	0.0320		0.3	0.0029	0.0160	0.0186	0.0344	0.0228	0.0495
	0.4	0.0001	0.0103	0.0072	0.0192	0.0153	0.0356		0.4	0.0034	0.0191	0.0153	0.0294	0.0248	0.0481
-0.3	-0.4	-0.0009	0.0162	0.0111	0.0337	0.0329	0.0787	0.2	-0.4	0.0005	0.0136	0.0071	0.0205	0.0196	0.0547
	-0.3	0.0006	0.0158	0.0072	0.0283	0.0225	0.0583		-0.3	-0.0007	0.0137	0.0084	0.0236	0.0199	0.0496
	-0.2	-0.0006	0.0158	0.0094	0.0309	0.0193	0.0593		-0.2	-0.0001	0.0140	0.0135	0.0332	0.0206	0.0467
	-0.1	0.0004	0.0148	0.0114	0.0323	0.0226	0.0580		-0.1	0.0009	0.0143	0.0130	0.0345	0.0256	0.0644
	0.1	0.0001	0.0141	0.0104	0.0252	0.0188	0.0429		0.1	0.0022	0.0155	0.0154	0.0352	0.0297	0.0663
	0.2	0.0007	0.0137	0.0127	0.0298	0.0228	0.0631		0.2	0.0027	0.0173	0.0185	0.0391	0.0258	0.0595
	0.3	0.0009	0.0135	0.0079	0.0202	0.0135	0.0354		0.3	0.0046	0.0210	0.0233	0.0407	0.0290	0.0537
	0.4	0.0001	0.0116	0.0078	0.0193	0.0112	0.0292		0.4	0.0072	0.0246	0.0202	0.0323	0.0237	0.0517
-0.2	-0.4	0.0003	0.0156	0.0093	0.0262	0.0267	0.0671	0.3	-0.4	-0.0002	0.0119	0.0064	0.0196	0.0134	0.0328
	-0.3	0.0000	0.0156	0.0106	0.0263	0.0328	0.0803		-0.3	0.0001	0.0130	0.0092	0.0225	0.0140	0.0359
	-0.2	-0.0007	0.0151	0.0126	0.0334	0.0257	0.0622		-0.2	-0.0004	0.0135	0.0082	0.0226	0.0149	0.0400
	-0.1	-0.0002	0.0156	0.0125	0.0323	0.0220	0.0585		-0.1	0.0018	0.0145	0.0133	0.0298	0.0184	0.0479
	0.1	0.0001	0.0141	0.0097	0.0264	0.0221	0.0511		0.1	0.0028	0.0176	0.0146	0.0336	0.0279	0.0550
	0.2	0.0000	0.0141	0.0178	0.0369	0.0206	0.0476		0.2	0.0052	0.0198	0.0178	0.0350	0.0242	0.0429
	0.3	0.0003	0.0126	0.0090	0.0223	0.0151	0.0351		0.3	0.0090	0.0273	0.0232	0.0398	0.0369	0.0691
	0.4	0.0002	0.0120	0.0074	0.0184	0.0156	0.0384		0.4	0.0138	0.0343	0.0287	0.0427	0.0319	0.0640
-0.1	-0.4	0.0005	0.0150	0.0096	0.0270	0.0251	0.0695	0.4	-0.4	0.0000	0.0107	0.0083	0.0217	0.0109	0.0296
	-0.3	-0.0006	0.0156	0.0123	0.0303	0.0160	0.0450		-0.3	0.0007	0.0112	0.0072	0.0186	0.0090	0.0270
	-0.2	-0.0006	0.0155	0.0096	0.0289	0.0286	0.0784		-0.2	0.0006	0.0124	0.0116	0.0250	0.0128	0.0299
	-0.1	0.0006	0.0156	0.0097	0.0289	0.0221	0.0499		-0.1	0.0007	0.0128	0.0081	0.0208	0.0157	0.0412
	0.1	0.0006	0.0141	0.0120	0.0314	0.0217	0.0543		0.1	0.0047	0.0194	0.0134	0.0279	0.0227	0.0432
	0.2	0.0011	0.0150	0.0093	0.0240	0.0231	0.0508		0.2	0.0088	0.0258	0.0210	0.0427	0.0307	0.0538
	0.3	0.0001	0.0136	0.0125	0.0304	0.0229	0.0515		0.3	0.0102	0.0356	0.0276	0.0453	0.0410	0.0719
	0.4	0.0016	0.0134	0.0091	0.0229	0.0168	0.0374		0.4	0.0133	0.0557	0.0369	0.0623	0.0511	0.0821

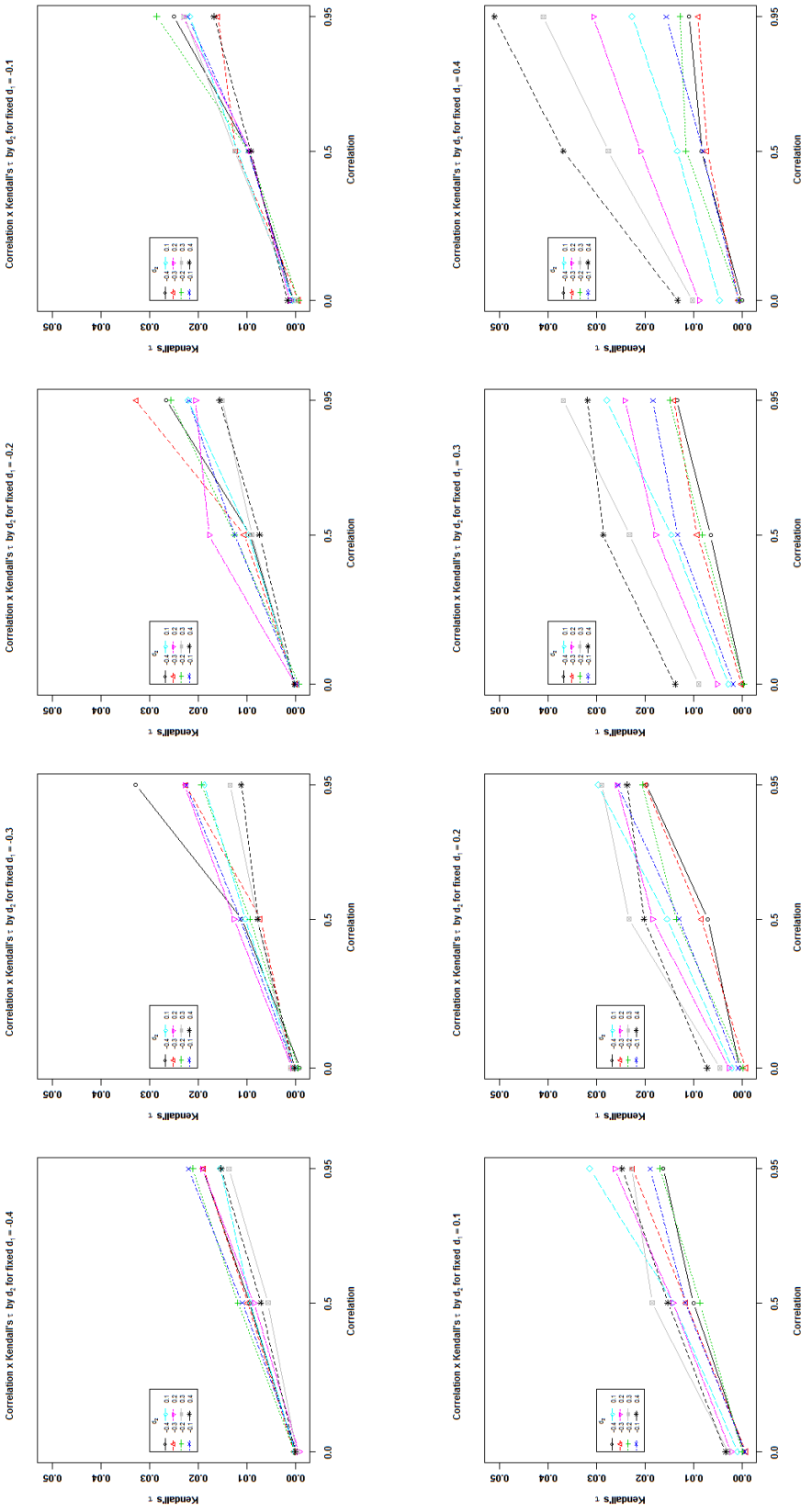


Figure 9: Plots of Kendall's τ by correlation for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.

Table 10: Simulation results for the Kendall's τ in the Gaussian case, $\boldsymbol{\sigma}^2 = (1, 3)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (1,3)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0011	0.0153	0.0085	0.0294	0.0203	0.0587	0.1	-0.4	-0.0006	0.0147	0.0119	0.0302	0.0151	0.0407
	-0.3	0.0004	0.0155	0.0083	0.0251	0.0197	0.0503		-0.3	0.0006	0.0137	0.0071	0.0241	0.0204	0.0511
	-0.2	-0.0003	0.0161	0.0128	0.0353	0.0230	0.0647		-0.2	0.0006	0.0150	0.0065	0.0238	0.0166	0.0505
	-0.1	-0.0008	0.0156	0.0128	0.0432	0.0234	0.0648		-0.1	0.0006	0.0148	0.0076	0.0269	0.0200	0.0533
	0.1	0.0000	0.0149	0.0093	0.0263	0.0135	0.0399		0.1	0.0013	0.0151	0.0117	0.0297	0.0222	0.0553
	0.2	-0.0003	0.0138	0.0096	0.0270	0.0179	0.0416		0.2	0.0025	0.0159	0.0108	0.0251	0.0213	0.0579
	0.3	0.0001	0.0126	0.0058	0.0199	0.0116	0.0326		0.3	0.0019	0.0165	0.0163	0.0326	0.0263	0.0564
	0.4	0.0004	0.0108	0.0076	0.0197	0.0146	0.0395		0.4	0.0040	0.0195	0.0136	0.0280	0.0213	0.0457
-0.3	-0.4	0.0006	0.0150	0.0143	0.0366	0.0247	0.0641	0.2	-0.4	0.0008	0.0133	0.0067	0.0225	0.0165	0.0589
	-0.3	-0.0003	0.0156	0.0102	0.0305	0.0188	0.0549		-0.3	0.0006	0.0137	0.0067	0.0221	0.0178	0.0480
	-0.2	-0.0002	0.0157	0.0085	0.0301	0.0201	0.0641		-0.2	-0.0004	0.0144	0.0098	0.0259	0.0156	0.0436
	-0.1	-0.0006	0.0152	0.0095	0.0294	0.0153	0.0517		-0.1	0.0010	0.0145	0.0089	0.0265	0.0256	0.0652
	0.1	-0.0003	0.0145	0.0089	0.0275	0.0227	0.0632		0.1	0.0023	0.0160	0.0154	0.0362	0.0238	0.0618
	0.2	0.0006	0.0136	0.0127	0.0291	0.0174	0.0469		0.2	0.0020	0.0174	0.0154	0.0347	0.0203	0.0587
	0.3	-0.0008	0.0128	0.0052	0.0199	0.0127	0.0389		0.3	0.0055	0.0203	0.0163	0.0333	0.0254	0.0526
	0.4	0.0001	0.0117	0.0071	0.0210	0.0100	0.0300		0.4	0.0055	0.0252	0.0166	0.0326	0.0236	0.0511
-0.2	-0.4	0.0007	0.0156	0.0095	0.0255	0.0147	0.0433	0.3	-0.4	-0.0011	0.0124	0.0054	0.0202	0.0120	0.0335
	-0.3	0.0001	0.0155	0.0086	0.0278	0.0292	0.0869		-0.3	0.0004	0.0129	0.0085	0.0235	0.0124	0.0366
	-0.2	-0.0002	0.0154	0.0085	0.0276	0.0277	0.0750		-0.2	0.0002	0.0130	0.0078	0.0220	0.0175	0.0464
	-0.1	0.0010	0.0161	0.0101	0.0329	0.0206	0.0574		-0.1	0.0017	0.0144	0.0100	0.0261	0.0172	0.0458
	0.1	0.0008	0.0142	0.0094	0.0278	0.0193	0.0478		0.1	0.0024	0.0171	0.0140	0.0303	0.0230	0.0487
	0.2	-0.0004	0.0142	0.0180	0.0357	0.0200	0.0497		0.2	0.0053	0.0214	0.0209	0.0412	0.0228	0.0492
	0.3	0.0002	0.0125	0.0083	0.0224	0.0135	0.0375		0.3	0.0085	0.0261	0.0229	0.0372	0.0301	0.0573
	0.4	0.0011	0.0116	0.0075	0.0202	0.0123	0.0389		0.4	0.0126	0.0376	0.0257	0.0440	0.0368	0.0685
-0.1	-0.4	0.0005	0.0150	0.0093	0.0275	0.0179	0.0529	0.4	-0.4	0.0000	0.0103	0.0072	0.0196	0.0088	0.0284
	-0.3	0.0002	0.0154	0.0116	0.0313	0.0174	0.0516		-0.3	-0.0001	0.0110	0.0062	0.0176	0.0103	0.0304
	-0.2	-0.0001	0.0156	0.0092	0.0282	0.0282	0.0823		-0.2	0.0004	0.0115	0.0092	0.0228	0.0145	0.0419
	-0.1	-0.0004	0.0150	0.0074	0.0270	0.0201	0.0532		-0.1	0.0009	0.0129	0.0106	0.0249	0.0137	0.0331
	0.1	-0.0005	0.0147	0.0110	0.0308	0.0239	0.0611		0.1	0.0035	0.0179	0.0115	0.0263	0.0210	0.0437
	0.2	0.0006	0.0134	0.0094	0.0239	0.0200	0.0473		0.2	0.0088	0.0243	0.0177	0.0361	0.0259	0.0483
	0.3	0.0003	0.0140	0.0101	0.0295	0.0199	0.0481		0.3	0.0097	0.0373	0.0257	0.0426	0.0411	0.0661
	0.4	0.0011	0.0131	0.0070	0.0212	0.0163	0.0381		0.4	0.0128	0.0513	0.0327	0.0610	0.0452	0.0733

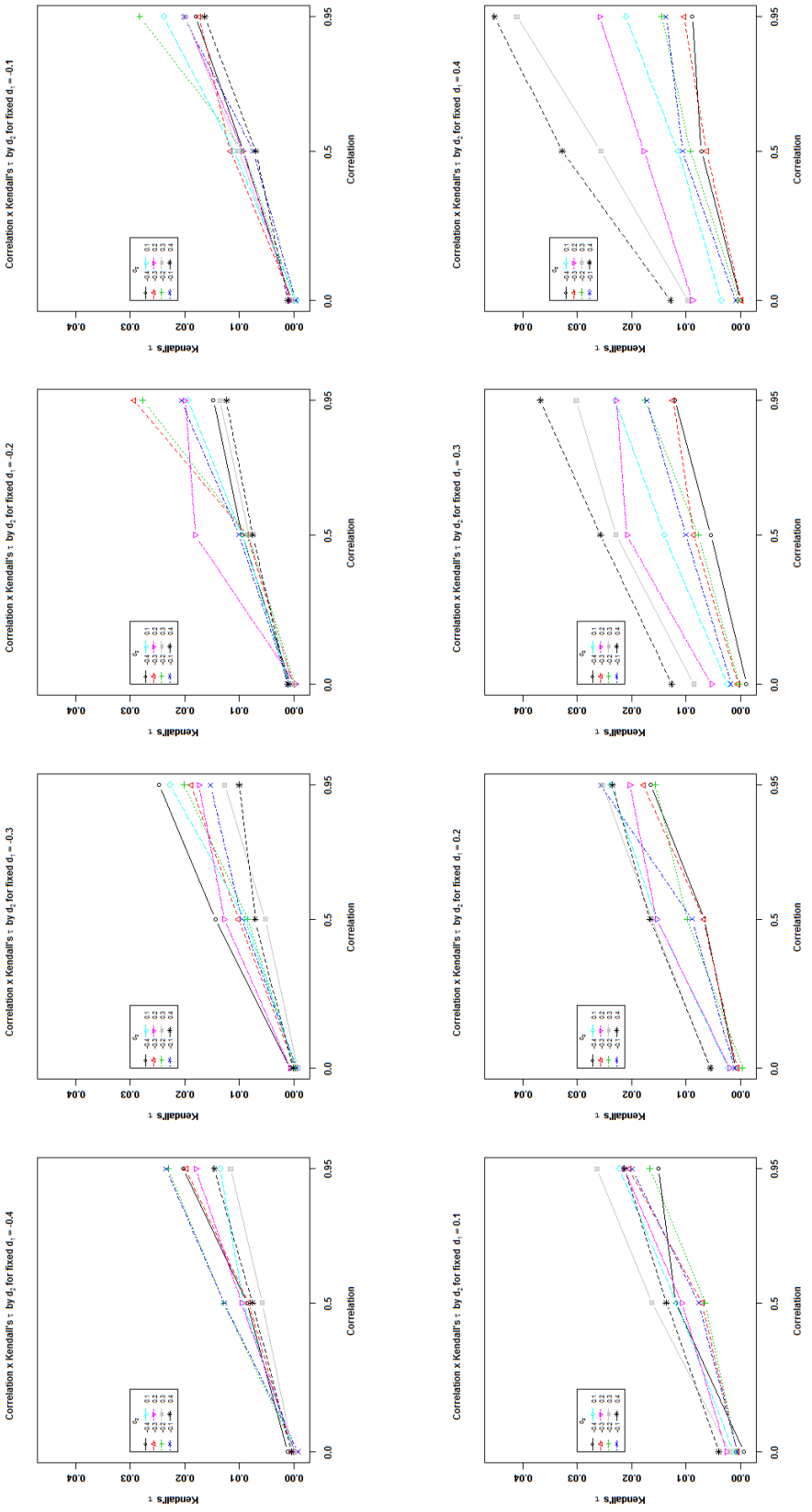


Figure 10: Plots of Kendall's τ by correlation for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

Table 11: Simulation results for the Kendall's τ in the Gaussian case, $\boldsymbol{\sigma}^2 = (2, 3)$.

$\boldsymbol{\varepsilon_t} \sim \mathcal{N}_\rho(\mathbf{0}, \boldsymbol{\sigma}^2)$, $\boldsymbol{\sigma}^2 = (2,3)$															
$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$		$\boldsymbol{d}$		$\rho = 0$		$\rho = 0.5$		$\rho = 0.95$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0007	0.0158	0.0089	0.0300	0.0199	0.0607	0.1	-0.4	0.0007	0.0141	0.0088	0.0274	0.0155	0.0445
	-0.3	0.0001	0.0157	0.0079	0.0290	0.0194	0.0552		-0.3	0.0004	0.0140	0.0074	0.0260	0.0185	0.0529
	-0.2	0.0004	0.0154	0.0113	0.0351	0.0210	0.0621		-0.2	0.0002	0.0145	0.0069	0.0261	0.0161	0.0483
	-0.1	0.0005	0.0153	0.0079	0.0272	0.0232	0.0691		-0.1	0.0009	0.0143	0.0095	0.0282	0.0173	0.0509
	0.1	0.0002	0.0147	0.0079	0.0267	0.0152	0.0459		0.1	0.0005	0.0151	0.0117	0.0309	0.0201	0.0535
	0.2	0.0002	0.0135	0.0087	0.0265	0.0171	0.0406		0.2	0.0027	0.0156	0.0108	0.0278	0.0297	0.0690
	0.3	0.0004	0.0117	0.0055	0.0200	0.0116	0.0354		0.3	0.0021	0.0166	0.0185	0.0355	0.0224	0.0533
	0.4	0.0002	0.0107	0.0049	0.0166	0.0143	0.0358		0.4	0.0037	0.0190	0.0134	0.0270	0.0228	0.0507
-0.3	-0.4	0.0001	0.0159	0.0096	0.0307	0.0221	0.0623	0.2	-0.4	0.0000	0.0135	0.0077	0.0238	0.0118	0.0358
	-0.3	0.0005	0.0160	0.0087	0.0310	0.0186	0.0574		-0.3	0.0001	0.0139	0.0072	0.0235	0.0178	0.0478
	-0.2	-0.0005	0.0162	0.0110	0.0330	0.0204	0.0687		-0.2	-0.0001	0.0140	0.0079	0.0261	0.0153	0.0447
	-0.1	0.0004	0.0158	0.0090	0.0301	0.0180	0.0525		-0.1	0.0009	0.0140	0.0100	0.0293	0.0209	0.0543
	0.1	0.0010	0.0151	0.0101	0.0285	0.0227	0.0546		0.1	0.0016	0.0161	0.0109	0.0278	0.0232	0.0644
	0.2	0.0007	0.0137	0.0115	0.0295	0.0164	0.0434		0.2	0.0032	0.0178	0.0139	0.0347	0.0227	0.0613
	0.3	0.0005	0.0125	0.0070	0.0224	0.0106	0.0372		0.3	0.0049	0.0206	0.0172	0.0310	0.0231	0.0514
	0.4	0.0004	0.0114	0.0063	0.0196	0.0107	0.0335		0.4	0.0070	0.0251	0.0183	0.0347	0.0242	0.0541
-0.2	-0.4	0.0000	0.0154	0.0095	0.0274	0.0192	0.0549	0.3	-0.4	0.0000	0.0127	0.0065	0.0200	0.0109	0.0318
	-0.3	-0.0003	0.0156	0.0094	0.0281	0.0190	0.0561		-0.3	0.0007	0.0123	0.0089	0.0240	0.0118	0.0359
	-0.2	0.0008	0.0158	0.0103	0.0309	0.0257	0.0636		-0.2	0.0003	0.0131	0.0079	0.0232	0.0163	0.0444
	-0.1	-0.0010	0.0151	0.0099	0.0298	0.0242	0.0620		-0.1	0.0023	0.0139	0.0107	0.0262	0.0147	0.0434
	0.1	0.0000	0.0145	0.0103	0.0277	0.0190	0.0539		0.1	0.0027	0.0168	0.0145	0.0317	0.0241	0.0522
	0.2	0.0000	0.0134	0.0150	0.0342	0.0209	0.0556		0.2	0.0043	0.0207	0.0191	0.0357	0.0237	0.0495
	0.3	-0.0005	0.0132	0.0084	0.0223	0.0151	0.0404		0.3	0.0081	0.0259	0.0222	0.0360	0.0306	0.0602
	0.4	0.0002	0.0116	0.0066	0.0202	0.0146	0.0408		0.4	0.0105	0.0370	0.0280	0.0437	0.0294	0.0646
-0.1	-0.4	-0.0002	0.0153	0.0117	0.0301	0.0168	0.0492	0.4	-0.4	-0.0002	0.0104	0.0064	0.0183	0.0095	0.0290
	-0.3	-0.0004	0.0153	0.0104	0.0304	0.0163	0.0546		-0.3	0.0005	0.0112	0.0047	0.0167	0.0100	0.0330
	-0.2	-0.0001	0.0151	0.0086	0.0283	0.0145	0.0464		-0.2	0.0002	0.0115	0.0085	0.0234	0.0117	0.0302
	-0.1	-0.0003	0.0145	0.0080	0.0303	0.0215	0.0553		-0.1	0.0012	0.0129	0.0090	0.0214	0.0129	0.0323
	0.1	0.0001	0.0148	0.0123	0.0356	0.0173	0.0518		0.1	0.0052	0.0184	0.0119	0.0268	0.0205	0.0435
	0.2	0.0011	0.0145	0.0098	0.0262	0.0215	0.0537		0.2	0.0076	0.0251	0.0189	0.0355	0.0273	0.0575
	0.3	0.0005	0.0140	0.0116	0.0297	0.0195	0.0462		0.3	0.0112	0.0367	0.0246	0.0439	0.0344	0.0641
	0.4	0.0009	0.0128	0.0089	0.0217	0.0149	0.0378		0.4	0.0133	0.0553	0.0306	0.0607	0.0515	0.0840

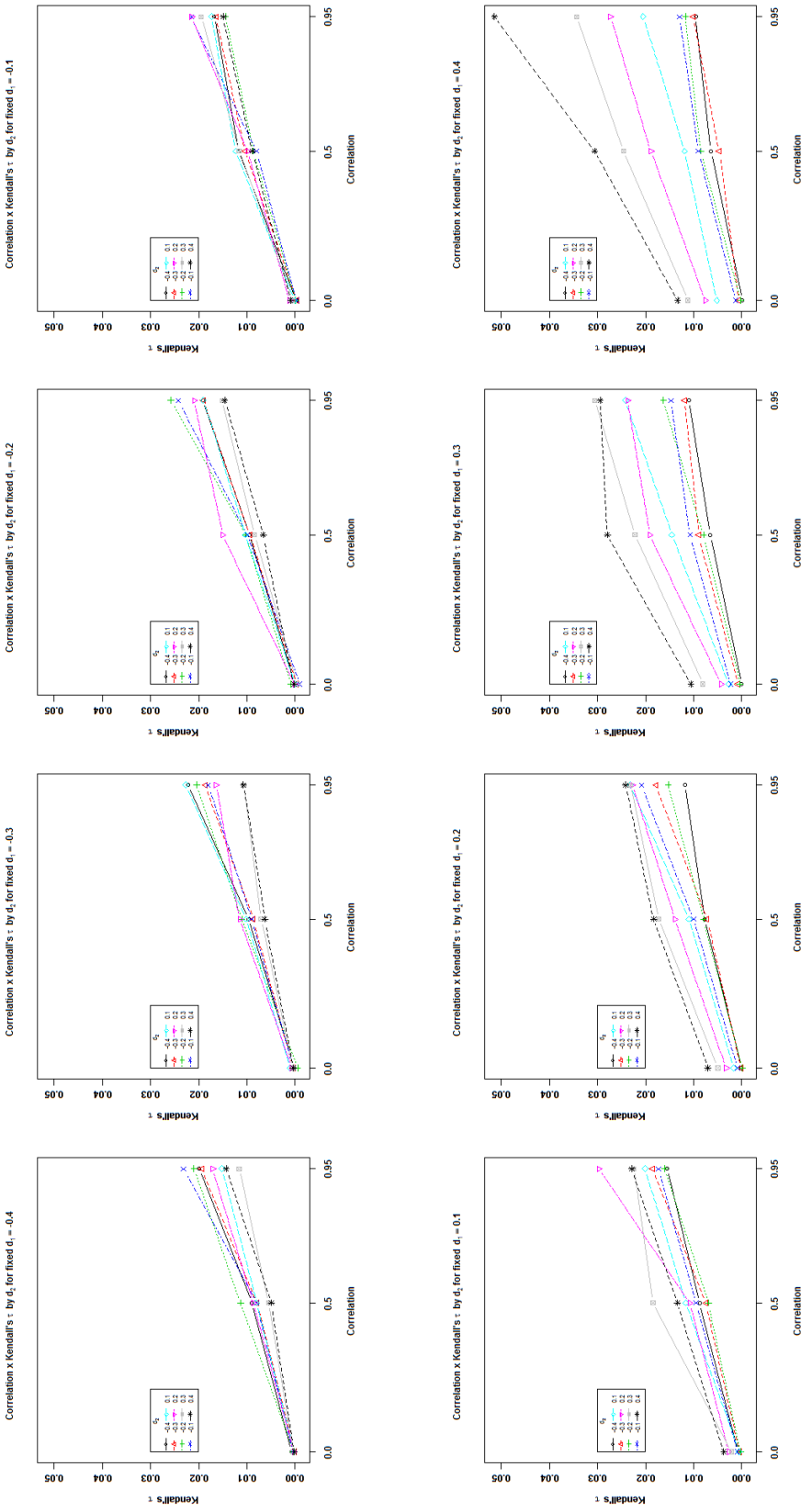


Figure 11: Plots of Kendall's τ by correlation for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.

Table 12: Simulation results for the Kendall's τ in the Frank-$\mathcal{N}(0,1)$ case

$\boldsymbol{\varepsilon_t} \sim$ Frank-$\mathcal{N}(0,1)$															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0059	0.0159	0.3238	0.0136	0.7770	0.0047	0.1	-0.4	0.0048	0.0140	0.2823	0.0123	0.6066	0.0071
	-0.3	0.0060	0.0157	0.3235	0.0134	0.7731	0.0046		-0.3	0.0050	0.0141	0.2963	0.0122	0.6493	0.0065
	-0.2	0.0061	0.0155	0.3200	0.0131	0.7542	0.0047		-0.2	0.0054	0.0141	0.3095	0.0121	0.6940	0.0057
	-0.1	0.0062	0.0152	0.3125	0.0128	0.7193	0.0051		-0.1	0.0058	0.0142	0.3208	0.0120	0.7387	0.0048
	0.1	0.0063	0.0143	0.2801	0.0121	0.6057	0.0072		0.1	0.0074	0.0146	0.3312	0.0123	0.7928	0.0038
	0.2	0.0063	0.0135	0.2522	0.0119	0.5288	0.0098		0.2	0.0085	0.0151	0.3243	0.0129	0.7637	0.0057
	0.3	0.0059	0.0125	0.2149	0.0123	0.4394	0.0141		0.3	0.0097	0.0162	0.3036	0.0144	0.6883	0.0125
	0.4	0.0052	0.0111	0.1692	0.0135	0.3408	0.0194		0.4	0.0106	0.0180	0.2662	0.0175	0.5796	0.0228
-0.3	-0.4	0.0059	0.0157	0.3238	0.0134	0.7730	0.0046	0.2	-0.4	0.0042	0.0133	0.2553	0.0121	0.5308	0.0094
	-0.3	0.0060	0.0155	0.3263	0.0132	0.7826	0.0043		-0.3	0.0044	0.0134	0.2713	0.0120	0.5733	0.0092
	-0.2	0.0061	0.0153	0.3257	0.0130	0.7772	0.0043		-0.2	0.0048	0.0136	0.2873	0.0119	0.6198	0.0088
	-0.1	0.0063	0.0151	0.3211	0.0127	0.7529	0.0045		-0.1	0.0053	0.0139	0.3028	0.0119	0.6698	0.0079
	0.1	0.0066	0.0143	0.2941	0.0121	0.6481	0.0067		0.1	0.0073	0.0152	0.3262	0.0128	0.7669	0.0049
	0.2	0.0067	0.0136	0.2681	0.0119	0.5711	0.0097		0.2	0.0089	0.0168	0.3284	0.0143	0.7836	0.0052
	0.3	0.0065	0.0126	0.2314	0.0125	0.4791	0.0147		0.3	0.0109	0.0196	0.3177	0.0169	0.7439	0.0094
	0.4	0.0059	0.0114	0.1849	0.0140	0.3757	0.0209		0.4	0.0126	0.0241	0.2895	0.0211	0.6508	0.0195
-0.2	-0.4	0.0057	0.0154	0.3206	0.0132	0.7541	0.0047	0.3	-0.4	0.0034	0.0122	0.2191	0.0122	0.4426	0.0133
	-0.3	0.0059	0.0153	0.3261	0.0130	0.7772	0.0042		-0.3	0.0036	0.0125	0.2360	0.0122	0.4827	0.0137
	-0.2	0.0061	0.0151	0.3288	0.0128	0.7884	0.0040		-0.2	0.0040	0.0128	0.2539	0.0123	0.5277	0.0140
	-0.1	0.0063	0.0149	0.3276	0.0125	0.7802	0.0040		-0.1	0.0045	0.0134	0.2725	0.0123	0.5782	0.0138
	0.1	0.0069	0.0143	0.3074	0.0120	0.6926	0.0060		0.1	0.0070	0.0162	0.3084	0.0139	0.6949	0.0105
	0.2	0.0071	0.0137	0.2841	0.0120	0.6171	0.0094		0.2	0.0091	0.0196	0.3210	0.0168	0.7495	0.0081
	0.3	0.0071	0.0129	0.2489	0.0127	0.5234	0.0152		0.3	0.0119	0.0253	0.3234	0.0219	0.7686	0.0094
	0.4	0.0067	0.0119	0.2024	0.0146	0.4158	0.0224		0.4	0.0147	0.0337	0.3090	0.0288	0.7193	0.0156
-0.1	-0.4	0.0055	0.0150	0.3135	0.0129	0.7193	0.0051	0.4	-0.4	0.0025	0.0108	0.1745	0.0130	0.3449	0.0184
	-0.3	0.0057	0.0149	0.3220	0.0127	0.7529	0.0044		-0.3	0.0026	0.0112	0.1909	0.0134	0.3804	0.0197
	-0.2	0.0060	0.0148	0.3282	0.0126	0.7804	0.0039		-0.2	0.0029	0.0118	0.2091	0.0137	0.4212	0.0209
	-0.1	0.0063	0.0147	0.3309	0.0124	0.7933	0.0037		-0.1	0.0034	0.0128	0.2292	0.0141	0.4686	0.0217
	0.1	0.0071	0.0143	0.3191	0.0120	0.7371	0.0052		0.1	0.0061	0.0178	0.2745	0.0165	0.5886	0.0202
	0.2	0.0076	0.0139	0.2996	0.0120	0.6666	0.0087		0.2	0.0087	0.0236	0.2972	0.0212	0.6604	0.0171
	0.3	0.0078	0.0134	0.2673	0.0129	0.5732	0.0153		0.3	0.0121	0.0331	0.3142	0.0297	0.7270	0.0154
	0.4	0.0077	0.0128	0.2218	0.0152	0.4622	0.0235		0.4	0.0160	0.0465	0.3176	0.0410	0.7510	0.0187

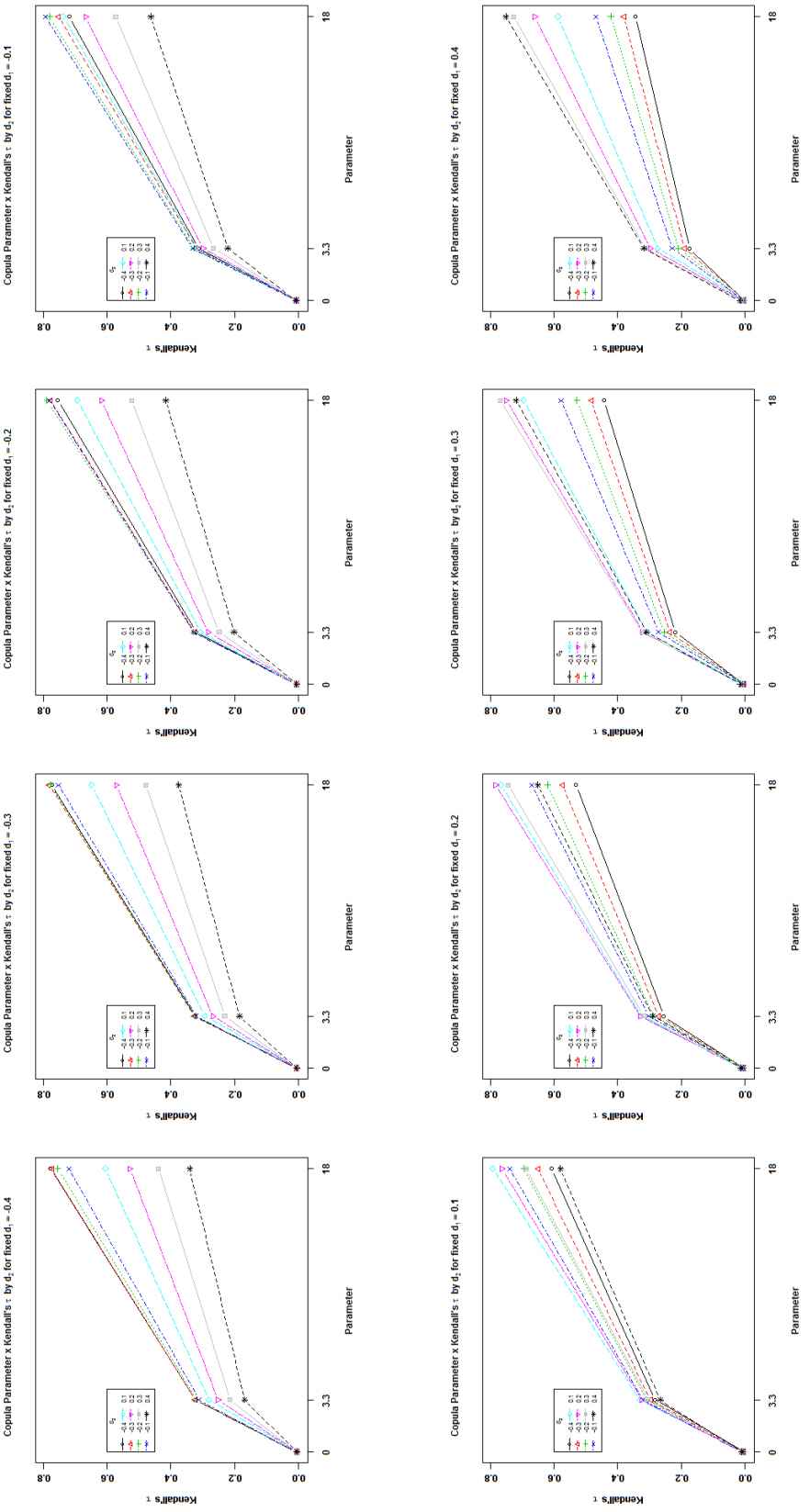


Figure 12: Plots of Kendall's τ by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be standard normal.

Table 13: Simulation results for the Kendall's τ in the Frank-t_3 case

$\boldsymbol{\varepsilon_t} \sim$ Frank-t_3															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0022	0.0572	0.3048	0.0729	0.7346	0.0954	0.1	-0.4	0.0013	0.0567	0.2598	0.0703	0.5530	0.0858
	-0.3	0.0023	0.0572	0.3058	0.0729	0.7340	0.0954		-0.3	0.0016	0.0567	0.2756	0.0711	0.5956	0.0880
	-0.2	0.0025	0.0572	0.3041	0.0728	0.7212	0.0947		-0.2	0.0020	0.0568	0.2924	0.0720	0.6454	0.0907
	-0.1	0.0027	0.0571	0.2977	0.0724	0.6897	0.0930		-0.1	0.0026	0.0568	0.3089	0.0728	0.7021	0.0936
	0.1	0.0029	0.0569	0.2571	0.0701	0.5523	0.0857		0.1	0.0042	0.0571	0.3227	0.0737	0.7724	0.0974
	0.2	0.0025	0.0566	0.2212	0.0682	0.4593	0.0811		0.2	0.0050	0.0573	0.3094	0.0732	0.7230	0.0952
	0.3	0.0019	0.0564	0.1783	0.0660	0.3608	0.0766		0.3	0.0055	0.0579	0.2805	0.0722	0.6287	0.0914
	0.4	0.0010	0.0561	0.1328	0.0637	0.2640	0.0723		0.4	0.0057	0.0586	0.2382	0.0710	0.5135	0.0875
-0.3	-0.4	0.0022	0.0572	0.3061	0.0730	0.7336	0.0954	0.2	-0.4	0.0006	0.0565	0.2242	0.0684	0.4617	0.0812
	-0.3	0.0024	0.0572	0.3098	0.0731	0.7441	0.0959		-0.3	0.0008	0.0565	0.2413	0.0693	0.5028	0.0834
	-0.2	0.0026	0.0571	0.3110	0.0731	0.7438	0.0959		-0.2	0.0013	0.0566	0.2606	0.0703	0.5523	0.0861
	-0.1	0.0029	0.0571	0.3079	0.0729	0.7237	0.0948		-0.1	0.0019	0.0567	0.2812	0.0714	0.6115	0.0892
	0.1	0.0032	0.0569	0.2730	0.0709	0.5949	0.0880		0.1	0.0038	0.0573	0.3109	0.0732	0.7281	0.0954
	0.2	0.0030	0.0567	0.2383	0.0691	0.5002	0.0834		0.2	0.0049	0.0582	0.3087	0.0737	0.7315	0.0958
	0.3	0.0024	0.0564	0.1951	0.0669	0.3977	0.0788		0.3	0.0061	0.0597	0.2910	0.0739	0.6745	0.0939
	0.4	0.0016	0.0562	0.1479	0.0647	0.2953	0.0744		0.4	0.0068	0.0621	0.2578	0.0742	0.5766	0.0911
-0.2	-0.4	0.0022	0.0571	0.3049	0.0729	0.7202	0.0946	0.3	-0.4	-0.0002	0.0562	0.1815	0.0662	0.3645	0.0766
	-0.3	0.0024	0.0571	0.3116	0.0732	0.7431	0.0959		-0.3	0.0000	0.0562	0.1986	0.0671	0.4019	0.0788
	-0.2	0.0027	0.0571	0.3163	0.0734	0.7586	0.0967		-0.2	0.0003	0.0563	0.2187	0.0682	0.4481	0.0815
	-0.1	0.0030	0.0571	0.3172	0.0734	0.7560	0.0965		-0.1	0.0009	0.0565	0.2415	0.0694	0.5051	0.0846
	0.1	0.0036	0.0569	0.2901	0.0718	0.6443	0.0906		0.1	0.0031	0.0577	0.2841	0.0723	0.6380	0.0917
	0.2	0.0035	0.0567	0.2576	0.0701	0.5492	0.0860		0.2	0.0046	0.0596	0.2934	0.0741	0.6815	0.0943
	0.3	0.0031	0.0565	0.2150	0.0681	0.4431	0.0814		0.3	0.0063	0.0631	0.2891	0.0763	0.6779	0.0953
	0.4	0.0024	0.0563	0.1664	0.0658	0.3349	0.0771		0.4	0.0077	0.0686	0.2690	0.0794	0.6197	0.0948
-0.1	-0.4	0.0021	0.0570	0.2992	0.0725	0.6884	0.0929	0.4	-0.4	-0.0011	0.0559	0.1360	0.0638	0.2681	0.0723
	-0.3	0.0023	0.0570	0.3091	0.0730	0.7225	0.0947		-0.3	-0.0010	0.0559	0.1516	0.0648	0.3000	0.0745
	-0.2	0.0027	0.0570	0.3179	0.0734	0.7552	0.0965		-0.2	-0.0007	0.0561	0.1706	0.0659	0.3405	0.0771
	-0.1	0.0031	0.0570	0.3236	0.0737	0.7771	0.0976		-0.1	-0.0002	0.0563	0.1934	0.0673	0.3921	0.0803
	0.1	0.0040	0.0569	0.3071	0.0727	0.7003	0.0936		0.1	0.0020	0.0583	0.2438	0.0713	0.5246	0.0880
	0.2	0.0041	0.0568	0.2784	0.0713	0.6074	0.0891		0.2	0.0038	0.0618	0.2629	0.0747	0.5875	0.0920
	0.3	0.0039	0.0567	0.2376	0.0693	0.4988	0.0845		0.3	0.0059	0.0683	0.2722	0.0799	0.6266	0.0957
	0.4	0.0034	0.0565	0.1887	0.0672	0.3850	0.0802		0.4	0.0079	0.0783	0.2676	0.0872	0.6214	0.0988

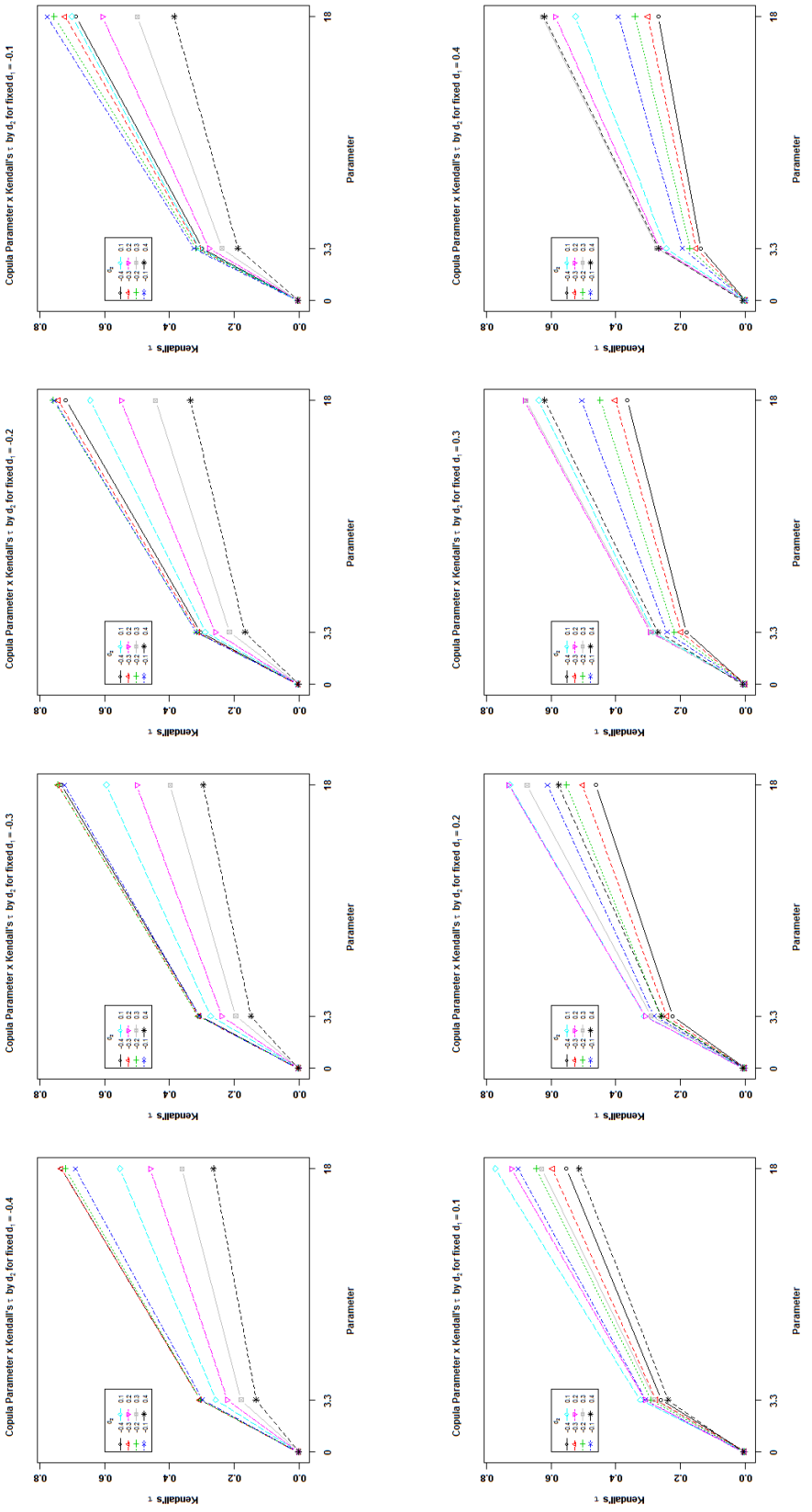


Figure 13: Plots of Kendall's τ by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be (standardized) t_3.

Table 14: Simulation results for the Kendall's τ in the Frank-t_7 case

$\boldsymbol{\varepsilon_t} \sim$ Frank-t_7															
$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$		$\boldsymbol{d}$		$\theta = 0$		$\theta = 3.3$		$\theta = 18$	
d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd	d_1	d_2	$\hat{\tau}$	sd	$\hat{\tau}$	sd	$\hat{\tau}$	sd
-0.4	-0.4	0.0026	0.0573	0.3151	0.0735	0.7596	0.0967	0.1	-0.4	0.0017	0.0567	0.2730	0.0710	0.5872	0.0874
	-0.3	0.0028	0.0572	0.3154	0.0734	0.7573	0.0966		-0.3	0.0019	0.0568	0.2879	0.0717	0.6303	0.0898
	-0.2	0.0029	0.0572	0.3125	0.0732	0.7406	0.0957		-0.2	0.0023	0.0568	0.3023	0.0725	0.6771	0.0922
	-0.1	0.0031	0.0571	0.3052	0.0728	0.7063	0.0938		-0.1	0.0028	0.0568	0.3151	0.0732	0.7255	0.0948
	0.1	0.0032	0.0569	0.2706	0.0708	0.5864	0.0874		0.1	0.0043	0.0570	0.3263	0.0738	0.7846	0.0980
	0.2	0.0030	0.0567	0.2406	0.0691	0.5053	0.0833		0.2	0.0053	0.0572	0.3179	0.0735	0.7501	0.0963
	0.3	0.0026	0.0564	0.2016	0.0671	0.4134	0.0789		0.3	0.0063	0.0576	0.2950	0.0726	0.6694	0.0926
	0.4	0.0018	0.0561	0.1561	0.0648	0.3157	0.0747		0.4	0.0070	0.0582	0.2565	0.0713	0.5593	0.0886
-0.3	-0.4	0.0026	0.0572	0.3158	0.0735	0.7571	0.0966	0.2	-0.4	0.0010	0.0565	0.2437	0.0693	0.5073	0.0833
	-0.3	0.0028	0.0572	0.3188	0.0736	0.7676	0.0971		-0.3	0.0012	0.0566	0.2604	0.0702	0.5499	0.0856
	-0.2	0.0030	0.0571	0.3189	0.0735	0.7641	0.0969		-0.2	0.0016	0.0566	0.2776	0.0711	0.5978	0.0882
	-0.1	0.0032	0.0571	0.3146	0.0733	0.7407	0.0957		-0.1	0.0022	0.0567	0.2945	0.0720	0.6507	0.0910
	0.1	0.0035	0.0569	0.2855	0.0716	0.6294	0.0897		0.1	0.0041	0.0572	0.3196	0.0736	0.7538	0.0964
	0.2	0.0035	0.0567	0.2571	0.0700	0.5476	0.0856		0.2	0.0056	0.0578	0.3208	0.0740	0.7677	0.0972
	0.3	0.0032	0.0565	0.2186	0.0681	0.4528	0.0812		0.3	0.0073	0.0589	0.3084	0.0739	0.7231	0.0952
	0.4	0.0025	0.0562	0.1720	0.0658	0.3500	0.0769		0.4	0.0087	0.0607	0.2790	0.0736	0.6291	0.0916
-0.2	-0.4	0.0025	0.0571	0.3133	0.0733	0.7402	0.0956	0.3	-0.4	0.0002	0.0562	0.2056	0.0673	0.4168	0.0790
	-0.3	0.0027	0.0571	0.3194	0.0736	0.7638	0.0969		-0.3	0.0004	0.0563	0.2229	0.0682	0.4565	0.0812
	-0.2	0.0030	0.0571	0.3228	0.0737	0.7769	0.0976		-0.2	0.0008	0.0564	0.2417	0.0693	0.5024	0.0837
	-0.1	0.0032	0.0571	0.3221	0.0736	0.7702	0.0972		-0.1	0.0014	0.0566	0.2616	0.0704	0.5551	0.0866
	0.1	0.0038	0.0569	0.3001	0.0724	0.6758	0.0922		0.1	0.0037	0.0575	0.2994	0.0727	0.6768	0.0928
	0.2	0.0039	0.0568	0.2743	0.0710	0.5950	0.0881		0.2	0.0057	0.0588	0.3113	0.0741	0.7292	0.0954
	0.3	0.0038	0.0566	0.2371	0.0691	0.4979	0.0837		0.3	0.0080	0.0613	0.3121	0.0756	0.7420	0.0963
	0.4	0.0033	0.0563	0.1901	0.0669	0.3905	0.0795		0.4	0.0104	0.0656	0.2963	0.0774	0.6907	0.0946
-0.1	-0.4	0.0024	0.0570	0.3064	0.0729	0.7059	0.0938	0.4	-0.4	-0.0007	0.0559	0.1608	0.0650	0.3197	0.0747
	-0.3	0.0026	0.0570	0.3156	0.0733	0.7403	0.0956		-0.3	-0.0006	0.0560	0.1772	0.0660	0.3547	0.0769
	-0.2	0.0029	0.0570	0.3227	0.0737	0.7701	0.0972		-0.2	-0.0003	0.0561	0.1959	0.0671	0.3960	0.0794
	-0.1	0.0032	0.0570	0.3263	0.0738	0.7859	0.0981		-0.1	0.0002	0.0563	0.2170	0.0683	0.4450	0.0823
	0.1	0.0041	0.0569	0.3133	0.0731	0.7239	0.0948		0.1	0.0027	0.0579	0.2638	0.0715	0.5690	0.0886
	0.2	0.0044	0.0568	0.2914	0.0719	0.6473	0.0909		0.2	0.0051	0.0603	0.2858	0.0740	0.6393	0.0919
	0.3	0.0046	0.0567	0.2568	0.0702	0.5497	0.0865		0.3	0.0081	0.0652	0.3008	0.0780	0.6983	0.0951
	0.4	0.0043	0.0566	0.2105	0.0682	0.4384	0.0823		0.4	0.0113	0.0735	0.3022	0.0835	0.7131	0.0968

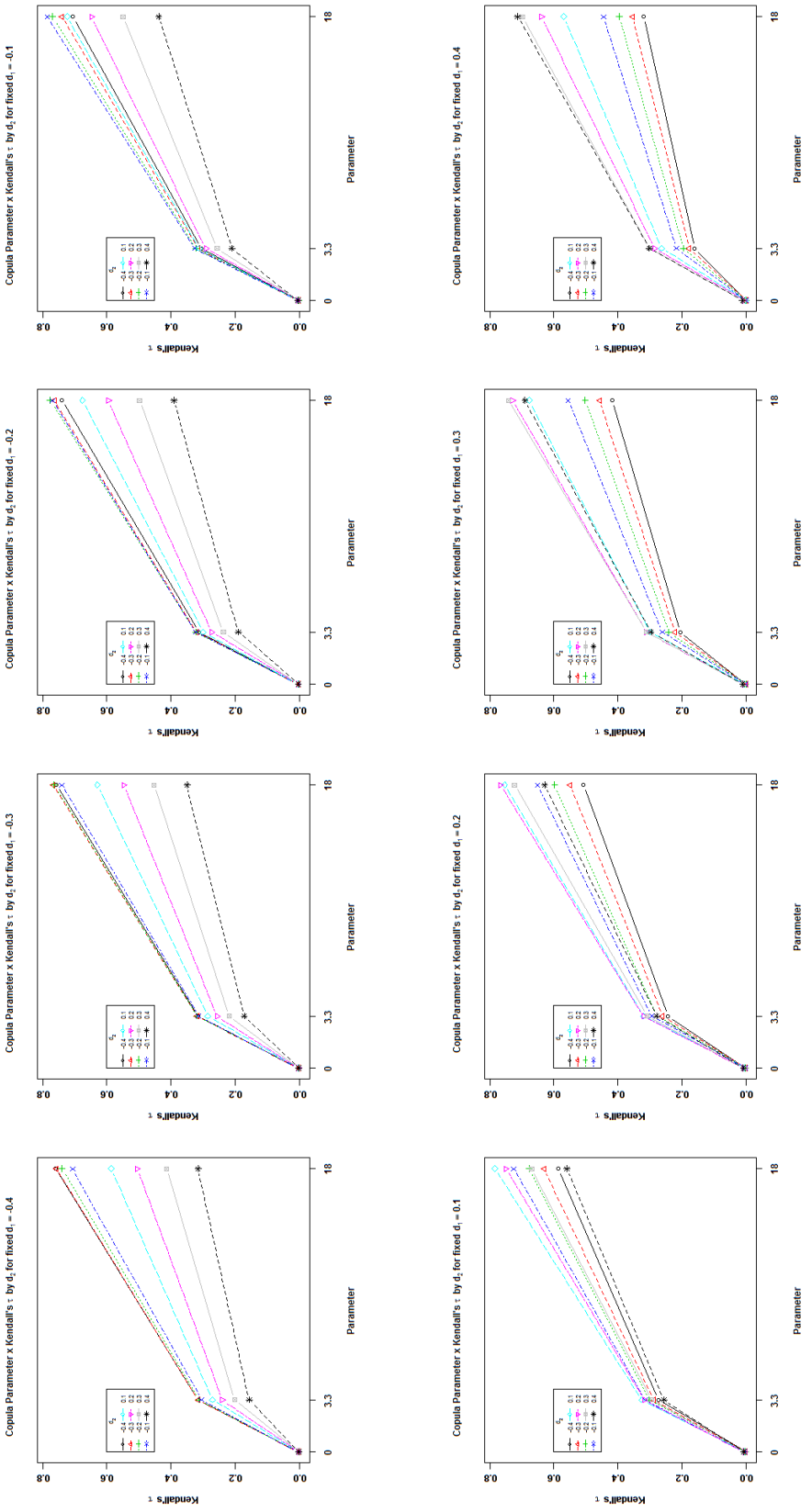


Figure 14: Plots of Kendall's τ by the Frank copula's parameter $\theta \in \{0, 3.3, 18\}$ for fixed d_1. Marginals were taken to be (standardized) t_7.

Anexo E

Artigo Pumi e Lopes (2011**c**)

Parameterization of Copulas and Covariance Decay of Stochastic Processes with Applications

Guilherme Pumi and Sílvia R.C. Lopes

Mathematics Institute
Federal University of Rio Grande do Sul

This version: 05/12/2011

Abstract

In this work we study the problem of constructing stochastic processes with a predetermined covariance decay by parameterizing its marginals and a given family of copulas. We present several examples to illustrate the theory, including the important Gaussian and Euclidean families of copulas. We associate the theory to common applied time series models and present a general methodology to estimate a given parameter of interest identifiable through the process' covariance decay. To exemplify the proposed methodology, we present simple Monte Carlo applications to parameter estimation in time series. The methodology is also applied to the S&P500 US stock market index.

Keywords: Copulas, decay of covariance, dependence structure, parameter estimation.

1 Introduction

Let $\{X_t\}_{t=0}^{\infty}$ be a weakly stationary process. Consider the widely studied problem of modeling the dependence structure among the variables X_t. From a strictly probabilistic/measure theoretical point of view, a stochastic process can be viewed as a measure on the space of sample paths. Within this probabilistic point of view, associated to any process, there is a sequence of distributions $\{F_n\}_{n=0}^{\infty}$, and to every pair of random variables (X_r, X_s) from the process, there is a copula $C_{r,s}$ (Nelsen, 2006) associated to it. In a weakly stationary process, the covariance between X_t and X_{t+h} depends only on the lag h, and, according to the copula version of Hoeffding's lemma (see Lemma 2.1), this covariance can be calculated by the knowledge of the copula and the marginals of X_t and X_{t+h} alone. Hence, studying the distributional properties among the variables is of vital importance in understanding the process' dependence structure.

Many classes of models are known to present certain covariance decay to zero as the lag h increases. In the class of ARFIMA(p, d, q) processes (Lopes, 2008; Taqqu, 2003), for instance, the covariance may decay exponentially fast to zero or very slowly, depending on the parameter d. For heteroskedastic models such as ARCH (Engle, 1982) and GARCH (Bollerslev, 1986), the process itself is uncorrelated, but the absolute value and/or the square of the process are not and often present slow decay of covariance. Estimation of the rate of such decay is generally of importance as it contains essential information about the long run structure of the process.

In this paper we are interested in studying the following problem. For a process $\{X_n\}_{n=0}^{\infty}$ with absolutely continuous marginal distributions $\{F_n\}_{n=0}^{\infty}$ and for a given parametric family of copulas $\{C_\theta\}_{\theta\in\Theta}$ satisfying some minor regularity conditions, we investigate how to obtain

a given decay of covariance by choosing a parameterization $\{\theta_n\}_{n=1}^{\infty}$ and by assuming that the copulas related to the pair (X_n, X_{n+h}) depends only on the lag h and are given by C_{θ_h}. We are especially interested in the case of slow decay of covariance, typical of long-range dependent processes, but the theory presented here is much more general than that and covers any arbitrary decay of covariance as well as for both, stationary and non-stationary processes.

Applying copulas as a tool to construct time series or other complex models have become routine. See, for instance, Lee and Long (2009). An interesting theory of Markov process based on a "product" operation for copulas is developed in Darsow et al. (1992) and investigated in Lagerås (2010). Some methods for constructing short memory time series based on conditional copulas are discussed in chapter 8 of Joe (1997). The author also discusses methods for constructing Markovian processes and short memory time series based on parameterization of distributions belonging to a convolution-closed infinitely divisible class. The method presented here is completely different from those and does not assume any special properties on the marginal distributions other than absolute continuity.

To illustrate the theory, we present examples including several families of copulas widely used in practice. Special attention is given to Gaussian processes. We also present two applications of the methodology. The first application is related to estimating a given parameter of interest identifiable through the covariance decay in weakly stationary processes. The second one is related to simulation of time series with a prescribed covariance decay by parameterizing a given family of bidimensional copulas. The methodology is also applied to the S&P500 stock market index data set.

The paper is organized as follows. In the next section we recall a few concepts and results useful in the paper as well as we introduce the idea of the paper by a simple example. In Sections 3 and 4, we consider the special cases of the Archimedean family of copulas and the Extreme Value copulas, respectively. In Section 5 we consider a general theory including arbitrary decay of covariance and present a discussion of the important case of Gaussian processes. In Section 6 we develop two applications based on the results from the previous sections. In the first one we develop a general methodology based on the theoretical results of the previous section to estimate a given parameter of interest in time series model, identifiable through the covariance decay. A Monte Carlo study is presented to exemplify the methodology and assess its finite sample performance. In the second one, we apply the theory to the problem of random variate generation of time series with a predetermined covariance decay. In Section 7 the methodology is applied to the S&P500 stock market index data set. Conclusions and final remarks are presented in Section 8.

2 Preliminaries

In this section we recall a few concepts and results we shall need in what follows. An n-dimensional copula is a distribution function defined in the n-dimensional hypercube I^n, where $I := [0, 1]$, and whose marginals are uniformly distributed. The interest in copulas has grown enormously in the last 15 years or so, especially because of its flexibility in applications. Copulas have been successfully applied and widely spread in several areas. In finances, copulas have been applied in major topics such as asset pricing, risk management and credit risk analysis among many others (see the book by Cherubini et al., 2004 for details). In hydrology the modeling

of rainfalls and storms often employ copulas to describe joint features between variables in the models (see, for instance, the recent work of Palynchuk and Guo, 2011 and references therein). In econometrics, copulas have been widely employed in constructing multidimensional extensions of complex models (see Lee and Long, 2009 and references therein). In statistics, copulas have been applied in all sort of problems, such as development of dependence measures, modeling, testing, just to cite a few (see Mari and Kotz, 2001, Nelsen, 2006 and references therein). Curiously, the main result in the theory, the celebrated Sklar's theorem recalled below, dates back to the late fifties.

Theorem 2.1 (Sklar's Theorem)**.** *Let $X_1, \cdots, X_n$ be random variables with joint distribution function H and marginals $F_1, \cdots, F_n$, respectively. Then, there exists a copula C such that,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all } \ (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

If the F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_n)$*. The converse also holds. Furthermore,*

$$C(u_1, \cdots, u_n) = H\big(F_1^{(-1)}(u_1), \cdots, F_n^{(-1)}(u_n)\big), \quad \textit{for all } \ (u_1, \cdots, u_n) \in I^n,$$

where for a function F, $F^{(-1)}$ denotes its pseudo-inverse given by $F^{(-1)}(x) := \inf\big\{u \in \mathrm{Ran}(F) : F(u) \geq x\big\}$ and, for a function f, $\mathrm{Ran}(f)$ *denotes the range of f.*

The usefulness of Sklar's Theorem (and of copulas for extension) as a tool for statistical analysis and modeling lies on allowing one to deal separately with the joint dependence structure, characterized by the copula, and the marginals of a given random vector. This flexibility has been extensively explored in the literature. Another important result which shall be frequently used here is the copula version of the Hoeffding's lemma.

Lemma 2.1 (Hoeffding's Lemma)**.** *Let X and Y be two continuous random variables with marginal distributions F and G, respectively, and copula C. Then, the covariance between X and Y is given by*

$$\mathrm{Cov}(X, Y) = \iint_{I^2} \frac{C(u, v) - uv}{F'\big(F^{(-1)}(u)\big)G'\big(G^{(-1)}(v)\big)}\, dudv. \tag{2.1}$$

More details on the theory of copulas can be found in the monographs by Nelsen (2006) and Joe (1997).

In this work, $\mathbb{N}$ denotes the set of natural numbers $\mathbb{N} := \{0, 1, 2, \cdots\}$, while $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. For a given set $A \subseteq \mathbb{R}$, $\overline{A}$ denotes the closure of A and A' denotes the set of all accumulation points. For a vector $x \in \mathbb{R}^k$, $x^\intercal$ denotes the transpose of x. The measure space behind the notion of measurable sets and functions is always assumed (without further mention) to be $\big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathfrak{m}\big)$ (or some appropriate restriction of it), where $\mathscr{B}(\mathbb{R}^n)$ denotes the Borel σ-field in $\mathbb{R}^n$ and $\mathfrak{m}$ is the Lebesgue measure in $\mathbb{R}^n$. Recall that a function $L : S \to \mathbb{R}$, for $S \subseteq \mathbb{R}$, is called *slowly varying at $a \in S'$* if L is measurable, limited on a bounded interval and satisfies $\lim_{x \to a} L(cx)/L(x) = 1$, for all $c > 0$. The set of slowly varying functions on the infinity will be denoted by

$$\mathscr{L} := \{L : L \text{ is measurable, limited on a bounded interval and } \lim_{x \to \infty} L(cx)/L(x) = 1, \forall c > 0\}.$$

We notice that if $L_1, L_2 \in \mathscr{L}$, $L_1 + L_2$, $L_1 L_2 \in \mathscr{L}$ and, for a constant $k \in \mathbb{R} \setminus \{0\}$, $kL_1 \in \mathscr{L}$. Moreover, for any $\beta \in (0, 1)$ and $L \in \mathscr{L}$, $L(n)n^{-\beta} \to 0$, fact usually explored in defining the

concept of long-range dependence. More details on slowly varying function can be found in Bingham et al. (1987). For two functions f and g, we denote $f(n) \sim g(n)$ if $f(n)/g(n) \to 1$, as n goes to infinity. For an arbitrary sequence of real numbers $\{a_n\}_{n\in\mathbb{N}^*}$, consider the function $\psi_{a_n} : (0,\infty) \to \mathbb{R}$ given by $\psi_{a_n}(x) := a_{\lceil x \rceil}$, where $\lceil\cdot\rceil$ denotes the ceiling function. For $L \in \mathscr{L}$, as a convention, by writing $a_n L \in \mathscr{L}$ we mean $\psi_{a_n} L \in \mathscr{L}$. We shall also write $a_n \in \mathscr{L}$ to mean $\psi_{a_n} \in \mathscr{L}$. If $a_n \to a \neq 0$, then always $a_n \in \mathscr{L}$. If $a_n \to 0$, then a_n may or may not belong to $\mathscr{L}$, depending on the particular convergence rate of the sequence to 0. For instance, $a_n = 1/\ln(n) \in \mathscr{L}$, but $a_n = 1/n \notin \mathscr{L}$.

As for long-range dependence, even though the literature on the matter is vast, there still no globally accepted definition for it. The most common definitions and their differences and similarities are discussed in section 4 of Taqqu (2003). In this work, we shall adopt the following general definition.

Definition 2.1. We say that a weakly stationary process with finite variance $\{X_t\}_{t\in\mathbb{N}}$ presents *long-range dependence* if

$$\mathrm{Cov}(X_t, X_{t+h}) \sim L(h)h^{-\beta}, \quad \text{as } h \to \infty,$$

for some $\beta \in (0,1)$ and $L \in \mathscr{L}$.

We start with an example to motivate the ideas we shall develop in the rest of the paper.

Example 2.1. Consider the *Farlie-Gumbel-Morgenstern* (FGM, for short) family of copulas which consists of parametric copulas of the form

$$C_\theta(u,v) = uv\big(1 + \theta(1-u)(1-v)\big),$$

for $|\theta| \leq 1$. As a particular case, we have the *independence copula* $\Pi(u,v) := uv$, for $\theta = 0$. Recall that a distribution function F belongs to the so-called *Type I Extreme Value* family with parameters $(a,b) \in \mathbb{R}\times(0,\infty)$, denoted by $\mathrm{EVI}(a,b)$, if

$$F(x) = e^{-e^{-(x-a)/b}}, \quad \text{for all } x \in \mathbb{R}.$$

Notice that $F'(F^{-1}(u)) = bu\ln(u)$, for all $u \in I$. Let $F_n \sim \mathrm{EVI}(a_n, b)$, for $\{a_n\}_{n\in\mathbb{N}}$ an arbitrary sequence of real numbers and $b > 0$. Consider the FGM family of copulas with parameterization $\theta_n := \kappa_0^{-1} n^{-\alpha}$, $n \in \mathbb{N}^*$, for $\alpha > 1$, where $\kappa_0 \geq \zeta(\alpha)$ and $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ is the so-called Riemman zeta function. For a fixed $n > 2$, consider the collection of $\binom{n}{2}$ copulas $C_{ij} := C_{\theta_{|i-j|}}$ for $i, j \in \{1,\cdots,n\}$, $i \neq j$. Now consider the n-dimensional copula (cf. example 3.2 in Dolati and Úbeda-Flores, 2005)

$$C_n(u_1,\cdots,u_n) := \sum_{1\leq i<j\leq n} C_{ij}(u_i,u_j) \prod_{\substack{k=1\\k\neq i,j}}^{n} u_k - \frac{(n-2)(n+1)}{2} \prod_{l=1}^{n} u_l \tag{2.2}$$

whose marginals are C_{ij}. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that X_n has distribution F_n, for each $n \in \mathbb{N}$. For any $n > 2$, let C_n given in (2.2) be the copula associated to $(X_0,\cdots,X_{n-1})$, which implies that the copula related to (X_r, X_s) is C_{rs}, $r, s \in \{0,\cdots,n-1\}$ and $r \neq s$. Furthermore, Hoeffding's lemma implies

$$\mathrm{Cov}(X_t, X_{t+h}) = \theta_h \left(\int_I \frac{1-u}{b\ln(u)}\, du \right)^2 = \frac{\ln(2)^2}{b^2}\,\theta_h = \frac{\ln(2)^2}{b^2\kappa_0} h^{-\alpha},$$

where the second equality follows from formula 4.267.8 in Gradshteyn and Ryzhik (2000). Since the construction (2.2) is valid for any $n > 2$, Sklar's theorem guarantees the existence of all finite

dimensional distribution functions with the marginals, bivariate copulas and n-dimensional copulas as specified in the construction above. Therefore, by the Kolmogorov's existence theorem, we have just constructed a weakly stationary process $\{X_n\}_{n\in\mathbb{N}}$ by reparameterizing a certain family of parametric copulas and its marginals.

The objective of this paper is to derive conditions for which a certain decay of covariance can be achieved by simply parameterizing a certain family of copula and its marginal. We shall be interested in a more general framework than the one presented in Example 2.1 in the sense that the theory covers both, weakly and strongly stationary processes as well as non-stationary ones.

Definition 2.2. Let $\{C_\theta\}_{\theta\in\Theta}$ be a parametric family of bidimensional copula, for $\Theta \subseteq \mathbb{R}$ a set with non-empty interior. If there exists a sequence $\{\theta_n\}_{n\in\mathbb{N}^*} \subseteq \overline{\Theta}$ and a sequence of absolutely continuous distributions $\{F_n\}_{n\in\mathbb{N}}$ such that

$$\iint_{I^2} \frac{C_{\theta_n}(u,v) - uv}{F_0'\big(F_0^{-1}(u)\big)F_n'\big(F_n^{-1}(v)\big)}\, dudv \sim L(n)n^{-\beta}, \tag{2.3}$$

for $L \in \mathscr{L}$ and $\beta \in (0,1)$, we say that (C_θ, F_n) is *compatible with long-range dependence structure.*

Notice that Definition 2.2 can be translated as follows. For a sequence $\{F_n\}_{n\in\mathbb{N}}$ of absolutely continuous distribution functions and for a sequence $\{C_n\}_{n\in\mathbb{N}^*}$ of copulas, consider $\{X_n\}_{n\in\mathbb{N}}$ a sequence of random variables such that X_n is distributed as F_n, for each $n \in \mathbb{N}$, and such that the copula associated to (X_0, X_n) is C_n, for $n \in \mathbb{N}^*$. This construction is always possible by Sklar's theorem (notice that we are not making any joint distributional assumption, other than the ones implied to (X_0, X_n), $n \in \mathbb{N}^*$). In this setting, Definition 2.2 is equivalent to ask that $\mathrm{Cov}(X_0, X_n) \sim L(n)n^{-\beta}$, for $L \in \mathscr{L}$ and $\beta \in (0,1)$. Also notice that a strongly stationary long-range dependent process always satisfies (2.3), but the construction alone is clearly not sufficient to specify a stochastic process, in which case we have to proceed analogously to Example 2.1.

To simplify the notation, for a sequence of absolutely continuous distribution functions $\{F_n\}_{n\in\mathbb{N}}$ and a parametric copula C_θ, we shall write

$$l_n(x) := F_n'\big(F_n^{(-1)}(x)\big) \qquad \text{and} \qquad \partial^k C_\theta(u,v) := \frac{\partial^k C_\theta(u,v)}{\partial\theta^k}\,.$$

3 Archimedean Family

Recall that the *Archimedean family* of copulas consists of copulas of the form

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

for some continuous decreasing convex possibly parametric function $\varphi : I \to [0,\infty]$, $I := [0,1]$, such that $\varphi(1) = 0$, called the *Archimedean generator* of the copula. We are interested in determining conditions in which a sequence of Archimedean copulas together with a sequence of absolutely continuous distribution functions is compatible with long-range dependence structure.

Proposition 3.1. *Let $\{\varphi_\theta\}_{\theta\in\Theta}$, for $\Theta \subseteq \mathbb{R}$ with non-empty interior, be a family of Archimedean generators and $\{C_\theta\}_{\theta\in\Theta}$ be the correspondent Archimedean family. Suppose that there exists $a \in \Theta'$ such that $\lim_{\theta\to a}\varphi_\theta(t) = -\ln(t)$, where the limit is to be understood as the adequate lateral limit if $a \notin \mathrm{int}(\Theta)$. Also assume that there exists a set $D \subseteq \Theta$ with non-empty interior*

such that $a \in D'$ and φ_θ, seen as a function of θ, is of class C^2 in D. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of absolutely continuous distribution functions and define the sequences

$$K_1(n) := \iint_{I^2} \lim_{\substack{\theta\to a\\ \theta\in D}} \frac{\partial C_\theta(u,v)}{l_0(u)l_n(v)} dudv \quad \text{and} \quad K_2(n) := \iint_{I^2} \lim_{\substack{\theta\to a\\ \theta\in D}} \frac{\partial^2 C_\theta(u,v)}{l_0(u)l_n(v)} dudv.$$

Let $\{\theta_n\}_{n\in\mathbb{N}^*}$ be a sequence in D converging to a. Suppose that, for some $\beta \in (0,1)$, $K_1(n)(\theta_n - a) \sim L_1(n)n^{-\beta}$ and $K_2(n)(\theta_n - a)^2 = o(L_2(n)n^{-\beta})$, as n goes to infinity, for $L_1, L_2 \in \mathscr{L}$. Then, (C_θ, F_n) is compatible with long-range dependence structure.

Proof. We present the proof for the case where $a \notin \operatorname{int}(\Theta)$ and assuming that $a > x$ for all $x \in D$. The other cases are proved analogously. Let $\{\alpha_m\}_{m\in\mathbb{N}^*}$ be a sequence of parameters in D converging from the left to a. Applying a Taylor's expansion with Lagrange's remainder in C_θ around $\theta = a$, we obtain

$$\begin{aligned} C_\theta(u,v) &= \lim_{m\to\infty} C_{\alpha_m}(u,v) + \lim_{m\to\infty} \partial C_{\alpha_m}(u,v)(\theta - a) + \frac{1}{2}\lim_{m\to\infty} \partial^2 C_{\alpha_m}(u,v)(\theta_0 - a)^2 \\ &= uv + \lim_{m\to\infty} \partial C_{\alpha_m}(u,v)(\theta - a) + \frac{1}{2}\lim_{m\to\infty} \partial^2 C_{\alpha_m}(u,v)(\theta_0 - a)^2, \end{aligned} \tag{3.1}$$

where $\theta_0 \in [\theta_n, a)$. Substituting θ by θ_n in (3.1), we obtain

$$C_{\theta_n}(u,v) = uv + \lim_{m\to\infty} \partial C_{\alpha_m}(u,v)(\theta_n - a) + \frac{1}{2}\lim_{m\to\infty} \partial^2 C_{\alpha_m}(u,v)\big(\theta_0(n) - a\big)^2, \tag{3.2}$$

where $\theta_0(n) \in [\theta_n, a)$ depends on n and satisfies $\lim_{n\to\infty} \theta_0(n) = a$. Under the notation of the enunciate, we obtain

$$\begin{aligned} \iint_{I^2} \frac{C_{\theta_n}(u,v) - uv}{l_0(u)l_n(v)}\, dudv &= \left(\iint_{I^2} \lim_{m\to\infty} \frac{\partial C_{\alpha_m}(u,v)}{l_0(u)l_n(v)}\, dudv\right)(\theta_n - a) + \\ &\quad + \frac{1}{2}\left(\iint_{I^2} \lim_{m\to\infty} \frac{\partial^2 C_{\alpha_m}(u,v)}{l_0(u)l_n(v)}\, dudv\right)\big(\theta_0(n) - a\big)^2 \\ &= K_1(n)(\theta_n - a) + \frac{1}{2}K_2(n)\big(\theta_0(n) - a\big)^2. \end{aligned} \tag{3.3}$$

Since $\theta_0(n) \in [\theta_n, a)$,

$$\big|K_2(n)\big|\big(\theta_0(n) - a\big)^2 \le \big|K_2(n)(\theta_n - a)^2\big| = o(L_2(n)n^{-\beta}), \tag{3.4}$$

by the hypothesis on K_2. The result now follows from (3.3) in view of (3.4) and the hypothesis on K_1. ∎

Example 3.1. The Ali-Mikhail-Haq (AMH, for short) family of copulas is Archimedean with generator $\varphi_\theta(t) = (1-\theta)^{-1}\ln\big([1+\theta(t-1)]t^{-1}\big)$, for $|\theta| \le 1$. The copulas of the AMH family have the form

$$C_\theta(u,v) = \frac{uv}{1-\theta(1-u)(1-v)}\,.$$

As a particular case, we have $C_0 = \Pi$. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions such that $F_n \sim \operatorname{Exp}(\lambda_n)$, for $\{\lambda_n\}_{n\in\mathbb{N}}$ a sequence of positive real numbers. Simple calculations show that $l_n(x) = (1-x)/\lambda_n$ and

$$K_1(n) = \lambda_0\lambda_n\left(\int_I u\,du\right)^2 = \frac{\lambda_0\lambda_n}{4} \quad \text{and} \quad K_2(n) = 2\lambda_0\lambda_n\left(\int_I u(1-u)du\right)^2 = \frac{\lambda_0\lambda_n}{18}. \tag{3.5}$$

From (3.5), there are many ways to take λ_n and θ_n in order to obtain the desired decay. For instance, for $\beta \in (0,1)$ and $L \in \mathscr{L}$, by taking $\lambda_n \in \mathscr{L}$ and $\theta_n \sim n^{-\beta}$, or $\lambda_n \sim n^{-\beta}$ and $\theta_n \to 0$ but such that $\theta_n \in \mathscr{L}$, then Proposition 3.1 applies and we conclude that (C_θ, F_n) is compatible with long-range dependence structure. If $\lambda_n = \lambda_0 > 0$, for all n, as it is the case for a strongly stationary process, then by taking $\theta_n \sim L(n)n^{-\beta}$, the same conclusion is obtained through Proposition 3.1.

Example 3.2. The *Gumbel-Barnett* family of copulas is Archimedean with generator $\varphi_\theta(x) = \ln(1-\theta\ln(x))$, for $\theta \in (0,1]$ and copulas given by

$$C_\theta(u,v) = uv\exp(-\theta\ln(u)\ln(v)).$$

As a limiting case we have $\lim_{\theta\to 0^+} C_\theta = \Pi$. Let F_n be distributed as EVI(a_n, b_n), $n \in \mathbb{N}$, for $\{a_n\}_{n\in\mathbb{N}}$ an arbitrary sequence of real numbers and $\{b_n\}_{n\in\mathbb{N}}$ a sequence of positive real numbers. A simple calculation shows that

$$K_1(n) = \frac{1}{b_0 b_n}\left(\int_I du\right)^2 = \frac{1}{b_0 b_n} \quad \text{and} \quad K_2(n) = \frac{1}{b_0 b_n}\left(\int_I \ln(u)du\right)^2 = \frac{1}{b_0 b_n}.$$

Therefore, the situation is similar to Example 3.1. For instance, for $\beta \in (0,1)$ and $L \in \mathscr{L}$, taking $1/b_n \in \mathscr{L}$ and $\theta_n \sim n^{-\beta}$, or $b_n \sim n^\beta$ and $\theta_n \to 0$ but such that $\theta_n \in \mathscr{L}$, then Proposition 3.1 applies and we conclude that (C_θ, F_n) is compatible with long-range dependence structure. If $1/b_n = k_0 > 0$ for all n, as it is the case for a strongly stationary process, then taking $\theta_n \sim L(n)n^{-\beta}$, the same conclusion is obtained through Proposition 3.1.

Example 3.3. Consider the *Frank* family of copulas with Archimedean generator $\varphi_\theta(x) = -\ln\left((e^{-\theta t}-1)(e^\theta - 1)^{-1}\right)$, for $\theta \in \mathbb{R}\setminus\{0\}$, with corresponding copula given by

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right).$$

This is one of the most applied Archimedean copula in statistics. Details of its nice properties can be found in Nelsen (2006) and references therein. As a particular case we have $C_0 = \Pi$. Let F_n be distributed as EVI(a_n, b_n), $n \in \mathbb{N}$, for $\{a_n\}_{n\in\mathbb{N}}$ an arbitrary sequence of real numbers and $\{b_n\}_{n\in\mathbb{N}}$ a sequence of positive real numbers. Upon applying formula 4.267.8 in Gradshteyn and Ryzhik (2000), it is routine to show that

$$K_1(n) = \frac{1}{2b_0 b_n}\left(\int_I \frac{1-u}{\ln(u)}du\right)^2 = \frac{\ln(2)^2}{2b_0 b_n} \quad \text{and} \quad K_2(n) = \iint_{I^2} \frac{p(u,v)}{b_0 u\ln(u) b_n v\ln(v)}dudv = \frac{k_0}{b_0 b_n}, \quad (3.6)$$

where

$$p(u,v) := -u^2v^3 - u^3v^2 + \frac{2u^3v^3}{3} - \frac{uv^2}{2} - \frac{u^2v}{2} + \frac{3u^2v^2}{2} + \frac{uv}{6} + \frac{u^3v}{3} + \frac{uv^3}{3}$$

and $k_0 = 2\ln(3)^2/3 - 2\ln(2)\ln(3) + 3\ln(2)^2/2$. In view of (3.6), for $\beta \in (0,1)$ and $L \in \mathscr{L}$, taking $1/b_n \in \mathscr{L}$ and $\theta_n \sim n^{-\beta}$, or $b_n \sim n^\beta$ and $\theta_n \to 0$ but such that $\theta_n \in \mathscr{L}$, then by Proposition 3.1 we conclude that (C_θ, F_n) is compatible with long-range dependence structure. If $1/b_n = k_0 > 0$ for all n, as it is the case for a strongly stationary process, then taking $\theta_n \sim L(n)n^{-\beta}$, the same conclusion holds.

4 Extreme Value Copulas

Recall that C is an *Extreme Value Copula* (EVC, for short) if there exists a copula C_0 such that

$$C(u,v) = \lim_{n\to\infty} C_0^n(u^{\frac{1}{n}}, v^{\frac{1}{n}}),$$

for all $(u,v) \in I^2$. A method to "parameterize" the EVC family is devised in Pickands (1981) (see also Nelsen, 2006, p.97). The construction show that a copula C belongs to the EVC family if it can be written as

$$C_A(u,v) = \exp\left(\ln(uv)A\left(\frac{\ln(u)}{\ln(uv)}\right)\right),$$

for some possibly parametric convex function $A : I \to [0.5, 1]$, called *dependence function*, satisfying $A(0) = A(1) = 1$ and $A(t) \in [\max\{t, 1-t\}, 1]$, for all $t \in I$. As a particular case, when $A \equiv 1$, $C_A = \Pi$. For a parametric dependence function A_θ, we shall denote the corresponding EVC family by $C_\theta := C_{A_\theta}$. Under some conditions, the EVC family is also compatible with long-range dependence structure.

To simplify the notation, let

$$\partial^k A_\theta(x) := \frac{\partial^k}{\partial t^k} A_t(x)\bigg|_{t=\theta}, \quad \mathcal{A}_\theta(u,v) := A_\theta\left(\frac{\ln(u)}{\ln(uv)}\right), \quad \text{and} \quad \partial\mathcal{A}_\theta(u,v) := \frac{\partial \mathcal{A}_\theta(u,v)}{\partial\theta},$$

for $k = 1, 2$. Proposition 4.1 below presents conditions similar to Proposition 3.1 to obtain the compatibility of a EVC family with long-range dependence structure.

Proposition 4.1. *Let $\{A_\theta\}_{\theta\in\Theta}$, for $\Theta \subseteq \mathbb{R}$ with non-empty interior, be a family of dependence functions and let $\{C_\theta\}_{\theta\in\Theta}$ be the associated* EVC *family. Suppose that there exists $a \in \Theta'$ such that $\lim_{\theta\to a} A_\theta \equiv 1$ where the limit is to be understood as the adequate lateral limit if $a \notin \mathrm{Int}(\Theta)$. Also assume that there exists a set $D \subseteq \Theta$ with non-empty interior such that $a \in D'$ and φ_θ, seen as a function of θ, is of class C^2 in D. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of absolutely continuous distribution functions and define the sequences*

$$K_1(n) := \iint_{I^2} \lim_{\substack{\theta\to a\\ \theta\in D}} \frac{\partial C_\theta(u,v)}{l_0(u)l_n(v)} dudv \quad \text{and} \quad K_2(n) := \iint_{I^2} \lim_{\substack{\theta\to a\\ \theta\in D}} \frac{\partial^2 C_\theta(u,v)}{l_0(u)l_n(v)} dudv.$$

Let $\{\theta_n\}_{n\in\mathbb{N}^*}$ be a sequence in D converging to a. Suppose that, for some $\beta \in (0,1)$, $K_1(n)(\theta_n - a) \sim L_1(n)n^{-\beta}$ and $K_2(n)(\theta_n - a)^2 = o(L_2(n)n^{-\beta})$, as n goes to infinity, for $L_1, L_2 \in \mathscr{L}$. Then, (C_θ, F_n) is compatible with long-range dependence structure.

Proof. The proof goes along the same lines as the proof of Proposition 3.1. ∎

It is easy to see that equivalent expressions to K_1 and K_2 in Proposition 4.1 are as follows:

$$K_1(n) = \iint_{I^2} \left[\lim_{\substack{\theta\to a\\ \theta\in D}} \partial\mathcal{A}_\theta(u,v)\right] \frac{uv\ln(uv)}{l_0(u)l_n(v)}\, dudv \tag{4.1}$$

and

$$K_2(n) = \iint_{I^2} \left(\ln(uv)\left[\lim_{\substack{\theta\to a\\ \theta\in D}} \partial\mathcal{A}_\theta(u,v)\right]^2 + \lim_{\substack{\theta\to a\\ \theta\in D}} \partial^2\mathcal{A}_\theta(u,v)\right) \frac{uv\ln(uv)}{l_0(u)l_n(v)}\, dudv. \tag{4.2}$$

The above expressions seem cumbersome at first glance, they are often simpler to calculate because the limits inside the integrals usually result in simple expressions.

Example 4.1. Consider the following dependence function and the associated EVC family

$$A_\theta(t) = 1 - \theta t(1-t) \quad \text{and} \quad C_\theta(u,v) = uv \exp\left(\frac{\theta \ln(v)\big(\ln(v) - \ln(uv)\big)}{\ln(uv)}\right),$$

for $\theta \in I$. This is known as the generator of the Tawn mixed model (cf. Mari and Kotz, 2001, p.96). Notice that $A_0(t) \equiv 1$, $\partial A_0(t) = -t(1-t)$ and $\partial^2 A_0(t) = 0$. Let F_n be distributed as EVI(a_n, b_n), $n \in \mathbb{N}$, for $\{a_n\}_{n\in\mathbb{N}}$ an arbitrary sequence of real numbers and $\{b_n\}_{n\in\mathbb{N}}$ a sequence of positive real numbers. By using (4.1) and (4.2),

$$K_1(n) = \iint_{I^2} \frac{\ln(v)\big(\ln(v) - \ln(uv)\big)}{b_0\, b_n \ln(uv)^2}\, dudv = -\frac{1}{b_0\, b_n} \iint_{I^2} \frac{\ln(u)\ln(v)}{\big(\ln(u) + \ln(v)\big)^2}\, dudv = -\frac{1}{6\, b_0\, b_n},$$

and

$$K_2(n) = \iint_{I^2} \frac{\big(\ln(v)\big(\ln(v) - \ln(uv)\big)\big)^2}{b_0\, b_n \ln(uv)^3}\, dudv = -\frac{1}{b_0\, b_n} \iint_{I^2} \frac{\ln(u)\ln(v)}{\big(\ln(u) + \ln(v)\big)^3}\, dudv = -\frac{1}{15\, b_0\, b_n},$$

where the integrals in both expressions are calculated by changing variables to $u = e^{-t}$, by applying integral by parts and by using formula 3.353.1 in Gradshteyn and Ryzhik (2000). Therefore, for $\beta \in (0,1)$ and $L \in \mathscr{L}$, taking $1/b_n \in \mathscr{L}$ and $\theta_n \sim n^{-\beta}$, or $b_n \sim n^{\beta}$ and $\theta_n \to 0$ but such that $\theta_n \in \mathscr{L}$, then Proposition 4.1 applies and we conclude that (C_θ, F_n) is compatible with long-range dependence structure. If $1/b_n = k_0 > 0$ for all n, by taking $\theta_n \sim L(n)n^{-\beta}$ the same conclusion holds.

5 General Theory

In the previous sections we studied two particular cases of a more general theory to be developed in this section. The focus of the previous section was on the compatibility with long-range dependence structure. However, as it shall become clear in Theorem 5.1 below, with a minimum effort we can extend the theory to cover any arbitrary decay of covariance in a wide class of parametric families of absolutely continuous copulas satisfying minimal regularity conditions.

Theorem 5.1. *Let $\{C_\theta\}_{\theta\in\Theta}$ be a family of parametric copulas, for $\Theta \subseteq \mathbb{R}$ with non-empty interior. Suppose that there exists a point $a \in \Theta'$ such that $\lim_{\theta\to a} C_\theta = \Pi$, where the limit is to be understood as the adequate lateral limit if $a \notin \mathrm{int}(\Theta)$. Also assume that there exist a set $D \subseteq \Theta$ with non-empty interior such that $a \in D'$ and C_θ, seen as a function of the parameter θ, is of class C^2 in D. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of absolutely continuous distribution functions and define the sequences*

$$K_1(n) := \iint_{I^2} \lim_{\substack{\theta\to a \\ \theta\in D}} \frac{\partial C_\theta(u,v)}{l_0(u) l_n(v)}\, dudv \quad \text{and} \quad K_2(n) := \iint_{I^2} \lim_{\substack{\theta\to a \\ \theta\in D}} \frac{\partial^2 C_\theta(u,v)}{l_0(u) l_n(v)}\, dudv.$$

Let $\{\theta_n\}_{n\in\mathbb{N}^*}$ be a sequence in D converging to a and let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $X_n \sim F_n$, $n \in \mathbb{N}$ and the copula associated with (X_0, X_n) be C_{θ_n}. Given a measurable function $R : \mathbb{R} \to \mathbb{R}$ satisfying $\lim_{n\to\infty} R(n) = 0$, a sufficient condition for $\mathrm{Cov}(X_0, X_n) \sim R(n)$, as n tends to infinity, is that $K_1(n)(\theta_n - a) \sim R(n)$ and $K_2(n)(\theta_n - a)^2 = o(R(n))$.

Proof. We present the proof for the case where $a \notin \mathrm{int}(\Theta)$ and assuming that $a > x$ for all

$x \in D$. The other cases are dealt analogously. Let $\{\alpha_m\}_{m\in\mathbb{N}^*}$ be an arbitrary sequence of parameters in D converging from the left to a. Applying a Taylor's expansion with Lagrange's remainder in C_θ around $\theta = a$, and proceeding as in the proof of Proposition 3.1, we obtain

$$C_{\theta_n}(u,v) = uv + \lim_{m\to\infty} \partial C_{\alpha_m}(u,v)(\theta_n - a) + \frac{1}{2}\lim_{m\to\infty} \partial^2 C_{\alpha_m}(u,v)\big(\theta_0(n) - a\big)^2, \tag{5.1}$$

with the notation as in the enunciate. By Hoeffding's lemma it follows that

$$\begin{aligned}\operatorname{Cov}(X_0, X_n) &= \iint_{I^2} \frac{C_{\theta_n}(u,v) - uv}{l_0(u)l_n(v)}\,dudv = \left(\iint_{I^2} \lim_{m\to\infty} \frac{\partial C_{\alpha_m}(u,v)}{l_0(u)l_n(v)}\,dudv\right)(\theta_n - a) + \\ &\quad + \frac{1}{2}\left(\iint_{I^2} \lim_{m\to\infty} \frac{\partial^2 C_{\alpha_m}(u,v)}{l_0(u)l_n(v)}\,dudv\right)\big(\theta_0(n) - a\big)^2 \\ &= K_1(n)(\theta_n - a) + \frac{1}{2}K_2(n)\big(\theta_0(n) - a\big)^2,\end{aligned}$$

where $\theta_0(n) \in [\theta_n, a)$ depends on n and satisfies $\lim_{n\to\infty} \theta_0(n) = a$, provided the integrals exist. By hypothesis

$$\big|K_2(n)\big|\big(\theta_0(n) - a\big)^2 \le \big|K_2(n)(\theta_n - a)^2\big| = o(R(n)),$$

and the result follows from the hypothesis on $K_1(n)$. ∎

Remark 5.1. We observe that calculating K_1 and K_2 in closed form may be cumbersome. In practice, however, numerical integration can be applied with a high degree of precision, since K_1 and K_2 are usually very smooth functions, when they exist. When they do not exist, the theory does not apply, but the method of direct parameterization as in Example 2.1 may be used instead, although finding a suitable parameterization in this case may be a non-trivial task.

Example 5.1 (*Covariance Decay on Gaussian Processes*)**.** Let ϕ, Φ and Φ^{-1} denote the density, the distribution function and the quantile function of a standard normal random variable, respectively. Also let Φ_ρ denote the distribution function of a bivariate normal distribution with mean $(0,0)^\mathsf{T}$ and variance-covariance matrix given by $\Omega := \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. The so-called *Gaussian family of copulas* comprehend the copulas given by

$$C_\rho(u,v) = \Phi_\rho\big(\Phi^{-1}(u), \Phi^{-1}(v)\big),$$

for $\rho \in [-1,1]$. As a particular case we have $C_0 = \Pi$. Let $F_n = \Phi$, for all $n \in \mathbb{N}$. For simplicity, for $u \in I$, we shall denote $q_u := \Phi^{-1}(u)$. The chain rule yields

$$\frac{\partial C_\rho(u,v)}{\partial\rho} = \frac{\rho C_\rho(u,v)}{2\pi(1-\rho)^{\frac{3}{2}}} + \frac{1}{2\pi\sqrt{1-\rho}} \iint_S \left[\frac{xy}{1-\rho^2} - \frac{(x^2+y^2-2\rho xy)}{(1-\rho^2)^2}\right] e^{-\frac{\rho(x^2+y^2-2\rho xy)}{2(1-\rho^2)}}\,dxdy,$$

where $S := S(u,v) = (-\infty, q_u] \times (-\infty, q_v]$. Thus,

$$\lim_{\rho\to 0} \frac{\partial C_\rho(u,v)}{\partial\rho} = \frac{1}{2\pi}\left(\int_{-\infty}^{q_u} y e^{-\frac{y^2}{2}}\,dy\right)\left(\int_{-\infty}^{q_v} y e^{-\frac{y^2}{2}}\,dy\right) = \phi(q_u)\phi(q_v),$$

where the last equality follows from integration by parts. Hence $K_1(n) = 1$. We move to calculate K_2. Elementary calculus yields

$$\begin{aligned}\lim_{\rho\to 0} \frac{\partial^2 C_\rho(u,v)}{\partial\rho^2} &= uv - \iint_S x^2(1-y^2)\phi(x)\phi(y)dxdy \\ &= uv - \Big[uv - q_u q_v \phi(q_u)\phi(q_v)\Big] = q_u q_v \phi(q_u)\phi(q_v).\end{aligned}$$

Now, since $F_n \sim \mathcal{N}(0,1)$,

$$K_2(n) = \left(\int_I q_u du\right)^2 = \left(\int_{\mathbb{R}} y\phi(y)dy\right)^2 = 0,$$

where the second equality follows by changing variables to $y = q_u$. Therefore, if $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of standard normal random variables and the copula associated with (X_0, X_n) is C_{ρ_n}, the calculations above show that the parameterization we choose for the sequence $\{\rho_n\}_{n\in\mathbb{N}^*}$ will ultimately determine the decay of covariance (and the decay of correlation for that matter, since $X_n \sim \mathcal{N}(0,1)$) of (X_0, X_n). Consider the strongly stationary Gaussian process $\{X_n\}_{n\in\mathbb{N}}$ where $X_n \sim \mathcal{N}(0,1)$ and the copula of (X_r, X_s) is $C_{\rho_{|r-s|}}$ (such a construction is always possible for Gaussian processes, since all finite dimensional copulas can be taken n-dimensional Gaussian copulas with the appropriate covariance structure). Some examples are as follows.

- For $q \in \mathbb{N}^*$, and $\{\vartheta_n\}_{n=0}^q$, the parameterization

$$\rho_n = \frac{1}{1+\vartheta_1^2}\sum_{j=0}^{q-|n|} \vartheta_j\vartheta_{j+|n|}\delta(|n| \le q)$$

  determines a Gaussian MA(q) process (cf. Brockwell and Davis, 1991, p.78 with the adequate adaptations).

- For $|\varphi| < 1$, the parameterization $\rho_n = \varphi^{|n|}$ determines a Gaussian AR(1) process (as noted in Joe, 1997).

- The parameterization $\rho_n = 2^{-n}(1 + 0.75n)$, for all $n \in \mathbb{N}^*$, determines a Gaussian ARMA$(2,1)$ process defined by $(1 - \mathcal{B} + 0.25\mathcal{B}^2)X_t = (1 + \mathcal{B})Z_t$, $t \in \mathbb{N}$, for $\{Z_t\}_{t\in\mathbb{N}}$ i.i.d. $\mathcal{N}(0, 32/3)$, where $\mathcal{B}$ denotes the backward shift operator (cf. Brockwell and Davis, 1991, p.92 with the adequate adaptations).

- For $d \in (-0.5, 0.5)$, the parameterization

$$\rho_n = \prod_{k=1}^{n} \frac{k-1+d}{k-d} \sim \frac{\Gamma(1-d)}{\Gamma(d)}\, n^{2d-1}, \tag{5.2}$$

  for $n \in \mathbb{N}^*$, determines a Gaussian ARFIMA$(0, d, 0)$ process (see Lopes, 2008), which is a long-range dependent process.

- For $\{c_k\}_{k\in\mathbb{N}}$ a sequence of real numbers satisfying $\sum_{k\in\mathbb{N}} c_k^2 < \infty$, the parameterization $\rho_n = \frac{\sum_{j\in\mathbb{N}} c_j c_{j+|n|}}{\sum_{k\in\mathbb{N}} c_k^2}$ determines a general Gaussian linear process with coefficients $\{c_k\}_{k\in\mathbb{N}}$.

Remark 5.2. We notice that in Example 5.1, we assume that $X_n \sim \mathcal{N}(0,1)$. But in the usual definition of an ARFIMA$(0, d, 0)$ process with standard normal innovations, for instance, the marginals are distributed as $\mathcal{N}\big(0, \sum_{k\in\mathbb{N}} c_k^2\big)$, where the sequence $\{c_k\}_{k\in\mathbb{N}}$ are the coefficients of the MA(∞) representation of the process. So that, in practice, one has to be careful to distinguish the meaning of the parameter ρ_n, which can be misleading.

The general framework of Theorem 5.1 can be extended to cover the case of parametric families of copulas for which the parameter space $\Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}^*$. Let $\{C_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$, $\Theta \subseteq \mathbb{R}^k$, be a family of copulas for which $C_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ on an open neighborhood $U \subseteq \Theta$ of a point $\boldsymbol{a} = (a_1, \cdots, a_k)^\intercal \in \text{int}(\Theta)$. Recall that the *differential* of $C_{\boldsymbol{\theta}}$

with respect to $\boldsymbol{\theta}$ at $\boldsymbol{a} \in \mathbb{R}^k$ is the linear functional $d_{\boldsymbol{\theta}} C_{\boldsymbol{a}}(u,v) : \mathbb{R}^k \to \mathbb{R}$ whose value at a point $\boldsymbol{b} = (b_1, \cdots, b_k)^\mathsf{T} \in \mathbb{R}^k$ is

$$d_{\boldsymbol{\theta}} C_{\boldsymbol{a}}(u,v) \cdot \boldsymbol{b} = \sum_{i=1}^{k} \frac{\partial}{\partial \theta_i} C_{\boldsymbol{\theta}}(u,v) b_i \bigg|_{\boldsymbol{\theta}=\boldsymbol{a}}.$$

The *second differential* of $C_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{a} \in \mathbb{R}^k$ applied to $\boldsymbol{b} = (b_1, \cdots, b_k)^\mathsf{T} \in \mathbb{R}^k$ is given by

$$d^2_{\boldsymbol{\theta}} C_{\boldsymbol{a}}(u,v) \cdot \boldsymbol{b}^2 = \sum_{i,j=1}^{k} \frac{\partial^2}{\partial \theta_i \partial \theta_j} C_{\boldsymbol{\theta}}(u,v) b_i b_j \bigg|_{\boldsymbol{\theta}=\boldsymbol{a}}.$$

With this formulation, in the following Theorem 5.2 we consider general decay of covariance in parametric families of copulas for which the space of parameter is a subset of $\mathbb{R}^k$, $k \geq 2$, allowing for some of the parameters to remain fixed.

Theorem 5.2. *Let $\{C_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \Theta}$, for $\Theta \subseteq \mathbb{R}^{k+s}$ with non-empty interior, $k \in \mathbb{N}^*$ and $s \in \mathbb{N}$, be a family of parametric copulas for which there exists $\boldsymbol{a} \in \Theta'$ such that $\lim_{\boldsymbol{\theta} \to \boldsymbol{a}} C_{\boldsymbol{\theta}}(u,v) = uv$, for all $u, v \in I$. The limit is to be understood as the coordinatewise adequate lateral limits in case $a \notin \operatorname{int}(\Theta)$, also allowing for s coordinates to remain fixed, that is, we allow for*

$$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k, \theta^0_{k+1}, \cdots, \theta^0_{k+s}) \longrightarrow (a_1, \cdots, a_k, \theta^0_{k+1}, \cdots, \theta^0_{k+s}) = \boldsymbol{a}.$$

Also assume that there exists a set $D \subseteq \Theta$ with non-empty interior such that $\boldsymbol{a} \in D'$ and $C_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\{\theta_1, \cdots, \theta_k\}$ in D. Let $\{F_n\}_{n \in \mathbb{N}}$ be a sequence of absolutely continuous distribution functions and define the sequences

$$K_1^{(i)}(n) = \iint_{I^2} \frac{1}{l_0(u) l_n(v)} \lim_{\boldsymbol{\theta} \to \boldsymbol{a}} \frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial \theta_i} \, du dv, \quad i = 1, \cdots, k,$$

and

$$K_2^{(i,j)}(n) = \iint_{I^2} \frac{1}{l_0(u) l_n(v)} \lim_{\boldsymbol{\theta} \to \boldsymbol{a}} \frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial \theta_i \partial \theta_j} \, du dv, \quad i, j = 1, \cdots, k.$$

Let $\{\boldsymbol{\theta}_n\}_{n \in \mathbb{N}^*}$ be a sequence in D converging to $\boldsymbol{a}$, with possibly s fixed coordinates and let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables such that $X_n \sim F_n$, $n \in \mathbb{N}$, and the copula associated with (X_0, X_n) is $C_{\boldsymbol{\theta}_n}$. Given a measurable function $R : \mathbb{R} \to \mathbb{R}$ satisfying $\lim_{n \to \infty} R(n) = 0$, suppose that

$$\sum_{i=1}^{k} K_1^{(i)}(n)(|\theta_n^{(i)} - a_i|) \sim R(n) \quad \text{and} \quad \sum_{i,j=1}^{k} K_2^{(i,j)}(n)\big(\big|\theta_n^{(i)} - a_i\big|\big)\big(\big|\theta_n^{(j)} - a_j\big|\big) = o\big(R(n)\big). \tag{5.3}$$

Then, $\operatorname{Cov}(X_0, X_n) \sim R(n)$ as n goes to infinity.

Proof. We present the proof for the case where $\boldsymbol{a} \notin \operatorname{int}(\Theta)$. The other cases are dealt analogously. Let $\{\boldsymbol{\alpha}_m\}_{m \in \mathbb{N}^*}$ be an arbitrary sequence of parameters in D such that $\boldsymbol{\alpha}_m \to \boldsymbol{a}$ (assuming the adequate lateral limit when necessary, allowing for s coordinates to remain fixed). Applying Taylor's formula with Lagrange's remainder in $(\theta_1, \cdots, \theta_k)^\mathsf{T}$ around $(a_1, \cdots, a_k)^\mathsf{T}$ (the other parameters are fixed), there exists $\boldsymbol{\theta}_0 = (\theta_0^{(1)}, \cdots, \theta_0^{(k)})^\mathsf{T}$ such that $|\theta_0^{(i)} - a_i| \leq |\theta_i - a_i|$, for all $i = 1, \cdots, k$, and

$$\begin{aligned} C_{\boldsymbol{\theta}}(u,v) &= \lim_{m \to \infty} C_{\boldsymbol{\alpha}_m}(u,v) + \lim_{m \to \infty} d_{\boldsymbol{\theta}} C_{\boldsymbol{\alpha}_m}(u,v)(|\boldsymbol{\theta} - \boldsymbol{a}|) + \frac{1}{2} \lim_{m \to \infty} d^2_{\boldsymbol{\theta}} C_{\boldsymbol{\alpha}_m}(u,v)(\boldsymbol{\theta}_0 - \boldsymbol{a})^2 \\ &= uv + \sum_{i=1}^{k} \lim_{m \to \infty} \left[\frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial \theta_i} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m} \right] (|\theta_i - a_i|) + \end{aligned}$$

$$+\frac{1}{2}\sum_{i,j=1}^{k}\lim_{m\to\infty}\left[\frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m}\right](|\theta_0^{(i)}-a_i|)(|\theta_0^{(j)}-a_j|). \tag{5.4}$$

Substituting $\boldsymbol{\theta}_n$ as in the enunciate in (5.4), we have

$$\begin{aligned}C_{\boldsymbol{\theta}_n}(u,v) = uv + \sum_{i=1}^{k}\lim_{m\to\infty}\left[\frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial\theta_i}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m}\right](|\theta_n^{(i)}-a_i|)+\\ +\frac{1}{2}\sum_{i,j=1}^{k}\lim_{m\to\infty}\left[\frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m}\right]\big(|\theta_0^{(i)}(n)-a_i|\big)\big(|\theta_0^{(j)}(n)-a_j|\big),\end{aligned} \tag{5.5}$$

where $\theta_0^{(i)}(n) \in \big(\min\{\theta_n^{(i)},a_i\},\max\{\theta_n^{(i)},a_i\}\big)\cup\{\theta_n^{(i)}\}$, for $i=1,\cdots,k$. Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{F_n\}_{n\in\mathbb{N}}$ be as in the enunciate. Hoeffding's lemma combined with (5.5) yields

$$\begin{aligned}\mathrm{Cov}(X_0,X_n) &= \sum_{i=1}^{k}\left[\iint_{I^2}\frac{1}{l_0(u)l_n(v)}\lim_{m\to\infty}\frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial\theta_i}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m}dudv\right](|\theta_n^{(i)}-a_i|)+\\ &\quad+\frac{1}{2}\sum_{i,j=1}^{k}\left[\iint_{I^2}\frac{1}{l_0(u)l_n(v)}\lim_{m\to\infty}\frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\alpha}_m}dudv\right]\big(|\theta_0^{(i)}(n)-a_i|\big)\big(|\theta_0^{(j)}(n)-a_j|\big)\\ &= \sum_{i=1}^{k}K_1^{(i)}(n)(|\theta_n^{(i)}-a_i|)+\frac{1}{2}\sum_{i,j=1}^{k}K_2^{(i,j)}(n)\big(|\theta_0^{(i)}(n)-a_i|\big)\big(|\theta_0^{(j)}(n)-a_j|\big)\\ &\sim R(n)+o\big(R(n)\big)\sim R(n),\end{aligned}$$

by the hypothesis on $K_1^{(i)}$ and $K_2^{(i,j)}$. ∎

Conditions (5.3) are general ones. Observe that if there exists $i_0\in\{1,\cdots,k\}$ such that

$$K_1^{(i_0)}(n)(|\theta_n^{(i_0)}-a_{i_0}|)\sim R(n) \quad\text{and}\quad K_1^{(i)}(|\theta_n^{(i)}-a_i|)=O\big(R(n)\big),\ \text{for all } i\neq i_0,$$

and

$$K_2^{(i,j)}(n)\big(|\theta_n^{(i)}-a_i|\big)\big(|\theta_n^{(j)}-a_j|\big)=o\big(R(n)\big),\ \text{for all } i,j=1,\cdots,k,\ i\le j,$$

then the conclusion of Theorem 5.2 holds. These conditions, which imply (5.3), are usually simpler to verify.

Recall that the Frèchet-Hoeffding lower bound copula and upper bound copulas are given respectively by, $W(u,v)=\max\{u+v-1,0\}$ and $M(u,v)=\min\{u,v\}$. The importance of the copulas W and M relies in two facts. First, for any copula C, $W(u,v)\le C(u,v)\le M(u,v)$, for all $u,v\in I$. Second, the copula related to the random vector $\big(X,h(X)\big)$ is W if, and only if, h is decreasing and it is M if, and only if, h is increasing. A family of copulas for which the copulas Π, W and M are particular (or limiting) cases is said to be *comprehensive*.

Comprehensiveness is a highly desirable property for a family of copulas because this means that the family can model a broad range of dependence structures. The Frank copula of Example 3.3 and the Gaussian copula of Example 5.1 are examples of comprehensive families. The FGM family from Example 2.1 is not, but due to the simple analytical form of the copulas in this family, it is widely employed in the literature in modeling, in testing association and in studying efficiency of nonparametric procedures (cf. Nelsen, 2006, p.78). However, this family is arguably too restrictive for most applications as, for instance, the Kendall's τ dependence coefficient for

this family ranges in $[-2/9, 2/9]$. Since the space of all bidimensional copulas is a convex space, a simple solution for this problem is to consider a new copula obtained from the convex combination of the FGM copula and another copula(s) presenting the desired complementary characteristics.

Example 5.2. Let $\{C_\gamma\}_{\gamma\in[-1,1]}$ denote the FGM family of Example 2.1 and let $\{C_\delta\}_{\delta\in[1,\infty)}$ be the Euclidean family of copulas with generator $\varphi_\delta(t) = (1-t)^\delta$. This family comprehend the copulas of the form

$$C_\delta(u,v) := \max\left\{1 - \left[(1-u)^\delta + (1-v)^\delta\right]^{\frac{1}{\delta}}, 0\right\}, \quad \text{for } \delta \geq 1.$$

Particular cases of this family are $C_1 = W$ and $C_\infty = M$. Define a new three-parameter comprehensive family of copulas by setting

$$C_{\boldsymbol{\theta}}(u,v) = \alpha C_\gamma(u,v) + (1-\alpha)C_\delta(u,v), \tag{5.6}$$

where $\boldsymbol{\theta} := (\gamma, \alpha, \delta)^\mathsf{T} \in [-1,1] \times [0,1] \times [1,\infty)$. Notice that $C_{\boldsymbol{\theta}} = \Pi$ when $\boldsymbol{\theta} = (0,1,\delta)^\mathsf{T}$, for all $\delta \in [1,\infty)$, so that, in the notation of Theorem 5.2, $\boldsymbol{a} := (0,1,\delta)^\mathsf{T}$. In order to exemplify the use of Theorem 5.2, we shall analyze the compatibility of the family (5.6) with long-range dependence. For simplicity, let us fix $\delta_0 = 1$ and consider the triple $\boldsymbol{\theta} = (\gamma, \alpha, \delta_0)^\mathsf{T}$. Consider the family of *triangular distribution functions* in $[a,b]$, denoted by $\mathrm{Tr}(a,b)$, whose distribution function and density are given, respectively, by

$$F(x;a,b) = \left(\frac{x-a}{b-a}\right)^2 \quad \text{and} \quad f(x;a,b) = \frac{2(x-a)}{(b-a)^2},$$

for all $x \in [a,b]$. For two bounded sequences of real numbers $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$, with $b_n > a_n$ for all $n \in \mathbb{N}$, let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions such that F_n is distributed as $\mathrm{Tr}(a_n, b_n)$, for each $n \in \mathbb{N}$. In this case $l_n(x) = \frac{2\sqrt{x}}{b_n - a_n}$. Let us denote $\partial_\gamma C_{\boldsymbol{\theta}}(u,v) := \frac{\partial C_{\boldsymbol{\theta}}(u,v)}{\partial\gamma}$ and similarly for the derivative with respect to α. The first derivative of $C_{\boldsymbol{\theta}}$ with respect to γ and α are

$$\partial_\gamma C_{\boldsymbol{\theta}}(u,v) = \alpha uv(1-u)(1-v) \quad \text{and} \quad \partial_\alpha C_{\boldsymbol{\theta}}(u,v) = C_\gamma(u,v) - C_{\delta_0}(u,v),$$

so that

$$K_1^{(1)}(n) = \frac{1}{4}\left(\int_I \sqrt{u}(1-u)du\right)^2 (b_0 - a_0)(b_n - a_n) = \frac{4}{15^2}(b_0 - a_0)(b_n - a_n)$$

and

$$K_1^{(2)}(n) = \frac{1}{4}\left[\left(\int_I \sqrt{u}\,du\right)^2 - \iint_{I^2} \frac{W(u,v)}{\sqrt{uv}}\,dudv\right](b_0 - a_0)(b_n - a_n) = k_0(b_0 - a_0)(b_n - a_n),$$

where $k_0 = \frac{1}{9} - \frac{\pi}{8}$. As for the second derivative, since $C_{\boldsymbol{\theta}}$ is a linear function of (γ, α),

$$\frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial\gamma^2} = \frac{\partial^2 C_{\boldsymbol{\theta}}(u,v)}{\partial\alpha^2} = 0, \quad \text{and} \quad \frac{\partial^2_{\gamma,\alpha} C_{\boldsymbol{\theta}}(u,v)}{\partial\gamma\partial\alpha} = uv\ln(u)\ln(v),$$

so that

$$K_2^{(1,2)}(n) = \frac{1}{9}(b_0 - a_0)(b_n - a_n) \qquad \text{and} \qquad K_2^{(1,1)}(n) = K_2^{(2,2)}(n) = 0.$$

Therefore, if we choose $b_n - a_n \in \mathscr{L}$, with $b_n - a_n \nrightarrow 0$, $\gamma_n \sim L(n)n^{-\beta}$ and $\alpha_n - 1 \sim o(n^{-\beta})$ (or vice-versa), for $\beta \in (0,1)$ and $L \in \mathscr{L}$, then $(C_{\boldsymbol{\theta}}, F_n)$ is compatible with long-range dependence. If we take $a_n = a_0$ and $b_n = b_0$, $a_0 < b_0$, instead, we still obtain the result.

Example 5.3. Recall that the *Archimax* family of copulas (Capéraà et al., 2000; Mari and Kotz, 2001) comprehend copulas of the form

$$C_{\varphi,A}(u,v) = \varphi^{-1}\left(\big(\varphi(u)+\varphi(v)\big)A\left(\frac{\varphi(u)}{\varphi(u)+\varphi(v)}\right)\right),$$

where φ is an Archimedean generator and A is a dependence function of an EVC family. Notice that when $\varphi \equiv 1$ we obtain the Euclidean family with generator φ and when $\varphi(t) = -\ln(t)$, we obtain the EVC family with dependence function A. Consider

$$\varphi_\theta(t) = \ln\left(\frac{1-\theta(1-t)}{t}\right) \quad \text{and} \quad A_{\alpha,\beta}(t) = 1 - \min\{\alpha t, \beta(1-t)\},$$

for $\theta \in [-1,1)$ and $\alpha, \beta \in I$. Notice that φ_θ generates the AMH family from Example 3.1 while $A_{\alpha,\beta}$ generates the *Marshall-Olkin* family of copulas (Nelsen 2006; Mari and Kotz, 2001), comprehending copulas of the form

$$C_{\alpha,\beta}(u,v) = \min\{uv^{1-\alpha}, vu^{1-\beta}\}.$$

Also notice that $A_{0,\beta} = A_{\alpha,0} \equiv 1$. Fixed $\beta_0 \in (0,1)$, the Archimax copula generated by φ_θ and A_{α,β_0} can be written as

$$C_{\varphi_\theta,A_{\alpha,\beta_0}}(u,v) = \begin{cases} \frac{(\theta-1)uv^{1-\alpha}}{\theta uv^{1-\alpha} - [1-\theta(1-u)][1-\theta(1-v)]^{1-\alpha}}, & \text{if } u \leq \frac{(\theta-1)v^{\frac{\beta_0}{\alpha}}}{v^{\frac{\beta_0}{\alpha}} - [1-\theta(1-v)]^{\frac{\beta_0}{\alpha}}}; \\ \frac{(\theta-1)vu^{1-\beta_0}}{\theta uv^{1-\beta_0} - [1-\theta(1-v)][1-\theta(1-u)]^{1-\beta_0}e^{\beta_0}}, & \text{if } u > \frac{(\theta-1)v^{\frac{\beta_0}{\alpha}}}{v^{\frac{\beta_0}{\alpha}} - [1-\theta(1-v)]^{\frac{\beta_0}{\alpha}}}. \end{cases}$$

Hence, Theorem 5.2, does not apply since there is no set where the copula is twice differentiable on (α,θ) for all $(u,v) \in I^2$. Notice that the support of the copula depends on both, the parameters (α,θ) and the argument (u,v).

6 Applications

In this section we present two applications of the theory developed in the previous sections. The first one is related to the parameter estimation in Gaussian time series models and the other is related to the simulation of time series with a given decay of covariance.

We start by setting the mathematical framework. Let $\{C_\theta\}_{\theta\in\Theta}$ be a family of parametric copulas, for $\Theta \subseteq \mathbb{R}$ with non-empty interior. Assume that there exists a point $a \in \Theta'$ such that $\lim_{\theta\to a} C_\theta = \Pi$, where the limit is to be understood as the adequate lateral limit if $a \notin \text{int}(\Theta)$. Also assume that there exist a set $D \subseteq \Theta$ with non-empty interior such that $a \in D'$ and C_θ, seen as a function of the parameter θ, is of class C^2 in D. Let $\{\theta_n\}_{n\in\mathbb{N}^*}$ be a sequence in D such that $\lim_{n\to\infty}\theta_n = a$. Let $\{X_n\}_{n\in\mathbb{N}}$ be a weakly stationary process for which X_n is identically distributed with common absolutely continuous distribution F, for all $n \in \mathbb{N}$. Notice that the choice of copulas and F implies that $K_1(n)$ and $K_2(n)$ in Theorem 5.1 are constants, provided they exist. Also assume that $\theta_n - a = R(n,\gamma)$, where $R(n,\gamma)$ is a given measurable function incorporating K_1 and satisfying $R(n,\gamma) \to 0$, as n goes to infinity, and $\gamma \in S \subseteq \mathbb{R}$ is some (identifiable) parameter of interest. These assumptions make the framework compatible with

the assumptions of Theorem 5.1, from which we conclude that $\mathrm{Cov}(X_t, X_{t+h}) \sim R(h, \gamma)$.

In this context, in principle, estimating the parameter γ is equivalent to estimating the particular parameterization $\{\theta_n\}$. Suppose we observe a realization (time series) $x_1, \cdots, x_n$ from $\{X_n\}_{n\in\mathbb{N}}$ as in the previous paragraph. Our goal is to estimate γ based on these observations. To estimate the parameter γ, the theory developed in the last sections suggests the following simple methodology.

1. We start by obtaining an estimate $\hat{F}$ of the underlying unknown marginal distribution function F. This can be achieved by any reasonable method. For instance, we could apply a parametric method by fitting a distribution to the data, or a non-parametric one, such as splines or any other kernel density estimator.

2. With the estimated distribution function $\hat{F}$ at hand, we can obtain K_1 and K_2, which must be finite and $K_1 \neq 0$. We then form a new time series by setting $y_i := \hat{F}(x_i)$, for $i = 1, \cdots, n$. Notice that y_i will lie on the unit interval.

3. Next we form a bivariate time series $\{\boldsymbol{u}_k^{(1)}\}_{k=1}^{n-1}$ by setting $\boldsymbol{u}_i^{(1)} := (y_i, y_{i+1})$, $i = 1, \cdots, n-1$ and by removing pairs containing 0 or 1, if any. By Sklar's theorem, $\{\boldsymbol{u}_k^{(1)}\}_{k=1}^{n-1}$ can be regarded as a correlated sample from C_θ. From these pseudo observations, θ can be estimated by using any reasonable method. For instance, a maximum likelihood procedure or a nonparametric approach such as the inversion of Kendall's τ or Spearman's ρ method could be applied. Let $\hat{\theta}_1$ denote the estimate obtained. Notice that $\hat{\theta}_1 - a$ is an estimate of $R(1, \gamma)$.

4. Proceeding analogously for each $s \in \{2, \cdots, m\}$, where $m < n$ is the maximum desired lag of estimation, we form the sequence $\{\boldsymbol{u}_k^{(s)}\}_{k=1}^{n-s}$, by setting $\boldsymbol{u}_i^{(s)} := (y_i, y_{i+s})$ (pairs containing 0 are removed, if any), from which we obtain the estimate $\hat{\theta}_s$. For each s, $\hat{\theta}_s - a$ is an estimate of $R(s, \gamma)$.

5. Let $\mathscr{D} : \mathbb{R}^k \times \mathbb{R}^k \to [0, \infty)$, $k > 1$, be a given function measuring the distance between two vectors. For $k \in \{1, \cdots, m\}$, let $\hat{\boldsymbol{\theta}}_k := (\hat{\theta}_1 - a, \cdots, \hat{\theta}_k - a)^\mathsf{T}$ and $\boldsymbol{R}_k(\gamma) := \big(R(1, \gamma), \cdots, R(k, \gamma)\big)^\mathsf{T}$. The estimator $\hat{\gamma}$ of γ is then defined as

$$\hat{\gamma} := \frac{1}{m}\sum_{k=1}^{m} \hat{\gamma}_k, \qquad \text{where} \qquad \hat{\gamma}_k := \operatorname*{argmin}_{\gamma\in S}\big\{\mathscr{D}\big(\hat{\boldsymbol{\theta}}_k, \boldsymbol{R}_k(\gamma)\big)\big\}, \quad k = 1, \cdots, m.$$

In Step 1, a histogram can be useful in determining a parametric distribution function candidate to fit the data. Alternatively, splines or some other kernel density estimator can be used to estimate F. In the case of a Gaussian process, a good estimate of F is by taking $\hat{F}$ normally distributed with mean and variance equal to the sample mean and sample variance, respectively. In Step 2, numerical integration can be applied to obtain K_1 and K_2 since, when they exist, they are usually very smooth functions.

Remark 6.1. The estimates $\hat{\theta}$ obtained from the bivariate samples in Steps 3 and 4 above may be highly biased depending on the strength of the dependence in the process. For instance, the maximum likelihood procedure for copulas is consistent under some regularity conditions, which include an i.i.d. condition on the sample. Small departures from it still yield reasonable

estimates, but in the presence of strong dependence, the procedure may result in underestimation. In that case, some bias correction method can be applied. See Subsections 6.1, 6.2 and 6.3 below.

In Step 4, the maximum desired lag m depends highly on the nature of the parameter γ. For instance, as we shall see in Subsections 6.1 and 6.2, for Gaussian MA(1) and AR(1) processes, $m = 1$ is a convenient choice for estimating the parameter of the process. As a rule of thumb, m should be a small fraction of n, since the bivariate correlated sample $\{\boldsymbol{u}_k^{(m)}\}_{k=1}^{n-m}$ has size $n - m$ so that, in order to yield reasonable estimates, it is desirable that the difference $n - m$ is reasonably large. Also removing pairs containing 0 and 1 from $\{\boldsymbol{u}_k^{(m)}\}$ is necessary because $C(0, u) = C(u, 0) = 0$, and $C(1, u) = C(u, 1) = u$ for all $u \in I$, for any copula C, so that the copula density function at these points is 0. From our simulations results, presented in the next subsections, when the memory parameter in a strongly dependent process is of interest, usually small m yields better results than large ones due to the cumulative bias in the estimation of $\hat{\theta}_k$. For processes with weaker levels of dependence, larger ones yield slightly better results for the memory parameter than smaller ones, since the estimation bias problem in $\hat{\theta}_k$ is not so substantial.

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^k$, $k > 1$, say $\boldsymbol{u} = (u_1, \cdots, u_k)^\intercal$ and $\boldsymbol{v} = (v_1, \cdots, v_k)^\intercal$, usual choices for the function $\mathscr{D}$ in Step 5 are

$$\mathscr{D}(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{k}\sum_{i=1}^{k} |u_i - v_i| \quad \text{and} \quad \mathscr{D}(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{k}\sum_{i=1}^{k} (u_i - v_i)^2.$$

In order to exemplify the methodology and assess its finite sample performance, we carry out some Monte Carlo experiments for common used time series models under a wide variety of dependence structures, from Markovian to strong long-range dependent ones. In all simulations we assume that the innovations are standard normally distributed. The marginal distribution F is always estimated by taking $\hat{F}$ to be normally distributed with mean and variance equal to the sample mean and sample variance of the time series, respectively. The underlying bivariate copulas are always assumed to belong to the Gaussian family and the parameter is estimated via the maximum likelihood procedure.

The main task of simulating the time series and performing Steps 1 through 4 are executed by using the computational resources from the (Brazilian) National Center of Super Computing (CESUP-UFRGS). The routines are all implemented in FORTRAN 95 language with OpenMP directives for parallel computing. Step 5 and post processing were performed by using the free statistical software R.

6.1 Estimation on Gaussian MA(1) Processes

Recall that a weakly stationary process $\{X_t\}_{t\in\mathbb{N}}$ is said to be an MA(q) process if, for all $t \in \mathbb{N}$,

$$X_t = \vartheta(\mathcal{B})Z_t, \qquad Z_t \sim N(0, \sigma_Z^2),$$

where $\vartheta(z) := 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$. When $q = 1$ we obtain an MA(1) process given by $X_t = Z_t + \vartheta Z_{t-1}$. We shall assume that $|\vartheta| < 1$ for identifiability purposes. In the simulations we take $\sigma_Z^2 = 1$.

In the framework of the methodology explained in the beginning of this section, the parameter of interest in this application is $\gamma = \vartheta$. Notice that in the notation of Example 5.1, $\rho_1 = \frac{\vartheta}{1+\vartheta^2}$ and $\rho_n = 0$, for $n > 1$, so that we set $m = 1$. Also observe that assuming $\vartheta \in (-1, 1)$, then $\rho_1 \in (-0.5, 0.5)$ and $\vartheta = \frac{1-\sqrt{1-4\rho_1^2}}{2\rho_1}$.

To assess the performance of the methodology, the following Monte Carlo experiment is carried out. We simulate Gaussian MA(1) processes of length $n = 1,000$ and coefficient $\vartheta \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$. As mentioned before, the estimation is performed for a maximum lag $m = 1$ so that there is no need to specify a function $\mathscr{D}$. With the estimated copula parameter $\hat{\rho}_1$, the natural idea is to estimate ϑ by $\widetilde{\vartheta} := \frac{1-\sqrt{1-4\hat{\rho}_1^2}}{2\hat{\rho}_1}$. However, for parameter values near ± 1, it may happen that the estimate $\hat{\rho}_1$ falls outside the region $(-0.5, 0.5)$ in which case $\widetilde{\vartheta} \notin \mathbb{R}$. A simple way to overcome this problem is the following. Given $\hat{\rho}_1$, we estimate ϑ by

$$\hat{\vartheta} := \frac{1 - \sqrt{1 - 4\psi(\hat{\rho}_1)^2}}{2\psi(\hat{\rho}_1)}, \quad \text{where} \quad \psi(x) := \text{sign}(x) \min\{0.5, |x|\},$$

where $\text{sign}(x) = -1$, if $x < 0$, and 1 otherwise. The experiment is replicated 1,000 times.

Table 6.1 reports the mean estimated value and its mean square error (mse, in parenthesis). From Table 6.1 we observe that for parameter values ± 0.1 and ± 0.5, the procedure performs well producing estimates with both small bias and small mse. For $\vartheta = \pm 0.9$ the estimates are still reasonable, but a greater bias and mse are observed compared to the other cases. This is due to many values of $\hat{\rho}_1$ falling outside the region $(-0.5, 0.5)$.

Table 6.1: Estimation results for Gaussian MA(1) and AR(1) processes upon applying the methodology of Subsections 6.1 and 6.2. Presented are the mean estimated value along with the corresponding mean square error in parenthesis

Parameter	Model		Parameter	Model	
	MA(1)	AR(1)		MA(1)	AR(1)
-0.9	-0.8777 (0.0168)	-0.8989 (0.0002)	0.1	0.0990 (0.0009)	0.0973 (0.0011)
-0.5	-0.5063 (0.0031)	-0.5011 (0.0008)	0.5	0.5048 (0.0031)	0.4970 (0.0007)
-0.1	-0.1014 (0.0011)	-0.1012 (0.0010)	0.9	0.8739 (0.0178)	0.8971 (0.0002)

6.2 Estimation on Gaussian AR(1) Processes

Recall that a weakly stationary process $\{X_t\}_{t\in\mathbb{N}}$ is a Gaussian AR(p) process if, for all $t \in \mathbb{N}$,

$$\varphi(\mathcal{B})X_t = Z_t, \qquad Z_t \sim N(0, \sigma_Z^2),$$

where $\varphi(z) := 1 - \varphi_1 z - \cdots - \varphi_p z^p$. When $p = 1$, we obtain the AR(1) process given by $X_t - \varphi X_{t-1} = Z_t$.

In this case the parameter of interest is $\gamma = \varphi$. The estimation is particularly simple because $\rho_1 = \varphi$ so that we can set the maximum lag as $m = 1$. We perform the following Monte Carlo experiment. We simulate a Gaussian AR(1) process of length $n = 1,000$ and coefficient $\varphi \in \{-0.9, -0.5, -0.1, 0.1, 0.5, 0.9\}$. As mentioned before, the estimation is performed for a maximum lag $m = 1$ so that there is no need to specify a function $\mathscr{D}$. The value obtained, $\hat{\rho}_1$, is taken as the estimate of φ. We replicate the experiment 1,000 times.

The mean estimated value and its mean square error (in parenthesis) are presented in Table

6.1. The results show that the methodology works very well, producing good estimates with fairly small mean square error. Notice that the higher the absolute value of the parameter, the more precise the estimation.

6.3 Estimation on Gaussian ARFIMA$(0, d, 0)$ Processes

Recall that a Gaussian ARFIMA(p, d, q) process $\{X_t\}_{t\in\mathbb{N}}$ is a weakly stationary solution of the difference equations

$$\varphi(\mathcal{B})(1-\mathcal{B})^d X_t = \vartheta(\mathcal{B})Z_t, \qquad Z_t \sim N(0, \sigma_Z^2), \tag{6.1}$$

where $\varphi(z) := 1 - \varphi_1 z - \cdots - \varphi_p z^p$ and $\vartheta(z) := 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$ are polynomials assumed to have no common roots. It can be shown that for $d \in (-1, 0.5)$, the solution of (6.1) is causal and invertible, provided that the polynomials φ and ϑ do not have roots for $|z| \leq 1$. In this case, $\{X_t\}_{t\in\mathbb{N}}$ will have an MA(∞) representation $X_t = \sum_{k\in\mathbb{N}} c_k Z_{t-k}$, for all $t \in \mathbb{N}$, where $\{c_k\}_{k\in\mathbb{N}}$ satisfies $\sum_{k\in\mathbb{N}} c_k^2 < \infty$ (see Palma, 2007, p.44 for more details). When $p = q = 0$, we obtain the ARFIMA$(0, d, 0)$ process given by $(1-\mathcal{B})^d X_t = Z_t$.

In the case of an ARFIMA$(0, d, 0)$ the interest lies on estimating the memory parameter d, that is, $\gamma = d$. The particular parameterization of the Gaussian copulas in this case is (5.2), as given in Example 5.1. In this case there is no obvious choice for the maximum lag m. Also, for large d, one can expect a significantly high bias in the parameter estimation of the copulas, which behaves cumulatively in performing Step 5 of the methodology as m increases. Therefore, a balance in the choice of m has to be searched for. Intuitively, one can expect to obtain better results with higher values of m when the dependence is weaker, and better results for small values of m in the presence of strong dependence.

To assess the performance of the methodology proposed, we perform a Monte Carlo study by simulating Gaussian ARFIMA$(0, d, 0)$ processes of length $n = 1,000$ for $d \in \{-0.1, 0.1, 0.2, 0.3, 0.4, 0.45\}$. The series were simulated by truncating the MA(∞) representation of the process. The truncation point is fixed at 50,000 for all cases. In the simulation we apply $m \in \{5, 10, 25, 50, 100\}$. For each simulated time series and each m, we estimate the copula parameters following Steps 3 and 4 to obtain, with the notation of Example 5.1, $\hat{\rho}_1, \cdots, \hat{\rho}_m$. For each $k \in \{1, \cdots, m\}$, we calculate

$$\widetilde{d}_k = \underset{d\in(-0.5,0.5)}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{h=1}^{m} \left| \hat{\rho}_h - \Gamma(1-d)\Gamma(d)^{-1} h^{2d-1} \right| \right\}. \tag{6.2}$$

The estimate of d is then $\widetilde{d}_{can} := m^{-1} \sum_{k=1}^{m} \widetilde{d}_k$. We replicate the experiment 1,000 times. Let us call the estimator obtained from (6.2) the *canonical estimator*. The mean estimated values along with their mean square errors are reported in the left side of Table 6.2.

In Table 6.2 we see that for $d \leq 0.2$, the methodology works well, but as d increases and the dependence among the variables become stronger, an increasingly significant bias in the estimates is observed. As for the value of m, we see that except for $d = -0.1$, $m = 25$ seems to produce overall the best results. However, for $d > 0.3$, the estimates are so biased that they are of no practical relevance. As we mentioned before, this high bias for high values of d is due to the bias on the estimation of the copula parameter based on strongly dependent data since, in this case, the maximum likelihood method greatly underestimates the parameter θ.

Table 6.2: Estimation results for ARFIMA$(0, d, 0)$ processes upon applying the methodology of Subsection 6.3 for lags $m \in \{5, 10, 25, 50, 100\}$ and by applying the canonical estimator (6.2) and the bias corrected estimator (6.3). Reported are the mean estimated value and its respective mean square error.

d	Canonical Estimator by Lag m					d	Bias Corrected Estimator by Lag m				
	5	10	25	50	100		5	10	25	50	100
-0.10	-0.1008 (0.0000*)	-0.1009 (0.0000)	-0.1010 (0.0000)	-0.1011 (0.0000)	-0.1011 (0.0000)	-0.10	-0.0948 (0.0000)	-0.0949 (0.0000)	-0.0950 (0.0001)	-0.0951 (0.0000)	-0.0951 (0.0000)
0.10	0.0979 (0.0003)	0.0982 (0.0003)	0.0986 (0.0001)	0.0985 (0.0000)	0.0983 (0.0000)	0.10	0.1045 (0.0007)	0.1046 (0.0006)	0.1047 (0.0002)	0.1044 (0.0000)	0.1040 (0.0000)
0.20	0.1867 (0.0013)	0.1874 (0.0007)	0.1878 (0.0000)	0.1875 (0.0000)	0.1866 (0.0002)	0.20	0.2105 (0.0031)	0.2093 (0.0020)	0.2072 (0.0003)	0.2050 (0.0001)	0.2024 (0.0000)
0.30	0.2636 (0.0004)	0.2652 (0.0001)	0.2664 (0.0006)	0.2661 (0.0008)	0.2640 (0.0014)	0.30	0.3118 (0.0022)	0.3080 (0.0013)	0.3021 (0.0000)	0.2971 (0.0001)	0.2913 (0.0003)
0.40	0.3307 (0.0006)	0.3325 (0.0011)	0.3337 (0.0044)	0.3333 (0.0048)	0.3314 (0.0056)	0.40	0.4058 (0.0002)	0.3966 (0.0002)	0.3856 (0.0015)	0.3774 (0.0017)	0.3688 (0.0023)
0.45	0.3599 (0.0024)	0.3617 (0.0034)	0.3628 (0.0084)	0.3624 (0.0089)	0.3603 (0.0092)	0.45	0.4468 (0.0001)	0.4355 (0.0006)	0.4220 (0.0036)	0.4124 (0.0040)	0.4028 (0.0042)

Note $*$**:** 0.0000 means that the mean square error is smaller than 5×10^{-5}.

To overcome this deficiency, we propose the following bias correction procedure. We follow the same steps as before but we substitute (6.2) by the following optimization procedure

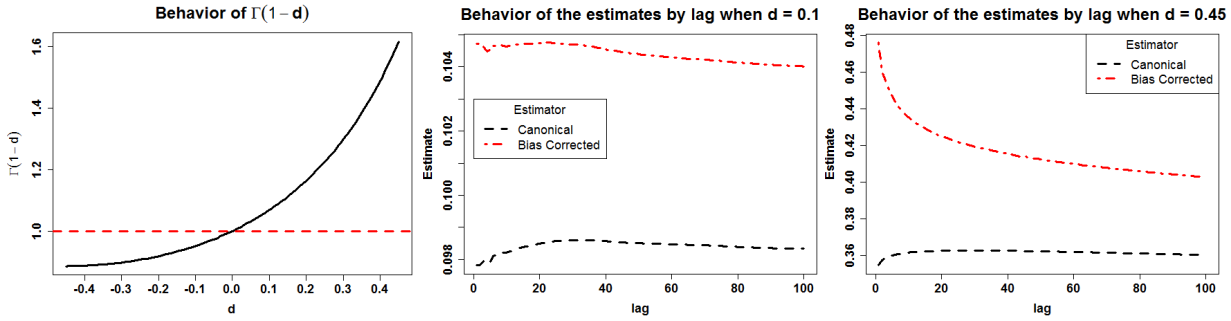
$$\hat{d}_k := \underset{d\in(-0.5,0.5)}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{h=1}^{m} \left| \hat{\rho}_h - \Gamma(d)^{-1} h^{2d-1} \right| \right\}, \tag{6.3}$$

and we set $\hat{d}_{cor} := m^{-1} \sum_{k=1}^{m} \hat{d}_k$ just like before. Let us call this the *bias corrected estimator*. Notice that the difference between (6.2) and (6.3) is the absence of $\Gamma(1-d)$ in the latter. The motivation for such correction lies on the behavior of $\Gamma(1-d)$ as shown in Figure 6.1(a). For small values of d, $\Gamma(1-d)$ assumes values close to 1, so that the good canonical estimates should not change much, but for large values of d, omitting $\Gamma(1-d)$ produces better estimates.

The results for the bias corrected estimator are presented in the right side of Table 6.2. From Table 6.2, for $d \leq 0.2$, the results for the bias corrected estimator have slightly worse performance than the canonical ones, but are still quite good. For higher values of d, however, the bias corrected estimator is far superior, especially for small lags. In Figure 6.1(b) and Figure 6.1(c), we show the evolution of the estimated values for both estimators when $d = 0.1$ and $d = 0.45$, respectively. Notice that for small values of d, the estimated value improves as the lag increases while for higher values of d, the opposite happens. As explained before, this is mainly due to the cumulative effect of the bias in estimating the copula parameter under strong dependence.

6.4 Application to Simulation of Time Series

In this subsection we present an application of the results in the previous sections to simulation of time series with a given decay of covariance. The main idea is to observe that a realization $x_1, \cdots, x_n$ from any process $\{X_t\}_{t\in\mathbb{N}^*}$ can be seen as a random variate from the joint distribution of $(X_1, \cdots, X_n)$. That is, if for a given process, $\{X_t\}_{t\in\mathbb{N}^*}$, we are able to determine the joint



(a) (b) (c)

Figure 6.1: (a) Plot of $\Gamma(1-d)$ for $d \in (-0.45, 0.45)$. Behavior of the estimated values of d by the canonical and bias corrected estimators for (b) $d = 0.1$ and (c) $d = 0.45$.

distribution of $(X_1, \cdots, X_n)$, say H, and we are able to generate random variates from H, then each single variate from H is a size n realization of the respective process.

With this in mind, consider the problem of simulating a time series, say $x_1, \cdots, x_n$, with a given decay of covariance, say $n^{-\alpha}$, for a prescribed $\alpha > 1$, with fat tailed marginals which vary with the time. This is certainly a non-trivial task, but we shall show how to approach the problem by using the theory developed in the previous sections.

Consider the following particular case of the construction presented in Example 2.1. Let $\theta_k := \kappa_0^{-1} k^{-\alpha}$, $k \in \mathbb{N}^*$, for $\alpha > 1$ and $\kappa_0 \geq \zeta(\alpha)$, and consider the sequence of copulas $\{C_{\theta_k}\}_{k\in\mathbb{N}^*}$ from the FGM family. Let $\{a_k\}_{k\in\mathbb{N}}$ be an arbitrary sequence of real numbers and let $b := \ln(2)\kappa_0^{-1/2}$. Let $\{F_k\}_{k\in\mathbb{N}}$ be a sequence of distribution functions distributed as $\mathrm{EVI}(a_k, b)$, for each $k \in \mathbb{N}$. From Example 2.1, if the copula related to X_t and X_{t+h} is C_{θ_h}, and X_t is distributed according to F_t, then

$$\mathrm{Cov}(X_t, X_{t+h}) = h^{-\alpha}, \quad \text{for all } t \in \mathbb{N} \text{ and } h \in \mathbb{N}^*.$$

From this point we observe that the task will be accomplished if we can determine an n-dimensional copula for which the marginals are the $\binom{n}{2}$ copulas $C_{ij} := C_{\theta_{|i-j|}}$, for $i, j \in \{1, \cdots, n\}$, $i \neq j$. For instance, proceeding as in Example 2.1, one can construct the n-dimensional copula (2.2). If we are able to produce a random variate from this n dimensional copula, say $u_1, \cdots, u_n$, by Sklar's theorem the desired time series is obtained by setting $x_k := F_k^{-1}(u_k)$, $k = 1, \cdots, n$.

Remark 6.2. A word of caution is in time. Because of the compatibility problem, an n-dimensional copula satisfying the conditions imposed in the Example 2.1, may not exist, so one has to be careful in the calculations.

Simulation of commonly applied Gaussian time series with a given decay of covariance is an easy task through the theory developed in the previous sections. Recall that the n-dimensional Gaussian copula is given by

$$C_\Omega(u_1, \cdots, u_n) := \Phi_\Omega\big(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_n)\big),$$

where $\Omega = (\Omega_{ij})_{i,j=1}^n$ is a symmetric positive definite real matrix with $\mathrm{diag}(\Omega) = (1, \cdots, 1)^\mathsf{T}$ and Φ_Ω denotes the multivariate normal distribution function with variance-covariance matrix Ω. It is clear that all bivariate marginals are also Gaussian copulas. Let $X_1, \cdots, X_n$ be random

variables for which the associated copula is an n-dimensional Gaussian copula C_Ω. Then the copula related to X_r and X_s is a Gaussian copula with parameter $\rho = \Omega_{rs}$. In order to simulate a Gaussian time series, therefore, we basically need to correctly specify the matrix Ω. Once Ω is specified, we obtain a single random variate from the respective Gaussian copula and apply Φ^{-1} to it. The result is the desired time series.

For instance, referring to Example 5.1, to simulate a sample of size n from an MA(q) process with coefficients $\{\vartheta_k\}_{k=0}^{q}$, we construct the matrix Ω by taking

$$\Omega_{ij} = \frac{1}{1+\vartheta_1^2} \sum_{k=0}^{q-|i-j+1|} \vartheta_k \vartheta_{k+|i-j+1|} \delta(|i-j+1| \leq q), \quad i,j = 1, \cdots, n, \ i \neq j.$$

A sample of size 400 observations from an MA(1) process with polynomial $\vartheta(z) = 1 + 0.3z$ is shown in Figure 6.2(a) along with the estimated autocorrelation function in Figure 6.2(d). Applying the methodology of Subsection 6.1, we obtain the estimated value $\hat{\vartheta} = 0.3052$.

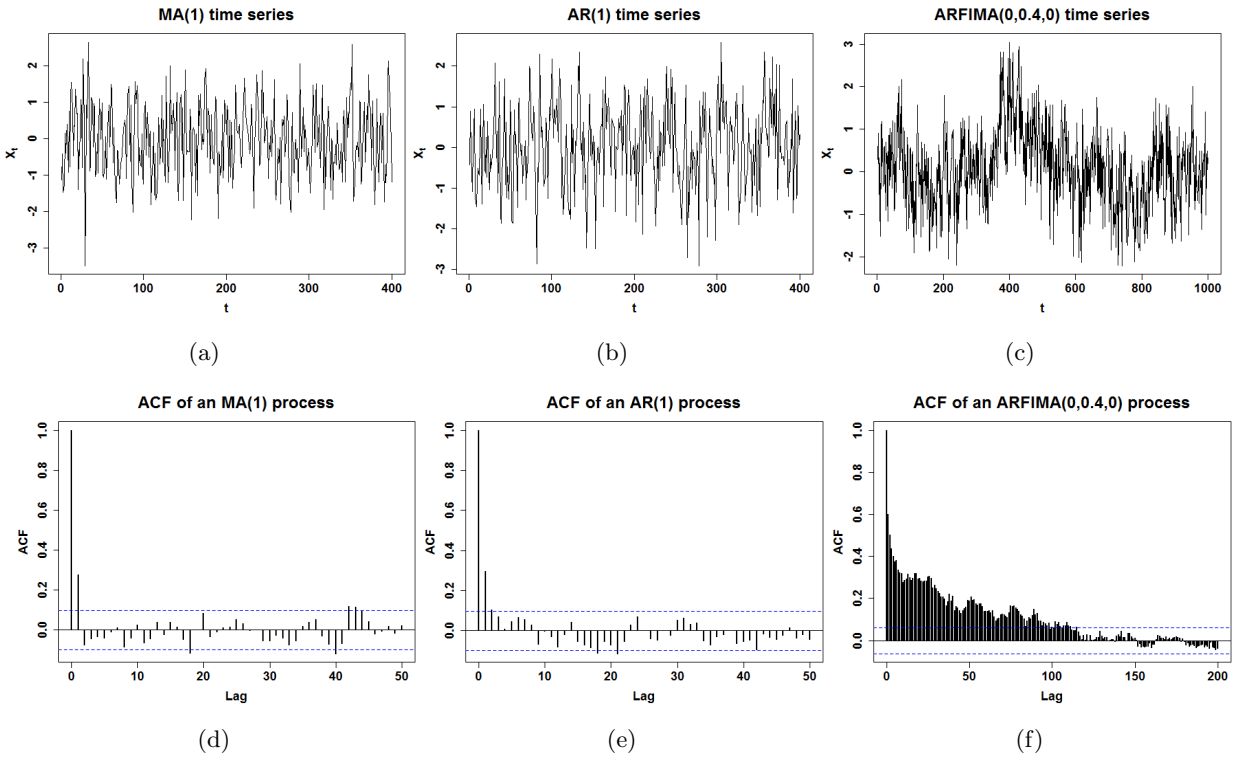


Figure 6.2: Simulated time series $\{x_t\}_{t=1}^{n}$ obtained by using the method of parameterization corresponding to an (a) MA(1); (b) AR(1), both with $n = 400$; and (c) ARFIMA(0,0.4,0) with $n = 1,000$. Estimated autocorrelation functions for lag 1 to 50 for (d) the MA(1) and (e) the AR(1) processes and for lag 1 to 200 for (f) the ARFIMA(0,0.4,0).

Referring to Example 5.1, to obtain a realization of an AR(1) process, with coefficient φ, we construct the matrix Ω by taking $\Omega_{ij} = \varphi^{|i-j|}$. A sample of size 400 observations from an AR(1) process with polynomial $\varphi(z) = 1 - 0.3z$ is shown in Figure 6.2(b) along with the estimated autocorrelation function in Figure 6.2(e). Applying the methodology of Subsection 6.2, we obtain the estimated value $\hat{\varphi} = 0.2963$.

Still referring to Example 5.1, to obtain a sample of size n observations from a Gaussian

ARFIMA$(0, d, 0)$ process, we set

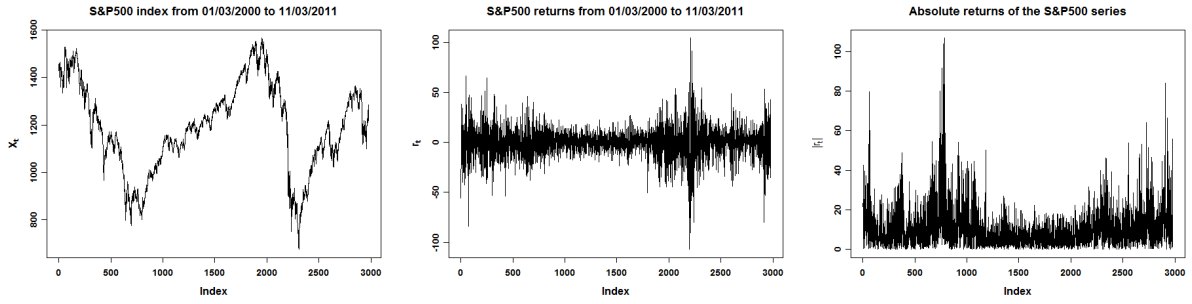
$$\Omega_{ij} = \prod_{k=1}^{|i-j|} \frac{k-1+d}{k-d}, \quad i, j = 1, \cdots, n, \ i \neq j.$$

The other expression in (5.2) works as well. A sample of size $n = 1,000$ observations from an ARFIMA$(0, d, 0)$ process for $d = 0.4$ is shown in Figure 6.2(c) along with the estimated autocorrelation function in Figure 6.2(f). Applying the methodology of Subsection 6.3, for lag 25, the canonical estimator (can) yields $\hat{d}_{can} = 0.3514$ and the bias corrected estimator (cor) yields $\hat{d}_{cor} = 0.4080$.

7 Real Data Application

In this section we apply the estimation methodology of Section 6 to the daily returns of the S&P500 US stock market index in the period from 01/03/2000 to 11/03/2011, which gives a sample size $n = 2,980$. Figure 7.1(a) to Figure 7.1(c) present the S&P500 time series, the correspondent returns $\{r_t\}_{t=1}^{2979}$ and the absolute returns $\{|r_t|\}_{t=1}^{2979}$, respectively.

Among the stylized facts about financial time series related to stock market indexes, one often observes that the returns are uncorrelated while the absolute and squared returns are correlated and present slow decay of covariance, usually proportional to $n^{-\beta}$ for $\beta \in [0.2, 0.4]$ in both cases (see Cont, 2001 and references therein). The study of the absolute and square returns are of great importance since they contain information about the (unobservable) volatility. Our goal is to apply the methodology developed in Section 6 to estimate β.



(a) (b) (c)

Figure 7.1: S&P500 (a) original time series; (b) return time series; and (c) absolute return time series.

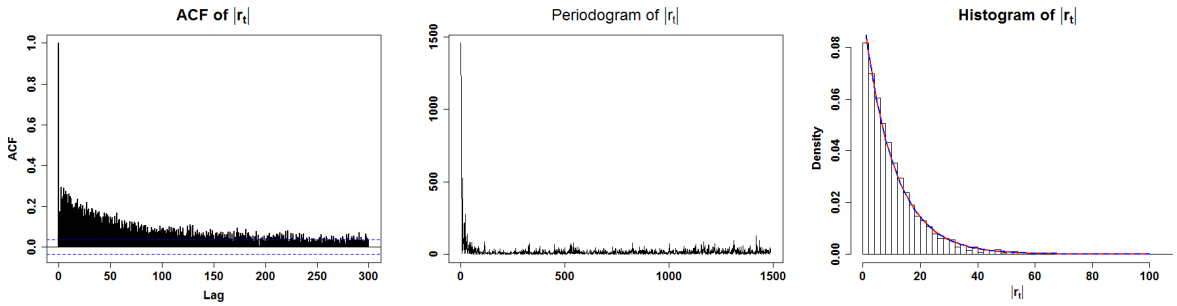
We shall work with the absolute return time series only, since due to the invariance of copulas to increasing transformations (notice that x^2 is a strictly increasing function for $x \in [0, \infty)$), the underlying bidimensional copulas are the same in both cases, and, as long as F is well estimated, the conclusions in both cases should be basically the same. Notice that this is coherent with the stylized facts. We shall assume that the underlying process is strongly stationary and ergodic and that all underlying bivariate copulas belong to the same one parameter copula family parameterized in the spirit of the theory developed in Section 5.

Figure 7.2(a) presents the autocorrelation function and the periodogram of the absolute return time series of the S&P500 index. The slow decay of the sample autocorrelation function and the pronounced peak at the zero frequency in the periodogram function both suggest long-range

dependence on the absolute return time series. Sometimes slow decay on the autocorrelation function may be due to non-stationarity. To rule this hypothesis out, a Phillips-Perron unit root test (Phillips and Perron, 1988) can be performed on $|r_t|$. The test rejects the null hypothesis of unit root at 5% confidence level, with p-value = 0.01. The conclusion reinforces the hypothesis of long-range dependence.

Given the behavior of the sample autocovariance function, it is often assumed that $\{|r_t|\}_{t=0}^{\infty}$ follows an ARFIMA(p, d, q) model, so that estimation of d is of interest. In this line, one has several estimator at disposal. See, for instance, the review in Lopes (2008) for classical and recent developments, Lopes et al. (2004) for the non-stationary ARFIMA(p, d, q) case and the recent method developed in Lopes et al. (2011). See also Palma (2007).

To apply the methodology of Section 6, we shall first estimate the underlying marginal distribution of the data. Figure 7.2(c) presents the histogram of $|r_t|$, which suggests an Exponential distribution.



(a) (b) (c)

Figure 7.2: (a) Autocorrelation function; (b) periodogram and (c) histogram of the S&P500 absolute return time series and the fitted Exponential density.

By using a maximum likelihood approach, we fit an Exp(10.688) distribution, whose density is shown in Figure 7.2(c). We then define $\hat{F} \sim$ Exp(10.688) as the estimator for the underlying marginals and set $y_i = \hat{F}(|r_t|)$, $t = 1, \cdots, 2979$.

Next step is to identify the underlying family of bidimensional copulas related to the process. One way to do this is by plotting $\{(y_i, y_{i+s})\}_{i=1}^{2979-s}$ for some values of s. The plots are a helpful tool to identify the underlying copulas by observing some key dependence features such as tail dependence, singular components, dependence strength, etc. In Figure 7.3 we have done so for lags $s \in \{1, 5, 25, 50, 100\}$. The plots suggest that the underlying family of copulas comprehend absolutely continuous ones, with relatively weak dependence for each lag and with no tail dependence. Furthermore, in view of Figure 7.3, it is reasonable to assume the bidimensional copulas to be Gaussian ones.

For the Gaussian copula with Exp(10.688) marginals, K_1 and K_2 can be shown to be finite with values approximately equal to 260.47 and 40.53, respectively. The next step is to estimate the copula parameter ρ from the pseudo samples $\{(y_i, y_{i+s})\}_{i=1}^{2979-s}$. Figure 7.4(a) presents the estimated copula parameter obtained via maximum likelihood method up to lag 400. Notice that the very first estimate appears distant from the cloud of estimates. Chosen a particular lag m, let $\hat{\rho}_1, \cdots, \hat{\rho}_m$ denote the estimates obtained in the previous step. For comparison purposes,

we perform the following three optimization procedures

$$\hat{\beta}_g(k) := \underset{\substack{\beta\in(0,1)\\ K\in\mathbb{R}\setminus\{0\}}}{\operatorname{argmin}}\left\{\frac{1}{k}\sum_{i=1}^{k}|\hat{\rho}_i - Ki^{-\beta}|\right\}, \tag{7.1}$$

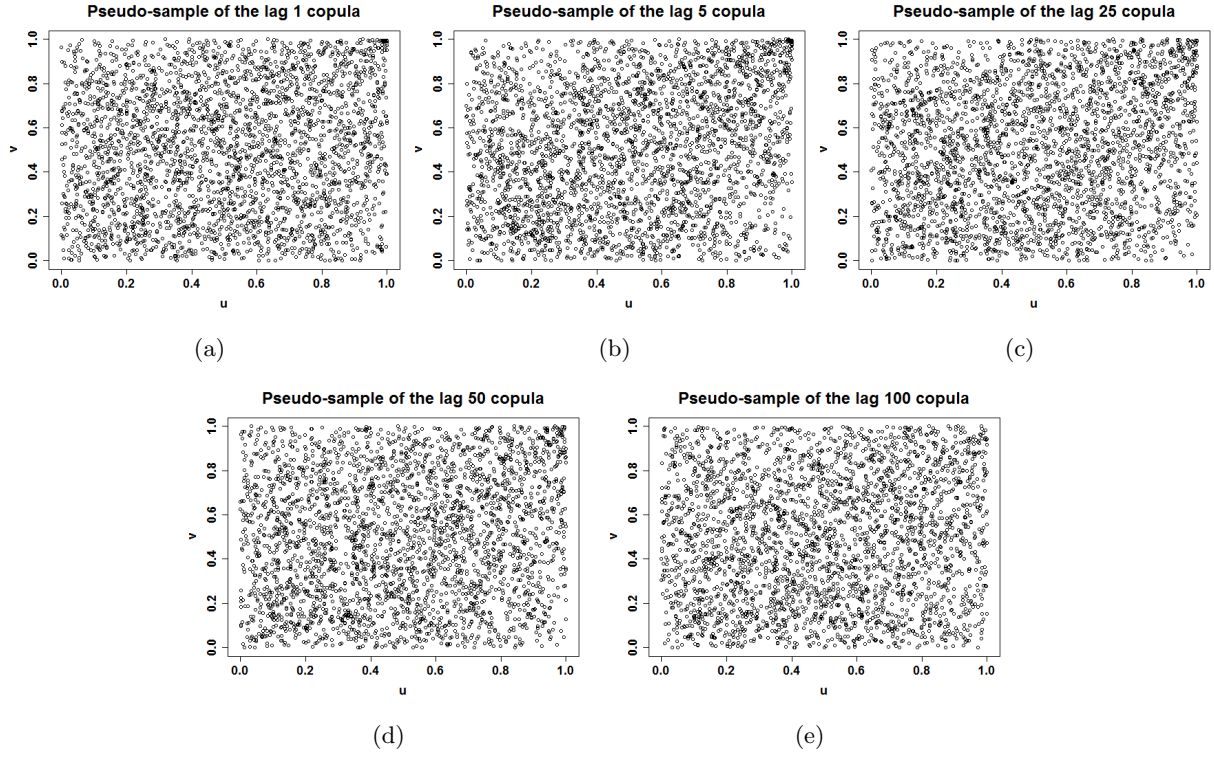
$$\hat{d}_{can}(k) := \underset{d\in(-0.5,0.5)}{\operatorname{argmin}}\left\{\frac{1}{k}\sum_{i=1}^{k}|\hat{\rho}_i - \Gamma(1-d)\Gamma(d)^{-1}i^{2d-1}|\right\}, \quad \hat{\beta}_{can}(k) := 1 - 2\hat{d}_{can}(k), \tag{7.2}$$

$$\hat{d}_{cor}(k) := \underset{d\in(-0.5,0.5)}{\operatorname{argmin}}\left\{\frac{1}{k}\sum_{i=1}^{k}|\hat{\rho}_i - \Gamma(d)^{-1}i^{2d-1}|\right\}, \quad \hat{\beta}_{cor}(k) := 1 - 2\hat{d}_{cor}(k), \tag{7.3}$$

for $k = 1, \cdots, m$. In all cases, the final estimate, denoted by $\hat{\beta}_g$, $\hat{\beta}_{can}$ and $\hat{\beta}_{cor}$, respectively, is defined as

$$\hat{\beta}_g := \frac{1}{m}\sum_{k=1}^{m}\hat{\beta}_g(k); \quad \hat{\beta}_{can} := \frac{1}{m}\sum_{k=1}^{m}\hat{\beta}_{can}(k) \quad \text{and} \quad \hat{\beta}_{cor} := \frac{1}{m}\sum_{k=1}^{m}\hat{\beta}_{cor}(k). \tag{7.4}$$

For estimating $\hat{\beta}_g$, the particular value of the constant K is of no interest, but it is necessary in the estimation since we are assuming that the decay of covariance is proportional to $n^{-\beta}$. The estimators $\hat{d}_{can}$ and $\hat{d}_{cor}$ are the canonical and the bias corrected estimators, respectively, introduced in Subsection 6.3 which are transformed in order to reflect a decay proportional to $n^{-\beta}$. In this case we are tacitly assuming an ARFIMA$(0, d, 0)$ model with exponentially distributed marginals.



(a) (b) (c)

(d) (e)

Figure 7.3: Plot of $\{(y_i, y_{i+s})\}_{i=1}^{2979-s}$ for (a) $s = 1$; (b) $s = 5$; (c) $s = 25$; (d) $s = 50$ and (e) $s = 100$.

Since in the literature of ARFIMA process, the covariance decay is taken proportional to n^{2d-1}, for $d \in (0, 0.5)$, and here we are studying a decay proportional to $n^{-\beta}$, for $\beta \in (0, 1)$,

when appropriate, we shall present the results in both scales. Notice that β and d are related by $\beta = 1 - 2d$.

Figure 7.4(b) presents the sequences $\hat{\beta}_g(k)$, $\hat{\beta}_{can}(k)$ and $\hat{\beta}_{cor}(k)$, given respectively by (7.1), (7.2) and (7.3), for $k = 1, \cdots, 400$. We notice that for small lags, all estimators vary rapidly, but become smooth after 50 lags or so. Figure 7.4(c) presents the estimators $\hat{\beta}_g$, $\hat{\beta}_{can}$ and $\hat{\beta}_{cor}$, given by (7.4), for $m = 1, \cdots, 400$. In Figure 7.4(b) and Figure 7.4(c), the horizontal lines represent the interval $[0.2, 0.4]$.

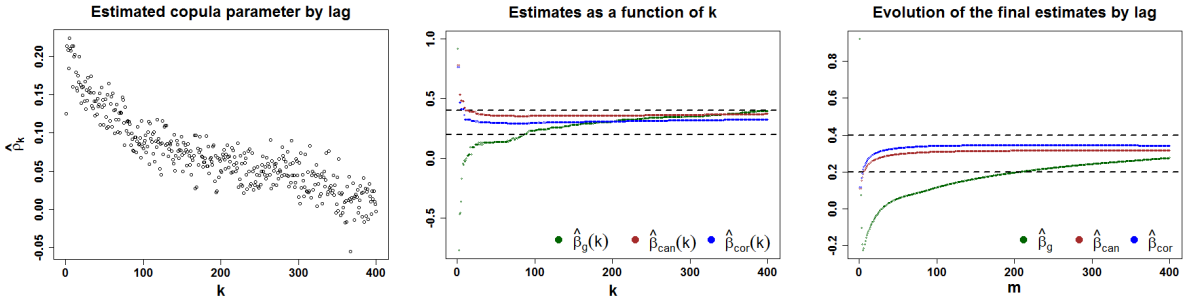


Figure 7.4: (a) Copula parameter estimated values; (b) Estimated $\hat{\beta}_g(k)$, $\hat{\beta}_{can}(k)$ and $\hat{\beta}_{cor}(k)$ given by (7.1), (7.2) and (7.3), respectively, for $k = 1, \cdots, 400$; (c) Evolution of the estimators $\hat{\beta}_g$, $\hat{\beta}_{can}$ and $\hat{\beta}_{cor}$ given by (7.4), for $m = 1, \cdots, 400$.

Table 7.1 presents the estimated value $\hat{\beta}_g$, $\hat{\beta}_{can}$ and $\hat{\beta}_{cor}$ for lags 5, 10, 25, 50, 100 and 400. The estimated value of $\hat{\beta}_g$ takes long to stabilize and, for small lags, this estimate is clearly off. This behavior is due to the flexibility in performing a two parameter optimization, which can lead to extremely bad estimates for small number of observations. However, as the lag increases, $\hat{\beta}_g$ stabilizes and takes values close to the other estimators.

Table 7.1: Estimated $\hat{\beta}_g$, $\hat{\beta}_{can}$ and $\hat{\beta}_{cor}$ values for lag $m \in \{5, 10, 25, 50, 100, 400\}$ for the absolute return time series of the S&P500 index. Also presented are $\hat{d}_{can}$ and $\hat{d}_{cor}$ (in parenthesis).

Estimator	Lag m					
	5	10	25	50	100	400
$\hat{\beta}_g$	-0.2266	-0.1442	-0.0200	0.0549	0.1151	0.2755
$\hat{\beta}_{can}$	0.6216	0.5397	0.4509	0.4081	0.3827	0.3682
($\hat{d}_{can}$)	(0.1892)	(0.2301)	(0.2746)	(0.2959)	(0.3087)	(0.3159)
$\hat{\beta}_{cor}$	0.5747	0.4798	0.3833	0.3424	0.3182	0.3152
($\hat{d}_{cor}$)	(0.2127)	(0.2601)	(0.3083)	(0.3288)	(0.3409)	(0.3424)

Comparing the values obtained for $\hat{d}_{can}$ and $\hat{d}_{cor}$ in Table 7.1 (in parenthesis) and the estimated values in Table 6.2, we expect considerable bias in $\hat{d}_{can}$. The most accurate estimate for $\hat{d}_{cor}$, according to Subsection 6.3, is obtained for $m = 25$, that is, $\hat{d}_{cor} = 0.3083$ with 95% confidence interval $(0.3038, 0.3140)$ which gives $\hat{\beta}_{cor} = 0.3834$ with 95% confidence interval $(0.3720, 0.3924)$. The quantiles and standard deviation necessary to obtain the confidence interval were obtained by using the so-called stationary bootstrap method (Politis and Romano, 1993) with random blocks of length geometrically distributed with mean 1,000. We perform 1,000 bootstrap replicates (see also Davison and Hinkley, 1997). Consistently, considering the value of $\hat{d}_{can}$ obtained, an inverse search on the results for the canonical estimator in Table 6.2

places d near 0.30, or equivalently, β close to 0.40.

8 Conclusions and Final Remarks

In this work, we study the problem of constructing stochastic processes with a predetermined decay of covariance by parameterizing a family of copulas and the process' marginals. Although the main interest in practice lies on stationary processes, the theory proposed here covers both stationary and non-stationary processes and allow for arbitrary decay of covariance. We present several examples to illustrate the theory, including the widely applied Euclidean, Extreme Value and Gaussian family of copulas. We show how the theory blends with some common applied time series models such as the large class of ARFIMA processes.

To show the usefulness of the theory, we apply it to the problem of estimation in time series models. We develop a general methodology of parameter estimation identifiable through the covariance decay of the process based on the theoretical results obtained in the paper. Examples are provided for better understanding the technique.

To assess the methodology's finite sample performance, we perform some Monte Carlo simulations. The simulations reinforce the competitiveness of the methodology, as well as illustrate its use. Also motivated by the theory developed in the paper, we present an application to simulation of time series with given decay of covariance and marginals by parameterizing the bidimensional copulas of the process. The method allows to obtain realizations of non-standard time series in a fast and relatively simple way.

We finally show how to apply the methodology by analyzing the S&P500 stock market index time series. We use the methodology to study the covariance decay of the absolute (and for consequence, of the squared) S&P500 return time series. We conclude that the methodology yield results coherent with the stylized facts of financial time series.

Acknowledgements

G. Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil. S.R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1 and also by INCT *em Matemática*. The authors are grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

References

[1] Bingham, N.H., Goldie, C.M and Teugels, J.L. (1987). *Regular Variation*. Cambridge: Cambridge University Press.

[2] Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, Vol. **31**, 307-327.

[3] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag. 2nd edition.

[4] Capéraà, P., Fougères, A-L. and Genest, C. (2000). "Bivariate Distributions with Given Extreme Value Atractor". *Journal of Multivariate Analysis*, Vol. **72**, 30-49.

[5] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester: Wiley.

[6] Cont, R. (2001). "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues". *Quantitative Finance*, **1**, 223-236.

[7] Darsow, W.F., Nguyen, B. and Olsen, E.T. (1992). "Copulas and Markov Processes". *Illinois Journal of Mathematics*, **36**(4), 600-642.

[8] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

[9] Dolati, A. and Úbeda-Flores, M. (2005). "A Method for Constructing Multivariate Distributions with Given Bivariate Margins". *Brazilian Journal of Probability and Statistics*, **19**, 85-92.

[10] Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of Variance of U.K. Inflation". *Econometrica*, Vol. **50**, 987-1008.

[11] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series and Products*. New York: Academic Press, 6th edition.

[12] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

[13] Lagerås, A.N (2010). "Copulas for Markovian Dependence". *Bernoulli*, **16**(2), 331-342.

[14] Lee, T-H. and Long, X. (2009). "Copula-based Multivariate GARCH Model with Uncorrelated Dependent Errors". *Journal of Econometrics*, **150**(2), 207-218.

[15] Lopes, S.R.C. (2008). "Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting". In Sidoravicius, V. and Vares, M.E. (editors), *In and out of equilibrium 2*. Basel: Birkhäuser, 497-525.

[16] Lopes, S. R. C., Olbermann, B. P. and Reisen, V. A. (2004). "A Comparison of Estimation Methods in Non-stationary Arfima Processes". *Journal of Statistical Computation and Simulation*, **74**(5), 339-347.

[17] Lopes, S.R.C., Pumi, G. and Zaniol, K. (2011). "Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes". *Communication in Statistics: Simulation and Computations*. Forthcomming. Preprint available at http://mat.ufrgs.br/∼slopes/selected_publications.htm.

[18] Mari, D.D. and Kotz, S. (2001). *Correlation and Dependence*. London: Imperial College Press.

[19] Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer Verlag. 2nd edition.

[20] Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. New Jersey: John Wiley.

[21] Palynchuk, B.A. and Guo, Y. (2011). "A Probabilistic Description of Rain Storms Incorporating Peak Intensities". *Journal of Hydrology*, **409**(1-2), 71-80.

[22] Phillips, P.C.B. and Perron, P. (1988). "Testing for a Unit Root in Time Series Regression". *Biometrika*, **75**(2), 335-346.

[23] Pickands, J. (1981). "Multivariate Extreme Value Distribution". Proceedings of the 43rd Session ISI, 859-878.

[24] Politis, D.N. and Romano, J.P. (1993). "The Stationary Bootstrap". *Journal of the American Statistical Association*, **89**, 1303-1313.

[25] Taqqu, M.S. (2003). "Fractional Brownian Motion and Long-Range Dependence". In Doukhan, P., Oppenheim, G. and Taqqu, M.S. (editors), *Theory and Applications of Long-Range Dependence*. Boston: Birkhäuser, 5-38.