

MAURÍCIO SILVA LACERDA

**ENSAIOS SOBRE CÓPULAS: APLICAÇÃO AO AGRONEGÓCIO E SELEÇÃO
DE MODELOS VIA AIC E BIC**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

Orientador: Paulo César Emiliano

Coorientadores: Julio Cesar A. Silva Junior
Luiz Alexandre Peternelli

**VIÇOSA - MINAS GERAIS
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

L131e
2022

Lacerda, Maurício Silva, 1991-
Ensaio sobre cópulas: aplicação ao agronegócio e seleção
de modelos via AIC e BIC / Maurício Silva Lacerda. – Viçosa,
MG, 2022.

1 tese eletrônica (77 f.): il. (algumas color.).

Orientador: Paulo César Emiliano.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Estatística, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.496>

Modo de acesso: World Wide Web.

1. Cópulas (Estatística matemática). 2. Açúcar - Preços.
I. Emiliano, Paulo César, 1982-. II. Universidade Federal de
Viçosa. Departamento de Estatística. Programa de
Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 519.535

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

MAURÍCIO SILVA LACERDA

ENSAIOS SOBRE CÓPULAS: APLICAÇÃO AO AGRONEGÓCIO E SELEÇÃO
DE MODELOS VIA AIC E BIC

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

APROVADA: 03 de junho de 2022.

Assentimento:

Maurício S. Lacerda

Maurício Silva Lacerda
Autor

Paulo César Emiliano

Paulo César Emiliano
Orientador

"Onde há problemas, há vida."
Aleksandr A. Zinovyev

Resumo

LACERDA, Maurício Silva, D.Sc., Universidade Federal de Viçosa, junho de 2022. **Ensaio sobre cópulas: Aplicação ao agronegócio e seleção de modelos via AIC e BIC.** Orientador: Paulo César Emiliano. Coorientadores: Julio Cesar Araujo Silva Junior e Luiz Alexandre Peternelli.

A compreensão acerca da dependência entre duas ou mais séries históricas é de grande importância para pesquisadores, produtores rurais, investidores e demais envolvidos nas séries de forma direta e indireta. Os modelos que vem ganhando destaque na mensuração da forma e da estrutura dessa dependência tem sido os diversos modelos de cópulas presentes na literatura, em especial as cópulas: gaussiana, t de Student, Clayton, Frank, Gumbel e Joe. Esta tese é composta por dois trabalhos, um envolvendo aplicações de modelos de cópulas em conjunto de dados reais e o outro metodológico. No primeiro analisou-se os impactos causados pela expansão da cana-de-açúcar no preço da soja e boi gordo para os estados de São Paulo, Goiás e Mato Grosso do Sul. O objetivo deste artigo foi analisar a associação dos movimentos conjuntos dos retornos dos preços regionais da cana-de-açúcar e das culturas concorrentes no uso da terra, verificando se a implantação da cana-de-açúcar vem gerando impacto nos preços dos alimentos cultivados nessa região. No segundo artigo analisou-se a precisão dos métodos de seleção de cópulas (AIC e BIC) por meio de simulações. O objetivo deste segundo trabalho foi verificar os possíveis erros que se incorre ao escolher um modelo de cópulas para um conjunto de dados reais, sobretudo para amostras pequenas, com menos de 200 observações, por exemplo. Os resultados encontrados no primeiro trabalho detectou-se dependência fraca entre as séries históricas das commodities analisadas, com destaque para a maior dependência encontrada, sendo 0,13 para o par Cana x Boi para o Estado de Goiás, indicando que a expansão da cana-de-açúcar para esses estados gerou baixo impacto nos preços das demais culturas tradicionais na região. Para o segundo artigo verificou-se taxas de acerto elevadas, acima de 90%, ainda para amostras pequenas ($n < 150$), quando as séries possuem medidas de dependência maiores ($\tau = 0,5$ e $0,75$). Ao passo que, para a medida de dependência menor ($\tau = 0,25$) foram necessários tamanhos amostrais maiores ($n > 200$), em alguns casos ($n > 400$), para se obter altas taxas de acertos, acima de 90%, por exemplo.

Palavras-chave: Setor sucroalcooleiro. Seleção de modelos. Tamanho amostral.

Abstract

LACERDA, Maurício Silva, D.Sc., Universidade Federal de Viçosa, June, 2022. **Studies on copulas: Application to agribusiness and selection of models via AIC and BIC.** Advisor: Paulo César Emiliano. Co-advisors: Julio Cesar Araujo Silva Junior and Luiz Alexandre Peternelli.

Understanding the dependence between two or more historical series is of great importance for researchers, rural producers, investors and others directly and indirectly involved in the series. The models that have gained prominence in the measurement of the form and structure of this dependence have been the various models of copulas present in the literature, especially the copulas: Gaussian, Student's t , Clayton, Frank, Gumbel and Joe. This thesis is composed of two works, one involving applications of copula models in real data sets and the other methodological. The first analyzed the impacts caused by the expansion of sugarcane on the price of soybeans and live cattle for the states of São Paulo, Goiás and Mato Grosso do Sul. The objective of this article was to analyze the association of the joint movements of the returns of the regional prices of the sugarcane and of the competing cultures in the use of the land, verifying if the implantation of the sugarcane has been generating impact in the prices of the foods cultivated in that area. region. In the second article, the accuracy of the copulas selection methods (AIC and BIC) was analyzed through simulations. The objective of this second work was to verify the possible errors that are incurred when choosing a copula model for a set of real data, especially for small samples, with less than 200 observations, for example. The results found in the first work detected a weak dependence between the historical series of the analyzed commodities, with emphasis on the greater dependence found, being 0.13 for the Cana x Boi pair for the State of Goiás, indicating that the expansion of sugarcane of sugar for these states had a low impact on the prices of other traditional crops in the region. For the second article, high hit rates were found, above 90%, even for small samples ($n < 150$), when the series have larger dependence measures ($\tau = 0.5$ and $0,75$). Whereas, for the smaller dependence measure ($\tau = 0.25$) larger sample sizes ($n > 200$), in some cases ($n > 400$) were needed to obtain high rates of hits, above 90%, for example.

Keywords: Sugar and alcohol sector. Selection of models. Sample size.

Lista de Figuras

1.4.1 Gráficos dos preços do boi, soja e cana-de-açúcar para São Paulo.	25
1.4.2 Gráficos dos preços do boi e soja para Goiás e cana-de-açúcar para São Paulo.	26
1.4.3 Gráficos dos preços do boi e soja para Mato Grosso do Sul e cana-de-açúcar para São Paulo.	27
1.4.4 Gráfico do log retorno das séries para os estados de Goiás, Mato Grosso do Sul e São Paulo.	27
2.4.1 Gráfico com taxas de erros e acertos da cópula gaussiana com diferentes medidas de dependência e tamanho amostral.	53
2.4.2 Gráfico com taxas de erros e acertos da cópula t de Student com diferentes medidas de dependência e tamanho amostral.	54
2.4.3 Gráfico com taxas de erros e acertos da cópula de Clayton com diferentes medidas de dependência e tamanho amostral.	55
2.4.4 Gráfico com taxas de erros e acertos da cópula de Gumbel com diferentes medidas de dependência e tamanho amostral.	56
2.4.5 Gráfico com taxas de erros e acertos da cópula de Frank com diferentes medidas de dependência e tamanho amostral.	57
2.4.6 Gráfico com taxas de erros e acertos da cópula de Joe com diferentes medidas de dependência e tamanho amostral.	58

Lista de Tabelas

1.4.1 Estatística descritiva dos log retornos dos preços da soja, boi e cana.	26
1.4.2 Resultados do ajuste dos modelos (ARMA-GARCH) para os log retornos.	28
1.4.3 Resultados da seleção das cópulas bivariadas.	30
1.4.4 Teste para qualidade do ajuste das cópulas bivariadas.	31
1.4.5 Resultado da seleção das cópulas trivariadas.	32
2.3.1 Taxas de VP, FN, FP e VN.	49
2.3.2 Características de estudos correlatos.	51
2.4.1 Artigos aplicados utilizando cópulas e tamanhos de amostras.	51
2.4.2 Sensibilidade e especificidade referente à cópula gaussiana.	59
2.4.3 Sensibilidade e especificidade referente à cópula t de Student.	60
2.4.4 Sensibilidade e especificidade referente à cópula de Clayton.	60
2.4.5 Sensibilidade e especificidade referente à cópula de Gumbel.	61
2.4.6 Sensibilidade e especificidade referente à cópula de Frank.	61
2.4.7 Sensibilidade e especificidade referente à cópula de Joe.	62
2.4.8 Medidas de EQM das famílias de cópulas estudadas AIC	62
2.4.9 Medidas de EQM das famílias de cópulas estudadas BIC	63

Sumário

Introdução geral	11
1 Dependência entre os preços regionais da cana-de-açúcar, boi e soja: uma investigação via cópulas	13
1.1 Introdução	14
1.2 Energia e alimento	18
1.3 Estratégias metodológicas dados e modelos	20
1.3.1 Dados	20
1.3.2 Modelo para distribuições marginais	21
1.3.3 Modelos de cópulas	22
1.3.4 Medidas de dependência	24
1.4 Resultados	25
1.4.1 Cópulas bivariadas	29
1.4.2 Cópulas trivariadas	32
1.5 Considerações finais	33
2 Análise de critérios de seleção para cópulas: limitações e melhores cenários	41
2.1 Introdução	42
2.1.1 Seleção das cópulas	43
2.2 Cópulas	44
2.2.1 Modelos de cópulas	45
2.2.2 Medidas de dependência	47
2.3 Critérios de informação	48
2.3.1 Sensibilidade e especificidade do critério	48
2.3.2 Estudos relacionados	50
2.4 Simulação e resultados	50
2.4.1 Parâmetros e cenários	50
2.4.2 Resultados e discussão	52
2.5 Considerações finais	63
Conclusão geral	68

Introdução geral

Em diversas áreas do estudo científico, pesquisadores estão interessados em compreender fenômenos relacionados ao comportamento conjunto de duas ou mais variáveis. Exemplos disso são encontrados nas áreas das finanças, da economia, da saúde, do turismo, entre outras. Compreender a relação, ou comovimentos, entre duas ou mais variáveis de interesse passa pela estimação de medidas estatísticas de associação ou ranking entre as variáveis.

Medidas comuns na literatura, e que medem o comportamento linear e simétrico entre as variáveis são os coeficientes de correlação, os quais apresentam limitações. Segundo Embrechets et al. (2002) o coeficiente de correlação de Pearson, por exemplo, mede a correlação linear entre duas variáveis captando da mesma forma os grandes e os pequenos movimentos de uma série, que se caracteriza como uma delimitação.

Diante das limitações encontradas nos coeficientes de correlação, as cópulas vêm se mostrando uma boa alternativa quando o interesse é medir estruturas de dependência. Em certas condições, os diferentes modelos de cópulas podem fornecer informações mais adequadas, quando comparadas a alguns métodos concorrentes. Conforme Nelsen (2006) por meio das cópulas, é possível modelar diferentes tipos de comportamento entre duas, ou mais, variáveis, como dependências simétricas e assimétricas, dependências de caudas positivas ou negativas, além de obter distribuição conjunta.

Dadas as vantagens elencadas, quando verificados problemas na área de ciências agrárias, notou-se uma lacuna nos estudos empíricos que poderiam ser melhor analisados por meio da metodologia de cópulas. Nesse sentido, desenvolveu-se o primeiro artigo desta tese (que consiste no capítulo 1). Nele foi analisada a estrutura de dependência entre os log-retornos dos preços do boi gordo e da soja confrontados com o preço da cana-de-açúcar, que é uma cultura que vem se expandindo para Estados de Goiás, Mato Grosso do Sul e São Paulo, que antes da chegada da cana-de-açúcar tinham a soja e o boi gordo como produções predominantes (Granco et al., 2017, Caldas et al., 2017).

Para o primeiro ensaio, os Estados escolhidos foram: São Paulo (SP), Goiás (GO) e Mato Grosso do Sul (MS), pelo fato deles terem se tornado os maiores produtores de cana-de-açúcar do país (Conab, 2019). Na aplicação foram analisadas as cópulas gaussiana, t de Student, Clayton, Gumbel e Joe-Clayton, que foram selecionadas para

a análise a partir dos resultados do AIC - critério de informação de Akaike. Além disso, foi aplicado o teste *goodness-of-fit*, o qual se mostrou bastante conservador, não rejeitando nenhuma cópula selecionada.

Ao realizar a seleção dos modelos de cópulas, que foram utilizados no primeiro artigo, por meio do critério de AIC, notou-se uma dificuldade em verificar se os modelos escolhidos estavam bem ajustados, em especial para amostras de tamanhos pequenos, abaixo de 200 observações, e para medidas de dependências baixa, $\tau < 0,25$. Com isso foi identificada uma lacuna para a proposição do segundo artigo da tese (que consiste do capítulo 2). Nele é dado enfoque nas taxas de acerto e de erro que se comete ao selecionar uma cópula por meio dos critérios de informação de Akaike (AIC) e Bayesiano (BIC) em diferentes cenários, com diferentes tamanhos amostrais e medidas de dependência entre as séries.

Para organizar as ideias da tese apresentam-se além dessa introdução mais dois capítulos e uma conclusão geral dos trabalhos. O capítulo 1 consiste do primeiro artigo, intitulado: Dependência entre os preços regionais da cana-de-açúcar, boi e soja: uma investigação via cópulas. O capítulo 2 contém o segundo artigo, intitulado: Análise do critério de informação Akaike para cópulas: limitações e melhores cenários.

Referências Bibliográficas

- Caldas, M. M., Granco, G., Bishop, C., Kastens, J. e Brown, J. C. (2017). Effects of sugarcane ethanol expansion in the brazilian cerrado: Land use response in the new frontier. Bioenergy and Land Use Change, 231:39.
- Conab - Companhia Nacional de Abastecimento (2019). Acompanhamento da safra brasileira: Cana-de-açúcar. 6(3). Safra 19/20. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em 15 de agosto de 2020. Relatório técnico.
- Embrechts, P., McNeil, A. e Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond, 1:176–223.
- Granco, G., Caldas, M. M., Bergtold, J. S. e Sant'Anna, A. C. (2017). Exploring the policy and social factors fueling the expansion and shift of sugarcane production in the brazilian cerrado. GeoJournal, 82(1):63–80.
- Nelsen, R. B. (2006). An introduction to copulas. Springer, New York.

Dependência entre os preços regionais da cana-de-açúcar, boi e soja: uma investigação via cópulas

Resumo: A compreensão sobre as relações existentes entre os mercados de alimentos e de energia ainda é alvo de investigação e tema não esgotado na literatura. Estudos sobre a dependência entre esses mercados é fundamental para formuladores de políticas públicas, investidores e produtores. Nesse contexto, o aumento de investimentos ocorridos nos últimos anos no setor sucroalcooleiro brasileiro pode ter ocasionado movimentos conjuntos entre os preços regionais da cana-de-açúcar e outras culturas substitutas no uso da terra. O objetivo deste artigo é analisar a associação dos movimentos dos retornos dos preços regionais da cana-de-açúcar e das culturas concorrentes no uso da terra nos estados de SP, MS e GO. Para tanto, utilizaram-se modelos de cópulas, bi e trivariadas, avaliando os comovimentos dos preços mensais da cana-de-açúcar, boi e soja de 2004 a 2019. Os resultados indicaram que a maior medida de dependência (0,13) foi encontrada para o par Cana x Boi no estado de Goiás. Os demais modelos apresentaram medidas de dependência menores e outras não significativas. Isso sugere que o efeito de expansão do plantio de cana do Sudeste para o Centro-Oeste não interferiu, de modo geral, significativamente nos preços dos produtos teoricamente concorrentes no uso das terras.

Palavras chaves: Cópula; Setor sucroalcooleiro brasileiro; Neutralidade; Cana-de-açúcar.

Abstract: The relationships between food and energy markets still remains a topic that is not exhausted in the Brazilian literature. Well-designed research on the dependency between these markets is essential for public policy makers, investors and producers. In recent years the increase in investments in the Brazilian sugar and alcohol sector may have caused joint movements between the regional prices of sugar cane and other

substitute crops in land use. The main goal of this article is to analyze the association between regional sugar cane prices and competing crops in land use, in the states of SP, MS and GO. Bivariate and trivariate Copulas were applied to model and evaluate the movements in the monthly prices of sugarcane, cattle and soy beans from 2004 to 2019. The results indicated that the highest dependence measure (0.13) was found for the pair sugarcane x cattle - GO. The other models presented smaller measures of dependence and others that were not significant. This suggests that the effect of expanding sugarcane plantation from the Southeast to the Midwest regions in Brazil, did not, in general, significantly affect the prices of products theoretically competing in land use.

Key words: Copulation; Sugar and alcohol sector Brazilian; Neutrality; Sugarcane.

1.1 Introdução

A produção da cana-de-açúcar no Brasil passou por um processo de expansão e mudanças, tais como na legislação, nos incentivos, na demanda e nas inovações tecnológicas, com destaque ao Estado de São Paulo (SP), o maior produtor nacional (Sant'Anna et al. 2016). Segundo Cherubin et al. (2015), o Brasil é o maior produtor de cana-de-açúcar do mundo, sendo responsável pela produção de um terço da safra global. A cana-de-açúcar é a terceira cultura do Brasil em termos de uso da terra, ficando atrás da produção de soja e do milho (Marin, 2016). No início do século XXI ocorreu uma expansão da produção desta commodity para o Centro-Oeste do país, em especial para os estados de Goiás (GO) e Mato Grosso do Sul (MS) (Granco et al., 2017, Caldas et al., 2017). Segundo Adami et al. (2012) o estado de SP tem sido há muito tempo o maior produtor de cana-de-açúcar, mas recentemente outros estados, como MS e GO aumentaram significativamente suas áreas de cana-de-açúcar.

Outros fatores também incentivaram a produção da cana-de-açúcar no Brasil, tais como: os preços elevados do petróleo por volta dos anos 2000; as crescentes críticas ao impacto dos combustíveis fósseis para o meio ambiente; o surgimento dos carros flex em 2003; as mudanças nas políticas públicas (pesquisas das instituições governamentais, assistência ao produtor, regulamentação ambiental, linhas de crédito e redução de impostos); as novas tecnologias para produção de etanol; concessão de crédito, entre outros fatores que contribuíram para o aumento da demanda e produção de biocombustíveis não somente no Brasil, mas numa esfera global (Goldemberg et al. (2014), Strassburg et al. (2014), Dhillon e Wuehlisch (2013) e Nuñez et al. (2013), Sant'Anna et al. (2016)).

O crescimento na produção da cana-de-açúcar no Brasil se deve, principalmente, à expansão territorial do plantio segundo Adami et al (2012). No entanto, essa ex-

pansão geográfica da produção de cana-de-açúcar gera demanda por terras e mão de obra, o que pode influenciar a produção de outras culturas na região (Lourenzani e Caldas, 2014). A área plantada na safra 2018/19 ultrapassou os 4,3 milhões, 950 e 660 mil hectares nos estados de SP, GO e MS respectivamente, o que, em consequência, possibilitou uma produção de mais de 340,8; 75,8 e 48,6 milhões de toneladas de cana-de-açúcar nos respectivos estados. Isso coloca esses estados como os maiores produtores de etanol do Brasil, país que está entre os maiores produtores do mundo (Conab, 2019).

Um importante motivador da produção são os incentivos do governo brasileiro, que vão desde oferecimento de empréstimos e redução nos impostos até controle da porcentagem de incremento de etanol na gasolina. Segundo Figueira (2020) a concessão de crédito está diretamente ligada à expansão territorial do plantio de cana-de-açúcar na região analisada. Além disso, existem políticas de controle do preço da gasolina, para manter o etanol competitivo em tempos de baixa no preço do petróleo internacional (Goldemberg, 2006; Granco et al., 2017). Por outro lado, as políticas de controle do preço da gasolina, utilizadas para controlar a inflação, também prejudicaram o consumo do etanol (Sant'Anna et al., 2016). Segundo Manoel et al. (2018) em períodos de petróleo em alta os subsídios do governo mantiveram o preço da gasolina baixo o que causou redução no caixa e até falência de algumas usinas do setor.

Em tempos que o preço internacional do petróleo estava em alta e a gasolina era subsidiada pelo governo, o preço do etanol precisava também ser mantido baixo, para seguir competitivo com a gasolina nas proporções de mercado. Isso, enquanto os custos de produção do etanol subiam 11,5% ao ano (Valdes et al., 2016), diminuindo a rentabilidade do setor, visto que a produção do etanol utiliza diversos derivados do petróleo desde as etapas do plantio da cana-de-açúcar até a moagem e transportes do mesmo (Sant'Anna et al., 2016).

A tecnologia é outro fator que incentivou a produção de etanol no Brasil e ela não se limita apenas aos carros flex, que foi o grande impulsionador da cana no início do século XXI. Como exemplos da evolução tecnológica que a produção atingiu podem-se mencionar: a chegada de novos utensílios e defensivos agrícolas, as novas espécies de cana-de-açúcar que melhor se adaptaram ao cerrado brasileiro, as novas máquinas utilizadas no plantio e na manutenção, por exemplo, além do maquinário empregado na colheita, que antes era muito criticada por se valer de empregos temporários em péssimas condições, passando atualmente para colheita mecanizada gerando empregos melhores, com salários mais altos e especializados segundo Defante et al. (2020). Mais detalhes acerca dos impactos causados com a redução de empregos na produção da cana-de-açúcar com o advento da tecnologia podem ser vistos em Gandolfi et al. (2019).

O mercado também apresentou novas formas de receita, como a produção de

fertilizantes a partir dos resíduos da produção do etanol, a possibilidade de vender os créditos de carbono e a produção de energia a partir da queima do bagaço (Neves et al., 2011). Segundo Trombeta e Caixeta Filho (2017) os subprodutos da produção de etanol eram descartados, no entanto atualmente têm se tornado potenciais matérias-primas à cogeração de energia elétrica e à produção de etanol de segunda geração, com isso a energia elétrica gerada a partir da queima das biomassas de cana-de-açúcar tem se mostrado importante na geração de receitas para as usinas. Araújo e Oliveira (2021) apresentam um estudo das possibilidades para tratar a vinhaça, um dos piores resíduos resultante da extração do etanol a partir do caldo da cana, que pode gerar biogás, eletricidade e crédito de carbono com suporte de biodigestores.

A preocupação com o meio ambiente é outro fator que gera impactos na produção do etanol, que podem ocorrer de forma direta e indireta. O etanol é menos poluente que a gasolina, fator que incentiva muitos consumidores a optarem por ele, principalmente quando ambos estão em mesma condição de rendimento, ou seja, o valor pago por quilômetro percorrido indefere entre álcool e gasolina. Sant'Anna et al. (2016) citam a regra dos 70%, sendo o etanol mais vantajoso quando seu preço é menor ou igual a 70% da gasolina. Tem-se ainda que a produção de etanol a partir da cana-de-açúcar emite menos gases que causam o efeito estufa se comparado ao etanol extraído a partir do milho, que é a principal fonte de produção de etanol dos Estados Unidos (Crago et al., 2010). Caldas et al. (2017) apontam a intensificação da produção bovina de forma positiva, visto que alguns estudos têm demonstrado que a conversão de pastagens degradadas em cana-de-açúcar pode ter um déficit de carbono negativo (Cohn et al. (2014); Mann et al. (2014); Graesser et al. (2015)).

A junção da tecnologia e a preocupação com o meio ambiente gera desafios, mas também produzem bons resultados na produção da cana-de-açúcar. Em um estudo que analisou a qualidade do ar após a implantação da colheita mecanizada, Saiani e Perosa (2016) encontraram resultados que sugerem uma relação negativa e significativa entre a mecanização e as internações por problemas respiratórios nas regiões produtoras, evidenciando o impacto positivo da mecanização da colheita sobre a saúde respiratória da população local, além de gerar emprego qualificado e melhor remunerado.

No entanto, outros fatores se mostram preocupantes a partir da crescente produção de etanol, como a segurança alimentar e o desmatamento. Com os incentivos que alavancaram a produção de etanol, a cana-de-açúcar vem ocupando o lugar da soja e de outros cultivos no uso da terra, o que causa uma competição por espaço para plantio. Segundo Hausman (2012) onde a produção de biocombustíveis compete com a cultura alimentar, pode ocorrer um aumento dos preços dos alimentos. Segundo Valdes et al. (2016) essa substituição de culturas impacta na oferta desses produtos, aumentando seus preços. Os incentivos que tornam a cana-de-açúcar mais interes-

sante que a soja e boi no estado de SP e principalmente nos estados de GO e MS precisam ser avaliados com cuidado. Eles podem gerar impactos nas ofertas de soja e carne bovina, que ao perderem espaço nos estados citados estão se deslocando para o Norte do país, causando desmatamento (Sant'Anna et al., 2016; Walker et al., 2009; Morton et al., 2006; McManus et al., 2016).

O Zoneamento Agroecológico da cana-de-açúcar (ZAE CANA), criado¹ em 2009 e extinto² em 2019, foi, por uma década, o responsável por direcionar a expansão da cana-de-açúcar no Brasil. Este mapeamento indicou quais regiões e Estados são mais adequados para produção da cana-de-açúcar levando em consideração diversos fatores, como o clima, o relevo, aspectos sociais e econômicos (Manzatto, 2009). O ZAE CANA indicou que os estados de SP, GO e MS possuem respectivamente 10,6; 12,5 e 10,8 milhões de hectares de áreas aptas para o plantio da cana-de-açúcar, colocando estes três estados dentre aqueles com mais terras propícias para tal cultivo, o que poderia justificar, em partes, uma expansão da produção nesses três Estados. No entanto, a produção de etanol, ao expandir para esses três estados pode estar tomando o lugar da soja e boi gordo, que são as commodities de maior produção desses Estados (Granco et al., 2017; Ferreira et al. (2013, 2016); Rodrigues e Miziara, 2008; Sano et al. (2010); Martinelle e Filoso, 2008).

O objetivo deste artigo é analisar o comportamento dos movimentos conjuntos dos preços da cana-de-açúcar com a soja e com o boi gordo nessas regiões em que pode haver competição pelo uso das terras, com a produção intensificada de uma nova commodity na região. O estudo regional permite identificar se a expansão da produção de cana-de-açúcar desloca as outras produções tradicionais da região, medido por variações nos preços. Para tanto, utiliza-se de cópulas, com distribuições marginais calculadas por meio de modelos ARMA-GARCH, e de dados mensais dos preços dos produtos de 2004 até 2019. Estudos regionais que medem estruturas de dependência via cópulas ainda não foram identificados na literatura para essa região e produtos, e é esta lacuna que se pretende preencher com este estudo.

Ao identificar se há neutralidade ou dependência entre os mercados de produção de energia e alimento, espera-se contribuir com informações que possam servir para um maior debate para a elaboração de políticas públicas do país, garantindo a produção energética (etanol) bem como a produção e segurança alimentar. Além disso, o estudo pode servir como base para orientar o produtor acerca dos movimentos conjuntos dos preços das variáveis investigadas, para que eles possam adotar estratégias de intensificação ou diversificação de suas culturas.

Quando se refere a políticas públicas e a produção agrícola no Brasil tem-se ainda um longo caminho a trilhar. Conforme Belik (2015) é necessária uma política de

¹Decreto 6961/2009 (Brasil, 2009)

²Decreto 10084/2019 (Brasil, 2019)

aumento de renda rural de forma sustentável em termos econômicos, sociais e ambientais, com enfoque na desigualdade que ainda há no meio rural. O reconhecimento do caráter heterogêneo do campo no Brasil e a adaptação de programas para as diferentes realidades é o caminho mais seguro para a promoção do desenvolvimento rural (Belik, 2015).

Além desta introdução, o artigo está estruturado em mais quatro seções. Na segunda seção tem-se uma revisão acerca de energia e alimento, na terceira tem-se uma descrição dos dados e das estratégias metodológicas com breve apresentação de seus modelos, na quarta apresenta-se os resultados. Por fim, são apresentadas as conclusões na quinta seção.

1.2 Energia e alimento

Estudos que investigam diferentes fontes de energia e os comparam com a produção agrícola, de biocombustíveis e alimento, são recentes e estão em destaque na literatura, como por exemplo, Kaltalioglu e Soytaş (2009), Chen et al. (2010), Mueller et al. (2011), Mutuc et al. (2011) e Nazlioglu e Soytaş (2011), Goldemberg et al. (2014), Lourenzani e Caldas (2014), Gilio e Moraes (2016), entre outros. As razões para o crescimento de investigações sobre o tema estão relacionadas a fatores como a necessidade de obtenção de combustíveis menos poluentes e de menores custos relativos de produção e operação do que os de origem fóssil, de modo a garantir a produção de energia e manter a segurança alimentar.

A expansão da cana-de-açúcar para os estados de SP, GO e MS é um desafio, principalmente no centro-oeste (Granco et al., 2018b), pois além do fato desses estados serem tradicionais produtores de alimentos, soja e boi, por exemplo³, a construção de uma usina demanda tempo e investimento. Segundo Manoel et al. (2018), uma usina com capacidade de esmagamento de 3 milhões de toneladas de cana-de-açúcar por ano exige um investimento inicial de cerca de US\$ 400 milhões e a geração de caixa começa apenas dois anos após o investimento. Mesmo assim, segundo Granco et al. (2018b), surgiram 80 usinas do setor sucroalcooleiro no Brasil entre 2006 e 2012, a maioria delas nos estados de SP, GO e MS. Isso faz com que essa região esteja mais voltada para a produção de cana-de-açúcar, de modo a suprir a demanda das usinas.

Conforme Conab (2019) esses três Estados, na safra de 2019/2020, representam juntos 70% da produção de etanol do país. Além dos altos investimentos a produção de cana-de-açúcar requer rápido processamento, fazendo com que a usina dependa de matéria-prima por perto. Capaz et al. (2013) e Neves et al. (1998) recomendam que as usinas devem ser situadas entre 45 e 50 km da produção, logo a expansão

³Ver Rada (2013)

da produção causa substituição de outras culturas que eram tradicionais antes da chegada das usinas.

O cultivo da cana-de-açúcar no Brasil preocupa também pelo fato desta produção não apresentar ganhos em produtividade, pelo contrário, conforme Nyko et al. (2013) a produtividade por hectare vem caindo. Bernardo et al. (2019) mostram ainda que entre os anos de 1990 e 2014 a cana-de-açúcar teve ganhos de produtividade inferiores em comparação com boi, milho e soja, no mesmo período. Para compensar essa queda de produtividade e suprir a demanda por matéria-prima, a expansão territorial acabou sendo a alternativa possível no curto prazo, gerando competição por terras. Os resultados apresentados por Figueira (2020) não constataram impacto do crédito rural sobre ganho em produtividade da cana-de-açúcar, encontrando relação apenas com ampliação territorial da produção.

É evidente a mudança na produção dos estados de SP, GO e MS após a chegada das usinas do setor sucroalcooleiro. Segundo Adami et al. (2012), da área total de cana-de-açúcar expandida no Estado de São Paulo desde 2005 e no restante da região Centro-Sul desde 2007, 69,7% eram pastagem e 25,0% eram áreas destinadas à cultivos de safra anual. Nassar et al. (2008) utilizaram imagens de satélites para avaliar a mudança na produção com a expansão da cana-de-açúcar nos anos de safra 2006/07 e 2008/09 e observaram que 51% se expandiram em pastagens; 48% em culturas anuais ou perenes e 1% em floresta cultivada ou natural.

A preocupação com o fornecimento adequado de alimentos e fontes energéticas não é apenas do Brasil. Em um estudo feito nos Estados Unidos Katchova e Sant'Anna (2019) detectaram que nas cidades em que há usinas de etanol o preço do milho, principal matéria-prima do etanol no país estudado, é mais elevado que em cidades que não contém as usinas, o que acaba por motivar plantação de milho nessas regiões, causando competição por terras. Reboredo (2012) analisou movimentos conjuntos entre o preço do petróleo e o preço das commodities milho, soja e trigo no mercado mundial. Seus resultados indicaram neutralidade entre os mercados.

Pesquisas que tratam os dados para os mercados brasileiros, em específico para região produtora de cana-de-açúcar para fins energéticos, estão interessadas geralmente em analisar os motivos que levaram tal produção para essa região e os impactos por essa expansão causadas na produção alimentar da região, como em Adami et al. (2012), Hausman (2012), Caldas et al. (2017), Granco et al. (2018a). A estrutura de dependência entre esses mercados foi analisada por Resende e Cândido (2015), porém esses autores pesquisaram a estrutura de dependência entre o setor sucroalcooleiro dessa região com preços do petróleo, câmbio e mercado acionário, não se atentando ao mercado de alimento.

Esta breve revisão mostra a preocupação que se tem em relação a produção de biocombustíveis causando impactos na produção alimentar regional, nacional e até

mundial. No entanto, não foi encontrado estudos que fazem uma análise da estrutura de dependência entre as commodities produzidas nos estados de SP, GO e MS que é a região mais crítica do Brasil, quando o assunto é competição de terras para produção alimentar e energética. Espera-se com este estudo contribuir para preencher essa lacuna.

1.3 Estratégias metodológicas dados e modelos

Para calcular a dependência entre duas ou mais variáveis aleatórias é necessário que se tenha conhecimento de sua distribuição conjunta, trabalho que se tornou mais facilitado com a utilização de modelos de cópulas, que será mais explorado adiante. Seu uso se mostra vantajoso, por ser uma metodologia eficiente para se obter distribuições conjuntas e, conseqüentemente, extrair medidas de dependência.

As cópulas podem ser estimadas a partir de processos que envolvem duas etapas. Na primeira, a determinação das distribuições marginais, e em seguida a estimação da cópula Nelsen (2006). Os modelos utilizados para a determinação das marginais, dos log retornos, foram os autorregressivos de médias móveis (ARMA) com heteroscedasticidade condicional autorregressiva generalizada (GARCH), com erros seguindo uma distribuição normal padrão ou t -Student assimétrica, escolhidos de acordo com sua capacidade de captar melhor o padrão serial.

Assim, adotaram-se as etapas de estimação das cópulas: determinação das distribuições marginais, a partir de modelos $ARMA(m, n) - GARCH(p, q)$ e a estimação das cópulas Gaussiana, t -Student, de Clayton, de Gumbel e Joe-Clayton simetrizada. Todas foram posteriormente avaliadas pelo teste de bondade de ajuste para cópulas, mais detalhes acerca do teste podem ser obtidos em Genest et al. (2009).

1.3.1 Dados

As séries históricas utilizadas neste estudo são o preço mensal do boi gordo e da soja, disponível no IPEADATA (2019), para os estados de SP, GO e MS, que são comparadas com o preço da cana-de-açúcar, disponível na UDOP (2019), no mesmo período, de março de 2004 a setembro de 2019, totalizando 187 observações. Para o preço da cana-de-açúcar utilizou-se a série do preço para o estado de SP para comparação, visto que os outros Estados não possuem tais registros que datam da mesma época. Os dados sofreram uma transformação de diferença logarítmica para atender condições de estacionariedade, e nos modelos de distribuição marginais foram utilizados os resultados dos log retornos mensais das séries de preços das commodities, que também são atrativos para serem analisados por parte de investidores. Geralmente utiliza-se log retornos no lugar do preço por se tratar de uma série com proprieda-

des estatísticas desejáveis nas análises e também pelo interesse do estudo, uma vez que um analista geralmente está interessado no retorno histórico e não somente no preço em si, pois cada item analisado pode ter preços diferentes e apresentar retornos parecidos.

Optou-se por utilizar o preço pago pela cana-de-açúcar no campo por considerar que este seria o preço que mais se assemelha com o preço do boi gordo e da soja, pois o preço do etanol envolve outros fatores e custos de processamento. Se fosse considerar o preço do etanol, por exemplo, este valor acabaria sendo contaminado pelos custos de transportes e produção.

Utilizou-se o preço das determinadas commodities para comparação por acreditar que o preço pago ao produtor é o maior incentivador para o plantio de uma cultura e também por ser um fator de mais fácil mensuração quando se analisa três estados diferentes. Fatores como disponibilidade e custo de mão de obra também podem ser levados em consideração na hora da escolha para produzir determinada cultura, no entanto tais medidas são mais complexas e subjetivas.

1.3.2 Modelo para distribuições marginais

Quando o interesse é analisar e modelar uma série temporal, a metodologia de Box e Jenkins se mostra bastante atrativa. Inicialmente, os modelos autorregressivos $AR(m)$ foram criados para modelar uma observação Z_t no instante t da série dependente apenas de uma combinação linear de observações anteriores, conforme definido em Box et al. (2015):

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_m Z_{t-m} + a_t, \quad (1.1)$$

em que Z_t é a observação da série no instante t , μ é o intercepto, ϕ_i , com $i = 1, 2, \dots, m$, são os parâmetros de estrutura e a_t é o ruído branco do modelo.

Posteriormente, o modelo $ARMA(m, n)$ foi sugerido, com a observação Z_t da série no instante t dependente de uma combinação linear das observações passadas e do ruído branco. Este modelo também está definido em Box et al. (2015) como:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_m Z_{t-m} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_n a_{t-n}, \quad (1.2)$$

em que Z_t é a observação da série no instante t , μ é o intercepto, ϕ_i com $i = 1, 2, \dots, m$ e θ_j com $j = 1, 2, \dots, n$, são os parâmetros de estrutura e a_t é o ruído branco do modelo.

Além do modelo $ARMA(m, n)$ para captar movimentos da série, pode se fazer necessário a utilização de um outro modelo para ajustar a sua variância, no caso o

$GARCH(p, q)$, o que é bastante comum quando se trabalha com dados de retornos e log-retornos.

Neste caso os modelos $ARMA(m, n) - GARCH(p, q)$ foram utilizados como distribuição marginal no processo de estimação das cópulas e são definidos em Morettin e Toloi (2006) como:

$$r_t^b = \delta_0 + \sum_{i=1}^m \psi_i r_{t-i}^b + a_t - \sum_{j=1}^n \Theta_j a_{t-j}, \quad (1.3)$$

$$a_t = \sigma_t \epsilon_t, \quad (1.4)$$

$$\sigma_t^2 = \omega + \sum_{k=1}^p \alpha_k a_{t-k}^2 + \sum_{l=1}^q \beta_l \sigma_{t-l}^2, \quad (1.5)$$

em que r_t^b é o retorno do ativo b (boi, cana e soja) no período t , δ_0 é a média, a_t é o erro aleatório do modelo no instante t , ψ_i e Θ_j são os parâmetros do modelo ARMA; σ_t é o desvio padrão condicional ou volatilidade no instante t , $\omega > 0$ é a média do processo de volatilidade e α_k e $\beta_l > 0$ são os parâmetros do modelo GARCH, para $k = 1, \dots, p$ e $l = 1, \dots, q$ e $\epsilon_t \stackrel{iid}{\sim} D(\cdot)$ com $t = 1, \dots, T$; $D(\cdot)$ representa uma função densidade de probabilidade (f.d.p.) qualquer e tem-se ainda que ϵ_t são observações não correlacionadas, também chamado de ruído branco, ou seja, erros não correlacionados.

1.3.3 Modelos de cópulas

De acordo com Nelsen (2006), cópulas são funções de distribuição multivariadas, cujas marginais são uniformes no intervalo $[0,1]$, que unem ou copulam funções de distribuições acumuladas às suas funções de distribuição conjunta. O teorema de Sklar (1959) liga a definição de cópula com uma função de distribuição n -dimensional e suas distribuições marginais⁴. De acordo com esse teorema, para variáveis aleatórias Y_1, \dots, Y_n com marginais F_1, \dots, F_n e F a função de distribuição conjunta, existe uma n -cópula C tal que para todo $\mathbf{y} \in \mathbb{R}$:

$$F(\mathbf{y}) = C(F_1(y_1), \dots, F_n(y_n)). \quad (1.6)$$

Se as marginais F_1, \dots, F_n são contínuas, C é unicamente definida. Se C é uma n -cópula, então a função F definida em (1.6) é uma distribuição conjunta com n marginais F_1, \dots, F_n .

Considerando as funções de cópulas que captam diferentes padrões de dependência entre cauda, pode-se obter as cópulas pelo método da inversão, que substitui os argumentos da distribuição conjunta pelas funções quantílicas marginais para obter

⁴Ver Nelsen (2006) para informações sobre C-volume e detalhes sobre cópulas.

os efeitos da distribuição marginal na dependência (Reboredo, 2012):

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad (1.7)$$

em que $F_1^{-1}, \dots, F_n^{-1}$ são as funções quantílicas marginais de u_1, \dots, u_n .

Os parâmetros de uma cópula podem ser estimados a partir de formas funcionais paramétricas, de modo geral eles expressam medidas de associação entre duas variáveis aleatórias. A estimação das cópulas, a partir das marginais, por meio da função *fitCopula* do pacote *Copula* (Hoeft et al., 2020), do R, pode ser via máxima verossimilhança ou pseudo máxima verossimilhança. Este segundo método é utilizado nos casos em que a matriz hessiana, utilizada na maximização da função log verossimilhança, não é invertível⁵. Alguns dos modelos de cópulas mais vistos em trabalhos da área são: gaussiana, *t* de Student, Gumbel, Clayton e Joe-Clayton simetrizada (SJC), por isso estes foram os escolhidos para este trabalho.

Dadas duas marginais com distribuição uniforme no intervalo $(0, 1)$ u e v , pode-se definir algumas cópulas bivariadas conforme em Nelsen (2006): Cópula gaussiana (1.8), *t* de Student (1.9), Clayton (1.10), Gumbel (1.11) e Joe-Clayton simetrizada (1.12):

$$C_G(u, v; \theta) = \Phi\left(\phi^{-1}(u), \phi^{-1}(v)\right) \quad (1.8)$$

$$C_T(u, v; \theta, \kappa) = T\left(t^{-1}(u), t^{-1}(v)\right), \quad (1.9)$$

$$C_C(u, v; \theta) = \max\left((u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, 0\right) \quad (1.10)$$

$$C_G(u, v; \theta) = \exp\left(-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{\frac{1}{\theta}}\right), \quad (1.11)$$

$$C_{SJC}(u, v; \lambda_U, \lambda_L) = 0,5(C_{JC}(u, v; \lambda_U, \lambda_L) + C_{JC}(1 - u, 1 - v; \lambda_U, \lambda_L) + u + v - 1), \quad (1.12)$$

em que θ é a correlação linear de Spearman entre u e v , κ é o grau de liberdade da distribuição *t*, λ_U e λ_L são as medidas de dependência de cauda inferior e superior, respectivamente, Φ é a acumulada da distribuição normal padrão bivariada e ϕ é a função densidade de probabilidade da normal padrão, T é a distribuição *t* de Student acumulada bivariada, t é a função densidade de probabilidade da distribuição *t* de Student e C_{JC} é a cópula de Joe-Clayton⁶. Estas e outras funções de cópulas podem ser encontradas com mais detalhes em Nelsen (2006) e Patton (2006) assim como suas extensões até n dimensões.

⁵ver Genest et al. (1995) para mais detalhes.

⁶ $C_{JC}(u, v; \kappa, \gamma) = 1 - \left(1 - \left\{[1 - (1 - u)^\kappa]^{-\gamma} + [1 - (1 - v)^\kappa]^{-\gamma} - 1\right\}^{-1/\gamma}\right)^{1/\kappa}$

A utilização destes modelos de cópulas (equações de (1.8) a (1.12)) em estudos acerca de dependências entre séries temporais são diversas. Conforme Nelsen (2006) as cópulas possuem como características a invariância a transformações crescentes e fornecem informações tanto das quantidades de dependência quanto da própria estrutura de dependência, indicando se ela é simétrica, assimétrica, de cauda superior e/ou inferior, além da facilidade em obter a distribuição conjunta entre duas ou mais variáveis aleatórias.

1.3.4 Medidas de dependência

Embrechts et al. (2002) afirmam que a medida tradicional de interdependência, o coeficiente de correlação de Pearson, indica a média dos desvios das médias dos retornos, não captando de forma distinta os grandes e pequenos movimentos do mercado. Com isso alguns autores (Reboredo (2012); Ning (2010); Aloui et al. (2011); Cholle et al. (2011); Patton (2006); Breymann et al. (2003)) tem utilizado estruturas de dependências que as famílias de cópulas permitem analisar, tais como dependência simétrica, assimétrica de caudas superior e inferior.

O τ de Kendall e ρ de Spearman são as medidas de dependência ou de associação mais conhecidas na literatura e são definidas em Nelsen (2006) como:

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (1.13)$$

$$\rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]), \quad (1.14)$$

em que (X_1, Y_1) , (X_2, Y_2) e (X_3, Y_3) são amostras independentes e identicamente distribuídas de um vetor (X, Y) de variáveis aleatórias.

No entanto, para se obter tais medidas é preciso de informações acerca da distribuição conjunta, o que nem sempre é possível ou fácil de se obter. Com isso as cópulas podem ser utilizadas para obter tais medidas, conforme em Nelsen (2006), da seguinte forma:

$$\tau = 4 \int C(u, v) dC(u, v) \quad (1.15)$$

$$\rho = 12 \int uv dC(u, v) - 3 \quad (1.16)$$

em que $C(u, v)$ é uma cópula, definida em (1.7).

A depender da natureza do problema, diferentes medidas de dependência de cópulas podem ser obtidas a partir das famílias escolhidas, inclusive uma representação de independência pode ser expressa. Por exemplo, para n variáveis aleatórias independentes são expressas em Nelsen (2006) por meio da seguinte representação:

$$C(\mathbf{u}) = \prod_{j=1}^n u_j \quad (1.17)$$

em que $\mathbf{u} = (u_1, \dots, u_n) \in [0, 1]^n$.

1.4 Resultados

As séries históricas dos preços de soja, cana-de-açúcar e boi gordo, para os estados de SP, GO e MS, apresentam valores crescentes ao longo do tempo, como pode ser observado nas Figuras 1.4.1, 1.4.2 e 1.4.3 em que se apresenta uma visão geral das informações analisadas, ao passo que na Figura 1.4.4 tem-se os gráficos dos log retornos das séries. É possível notar uma evolução parecida nos preços das três commodities, com exceção de alguns períodos. Porém, uma análise meramente visual não é suficiente para conclusões acerca de um movimento conjunto ou independente.

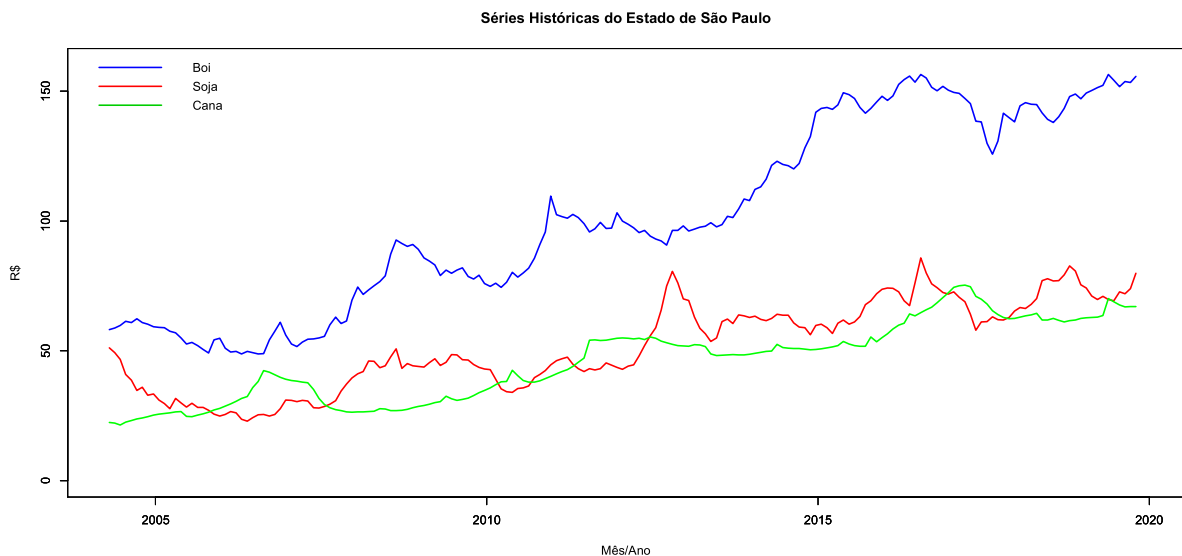


Figura 1.4.1: Gráficos dos preços do boi, soja e cana-de-açúcar para São Paulo.

A análise descritiva dos log retornos das séries históricas do preço da soja, do boi e da cana-de-açúcar, para SP, GO e MS, é apresentada na Tabela 1.4.1. Nota-se que as médias são todas positivas, com a cana-de-açúcar apresentando maior valor, indicando que tal produto teve mais valorização média no período estudado. No entanto, todas as médias são baixas se comparadas com os respectivos desvios, o que sugere resultados muito voláteis. Para confirmar tal afirmação realizou-se o teste LM-ARCH (Engle, 1982), que apresentou significância para quase todas as séries, exceto soja em GO, indicando que há efeito ARCH (GARCH) nas demais séries.

As medidas de curtose apresentaram valores próximos a 3 nas séries de soja, o que sugerem possibilidade de normalidade conforme Casella e Berger (2010), resul-

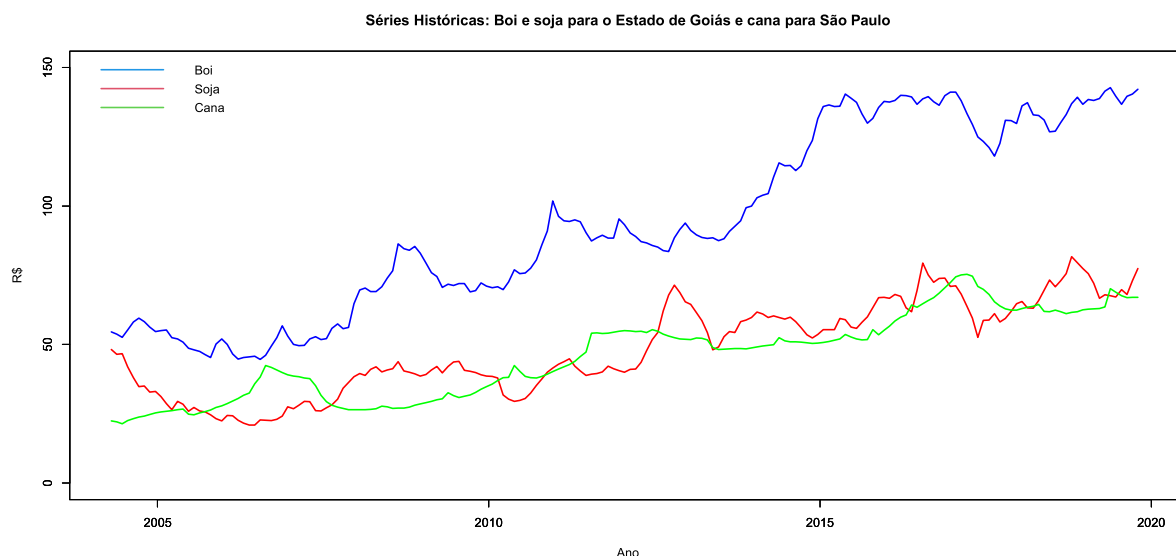


Figura 1.4.2: Gráficos dos preços do boi e soja para Goiás e cana-de-açúcar para São Paulo.

Tabela 1.4.1: Estatística descritiva dos log retornos dos preços da soja, boi e cana.

	Soja GO	Soja SP	Soja MS	Boi GO	Boi SP	Boi MS	Cana SP
Média	0,0025	0,0024	0,0024	0,0052	0,0053	0,0050	0,0059
Min	-0,1739	-0,1593	-0,2070	-0,0710	-0,0844	-0,1130	-0,1043
Max	0,1347	0,1348	0,1850	0,1422	0,1340	0,1291	0,1357
DP	0,0531	0,0520	0,0644	0,0341	0,0332	0,0350	0,0300
Ass	-0,0748	0,0088	-0,1621	0,8850	0,7996	0,7730	0,5452
Curt	3,2361	3,2420	3,5307	4,6175	4,9712	5,6946	3,8278
JB	0,6058	0,4561	2,9971	44,5540*	49,9360*	74,7950*	127,3500*
Q(36)	84,4130*	71,1400*	105,8000*	75,3900*	51,3990*	65,8760*	93,9050*
LM	18,0200	20,7900*	20,9400*	25,1600*	30,7900*	26,4100*	36,3100*
ADF	-7,8188*	-8,0784*	-7,7399*	-9,2797*	-9,2045*	-9,1907*	-6,2456*
PP	-9,7624*	-9,9293*	-9,5854*	-9,4180*	-10,7420*	-10,0250*	-9,4308*
KPSS	0,1626	0,1547	0,1040	0,0628	0,0669	0,0660	0,0874

Min - Mínimo. Max - Máximo. DP - Desvio padrão. Ass - Assimetria. Curt - Curtose. JB - teste de Jarque e Bera (1987). LM - Teste multiplicador de Lagrange, Engle (1982) com 12 defasagens. ADF - teste de Dickey e Fuller (1979). PP - teste Phillips e Perron (1988). KPSS - teste de Kwiatkowski et al. (1992). $Q(k)$ - estatística de Ljung e Box (1978), com k defasagens. * indica rejeição da hipótese nula ao nível de 5% de significância.

tado corroborado pelo teste de Jarque-Bera (1987) (que possui como hipótese H_0 : A distribuição é normal), ao passo que para as demais séries têm-se valores distantes de 3, sugerindo caudas pesadas, comum em séries de preços de commodities. Valores positivos de assimetria sugerem probabilidade de altas mais pronunciadas, exceto a soja em GO e MS, em que esse comportamento se relaciona com perdas.

Nas séries que o teste de Jarque-Bera rejeitou a hipótese de normalidade foram utilizadas a distribuição t de Student assimétrica na estimativa das marginais e naquelas que a hipótese não foi rejeitada, foram utilizadas a distribuição normal padrão

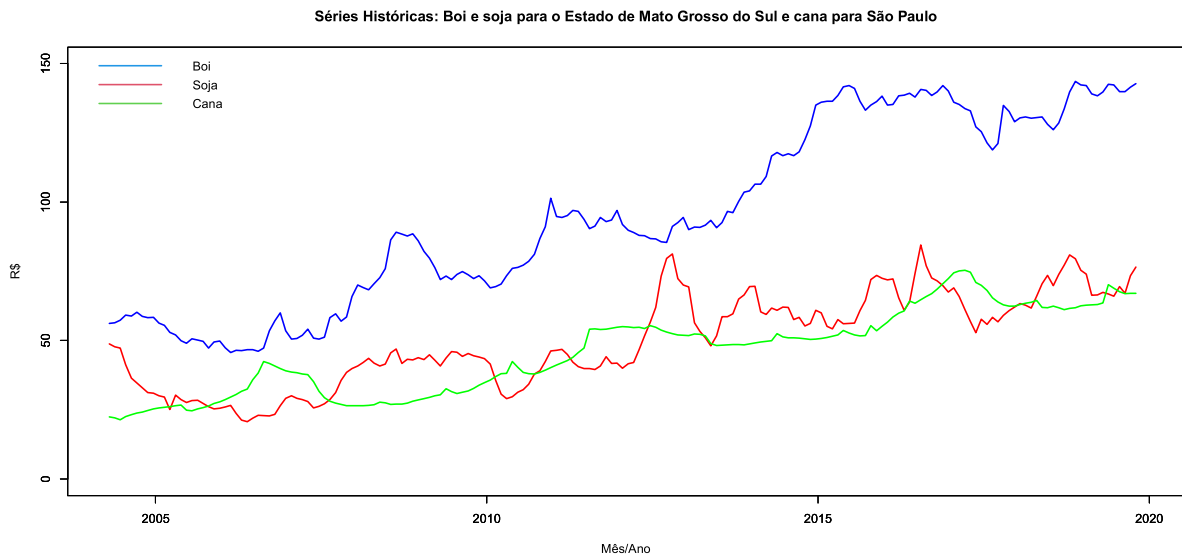


Figura 1.4.3: Gráficos dos preços do boi e soja para Mato Grosso do Sul e cana-de-açúcar para São Paulo.

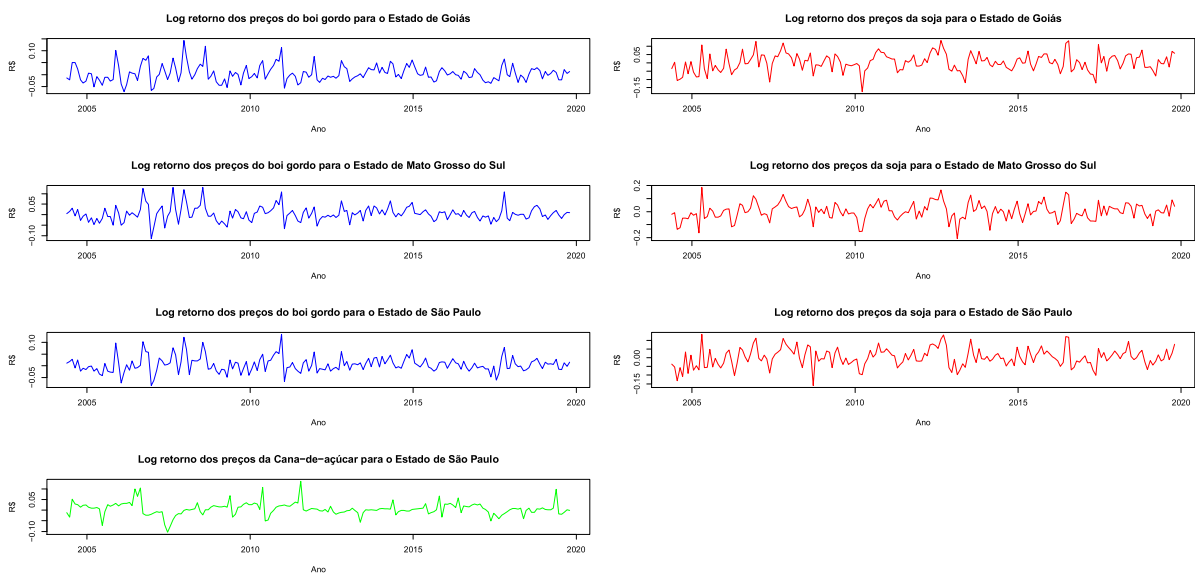


Figura 1.4.4: Gráfico do log retorno das séries para os estados de Goiás, Mato Grosso do Sul e São Paulo.

no modelo ARMA-GARCH. A estatística do teste de Ljung-Box (Ljung e Box, 1978) (que possui como hipótese H_0 : Não há correlação serial) indicou correlação serial para todas as séries. O teste LM-ARCH do Multiplicador de Lagrange (que possui como hipótese H_0 : Não há estrutura ARCH na série) para a estrutura ARCH rejeitou a hipótese de nulidade, apontando que há uma estrutura do tipo ARCH nas séries, exceto para soja GO. Os testes de Dickey e Fuller (1979), de Phillips e Perron (1988) e Kwiatkowski et al. (1992) foram utilizados para testar a presença de raiz unitária, nos quais todas as séries se mostraram estacionárias.

Os modelos marginais, $ARMA(m, n) - GARCH(p, q)$, foram selecionados com base nos critérios de informação de Akaike (1974) escolhendo as ordens m, n, p e q que apresentavam menor valor para o AIC. Além disso, a seleção da especificação destes modelos levou em conta as condições necessárias para proceder o passo posterior de estimação das cópulas. Para tanto, os modelos selecionados foram avaliados pelo teste de Ljung-Box, aplicado aos resíduos, para verificar a existência de dependência serial. Posteriormente, os parâmetros $\psi_i, \Theta_j, \alpha_k$ e β_l do modelo $ARMA(m, n) - GARCH(p, q)$ selecionados foram estimados pelo método da máxima verossimilhança.

Encontram-se na Tabela 1.4.2 as estimativas dos parâmetros dos modelos marginais. As ordens destes modelos foram selecionadas após combinações de 0 a 2 para m, n, p e q . As séries de soja para os estados de SP, GO e MS foram ajustadas utilizando a normal padrão como distribuição dos resíduos, ao passo que para as demais séries foram utilizadas uma t de Student assimétrica. A escolha dessas distribuições foi realizada com base no resultado do teste de Jarque-Bera e valores de assimetria e curtose apresentados na Tabela 1.4.1.

Tabela 1.4.2: Resultados do ajuste dos modelos (ARMA-GARCH) para os log retornos.

	Soja GO [#]	Soja SP [#]	Soja MS [#]	Boi GO	Boi SP	Boi MS	Cana SP
Especificação ARMA							
δ_0	0,0025	0,0012	0,0020	0,0037	0,0045	0,0029	0,0039*
	-0,0036	-0,0039	-0,0059	-0,0022	-0,0024	-0,0022	-0,0039
	0,3291*	-	-	0,2939*	-	0,2717*	0,6991*
ψ_1	-0,0738	(-)	(-)	-0,0936	(-)	-0,0654	-0,0409
	-	-	0,3375*	-	-	-	-
Θ_0	(-)	(-)	-0,0699	(-)	(-)	(-)	(-)
Especificação GARCH							
ω	0,0021*	0,0019*	0,0032*	0,0002	0,0008*	0,0001	0,0002*
	-0,0003	-0,0003	-0,0004	-0,0001	-0,0002	-0,0003	-0,0001
	0,1448	0,3115*	0,1212	0,0836	0,3575*	0,038	0,9999*
α_1	-0,1040	-0,1450	-0,0731	-0,0764	-0,1653	-0,0354	-0,3485
	-	-	-	0,8883*	-	0,9471*	-
β_1	(-)	(-)	(-)	-0,0936	(-)	-0,0504	(-)
	-	-	-	1,3780*	1,3872*	1,0660*	1,0448*
Skew	(-)	(-)	(-)	-0,2719	-0,1442	-0,1300	-0,0760
	-	-	-	8,8130	4,4861*	3,9900*	2,4168*
Shape	(-)	(-)	(-)	-6,1180	-1,4309	-1,0940	-0,1755
Testes							
Q(20)	10,3997	15,8657	18,3288	21,6798	23,5308	24,1895	20,9954
	[0,9603]	[0,7249]	[0,5657]	[0,3581]	[0,2635]	[0,2342]	[0,3974]
LM ARCH	6,1089	10,1826	17,8313	20,3730	11,6431	15,6503	15,2156
	[0,9105]	[0,5999]	[0,1209]	[0,0604]	[0,4747]	[0,2078]	[0,2299]
KS	0,0439	0,0409	0,0582	0,0390	0,0591	0,0441	0,0781
	[0,8655]	[0,9143]	[0,5551]	[0,9395]	[0,5341]	[0,8618]	[0,2061]

[#]Indica ARMA-GARCH considerando distribuição normal. * Indica rejeição da hipótese nula no nível de 5% de significância. Valores entre parênteses, (-), são os erros padrão. *Skew* e *Shape* são assimetria e forma, respectivamente. Valores entre colchetes, [], indicam o valor p para cada teste. $Q(k)$ - estatística de Ljung e Box (1978), com k defasagens. LM ARCH - teste multiplicador de Lagrange. KS - teste de Kolmogorov-Smirnov.

Para o teste de Ljung - Box aplicado aos resíduos, com *lag* até 20, o valor p acima

de 0,05 indica que os modelos selecionados estão especificados de forma adequada para descrever a dependência linear entre os retornos. O teste LM-ARCH, que verifica o efeito do modelo $GARCH(m, n)$ nos resíduos até a 20ª ordem, teve sua hipótese de nulidade não rejeitada, indicando que não há efeito ARCH nos resíduos. Dito de outra forma, os modelos da classe $GARCH(m, n)$ selecionados captaram os movimentos da volatilidade das séries.

Para cumprir os requisitos de estimação das cópulas (todas as distribuições marginais devem ser uniformes no intervalo $[0,1]$), foi realizada uma padronização nos resíduos dos modelos $ARMA(m, n) - GARCH(p, q)$. Após a padronização aplicou-se o teste KS⁷ de Kolmogorov-Smirnov nesses resíduos padronizados, indicando aderência à distribuição uniforme, conforme visto na Tabela 1.4.2.

1.4.1 Cópulas bivariadas

As estimativas para os parâmetros das cópulas⁸, para cada par de séries, encontram-se na Tabela 1.4.3. Os valores de AIC em negrito indicam o menor (melhor modelo) valor entre os modelos de cópulas de um mesmo par, representando as cópulas selecionadas em cada caso. Tanto a forma de seleção, por meio do AIC, quanto os resultados não significativos dos parâmetros indicam, exceto para “Cana x Boi GO”, uma relação de neutralidade entre mercados de energia e alimentos nas regiões avaliadas. Embora as cópulas se mostrem bem ajustadas, a independência é corroborada com o valor igual a 1 estimado para o parâmetro da cópula Gumbel, assim como os demais parâmetros das famílias ajustadas. Perceba que com $\theta = 1$ para essa cópula, sugere-se independência entre variáveis, pois se $\theta = 1$ na equação (1.11), recai-se no caso da equação (1.17).

Uma causa possível para esses resultados de independência pode ser, por exemplo, o aumento territorial da produção da cana-de-açúcar proporcionado pelo ganho de produtividade de outras culturas, como evidenciado por Adami et al. (2012). Os autores afirmam que embora a soja e as pastagens venham sendo convertidas em plantações de cana-de-açúcar, essas culturas têm apresentado ganho elevado em produtividade, o que faz com que seja possível produzir a mesma quantidade em menos terras. Por outro lado, esses resultados também podem sugerir que regionalmente, no Brasil, os comovimentos entre bioenergia e alimentos são independentes, como visto em outros estudos internacionais (Mueller et al., 2011; Reboledo, 2012; entre outros).

Embora tais produtos concorram por uso das terras, para suprir a produção de boi e soja, que vem perdendo espaço para a cana-de-açúcar, os produtores estão se valendo de tecnologias para elevar suas produções. Segundo IBGE (2011), no estado

⁷Mais detalhes em Conover (1999).

⁸Cópulas estimadas pela função `fitCopulaCopula` do software R 4.1.2 (R Core Team, 2022).

Tabela 1.4.3: Resultados da seleção das cópulas bivariadas.

Cópula	Cana X Soja GO	Cana X Soja SP	Cana X Soja MS	Cana X Boi GO	Cana X Boi SP	Cana X Boi MS
Gaussiana						
θ	0,0646 (0,0700)	-0,0091 (0,0710)	0,0109 (0,0710)	0,2059* (0,0670)	0,0902 (0,0680)	0,0897 (0,0690)
AIC	1,163	1,9832	1,9761	-6,5979	0,2933	0,3463
<i>t</i> -Student						
θ	0,0639 (0,0770)	-0,0049 (0,0770)	0,0109 (0,0710)	0,2054* (0,0670)	0,0898 (0,0690)	0,0898 (0,0690)
κ	14,1654 (13,2800)	14,0031 (13,5210)	7331,0000* (34,3840)	8242,284* (38,4420)	107,4397 (442,3200)	7117,0000 (76,8850)
AIC	1,8229	2,7516	3,9791	-4,5935	2,2639	2,3485
Clayton						
θ	0,1300 (0,0810)	0,0662 (0,0750)	0,0241 (0,0770)	0,2830* (0,1020)	0,0965 (0,0770)	0,1196 (0,0860)
AIC	-1,5470	1,3017	2,5724	-3,0224	0,0828	0,8828
Gumbel						
θ	1,0000* (0,0450)	1,0000* (0,0470)	1,0030* (0,0370)	1,0810* (0,0530)	1,0080* (0,0450)	1,0000* (0,0510)
AIC	2,0000	2,0001	1,9950	-0,7746	1,9645	2,0000
Joe						
θ	1,0000* (0,0670)	1,0000* (0,0710)	1,0050* (0,0490)	1,056* (0,0710)	1,0000* (0,0680)	1,0000* (0,0760)
AIC	2,0000	2,0001	1,9892	1,2989	2,0000	2,0000
SJC						
τ	0,0600	0,0300	0,0000	0,0900	0,0500	0,0400
λ_U	0,0000	0,0000	0,0100	0,0000	0,0000	0,0000
λ_L	0,0000	0,0000	0,0000	0,0300	0,0000	0,0000
AIC	6,9500	3,3500	4,0100	-1,6900	2,1100	2,7300

Valores entre parênteses () representam os erros padrão. Os parâmetros das cópulas foram estimados via máxima verossimilhança, e em alguns casos por pseudo máxima verossimilhança, (cópulas Gumbel e Joe para Cana X soja GO e SP e cópula Joe para Cana x Boi SP). *Um asterisco indica significância a 5% de significância.

de São Paulo a capacidade de produção de gado passou de 1,00 cabeça por hectare em 1975 para 1,85 cabeça por hectare em 2006, e embora a área de pastagem tenha sido significativamente reduzida em 4,4 milhões de hectares, o rebanho bovino aumentou em 1,4 milhão de cabeças.

Além disso, a região centro-sul experimentou tanto uma perda de 24,1 milhões de hectares de pastagens como um aumento de 45,3 milhões de cabeças de gado aumentando a capacidade de produção de 0,58 para 1,23 cabeça por hectare de 1975 a 2006, o que pode justificar a existência de independência entre os preços das commodities, mesmo elas competindo por terras. Conforme Macedo (2005) e Nassar (2009), a intensificação do uso da terra de pastagem, devido ao aumento da eficiência da pecuária, liberou terras para a produção agrícola.

Por outro lado, se a produtividade das outras commodities não forem suficientes para liberar terras, a expansão da cana-de-açúcar poderia levar, indiretamente, a um possível desmatamento nessa região (ver, Martinelli e Filoso, 2008). Contudo, essa possível associação da expansão do plantio de cana com desmatamento não foi

avaliada neste estudo.

Outro fator que dificulta uma resposta rápida à oscilação dos preços do mercado é o processo de produção da cana-de-açúcar que requer um pensamento de longo prazo e de exclusividade de produção, isso contextualiza com os resultados encontrados neste trabalho. Khanna et al. (2017) e Moraes e Zilberman (2014) afirmam que a natureza perene da cana-de-açúcar implica um compromisso mínimo de produção de 5 anos. Granco et al. (2018a) apontam alguns fatores que dificultam essa transformação de cultura: o ciclo de produção de longo prazo, sendo economicamente viável por 6 anos-safra (Goldemberg, 2006; Sant’Anna et al., 2015), a cana-de-açúcar cresce em um sistema soca (planta que produz novamente após corte) (Nogueira et al., 2013) e o maquinário envolvido na produção desta cultura é especialmente projetado para ela, o que reduz a capacidade dos agricultores de produzir outras culturas (Aguiar et al., 2011; Coelho et al., 2006; Van Den Wall Bake et al., 2009). Segundo Defante et al. (2020), este cultivo de longo prazo fez com que o setor sucroalcooleiro não sofresse, de imediato, tantos impactos na crise de 2008 e 2014, pois devido aos compromissos de produção firmados antes da crise.

Por fim, foi aplicado um teste para qualidade do ajuste para as cópulas⁹ (Tabela 1.4.4), o qual consiste em comparar uma cópula empírica com uma estimativa paramétrica. Os resultados para a qualidade de ajuste das cópulas, em que todos os modelos escolhidos pelo critério de AIC, se mostraram significativos para um bom ajuste (valor p maior que 0,05). Os valores p para a estatística deste teste foram ajustados por meio de *bootstrap* paramétrico¹⁰.

Tabela 1.4.4: Teste para qualidade do ajuste das cópulas bivariadas.

Cópula	Cana SP X Soja GO	Cana SP X Soja SP	Cana SP X Soja MS	Cana SP X Boi GO	Cana SP X Boi SP	Cana SP X Boi MS
Gaussiana	0,0321 [0,1238]	0,0486* [0,0049]	0,0182 [0,5891]	0,0264 [0,1535]	0,0248 [0,3416]	0,0249 [0,3020]
t -Student	0,0361 [0,1832]	0,0335 [0,0842]	0,0184 [0,5889]	0,0225 [0,3911]	0,0263 [0,2624]	0,0262 [0,2129]
Clayton	0,1111* [0,0049]	0,0315 [0,0941]	0,0187 [0,5990]	0,0257 [0,2426]	0,0355 [0,1040]	0,0276 [0,1337]
Gumbel	0,1174* [0,0049]	0,0522* [0,0247]	0,0173 [0,7970]	0,0274 [0,2525]	0,0300 [0,2525]	0,0261 [0,3020]
SJC	0,0492 [0,8000]	0,0761 [0,4600]	0,0322 [0,9700]	0,1131 [0,2200]	0,0684 [0,5200]	0,0581 [0,6600]

Valor p entre colchetes [], * Indica rejeição da hipótese de nulidade a 0,05.

⁹Mais detalhes sobre o teste em Genest e Remillard (2008) e Genest et al. (2009).

¹⁰Teste realizado pela função *gofCopula* do pacote *copula* do software R, versão 4.1.2 (R Core Team, 2022), ver Genest et al. (2009).

1.4.2 Cópulas trivariadas

Para cada estado em análise (GO, MS e SP) tem-se as informações acerca de três commodities, boi, soja e cana-de-açúcar. Com isso, analisar os movimentos conjuntos das variáveis por meio de cópulas bivariadas pode deixar de captar algum comovimento que possa vir a ser identificado quando se analisa todas as variáveis presentes no mesmo estado. Encontram-se na Tabela 1.4.5 os resultados das cópulas trivariadas.

Tabela 1.4.5: Resultado da seleção das cópulas trivariadas.

Cópula	Cana – Boi – Soja GO	Cana – Boi – Soja MS	Cana – Boi – Soja SP
<i>Gaussiana</i>			
ρ	0,1002* (0,0610)	0,0751* (0,0430)	0,0430* (0,0420)
AIC	-3,5518	-1,2104	0,9106
<i>t-Student</i>			
ρ	0,1014* (0,0460)	0,0751* (0,0450)	0,0436* (0,0440)
κ	47,4984 (81,5500)	1336,2145 (65,2410)	30,4402 (36,4970)
AIC	-1,9242	0,7939	2,1442
<i>Clayton</i>			
α	0,1252* (0,0520)	0,03247* (0,0400)	0,0579* (0,0450)
AIC	-5,9906	1,2484	-0,0541
<i>Gumbel</i>			
δ	1,0350* (0,0300)	1,0190* (0,0290)	1,0000* (0,0310)
AIC	0,8855	1,5244	2
<i>SJC</i>			
τ	1,001* (0,1100)	1,0200* (0,0900)	1,0300* (0,0700)
λ_U	0,0001	0,0300	0,0400
λ_L	0,0000	0,0000	0,0000
AIC	3,1500	2,5600	3,7100

Valores entre parênteses () representam os erros padrão. *Um asterisco indica significância a 5% de significância.

O teste para qualidade do ajuste das cópulas aplicado para o caso bivariado também foi utilizado no caso trivariado, apresentando significância para todas as cópulas estimadas. As cópulas que apresentaram menor AIC foram escolhidas, sendo a Clayton para o estado de Goiás e São Paulo e a gaussiana para o estado de Mato Grosso do Sul. As medidas de dependência estimadas com base nessas cópulas foram $\tau = 0,0589$; $\tau = 0,0159$; $\tau = 0,0478$ respectivamente.

A particularidade dos mercados existentes em cada estado analisado pode ter sido um dos motivos que ocasionou uma correlação baixa entre os preços das commodities. Nos estados de GO e MS é comum a venda de commodities por meio de contratos a preço futuro, o que causa menor oscilação. Abitante (2008) faz uma análise acerca da correlação entre o mercado futuro e o mercado *spot* (à vista) para algumas commodities e encontrou relação entre esses dois preços em grande parte do ano, ou seja, o preço no mercado *spot* acaba convergindo para o preço do mercado futuro em

diversos momentos.

1.5 Considerações finais

Neste trabalho foi analisada a dependência entre as séries de seis pares das cópulas bivariadas, confrontando os log retornos do preço da soja e do boi gordo com cana-de-açúcar para os estados de SP, GO e MS. No entanto, somente a cópula “Cana x Boi GO” apresentou dependência significativa, no entanto com um valor pequeno ($\tau = 0,13$), indicando uma dependência fraca. Para os demais casos, os resultados indicam que não há evidências para afirmar que há dependência em nenhuma direção, sugerindo neutralidade entre os mercados regionais de Cana X Soja para os estados de SP, GO e MS, além de Cana x Boi para os estados de SP e MS. Isso sugere que o preço por si só não é determinante suficiente para a mudança da cultura nessa região. Uma vez que a oscilação no preço, na região estudada, se mostra moderada, possivelmente relacionada aos contratos futuros de venda com preços preestabelecidos.

Embora exista concorrência direta por terras entre esses produtos agrícolas e gado, a expansão da cana-de-açúcar em direção ao oeste de SP e cerrado brasileiro, em especial os estados de GO e MS, aparentemente não gera dependência entre os preços. Com isso, a expansão da produção da cana-de-açúcar não parece ser o suficiente para influenciar no preço das demais culturas.

Segundo Nassar et al. (2008) apenas 1%, Silva e Miziara (2011) 6% e Adami et al. (2012) apenas 5% das terras em que a cana-de-açúcar vem se expandindo são de cultivos naturais do solo, não precisamente florestas, mostrando que o plano ZAE-CANA vinha surtindo efeito e não seria recomendável ser revogado, tal qual aconteceu em 2019. No entanto, a produção bovina no estado de Goiás merece um pouco de atenção, pois os modelos sugerem que há uma elevação no preço e que ele ocorre em conjunto com a produção da cana-de-açúcar neste Estado.

Não se sabe até quando a cana-de-açúcar precisará de terras e expansões. Valdes et al. (2016) orientam para dois cenários importantes, se houver alta no petróleo possivelmente haverá maiores demandas por etanol, se houver baixa no petróleo é provável que a demanda por etanol e por terras diminua. Outro fator que pode incentivar a produção de etanol é a demanda por energias mais limpas. Com base nos resultados de Pereira e Silveira (2016) espera-se que o avanço do setor somente seja possível com o forte investimento em novos processos e tecnologias, incluindo o etanol de segunda geração, além da melhoria do padrão de eficiência das usinas já existentes, para que o setor não se baseie somente na expansão da fronteira agrícola da cultura de cana-de-açúcar, mas utilize da melhor forma possível os recursos já existentes no setor.

Este trabalho leva em consideração apenas as séries dos log retornos dos preços

da soja, boi e cana-de-açúcar nos estados de SP, GO e MS. Acredita-se ser interessante analisar neste mesmo período a produtividade dessas commodities bem como áreas plantadas, além de analisar o nível de desmatamento nesses estados.

Referências Bibliográficas

- Abitante, K. G. (2008). Co-integração entre os mercados spot e futuro: evidências dos mercados de boi gordo e soja. Revista de Economia e Sociologia Rural, 46:75–96.
- Adami, M., Rudorff, B. F. T., Freitas, R. M., Aguiar, D. A., Sugawara, L. M. e Mello, M. P. (2012). Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in brazil. Sustainability, 4(4):574–585.
- Aguiar, D. A., Rudorff, B. F. T., Silva, W. F., Adami, M. e Mello, M. P. (2011). Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in são paulo state, brazil. Remote Sensing, 3(12):2682–2703.
- Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6):716–723.
- Aloui, R., Aïssa, M. S. B. e Nguyen, D. K. (2011). Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? Journal of Banking & Finance, 35(1):130–141.
- Belik, W. (2015). A heterogeneidade e suas implicações para as políticas públicas no rural brasileiro. Revista de Economia e Sociologia Rural, 53(1):9–30.
- Bernardo, R., Lourenzani, W. L., Satolo, E. G. e Caldas, M. M. (2019). Analysis of the agricultural productivity of the sugarcane crop in regions of new agricultural expansions of sugarcane. Gestão & Produção, 26.
- Box, G. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
- Brasil (2009). Decreto nº 6.961, de 17 de setembro de 2009. dispõe sobre aprovação do zoneamento agroecológico para o cultivo da cana-de-açúcar no território brasileiro, além de orientar financiamento da produção. Relatório técnico.
- Brasil (2019). Decreto nº 10.084, de 5 de novembro de 2019. revogou o decreto nº 6.961. Relatório técnico.
- Breymann, W., Dias, A. e Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance.
- Caldas, M. M., Granco, G., Bishop, C., Kastens, J. e Brown, J. C. (2017). Effects of sugarcane ethanol expansion in the brazilian cerrado: Land use response in the new frontier. Bioenergy and Land Use Change, 231:39.
- Capaz, R. S., Carvalho, V. S. B. e Nogueira, L. A. H. (2013). Impact of mechanization and previous burning reduction on ghg emissions of sugarcane harvesting operations in brazil. Applied Energy, 102:220–228.

- Casella, G. e Berger, R. L. (2010). *Inferência estatística*. São Paulo: Cengage Learning.
- Chen, S.-T., Kuo, H.-I. e Chen, C.-C. (2010). Modeling the relationship between the oil price and global food prices. *Applied Energy*, 87(8):2517–2525.
- Cherubin, M. R., Franco, A. L. C., Cerri, C. E. P., da Silva Oliveira, D. M., Davies, C. A. e Cerri, C. C. (2015). Sugarcane expansion in brazilian tropical soils—effects of land use change on soil chemical attributes. *Agriculture, Ecosystems & Environment*, 211:173–184.
- Chollete, L., de la Pena, V. e Lu, C.-C. (2011). International diversification: A copula approach. *Journal of banking & finance*, 35(2):403–417.
- Coelho, S. T., Goldemberg, J., Lucon, O. e Guardabassi, P. (2006). Brazilian sugarcane ethanol: lessons learned. *Energy for Sustainable Development*, 10(2):26–39.
- Cohn, A. S., Mosnier, A., Havlík, P., Valin, H., Herrero, M., Schmid, E., O’Hare, M. e Obersteiner, M. (2014). Cattle ranching intensification in brazil can reduce global greenhouse gas emissions by sparing land from deforestation. *Proceedings of the National Academy of Sciences*, 111(20):7236–7241.
- Conab - Companhia Nacional de Abastecimento (2019). Acompanhamento da safra brasileira: Cana-de-açúcar. 6(3). Safra 19/20. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em 15 de agosto de 2020. Relatório técnico.
- Conover, W. J. (1999). *Practical nonparametric statistics*, volume 350. John Wiley & Sons.
- Crago, C. L., Khanna, M., Barton, J., Giuliani, E. e Amaral, W. (2010). Competitiveness of brazilian sugarcane ethanol compared to us corn ethanol. *Energy Policy*, 38(11):7404–7415.
- de Araujo, G. J. F. e de Oliveira, S. V. W. B. (2021). Energy and environmental analysis of vinasse processing using internal circulation biodigesters and concentrators for different production ranges of ethanol and its potential impact on brazil. *Environment, Development and Sustainability*, 23(3):3130–3163.
- Defante, L. R., Vilpoux, O. F. e Sauer, L. (2020). Importance of the sugarcane industry in the formal employment in the state of mato grosso do sul during the period of 2008 to 2014. *Revista de Economia e Sociologia Rural*, 58.
- Dhillon, R. e von Wuehlisch, G. (2013). Mitigation of global warming through renewable biomass. *Biomass and bioenergy*, 48:75–89.
- Dickey, D. A. e Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.
- Embrechts, P., McNeil, A. e Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the econometric society, páginas 987–1007.
- Ferreira, M. E., Ferreira Jr, L. G., Latrubesse, E. M. e Miziara, F. (2016). Considerations about the land use and conversion trends in the savanna environments of central brazil under a geomorphological perspective. Journal of Land Use Science, 11(1):33–47.
- Ferreira, M. E., Ferreira Jr, L. G., Miziara, F. e Soares-Filho, B. S. (2013). Modeling landscape dynamics in the central brazilian savanna biome: future scenarios and perspectives for conservation. Journal of Land Use Science, 8(4):403–421.
- Figueira, S. R. F. (2020). Impactos dos preços e do crédito rural sobre a produção de cana-de-açúcar no estado de são paulo. Revista de Economia e Sociologia Rural, 58.
- Gandolfi, M. R. C., Jesus, C. M. e Gandolfi, P. E. (2019). Qualidade do emprego nas culturas de café, cana-de-açúcar e milho: estudo comparativo 2000/2010 entre mesorregiões de minas gerais. Revista de economia e sociologia rural, 57:63–80.
- Genest, C., Ghoudi, K. e Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3):543–552.
- Genest, C. e Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Em Annales de l’IHP Probabilités et statistiques, volume 44, páginas 1096–1127.
- Genest, C., Rémillard, B. e Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and economics, 44(2):199–213.
- Gilio, L. e de Moraes, M. A. F. D. (2016). Sugarcane industry’s socioeconomic impact in são paulo, brazil: A spatial dynamic panel approach. Energy Economics, 58:27–37.
- Goldemberg, J. (2006). The ethanol program in brazil. Environmental Research Letters, 1(1):014008.
- Goldemberg, J., Mello, F. F., Cerri, C. E., Davies, C. A. e Cerri, C. C. (2014). Meeting the global demand for biofuels in 2021 through sustainable land use change policy. Energy Policy, 69:14–18.
- Graesser, J., Aide, T. M., Grau, H. R. e Ramankutty, N. (2015). Cropland/pastureland dynamics and the slowdown of deforestation in latin america. Environmental Research Letters, 10(3):034017.
- Granco, G., Caldas, M., Featherstone, A., Sant’Anna, A. C. e Bergtold, J. (2018a). Farmers’ acreage responses to the expansion of the sugarcane ethanol industry: The case of goiás and mato grosso do sul, brazil. Em Land Allocation for Biomass Crops, páginas 103–123. Springer.

- Granco, G., Caldas, M. M., Bergtold, J. S. e Sant'Anna, A. C. (2017). Exploring the policy and social factors fueling the expansion and shift of sugarcane production in the brazilian cerrado. GeoJournal, 82(1):63–80.
- Granco, G., Sant'Anna, A. C., Bergtold, J. S. e Caldas, M. M. (2018b). Factors influencing ethanol mill location in a new sugarcane producing region in brazil. Biomass and bioenergy, 111:125–133.
- Hausman, C. (2012). Biofuels and land use change: sugarcane and soybean acreage response in brazil. Environmental and Resource Economics, 51(2):163–187.
- Hofert, M., Kojadinovic, I., Maechler, M. e Yan, J. (2020). copula: Multivariate Dependence with Copulas. R package version 1.0-1.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2011). Sistema IBGE de Recuperação Automática. Disponível em: <http://www.sidra.ibge.gov.br>. Acesso em 15 de junho de 2020. Relatório técnico.
- IPEA - Instituto de Pesquisa Econômica Aplicada - Ipeadata (2019). Dados macroeconômicos e regionais. Disponível em: <http://www.ipeadata.gov.br>. Acesso em 20 de outubro de 2019. Relatório técnico.
- Jarque, C. M. e Bera, A. K. (1987). A test for normality of observations and regression residuals. International Statistical Review/Revue Internationale de Statistique, páginas 163–172.
- Kaltalioglu, M. e Soytaş, U. (2009). Price transmission between world food, agricultural raw material, and oil prices. Em GBATA International Conference Proceedings, páginas 596–603.
- Katchova, A. L. e Sant'Anna, A. C. (2019). Impact of ethanol plant location on corn revenues for us farmers. Sustainability, 11(22):6512.
- Khanna, M., Louviere, J. e Yang, X. (2017). Motivations to grow energy crops: the role of crop and contract attributes. Agricultural Economics, 48(3):263–277.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P. e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of econometrics, 54(1-3):159–178.
- Ljung, G. M. e Box, G. E. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2):297–303.
- Lourenzani, W. L. e Caldas, M. M. (2014). Mudanças no uso da terra decorrentes da expansão da cultura da cana-de-açúcar na região oeste do estado de são paulo. Ciência Rural, 44:1980–1987.
- Macedo, L. O. B. (2005). Modernização da pecuária de corte bovina no brasil e a importância do crédito rural. Agroanalysis, Rio de Janeiro, 25(6):35–36.
- Mann, M. L., Kaufmann, R. K., Bauer, D. M., Gopal, S., Nomack, M., Womack, J. Y., Sullivan, K. e Soares-Filho, B. S. (2014). Pasture conversion and competitive cattle rents in the amazon. Ecological economics, 97:182–190.

- Manoel, A. A. S., Moraes, M. B. d. C., Santos, D. F. L. e Neves, M. F. (2018). Determinants of corporate cash holdings in times of crisis: insights from brazilian sugarcane industry private firms. International Food and Agribusiness Management Review, páginas 201–218.
- Manzatto, C. V., Assad, E. D., BACA, J. F. M., Zaroni, M. J. e Pereira, S. E. M. (2009). Zoneamento agroecológico da cana-de-açúcar: expandir a produção, preservar a vida, garantir o futuro. Embrapa Solos-Documentos (INFOTECA-E).
- Marin, F. R. (2016). Understanding sugarcane production, biofuels, and market volatility in brazil—a research perspective.
- Martinelli, L. A. e Filoso, S. (2008). Expansion of sugarcane ethanol production in brazil: environmental and social challenges. Ecological applications, 18(4):885–898.
- McManus, C., Barcellos, J. O. J., Formenton, B. K., Hermuche, P. M., Carvalho Jr, O. A. d., Guimarães, R., Gianezini, M., Dias, E. A., Lampert, V. d. N., Zago, D. et al. (2016). Dynamics of cattle production in brazil. PloS one, 11(1):e0147138.
- Moraes, M. A. F. D. e Zilberman, D. (2014). Production of Ethanol from Sugarcane in Brazil: from state intervention to a free market, volume 43. Springer Science & Business Media.
- Morettin, P. A. e Tolo, C. (2006). Análise de séries temporais. Em Análise de séries temporais, página 538.
- Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Arai, E., del Bon Espirito-Santo, F., Freitas, R. e Morissette, J. (2006). Cropland expansion changes deforestation dynamics in the southern brazilian amazon. Proceedings of the National Academy of Sciences, 103(39):14637–14641.
- Mueller, S. A., Anderson, J. E. e Wallington, T. J. (2011). Impact of biofuel production and other supply and demand factors on food price increases in 2008. Biomass and bioenergy, 35(5):1623–1632.
- Mutuc, M., Pan, S. e Hudson, D. (2011). Response of cotton to oil price shocks. Agricultural Economics Review, 12(389-2016-23468).
- Nassar, A. M. (2009). Brazil as an agricultural and agroenergy superpower. Brazil as an economic superpower, páginas 55–80.
- Nassar, A. M., Rudorff, B. F., Antoniazzi, L. B., Aguiar, D. d., Bacchi, M. R., Adami, M. et al. (2008). Prospects of the sugarcane expansion in brazil: impacts on direct and indirect land use changes. Sugarcane ethanol: contributions to climate change mitigation and the environment, páginas 63–93.
- Nazlioglu, S. e Soytas, U. (2011). World oil prices and agricultural commodity prices: Evidence from an emerging market. Energy Economics, 33(3):488–496.
- Nelsen, R. B. (2006). An introduction to copulas. Springer, New York.
- Neves, M., Trombin, V. e Consoli, M. (2011). The sugar energy map of brazil in ethanol and bioelectricity: sugarcane in the future of the energy matrix. So Paulo: Unica. Nicholson, FA, Smith, SR, Alloway, BJ, Carlton-Smith, C.

- Neves, M. F., Waack, R. S. e Marino, M. K. (1998). Sistema agroindustrial da cana-de-açúcar: caracterização das transações entre empresas de insumos, produtores de cana e usinas. Em Congresso Brasileiro de Economia e Sociologia Rural, volume 36, páginas 559–572.
- Ning, C. (2010). Dependence structure between the equity market and the foreign exchange market—a copula approach. Journal of International Money and Finance, 29(5):743–759.
- Nogueira, L. A. H., Moreira, J. R., Schuchardt, U. e Goldemberg, J. (2013). The rationality of biofuels. Energy Policy, 61:595–598.
- Nuñez, H. M., Önal, H. e Khanna, M. (2013). Land use and economic effects of alternative biofuel policies in brazil and the united states. Agricultural Economics, 44(4-5):487–499.
- Nyko, D., Valente, M. S., Milanez, A. Y., Tanaka, A. K. R. e Rodrigues, A. V. P. (2013). A evolução das tecnologias agrícolas do setor sucroenergético: estagnação passageira ou crise estrutural? BNDES Setorial, n. 37, mar. 2013, p 399-442.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International economic review, 47(2):527–556.
- Pereira, C. N. e Silveira, J. M. F. J. (2016). Análise exploratória da eficiência produtiva das usinas de cana-de-açúcar na região centro-sul do brasil. Revista de Economia e Sociologia Rural, 54:147–166.
- Phillips, P. C. e Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2):335–346.
- Rada, N. (2013). Assessing brazil’s cerrado agricultural miracle. Food Policy, 38:146–155.
- Reboredo, J. C. (2012). Do food and oil prices co-move? Energy policy, 49:456–467.
- Resende, A. G. e Candido, O. (2015). Dependence analysis of ethanol, sugar, oil, brl/usd exchange rate and bovespa: A vine copula approach. Economia Aplicada, 19:455–480.
- Rodrigues, D. M. T. e Miziara, F. (2008). Expansão da fronteira agrícola: a intensificação da pecuária bovina no estado de goiás.
- Saiani, C. C. S. e Perosa, B. B. (2016). Saúde respiratória e mecanização da colheita da cana-de-açúcar nos municípios paulistas: a importância do protocolo agroambiental. Revista de Economia e Sociologia Rural, 54:29–50.
- Sano, E. E., Rosa, R., Brito, J. L. e Ferreira, L. G. (2010). Land cover mapping of the tropical savanna region in brazil. Environmental monitoring and assessment, 166(1):113–124.
- Sant’Anna, A. C., Granco, G., Bergtold, J. e Caldas, M. M. (2015). Os desafios da expansão da cana-de-açúcar: como pensam e agem arrendatários e produtores?

- Sant'Anna, A. C., Shanoyan, A., Bergtold, J. S., Caldas, M. M. e Granco, G. (2016). Ethanol and sugarcane expansion in brazil: what is fueling the ethanol industry? International Food and Agribusiness Management Review, 19(1030-2017-2132):163–182.
- Silva, A. A. e Miziara, F. (2011). Sucroalcohol sector and agricultural frontier expansion in the goiás state, brazil. Pesquisa Agropecuária Tropical, 41(3).
- Sklar, A. (1959). Fonctions de répartition à n dimension et leurs marges. Université Paris, 8.
- Strassburg, B. B., Latawiec, A. E., Barioni, L. G., Nobre, C. A., Da Silva, V. P., Valentim, J. F., Vianna, M. e Assad, E. D. (2014). When enough should be enough: Improving the use of current agricultural lands could meet production demands and spare natural habitats in brazil. Global Environmental Change, 28:84–97.
- Team, R. C. et al. (2022). R: A language and environment for statistical computing.
- Trombeta, N. d. C. e Caixeta, J. V. (2017). Potencial e disponibilidade de biomassa de cana-de-açúcar na região centro-sul do brasil: indicadores agroindustriais. Revista de Economia e Sociologia Rural, 55:479–496.
- UDOP - União Nacional dos Produtores de Bioenergia (2019). Valores de ATR e o preço da tonelada de cana-de-açúcar - Consecana do Estado de São Paulo. Disponível em: https://udop.com.br/cana/tabela_consecana_saopaulo.pdf. Acesso em 04 de novembro de 2019. Relatório técnico.
- Valdes, C., Hjort, K. e Seeley, R. (2016). Brazil's agricultural land use and trade: effects of changes in oil prices and ethanol demand. Relatório técnico.
- Van den Wall Bake, J., Junginger, M., Faaij, A., Poot, T. e Walter, A. (2009). Explaining the experience curve: Cost reductions of brazilian ethanol from sugarcane. Biomass and bioenergy, 33(4):644–658.
- Walker, R., Moore, N. J., Arima, E., Perz, S., Simmons, C., Caldas, M., Vergara, D. e Bohrer, C. (2009). Protecting the amazon with protected areas. Proceedings of the National Academy of Sciences, 106(26):10582–10586.

Análise de critérios de seleção para cópulas: limitações e melhores cenários

Resumo: Os estudos aplicados com o uso da metodologia de cópulas têm apresentado crescimento nos últimos anos. Em geral, utiliza-se os critérios de informação de Akaike (AIC) e bayesiano (BIC) para determinar a função cópula que melhor se ajusta aos dados. Contudo, a maioria dos estudos não se preocupa com a limitação de tais critérios em diferentes tamanhos amostrais, distribuições de dados e medidas de dependência. Este artigo objetiva ampliar a discussão sobre os critérios de informação para cópulas apresentando resultados de desempenho do AIC e BIC frente à diferentes condições. Para isso, foram geradas amostras de diferentes tamanhos desde $n = 50$ até $n = 600$ com diversos intervalos entre os tamanhos amostrais, utilizando as famílias de cópulas gaussiana, t de Student, Clayton, Frank, Gumbel e Joe, visando encontrar o tamanho amostral que resulte em uma alta taxa de acerto, em selecionar o verdadeiro modelo simulado, usando os critérios de seleção de modelos AIC e BIC. Os resultados indicaram que para medidas de dependência maiores ($\tau = 0,5$ e $0,75$) as taxas de acerto foram elevadas, até mesmo para amostras pequenas. Já para a medida de dependência menor ($\tau = 0,25$) foram necessários tamanhos amostrais maiores ($n > 200$) para se obter altas taxas de acertos, maior que 90%, por exemplo. Outro resultado interessante é a diferença de taxas de acertos entre as famílias de cópulas, das quais a gaussiana e t apresentaram menores taxas de acertos, ao serem comparadas com as demais, em quase todos os cenários.

Palavras chaves: Cópulas; critério de informação de Akaike; critério de Schwarz; tamanho amostral; dependência.

Abstract: Applied studies using the copulation methodology have shown growth in recent years. In general, Akaike (AIC) and Bayesian (BIC) information criteria are used to determine the copula function that best fits the data. However, most studies are not concerned with limiting such criteria to different sample sizes, data distributions and dependence measures. This article aims to broaden the discussion on the information

criteria for copulas by presenting AIC and BIC performance results under different conditions. For this, samples of different sizes were generated from $n = 50$ to $n = 600$ with different intervals between the sample sizes, using the Gaussian, Student, Clayton, Frank, Gumbel and Joe copula families, aiming to find the sample size that results in a high hit rate, in selecting the true simulated model, using the AIC and BIC model selection criteria. For larger dependency measures ($\tau = 0.5$ and 0.75) the hit rates were high, even for small samples. As for the smaller dependence measure ($\tau = 0.25$), larger sample sizes ($n > 200$) were needed to obtain high hit rates, greater than 90%, for example. Another interesting result is the difference in hit rates between different copula families, of which the Gaussian and t presented lower hit rates, when compared to the others, in almost all scenarios.

Key words: Copula; Akaike information criterion; Schwarz criterion; sample size; dependency.

2.1 Introdução

Não é raro encontrar conjuntos de dados que possuem dependência entre algumas de suas variáveis analisadas. Segundo Tovar Cuevas et al. (2019) existem conjuntos de observações obtidas a partir de variáveis cujo comportamento natural tem alguma estrutura de dependência, seja ela alta ou baixa, e que este fenômeno pode ser encontrado em diversas áreas do conhecimento, tais como medicina, engenharia, economia, ecologia, dentre outras.

No entanto, quando se tem uma estrutura de dependência entre duas ou mais séries históricas, as medidas de dependência usuais mostram algumas fragilidades. Embrechts et al. (2002) afirmaram que a medida tradicional de interdependência, o coeficiente de correlação de Pearson, indica a média dos desvios das médias dos retornos, não captando de forma distinta os grandes e pequenos movimentos da série, o que prejudica os resultados quando se deseja analisar dependência entre duas, ou mais, séries de tempo.

Como as medidas de dependência tradicionais não captam bem as estruturas de dependência entre séries de tempo, tem sido comum analisar tal dependência por meio de cópulas. Segundo Nelsen (2006) as cópulas possuem como características a invariância à transformações crescentes e fornecem informações tanto das quantidades quanto da própria estrutura da dependência, seja ela simétrica ou não. Além da facilidade em obter a distribuição conjunta entre duas ou mais variáveis aleatórias.

Diante das vantagens citadas, as cópulas se mostram mais atrativas quando se deseja analisar estruturas de dependência entre séries temporais, não atoa encontra-se diversos estudos recentes que se valem das cópulas em diferentes áreas, tais como: finanças (Cherubini et al., 2004), climatologia (Schoelzel e Friederichs, 2008), ciências

ambientais (Kazianka e Pilz, 2011), medicina (Beaudoin e Lakhali-Chaieb, 2008), mercado de petróleo (Reboredo, 2012), física (Sato et al., 2010), turismo (Perez-Rodríguez et al., 2015) e (Tang et al., 2017) entre outras.

2.1.1 Seleção das cópulas

O processo de seleção de modelos para descrição de fenômenos e projeções é uma etapa fundamental em trabalhos de diversas áreas, sobretudo de estatística. Segundo Grønneberg e Hjort (2014), geralmente se têm vários modelos candidatos para descrever um comportamento específico, situação na qual deve-se utilizar algum critério para a escolha do mais adequado para o problema em questão. No entanto este “mais adequado” é subjetivo, dependendo dos objetivos do pesquisador. Pode ser que o interesse seja o modelo que apresenta melhores previsões ou aquele que apresenta menor distância entre as estimativas de seus parâmetros e os parâmetros do modelo verdadeiro (viés), por exemplo. Ko et al. (2019) reiteram que a seleção geralmente é feita por algum critério que minimiza a divergência de Kullback-Leibler (Kullback e Leibler, 1951), por exemplo o AIC - Critério de Informação de Akaike (1974).

Encontram-se na literatura alguns trabalhos que avaliam, desenvolvem e comparam formas e critérios de seleção para a escolha da melhor especificação, considerando diferentes famílias de cópulas, tais como: Huard et al. (2006), Grønneberg e Hjort (2014), Guloksuz (2016), Jordanger e Tjostheim (2014), Rossi et al. (2012), Tovar Cuevas et al. (2019), dentre outros.

No entanto, não foram encontrados trabalhos que verifiquem o tamanho mínimo de uma amostra para que ela apresente resultados mais assertivos (a partir de um limite de 90% de acertos, por exemplo) na comparação e seleção de famílias de cópulas. Com isso, não se sabe o tamanho amostral necessário para que um modelo de cópula possa se ajustar com bom índice de acerto ao conjunto de dados desejado, utilizando determinados critérios de seleção. Além disso, não se tem clareza do comportamento de algumas famílias de cópulas para amostras pequenas com diferentes medidas de dependência, o que torna interessante os estudos que acrescentam na discussão acerca do tamanho amostral e da precisão que se incorre ao selecionar determinado modelo.

Entre os diferentes critérios de informação utilizados, o AIC e o BIC carecem de uma investigação mais profunda sobre sua capacidade de indicar o melhor modelo de cópula. Considerando diferentes situações, como tamanhos amostrais e medidas de dependência, sobretudo para amostras pequenas. Desta forma, objetiva-se avaliar quais os tamanhos amostrais necessários para que o AIC e o BIC tenham boa performance na seleção de cópulas, cabendo neste caso uma comparação entre os critérios, bem como mensurar a taxa de acerto e de erros de cada um. Adicionalmente, deseja-se verificar o erro quadrático médio (EQM) que incorre ao selecionar os mode-

los de cópulas. As cópulas analisadas foram aquelas que são encontradas com mais frequência na literatura: gaussiana, t de Student, Clayton, Frank, Gumbel e Joe.

A contribuição deste estudo é fornecer informações importantes sobre as condições nas quais os critérios AIC e BIC apresentam bons resultados e, em quais situações, deve-se tomar cuidado ao utilizar tais critérios de seleção para selecionar um modelo de cópulas. Além disso, neste processo de avaliação deseja-se apontar o tamanho aceitável/desejável que uma amostra deve ter para que tais famílias de cópulas sejam utilizadas considerando alguns cenários diferentes para suas respectivas medidas de dependência. Adicionalmente, pretende-se apontar a sensibilidade e especificidade (taxas de acertos e erros vide seção 2.3.1) do critério em cada tamanho amostral estudado, que é um outro diferencial deste trabalho. Mais informações sobre sensibilidade e especificidade podem ser vistas em Dziak et al. (2020) e Swift et al. (2020).

Além desta introdução, o artigo está estruturado em mais cinco seções. Na segunda seção tem-se uma revisão acerca de cópulas; na terceira tem-se uma discussão acerca dos critérios AIC e BIC e do processo de simulação adotado; na quarta os resultados e discussão além de apresentar trabalhos com dados reais com diferentes tamanhos amostrais e medidas de dependência. Por fim, são apresentadas as conclusões na sexta seção.

2.2 Cópulas

Conforme Nelsen (2006) as cópulas são funções que unem, ou copulam, funções de distribuição multivariadas às suas funções de distribuição marginais unidimensionais. Dito de outra forma, as cópulas são funções de distribuição multivariadas cujas marginais unidimensionais são uniformes no intervalo $[0, 1]$. Uma cópula bivariada é definida na região $[0, 1] \times [0, 1]$ como:

$$C(u, v) = P[U \leq u; V \leq v], \quad (2.1)$$

em que U e V são uniformes no intervalo $[0, 1]$.

Fazendo uma ligação entre a definição de cópulas e o teorema de Sklar (1959) obtém-se a distribuição conjunta a partir das marginais e uma função de cópulas que descreve a relação entre as variáveis. Este teorema serviu de base para a teoria de cópulas.

Para o caso bivariado, o teorema de Sklar (1959) é definido da seguinte forma: Seja H uma função de distribuição conjunta com marginais F e G . Então existe uma cópula C tal que para todo $(x, y) \in \mathbb{R}^2$,

$$H(x, y) = C(F(x), G(y)). \quad (2.2)$$

Para o caso multivariado o mesmo teorema é definido como:

$$F(\mathbf{y}) = C(F_1(y_1), \dots, F_n(y_n)), \quad (2.3)$$

em que, se Y_1, \dots, Y_n são variáveis aleatórias com marginais F_1, \dots, F_n , respectivamente, e F é a função de distribuição conjunta das marginais, então existe uma n -cópula C para todo $\mathbf{y} \in \mathbb{R}^n$.

Quando se deseja obter as famílias de cópulas com base nas distribuições marginais e conjuntas, o teorema de Sklar pode ser analisado de forma inversa, dando origem a um corolário que expressa as cópulas em termos de uma função de distribuição conjunta e as inversas generalizadas¹ das marginais:

$$C(u_1, \dots, u_n) = F(F_1^{(-1)}(u_1), \dots, F_n^{(-1)}(u_n)), \quad (2.4)$$

em que $F_1^{(-1)}, \dots, F_n^{(-1)}$ são inversas generalizadas de F_1, \dots, F_n . No entanto, quando a função é estritamente crescente, caso da $F(u)$ quando $u \in [0, 1]$ a inversa generalizada coincide com a inversa (Nelsen, 2006). Desta forma, é possível obter alguns modelos das cópulas por meio da equação:

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad (2.5)$$

em que $F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)$ são as funções quantílicas marginais de u_1, \dots, u_n . Quando F_i são contínuas, o modelo apresentado na equação (2.5) fornece um método de obtenção de cópulas a partir de distribuições conjuntas.

2.2.1 Modelos de cópulas

A estrutura de dependência entre duas, ou mais, séries de tempo pode apresentar diferentes características, tais como: dependência simétrica, assimétrica e de cauda superior e/ou inferior. Para captar tais estruturas de dependência são utilizados vários modelos e famílias de cópulas diferentes, cada qual com suas características. As cópulas mais utilizadas em estudos aplicados são: gaussiana, t de Student, Clayton, Gumbel, Frank e Joe².

A cópula gaussiana ou normal é definida como:

$$C_G(u, v; \theta) = \Phi\left(\phi^{-1}(u), \phi^{-1}(v)\right) \quad (2.6)$$

em que Φ é a função de distribuição acumulada de uma normal padrão bivariada, θ é o coeficiente de correlação de Pearson, pertencendo ao intervalo $[-1, 1]$ que mede a

¹Definida em Morettin e Toloi (2006) como $F_i^{(-1)} = \sup_z \{F_i(z) \leq u\}$.

²Mais sobre cópulas e diferentes famílias pode ser visto em Joe (1997) e Nelsen (2007).

dependência entre u e v e $\phi^{-1}(u)$ e $\phi^{-1}(v)$ são funções quantílicas da normal padrão univariadas. Essa família capta dependência positiva e negativa, além do fato de que quando o parâmetro θ se aproxima de -1 e 1 a cópula tende para os limites inferior e superior de Fréchet-Hoeffding³, respectivamente.

A cópula t de Student é definida como:

$$C_T(u, v; \theta, \kappa) = T\left(t^{-1}(u), t^{-1}(v)\right), \quad (2.7)$$

em que T é a função de distribuição acumulada bivariada de uma t de Student, $t^{-1}(u)$ e $t^{-1}(v)$ são funções quantílicas de uma t de Student univariada, θ é o coeficiente de correlação de Pearson, também pertencente ao intervalo $[-1, 1]$ e κ são os graus de liberdade. Assim como ocorre com a cópula gaussiana, esta família de cópulas também é simétrica, porém com diferenças nos pesos das caudas, sendo uma boa alternativa à gaussiana em alguns casos.

A cópula de Clayton (1978) é definida por:

$$C_C(u, v; \theta) = \max\left(\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, 0\right) \quad (2.8)$$

em que $\theta \in (0, +\infty)$ é o parâmetro de dependência entre u e v . Quando θ se aproxima de zero indica independência entre as variáveis, ao passo que quando θ aumenta tendendo ao infinito a cópula converge para o limite superior de Fréchet-Hoeffding. Vale ressaltar ainda que esta cópula não é capaz de captar dependência negativa.

A cópula de Gumbel (1960) é definida como:

$$C_G(u, v; \theta) = \exp\left(-\left(\left(-\ln u\right)^\theta + \left(-\ln v\right)^\theta\right)^{\frac{1}{\theta}}\right), \quad (2.9)$$

em que $\theta \in [1, +\infty)$ é o parâmetro de dependência entre u e v . Assim, quando $\theta = 1$ tem-se o caso de independência entre as variáveis, ao passo que quando θ tende ao infinito a função converge para o limite superior de Fréchet-Hoeffding. Assim como a cópula Clayton, a cópula Gumbel não é capaz de captar dependência negativa.

A cópula de Frank (1979) é definida como:

$$C_F(u, v; \theta) = -\theta^{-1} \log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right) \quad (2.10)$$

em que $\theta \in \mathbb{R}^*$ é o parâmetro de dependência entre u e v . Esta família de cópulas é capaz de captar dependência positiva e negativa, de forma assimétrica. Quando θ tende a $-\infty$ e $+\infty$ a função converge para os limites inferiores e superiores de Fréchet-Hoeffding, respectivamente. Para $\theta = 0$ o modelo não está definido, pois para este

³Para mais detalhes acerca dos limites, ver Nelsen (2006).

valor tem-se o caso de independência entre as variáveis. No entanto, esta cópula não é recomendada quando se tem variáveis com grande dependência nas caudas.

A cópula de Joe (1993) é definida como:

$$C_J(u, v; \theta) = 1 - \left((1-u)^\theta + (1-v)^\theta - (1-u)^\theta (1-v)^\theta \right)^{\frac{1}{\theta}} \quad (2.11)$$

em que $\theta \in (1, +\infty)$ é o parâmetro de dependência entre u e v . Esta família de cópulas é capaz de captar apenas dependência positiva e quando $\theta = 1$ tem-se o caso de independência, ao passo que quando θ tende ao infinito a função converge para o limite superior de Fréchet-Hoeffding. Esta cópula apresenta uma dependência forte de cauda superior.

2.2.2 Medidas de dependência

Duas variáveis aleatórias são ditas concordantes quando se tem altos valores de uma combinado com altos valores de outra e baixos valores de uma com baixos valores de outra. E são ditas discordantes quando altos valores de uma se relacionam com baixos valores de outra e baixos valores de uma com altos valores de outra.

O τ de Kendall e ρ de Spearman são as medidas de dependência ou de associação mais conhecidas na literatura e são definidas em Nelsen (2006) como:

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (2.12)$$

$$\rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]), \quad (2.13)$$

em que (X_1, Y_1) , (X_2, Y_2) e (X_3, Y_3) são amostras independentes e identicamente distribuídas de um vetor (X, Y) de variáveis aleatórias.

No entanto, para se obter tais medidas é preciso de informações acerca da distribuição conjunta, o que nem sempre é possível ou fácil de se obter. Com isso as cópulas podem ser utilizadas para obter tais medidas, conforme em Nelsen (2006), da seguinte forma:

$$\tau = 4 \int C(u, v) dC(u, v) \quad (2.14)$$

$$\rho = 12 \int uv dC(u, v) - 3 \quad (2.15)$$

em que $C(u, v)$ é uma cópula, definida em (2.4).

2.3 Critérios de informação

O critério de informação de Akaike (AIC) é uma medida da qualidade do ajuste de um modelo estatístico e é extensivamente utilizado para comparar modelos candidatos ao ajuste de um conjunto de dados. Segundo Emiliano et al. (2014) esse critério baseia-se no conceito de informação e fornece uma medida relativa das informações perdidas quando um determinado modelo é usado para descrever um fenômeno real. Akaike (1974) descobriu uma relação entre a distância de Kullback-Leibler e a otimização da função suporte, o que tornou viável uma relação entre a teoria e a prática quando se deseja selecionar um modelo com base em algum conjunto de dados.

O AIC é definido por Akaike (1974) como:

$$AIC = -2 \log L(\hat{\theta}) + 2(p), \quad (2.16)$$

em que $L(\hat{\theta})$ é a função verossimilhança do modelo avaliado no estimador $\hat{\theta}$ que é o estimador de máxima verossimilhança e p é o número de parâmetros a serem estimados. Akaike (1974) demonstrou que o viés da função suporte $L(\hat{\theta})$ converge assintoticamente para o número de parâmetros.

O Critério de informação bayesiano (BIC) proposto por Schwarz (1978) é considerado um critério mais rigoroso, caracterizado por penalizar com mais rigor os modelos que levam em consideração muitos parâmetros em sua estimação. Além disso, este critério avalia modelos definidos em termos de probabilidade *a posteriori*, recebendo este nome porque o autor do critério se valeu de um argumento bayesiano para prová-lo. O BIC é definido por:

$$BIC = -2 \log L(\hat{\theta}) + p \log(n), \quad (2.17)$$

em que $L(\hat{\theta})$ é a função verossimilhança do modelo avaliado no estimador de máxima verossimilhança $\hat{\theta}$, p é o número de parâmetros utilizado na estimação e n é o tamanho amostral.

2.3.1 Sensibilidade e especificidade do critério

Como se trata de critérios de seleção de modelos e não de um teste de hipóteses, não cabe aqui falar em erro de tipo I e II e nem de poder do teste. No entanto, ao selecionar uma família de cópulas para um conjunto de dados o pesquisador se encontra sujeito a erros e acertos. Com isso, saber da taxa de acerto e de erro que se incorre ao selecionar uma família de cópulas para determinado conjunto de dados em diferentes cenários se torna interessante. Para tanto, muitas vezes, um processo de simulação computacional faz-se necessário.

Ao se deparar com um conjunto de dados e alguns modelos de cópulas candidatos tem-se quatro possibilidades. Suponha duas famílias de cópulas A e B e um conjunto de dados $(u, v) \in \mathbb{R}^2$. Pode-se ajustar a família A quando a amostra é realmente proveniente desta família, o que é chamado de verdadeiro positivo (VP). Não classificar o conjunto como sendo da família A quando de fato ele é desta família, o que é chamado de falso negativo (FN). Selecionar a família A para o conjunto de dados, quando na verdade ele é proveniente de outra família, neste caso tem-se o falso positivo (FP). Classificar o conjunto como não sendo da família A , quando de fato ele não é da família A , neste caso tem-se o verdadeiro negativo (VN). Estas possibilidades estão ilustrados na Tabela 2.3.1.

Tabela 2.3.1: Taxas de VP, FN, FP e VN.

	Família correta		
	A	B	
Família selecionada	A	VP	FP
	B	FN	VN

Assim a sensibilidade (S), ou taxa de acertos que informa a porcentagem de a cópula escolhida estar corretamente ajustada é dada por:

$$S = \frac{VP}{VP + FN}. \quad (2.18)$$

Ao passo que a especificidade (E), ou taxa de verdadeiro negativo que informa o “poder” de seleção do critério adotado é dada por:

$$E = \frac{VN}{VN + FP}. \quad (2.19)$$

Numa revisão de literatura acerca de diversos critérios utilizados na seleção de cópulas, não foi encontrado trabalho que apresenta estas medidas de precisão e taxas de erros cometidos pelos critérios durante o processo de seleção. Com isso este trabalho apresenta esse importante diferencial, que servirá de orientação quanto a precisão que cada cenário apresenta. Mais informações acerca de sensibilidade e especificidade podem ser obtidas em Swiff et al. (2020) e Dziak et al. (2020).

Adicionalmente, este trabalho apresenta o Erro Quadrático Médio (EQM), essa medida de erro foi obtida no processo de seleção dos modelos de cópulas, o que é outro diferencial neste trabalho se confrontado com os estudos relacionados. O EQM representa a diferença quadrática média entre um determinado estimador e seu parâmetro inicial, ou ainda a soma da variância do estimador e o viés do estimador elevado ao quadrado e é definido em Morettin e Bussab (2017) como:

$$\text{EQM} = E \left[(\hat{\theta} - \theta)^2 \right] = \text{Var}(\hat{\theta}) + [\text{Viés}(\hat{\theta})]^2. \quad (2.20)$$

O EQM é bastante útil para comparação de dois estimadores. Em comparação entre dois estimadores deve-se optar por aquele com menor valor de EQM.

2.3.2 Estudos relacionados

Este estudo se assemelha com o de Jordanger e Tjøstheim (2014), sobretudo com relação a famílias de cópulas e medidas de dependências utilizadas. No entanto, deseja-se analisar com mais detalhes os tamanhos amostrais menores que 100, além daquelas amostras com tamanhos entre 100 e 250, onde acredita-se ter um espaçamento grande, não abrangendo diversos estudos práticos que possuem tamanhos amostrais nesse intervalo. O presente estudo apresenta ainda as taxas de sensibilidade e especificidade e o EQM referente ao parâmetro estimado da cópula em cada processo.

Apesar de estudos da literatura possuírem características semelhantes ao que se encontra neste trabalho (tamanho de amostra, medidas de dependência e algumas famílias de cópulas) que estão destacadas na Tabela 2.3.2, ele contribui acrescentando na literatura as taxas de acerto e erros que o pesquisador pode estar cometendo ao selecionar determinada família em certas condições e apresenta ainda subintervalos menores para o tamanho amostral. O que será útil aos pesquisadores, pois seu tamanho amostral real pode estar distante dos tamanhos amostrais já apresentados em outros trabalhos que apresentam intervalos maiores entre os tamanhos amostrais.

As contribuições aqui apresentadas podem minimizar os erros na escolha inadequada de um modelo de cópulas, o que conduz o investigador a apresentar um valor de suas medidas de dependência super ou subestimado, acarretando em conclusões errôneas acerca de sua variável alvo.

2.4 Simulação e resultados

2.4.1 Parâmetros e cenários

Na prática encontram-se trabalhos que envolvem a teoria de cópulas em sua aplicação com diversos tamanhos amostrais, vide Tabela 2.4.1, com dados que apresentam diferentes frequências, desde intervalos de hora ou dia até os casos que se tem registros mensais ou trimestrais. Dadas diferentes frequências de registros, não é raro encontrar estudos com tamanho amostral acima de 500 observações, no entanto não se tem a mesma realidade para todas as séries, sendo possível encontrar estudos com tamanho amostral próximo de 50, 75 ou 100 observações. Encontram-se na Tabela

Tabela 2.3.2: Características de estudos correlatos.

Autor	Critérios	Características comuns			Cópulas
		Medida de dependência	Tamanho amostral	Replicações N	
Gülöksüz (2016)	AIC, BIC e DM	0,3; 0,5 e 0,7	50, 100 e 300	100	Gumbel, Clayton, Frank e Joe
Jordanger e Tjøstheim (2014)	AIC e CIC	0,25, 0,5 e 0,75	100, 250, 500 e 1000	5000	Clayton, Frank, Gumbel, gaussiana e t
Rossi et al. (2012)	DIC	1/3 e 2/3	200, 500 e 1000	500	gaussiana, t, Clayton, Gumbel e Heavy Tail
Ko e Hjort (2019)	AIC e CIC	0,56 e 0,60	1000	1000	Frank e Clayton
Tovar Cuevas (2019)	EMV, BIC e GOF	0,2; 0,5 e 0,8	50, 100, 500 e 1000	1000	Clayton, Frank, Gumbel Hougaard, Farlie-Gumbel-Morgenstern (FGM) e Gumbel Barnett.
Huard et al. (2006)	BIC	-0,2; 0,2; 0,5 e 0,7	30, 100, 300 e 600	1000	Clayton, AMH, Gumbel, Frank, A12, A14, FGM e gaussiana
Dissmann et al. (2013)	Método proposto	0,05; 0,10; 0,15; 0,20; 0,40 e 0,60	500, 1000 e 2000	1000	gaussiana, t, Gumbel, Frank e mixed
dos Santos Silva e Lopes (2008)	DIC	1/3; 2/3 e 0,9	100, 200 e 500	1000	Clayton, Frank, gaussiana, Gumbel, t de Student e Heavy Tail

AIC - Critério de informação Akaike, BIC - Critério de informação Bayesiano, DM - Distância mínima, CIC - Critério de informação para cópulas, critério de informação de desvio, GOF - Goodness-Of-Fit, EMV - Estimativa de máxima verossimilhança

2.4.1 alguns exemplos de estudos aplicados utilizando cópulas com bases de dados pequenas, tais como Beaudoin e Lakhal-Chaieb (2008) e Perez-Rodríguez et al. (2015) que utilizaram tamanhos amostrais menores que 100.

Tabela 2.4.1: Artigos aplicados utilizando cópulas e tamanhos de amostras.

Artigos	n	τ
Reboredo (2012)	670 (484 e 186)	-0,014; 0,08; 0,3 e 0,4
Resende e Cândido (2015)	639	-0,4 até 0,2
Schoelzel e Friederich (2008)	1218	0,07 até 0,15
Kazianka e Pilz (2011)	902	-
Beaudoin e Lakhal-Chaieb (2008)	97 e 293	0,336 e 0,082
Nabaei et al. (2019)	324	-
Serra e Gil (2013)	240	-
Tang et al. (2017)	240	0,4; 0,3; 0,2 e 0,1
Perez-Rodríguez et al. (2015)	130, 73 e 68	0,1 até 0,22

Para atingir o objetivo realizou-se um processo de simulação com as cópulas: gaussiana, t de Student, Clayton, Gumbel, Frank e Joe. Foram utilizadas ainda diferentes medidas de dependência pré-estabelecidas ($\tau = 0,25; 0,5$ e $0,75$ simbolizando dependência fraca, moderada e alta, respectivamente) entre as variáveis.

Para todos os casos foram analisadas amostras de tamanhos $n = 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500$ e 600 . O processo de simulação envolveu a gera-

ção⁴ de 1000 réplicas das amostras para cada cenário, variando as famílias das cópulas, os tamanhos amostrais e as diferentes medidas de dependência. Para evitar possíveis erros de estimação pontual, foram realizadas 20 repetições de cada processo e a média foi apresentada como resultado final, de onde foram contados os números de acertos e erros de seleção⁵.

O procedimento de simulação envolveu 234 cenários distintos, oscilando as 6 famílias de cópulas com 13 tamanhos e 3 medidas de dependências. A seleção se deu por meio dos critérios AIC ou BIC objetivando descobrir o menor tamanho amostral que apresentaria boa precisão (90% de acerto, por exemplo).

O tamanho da amostra e os diferentes valores para dependência serial foram os termos que oscilaram nas comparações por se tratar de condicionantes importantes. Em seus resultados Gülöksüz (2016) afirma que o tamanho da amostra e a medida de dependência entre as séries são características importantes, influenciando diretamente no número de acertos de cada processo.

2.4.2 Resultados e discussão

Após simulação dos dados, a análise dos resultados deste estudo corrobora com os resultados encontrados em Gülöksüz (2016), Jordanger e Tjøstheim (2014) e Ko e Hjort (2019), conforme pode ser observado nas Figuras 2.4.1 a 2.4.6, em que é possível notar que as taxas de acertos são menores para amostras de tamanhos menores e também para as medidas de dependências mais baixas. Para a maior medida de dependência analisada ($\tau = 0,75$) todas as famílias atingiram altas taxas de acerto, 90% ou mais, ainda para tamanhos amostrais pequenos, $n = 75$ ou $n = 100$, a exceção foi a família t de Student que precisou de $n = 200$ para atingir tal taxa. Com a medida de dependência moderada, $\tau = 0,5$, grande parte das famílias precisaram de amostras de tamanho $n = 200$ para atingir a taxa de acerto acima de 90%, apenas as famílias gaussiana e t de Student não atingiram tal taxa, necessitando tamanho amostral acima de 200.

Com relação a medida de dependência baixa, $\tau = 0,25$, foi necessário um tamanho amostral bem maior que os demais casos para que as cópulas t de Student, Clayton e Joe atingissem a taxa de 90% ou mais. Já as cópulas Gaussiana, Gumbel e Frank não atingiram a taxa esperada nem para o maior tamanho amostral em estudo, $n = 600$, no entanto é possível notar que a taxa de acerto segue crescente a medida que o tamanho amostral aumenta.

Os resultados apresentados nas Figuras 2.4.1 a 2.4.6, que ilustram as taxas de acertos e erros para cada cenário, mostram que em alguns casos é preciso tomar cuidado

⁴Por meio da função *BiCopSim* do pacote *VineCopula* de autoria de Nagler (2019)

⁵Por meio da função *BiCopEstList* do pacote *VineCopula*

para realizar modelagem via cópulas, sobretudo naqueles casos em que a medida de dependência se mostra baixa. Nesses casos é preciso uma amostra de tamanho acima de 500 para garantir um processo de modelagem adequado e que apresente menos erros. Já para os casos em que as medidas de dependência são maiores, o tamanho amostral não se torna um fator de preocupação e amostras menores podem ser utilizadas com boa precisão nas estimativas, na maioria dos casos.



Figura 2.4.1: Gráfico com taxas de erros e acertos da cópula gaussiana com diferentes medidas de dependência e tamanho amostral.

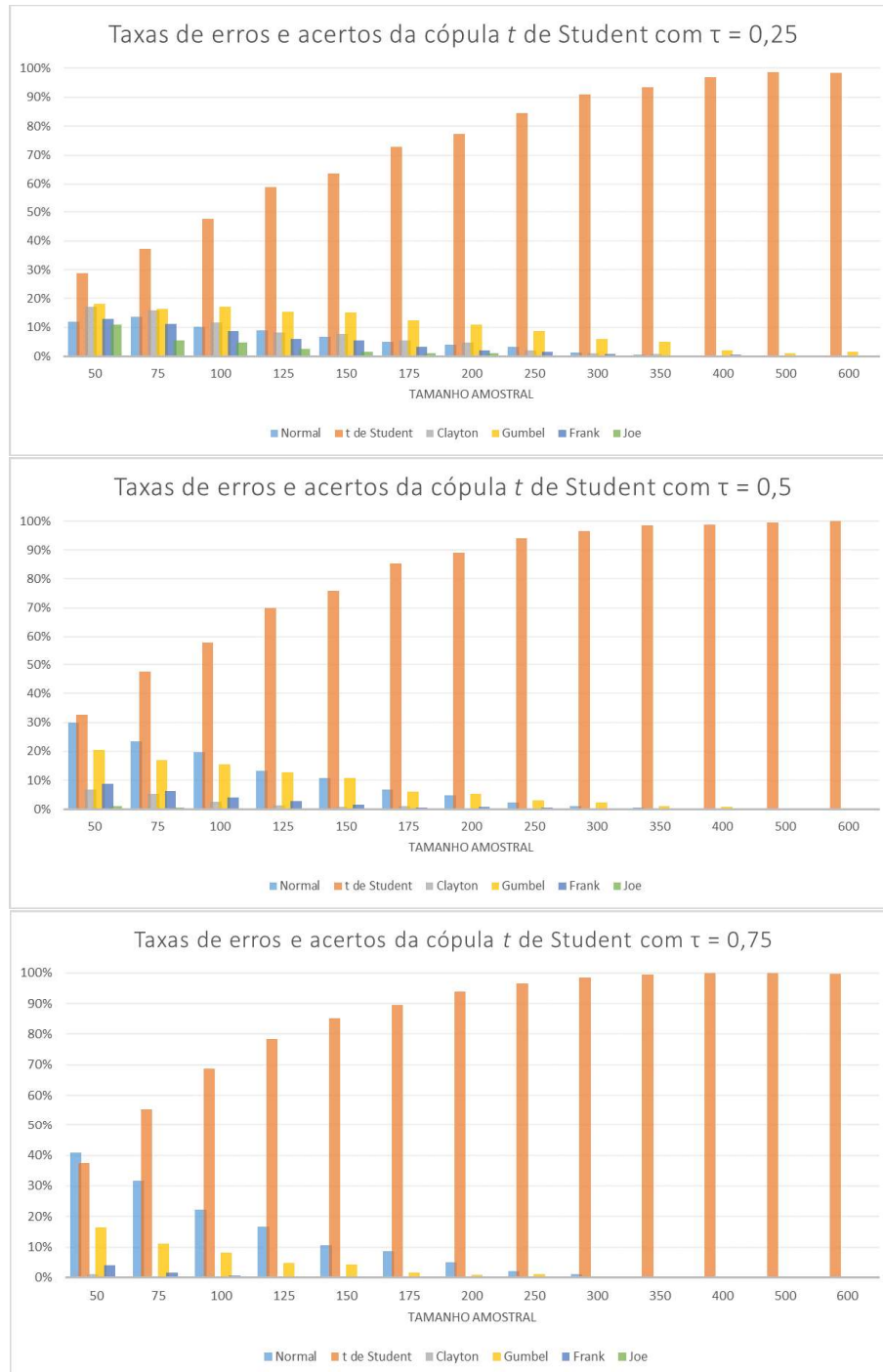


Figura 2.4.2: Gráfico com taxas de erros e acertos da cópula t de Student com diferentes medidas de dependência e tamanho amostral.

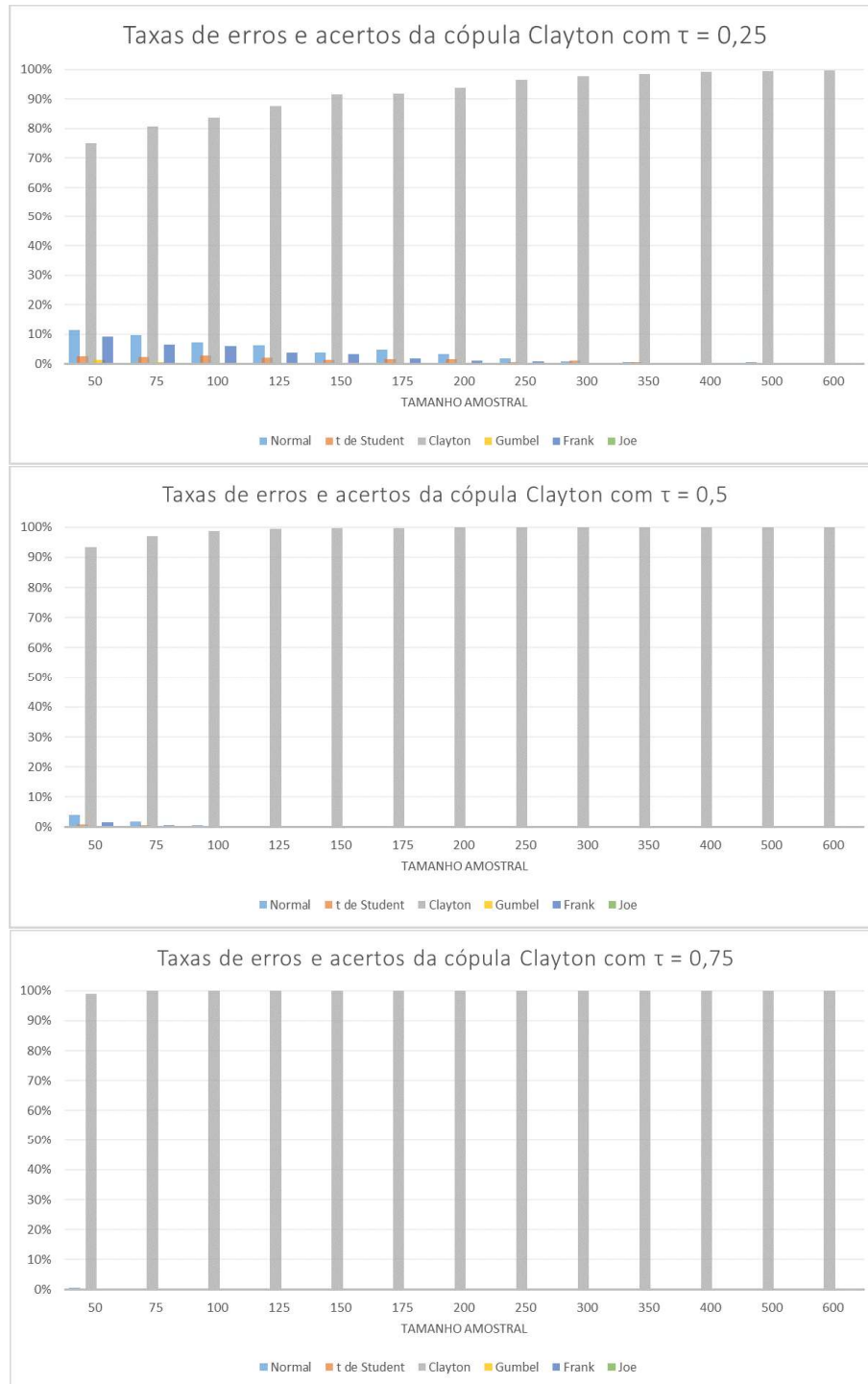


Figura 2.4.3: Gráfico com taxas de erros e acertos da cópula de Clayton com diferentes medidas de dependência e tamanho amostral.

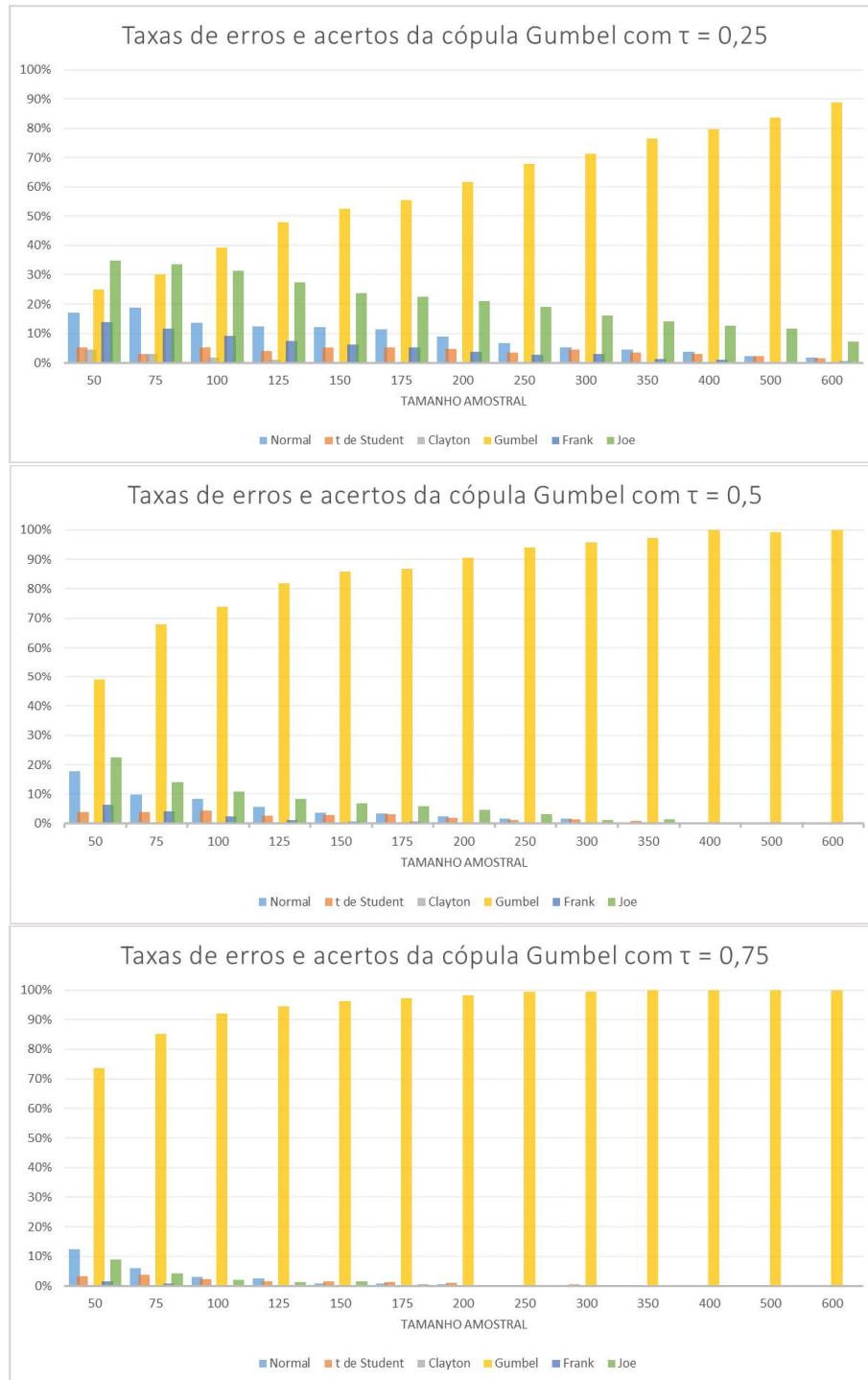


Figura 2.4.4: Gráfico com taxas de erros e acertos da cópula de Gumbel com diferentes medidas de dependência e tamanho amostral.

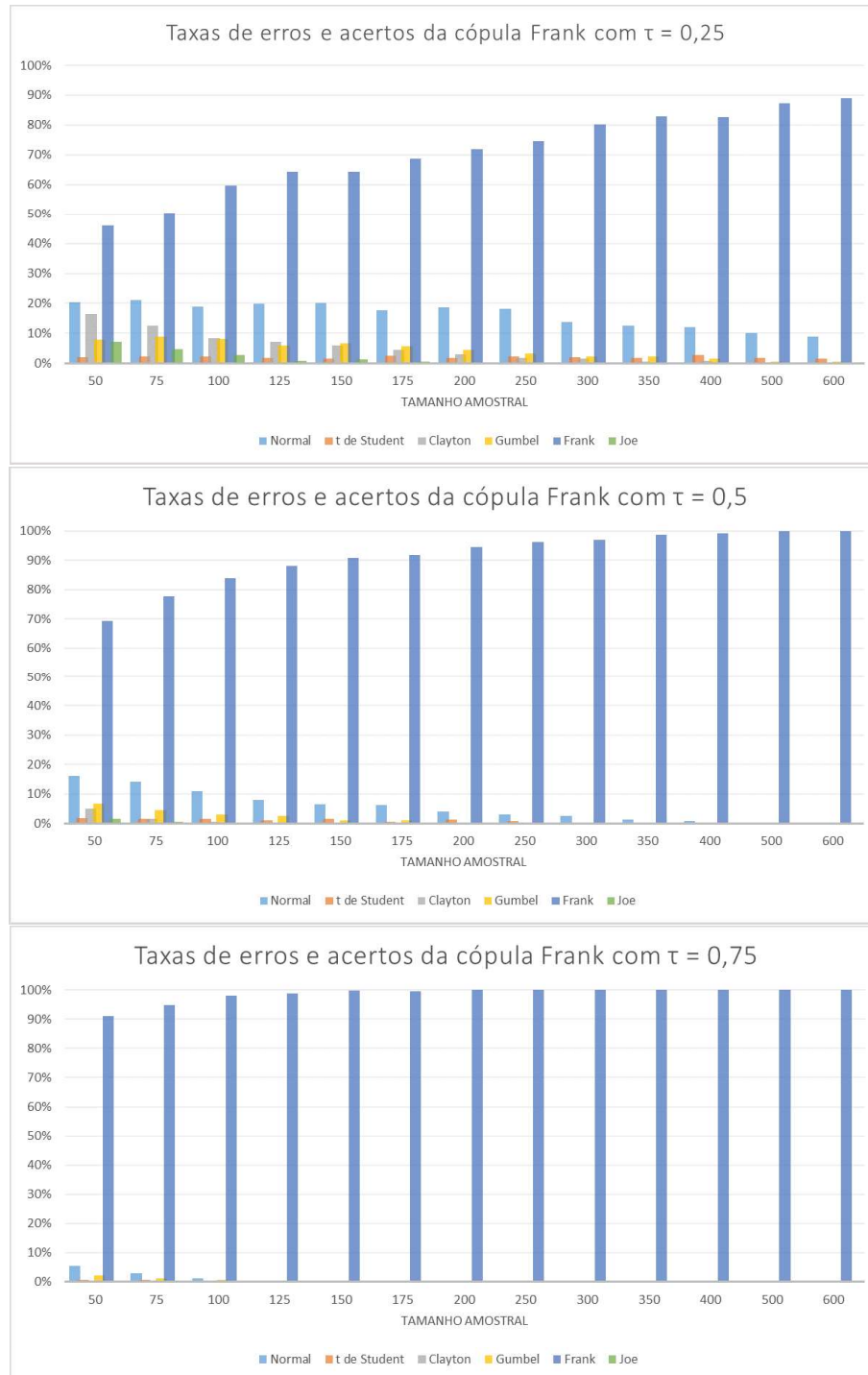


Figura 2.4.5: Gráfico com taxas de erros e acertos da cópula de Frank com diferentes medidas de dependência e tamanho amostral.

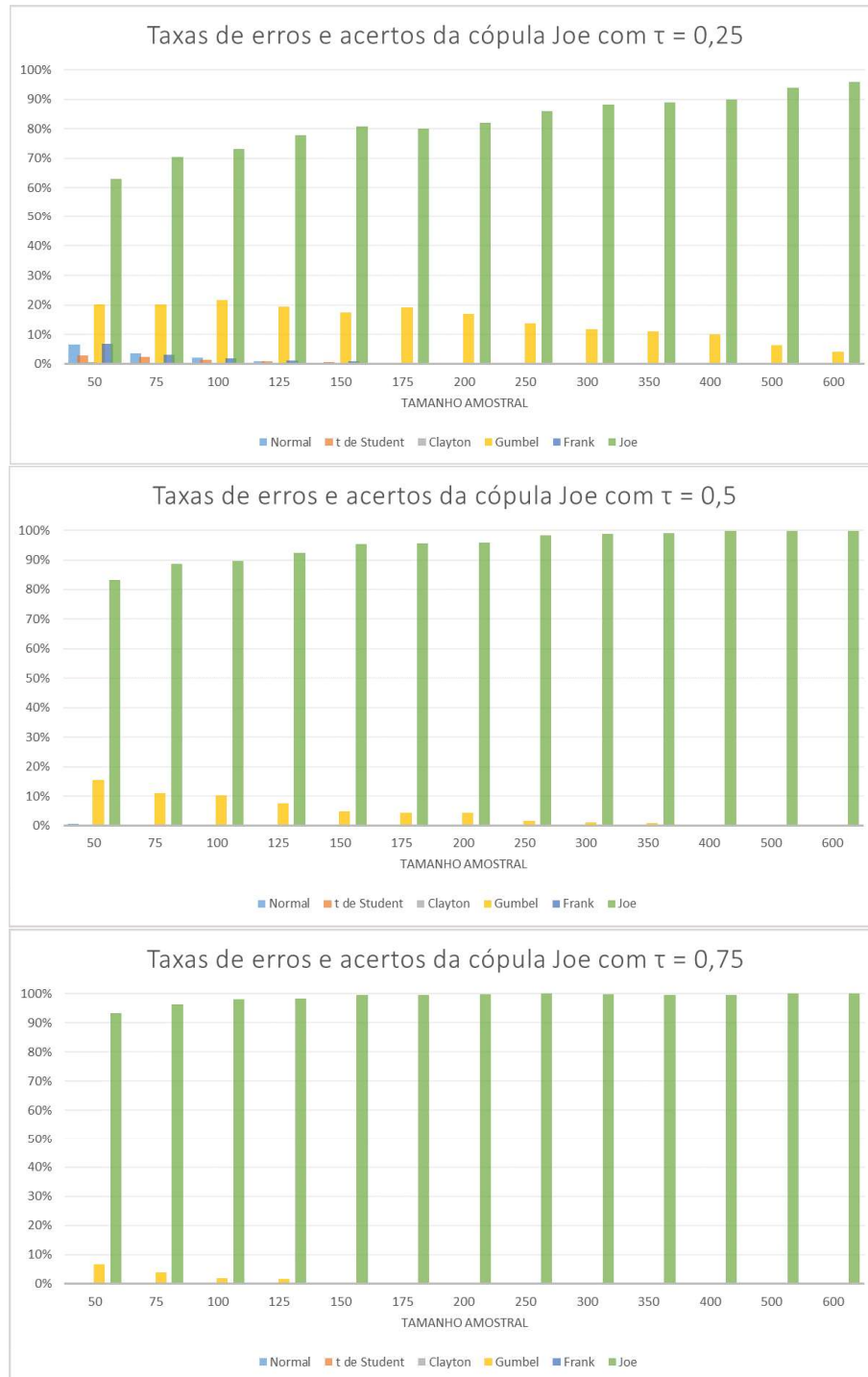


Figura 2.4.6: Gráfico com taxas de erros e acertos da cópula de Joe com diferentes medidas de dependência e tamanho amostral.

A precisão do processo de seleção dos modelos de cópulas realizadas pelos critérios AIC e BIC foram analisadas por meio das medidas de sensibilidade e especificidade, que é um dos diferenciais deste trabalho. Tais medidas para as cópulas gaussianas, t de Student, Clayton, Gumbel, Frank e Joe se encontram nas respectivas Tabelas 2.4.2, 2.4.3, 2.4.4, 2.4.5, 2.4.6 e 2.4.7, que corrobora com as informações que podem ser vistas nas respectivas Figuras 2.4.1, 2.4.2, 2.4.3, 2.4.4, 2.4.5 e 2.4.6, em que

pode ser destacado que tanto a sensibilidade quanto a especificidade aumentaram ao passo que o tamanho amostral aumenta, conforme o esperado.

Vale destacar que uma família de cópula que apresenta boa sensibilidade se mostra interessante para aplicações futuras, mas de nada vale se sua especificidade não se mostrar alta também. Afirmar que uma família A é adequada quando realmente é a família A que deveria ser escolhida parece razoável, mas de nada adianta se ao comparar as demais famílias de cópulas o modelo classificá-las também como A . No entanto, além da especificidade se mostrar alta, a sensibilidade se mostrou ainda maior na maioria dos casos.

Os piores resultados encontrados em grandes amostras ocorreram para a menor medida de dependência analisada, $\tau = 0,25$, mas se mostrando em linha com os resultados vistos em Jordanger e Tjøstheim (2014), apresentando taxas de acerto na casa do 83% para os casos em que se comete mais erros, chegando a acertar até 100% para os casos de maiores amostras, considerando $n \geq 500$.

As cópulas gaussiana e Gumbel apresentaram menores valores de especificidade e sensibilidade, indicando menor precisão no processo de seleção quando se escolhe esses modelos de cópulas, em especial nos casos com medidas de dependência menores. Comparando os critérios de seleção é possível ver uma pequena melhora quando se utiliza o critério bayesiano - BIC, no entanto não é uma melhora que possa ser generalizar para todos os casos.

Tabela 2.4.2: Sensibilidade e especificidade referente à cópula gaussiana.

$\tau = 0,25 - AIC$													
Tamanho amostral	50	75	100	120	150	175	200	250	300	350	400	500	600
Sensibilidade	0,3492	0,4331	0,4934	0,5593	0,5756	0,6297	0,6263	0,6812	0,7430	0,7620	0,7894	0,8393	0,8523
Especificidade	0,8652	0,8658	0,8962	0,9036	0,9142	0,9216	0,9296	0,9402	0,9578	0,9634	0,9676	0,9742	0,9778
$\tau = 0,5 - AIC$													
Sensibilidade	0,6400	0,7171	0,8131	0,8355	0,8682	0,8653	0,8872	0,9325	0,9237	0,9407	0,9412	0,9281	0,9361
Especificidade	0,8632	0,9012	0,9204	0,9458	0,9574	0,9666	0,9778	0,9858	0,9892	0,9954	0,9978	0,9992	1
$\tau = 0,75 - AIC$													
Sensibilidade	0,8560	0,9021	0,9182	0,9302	0,9394	0,9526	0,9319	0,9423	0,9641	0,9321	0,9350	0,9470	0,9436
Especificidade	0,8814	0,9188	0,9472	0,9606	0,9770	0,9806	0,9888	0,9956	0,9976	0,9994	0,9998	1	1
$\tau = 0,25 - BIC$													
Sensibilidade	0,3521	0,4380	0,5012	0,5734	0,5926	0,6391	0,6495	0,6991	0,7702	0,7988	0,8332	0,8721	0,9023
Especificidade	0,8618	0,8618	0,8896	0,8908	0,9024	0,9096	0,9146	0,9286	0,9494	0,9552	0,9598	0,9706	0,9754
$\tau = 0,5 - BIC$													
Sensibilidade	0,6510	0,7332	0,8321	0,8644	0,8998	0,9053	0,9281	0,9686	0,9690	0,9845	0,9883	0,9966	0,9958
Especificidade	0,8532	0,8838	0,9018	0,9180	0,9328	0,9452	0,9548	0,9678	0,9801	0,9882	0,9948	0,9982	0,9992
$\tau = 0,75 - BIC$													
Sensibilidade	0,8711	0,9252	0,9456	0,9667	0,9872	0,9842	0,9834	0,9955	0,9971	0,9973	0,9961	0,9982	0,9970
Especificidade	0,8618	0,8922	0,9164	0,9294	0,9464	0,9572	0,9688	0,9828	0,9902	0,9930	0,9966	0,9998	1

Tabela 2.4.5: Sensibilidade e especificidade referente à cópula de Gumbel.

$\tau = 0,25 - AIC$													
Tamanho amostral	50	75	100	125	150	175	200	250	300	350	400	500	600
Sensibilidade	0,2492	0,3123	0,3932	0,4785	0,5250	0,5542	0,6165	0,6795	0,7132	0,7652	0,7972	0,8354	0,8897
Especificidade	0,8796	0,8846	0,8788	0,8926	0,9010	0,9058	0,9162	0,9312	0,9452	0,9538	0,9654	0,9782	0,9848
$\tau = 0,5 - AIC$													
Sensibilidade	0,4901	0,6792	0,7422	0,8182	0,8572	0,8691	0,9065	0,9395	0,9566	0,9726	0,9987	0,9929	0,9987
Especificidade	0,8892	0,9132	0,9278	0,9450	0,9580	0,9708	0,9754	0,9874	0,9906	0,9956	0,9972	0,9986	0,9994
$\tau = 0,75 - AIC$													
Sensibilidade	0,7374	0,8520	0,9210	0,9442	0,9612	0,9725	0,9822	0,9932	0,9934	0,9990	1	1	1
Especificidade	0,9324	0,9580	0,9716	0,9836	0,9884	0,9944	0,9960	0,9974	0,9995	0,9988	0,9992	1	0,9996
$\tau = 0,25 - BIC$													
Sensibilidade	0,2680	0,3077	0,4258	0,4970	0,5510	0,5882	0,6433	0,7012	0,7372	0,7877	0,8182	0,8492	0,8985
Especificidade	0,8645	0,8626	0,8536	0,8660	0,8716	0,8770	0,8896	0,9086	0,9232	0,9356	0,9530	0,9701	0,9794
$\tau = 0,5 - BIC$													
Sensibilidade	0,5102	0,7010	0,7612	0,8355	0,8714	0,8893	0,9175	0,9466	0,9652	0,9802	0,9912	0,9965	0,9987
Especificidade	0,8756	0,8944	0,9082	0,9256	0,9426	0,9552	0,9618	0,9788	0,9840	0,9904	0,9944	0,9974	0,9986
$\tau = 0,75 - BIC$													
Sensibilidade	0,7565	0,8726	0,9322	0,9533	0,9669	0,978	0,9861	0,9965	0,9995	1	1	1	1
Especificidade	0,9218	0,9472	0,9618	0,9758	0,9804	0,9902	0,9918	0,9952	0,9984	0,9986	0,9982	0,9996	0,9996

Tabela 2.4.6: Sensibilidade e especificidade referente à cópula de Frank.

$\tau = 0,25 - AIC$													
Tamanho amostral	50	75	100	125	150	175	200	250	300	350	400	500	600
Sensibilidade	0,4592	0,5056	0,5967	0,6431	0,6442	0,6887	0,7187	0,7460	0,8032	0,8307	0,8260	0,8732	0,8917
Especificidade	0,8732	0,8916	0,9078	0,9282	0,9292	0,9456	0,9498	0,9570	0,9676	0,9718	0,9752	0,9832	0,9846
$\tau = 0,5 - AIC$													
Sensibilidade	0,6922	0,7765	0,8390	0,8815	0,9077	0,9191	0,9445	0,9626	0,9702	0,9867	0,9917	0,9997	0,9999
Especificidade	0,9375	0,9532	0,9708	0,9782	0,9878	0,9872	0,9908	0,9962	0,9966	0,9986	0,9992	0,9998	0,9996
$\tau = 0,75 - AIC$													
Sensibilidade	0,9127	0,9485	0,9801	0,9891	0,9980	0,9971	1	1	1	1	1	1	1
Especificidade	0,9822	0,9906	0,9958	0,9978	1	1	1	0,9998	1	1	1	1	1
$\tau = 0,25 - BIC$													
Sensibilidade	0,4656	0,5127	0,6016	0,6505	0,6522	0,6961	0,7292	0,7525	0,8122	0,8401	0,8391	0,8866	0,8978
Especificidade	0,8692	0,8846	0,9010	0,9202	0,9210	0,9378	0,9436	0,9526	0,9621	0,9696	0,9732	0,9812	0,9838
$\tau = 0,5 - BIC$													
Sensibilidade	0,6952	0,7801	0,8430	0,8881	0,9150	0,9222	0,9495	0,9655	0,9732	0,9861	0,9937	0,9999	1
Especificidade	0,9324	0,9464	0,9668	0,9746	0,9842	0,9836	0,9882	0,9946	0,9964	0,9986	0,9995	0,9998	0,9992
$\tau = 0,75 - BIC$													
Sensibilidade	0,9152	0,951	0,9820	0,9892	0,9981	0,9981	1	1	1	1	1	1	1
Especificidade	0,9804	0,9898	0,9951	0,9976	0,9998	1	1	0,9998	1	1	1	1	1

Tabela 2.4.7: Sensibilidade e especificidade referente à cópula de Joe.

$\tau = 0,25 - AIC$													
Tamanho amostral	50	75	100	125	150	175	200	250	300	350	400	500	600
Sensibilidade	0,6301	0,7051	0,7313	0,7782	0,8072	0,8012	0,8191	0,8601	0,8813	0,8894	0,8985	0,9380	0,9591
Especificidade	0,8728	0,9028	0,9182	0,9358	0,9454	0,9516	0,9542	0,961	0,9674	0,9718	0,9744	0,9770	0,9854
$\tau = 0,5 - AIC$													
Sensibilidade	0,8331	0,8871	0,8977	0,9230	0,9521	0,9567	0,9573	0,9841	0,9891	0,9914	0,9980	0,9971	0,9980
Especificidade	0,9484	0,9702	0,9774	0,9830	0,9862	0,9884	0,9908	0,9936	0,9974	0,9972	0,9990	0,9996	0,9996
$\tau = 0,75 - AIC$													
Sensibilidade	0,9331	0,9624	0,9821	0,9845	0,9975	0,9967	0,9987	1	0,9999	0,9979	0,9967	1	1
Especificidade	0,9822	0,9912	0,9956	0,9974	0,9968	0,9986	0,9994	0,9996	0,9998	1	1	1	1
$\tau = 0,25 - BIC$													
Sensibilidade	0,6397	0,7124	0,7377	0,7877	0,8070	0,8012	0,8222	0,8612	0,8826	0,8897	0,8988	0,9387	0,9590
Especificidade	0,8678	0,8988	0,9162	0,934	0,9444	0,9514	0,9538	0,9606	0,9668	0,9718	0,9744	0,9768	0,9856
$\tau = 0,5 - BIC$													
Sensibilidade	0,8340	0,8882	0,8975	0,9230	0,9521	0,9561	0,9572	0,9847	0,9899	0,9918	0,9987	0,997	0,998
Especificidade	0,9480	0,9698	0,9770	0,9833	0,9862	0,9884	0,9908	0,9936	0,9974	0,9972	0,9990	0,9996	1
$\tau = 0,75 - BIC$													
Sensibilidade	0,9930	0,9631	0,9821	0,9844	0,9971	0,9964	0,9982	1	0,9990	0,9970	0,9960	1	1
Especificidade	0,9820	0,9912	0,9956	0,9974	0,9968	0,9986	0,9994	0,9996	0,9998	1	1	1	1

Visando avaliar melhor as escolhas dos modelos de cópulas, além da sensibilidade e especificidade, utilizou-se também a medida de EQM referente ao parâmetro estimado da cópula, que apresentou resultados satisfatórios apenas para as cópulas normal e t . Para as demais cópulas o método de seleção apresentou valores altos para o EQM, indicando algum problema no processo de escolha dessas famílias, ainda que apresentassem bons índices de acertos se analisados os índices de sensibilidade e especificidade. As Tabelas 2.4.8 e 2.4.9 mostram os valores de EQM para as diferentes medidas de dependência analisadas.

Tabela 2.4.8: Medidas de EQM das famílias de cópulas estudadas AIC

Tamanho amostral														
	50	75	100	125	150	175	200	250	300	350	400	500	600	
Cópula Normal	$\tau = 0,25$	0,0129	0,0084	0,0061	0,0049	0,0042	0,0035	0,0031	0,0025	0,0021	0,0018	0,0015	0,0012	0,0011
	$\tau = 0,5$	0,0032	0,0022	0,0016	0,0013	0,0011	0,0010	0,0008	0,0007	0,0006	0,0005	0,0004	0,0003	0,0003
	$\tau = 0,75$	0,0002	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Cópula t de Student	$\tau = 0,25$	0,0293	0,0167	0,0121	0,0094	0,0077	0,0066	0,0058	0,0045	0,0037	0,0032	0,0029	0,0023	0,0019
	$\tau = 0,5$	0,0112	0,0064	0,0044	0,0034	0,0027	0,0024	0,0020	0,0015	0,0013	0,0011	0,0010	0,0008	0,0007
	$\tau = 0,75$	0,0013	0,0006	0,0004	0,0003	0,0003	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001	0,0000	0,0000
Cópula de Clayton	$\tau = 0,25$	0,0961	0,0916	0,0878	0,0853	0,0837	0,0829	0,0819	0,0810	0,0802	0,0798	0,0796	0,0791	0,0789
	$\tau = 0,5$	1,7456	1,7397	1,7340	1,7338	1,7324	1,7327	1,7323	1,7320	1,7316	1,7323	1,7319	1,7317	1,7321
	$\tau = 0,75$	26,1560	26,1529	26,1562	26,1522	26,1545	26,1531	26,1537	26,1537	26,1541	26,1567	26,1546	26,1560	26,1540
Cópula de Gumbel	$\tau = 0,25$	0,8194	0,8479	0,8618	0,8721	0,8760	0,8803	0,8806	0,8879	0,8890	0,8886	0,8906	0,8926	0,8935
	$\tau = 0,5$	1,6732	1,6799	1,6813	1,6851	1,6859	1,6872	1,6869	1,6876	1,6881	1,6878	1,6880	1,6885	1,6878
	$\tau = 0,75$	9,5036	9,5061	9,5059	9,5056	9,5052	9,5051	9,5062	9,5048	9,5052	9,5051	9,5051	9,5048	9,5046
Cópula de Frank	$\tau = 0,25$	4,2995	4,3013	4,2862	4,2884	4,2847	4,2826	4,2836	4,2787	4,2769	4,2777	4,2757	4,2728	4,2753
	$\tau = 0,5$	26,6485	26,6131	26,6001	26,5876	26,5813	26,5779	26,5699	26,5633	26,5602	26,5601	26,5618	26,5599	26,5569
	$\tau = 0,75$	177,3769	177,3285	177,3288	177,3192	177,3147	177,3071	177,3125	177,3081	177,1100	177,3147	177,3107	177,3111	177,3108
Cópula de Joe	$\tau = 0,25$	1,5139	1,5011	1,4926	1,4877	1,4839	1,4772	1,4750	1,4684	1,4662	1,4645	1,4636	1,4603	1,4592
	$\tau = 0,5$	4,7919	4,7666	4,7543	4,7512	4,7455	4,7375	4,7354	4,7341	4,7345	4,7329	4,7312	4,7306	4,7330
	$\tau = 0,75$	34,8185	34,8033	34,7940	34,7873	34,7882	34,7868	34,7858	34,7835	34,7827	34,7841	34,7852	34,7835	34,7840

Tabela 2.4.9: Medidas de EQM das famílias de cópulas estudadas BIC

		Tamanho amostral												
		50	75	100	125	150	175	200	250	300	350	400	500	600
Cópula Normal	$\tau = 0,25$	0,0132	0,0084	0,0061	0,0050	0,0042	0,0035	0,0030	0,0025	0,0021	0,0018	0,0016	0,0013	0,0011
	$\tau = 0,5$	0,0032	0,0022	0,0016	0,0013	0,0011	0,0010	0,0008	0,0007	0,0006	0,0005	0,0004	0,0003	0,0003
	$\tau = 0,75$	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Cópula t de Student	$\tau = 0,25$	0,0349	0,0190	0,0136	0,0102	0,0082	0,0070	0,0061	0,0046	0,0038	0,0032	0,0029	0,0023	0,0019
	$\tau = 0,5$	0,0142	0,0078	0,0051	0,0038	0,0029	0,0026	0,0021	0,0016	0,0013	0,0011	0,0010	0,0008	0,0007
	$\tau = 0,75$	0,0017	0,0008	0,0005	0,0004	0,0003	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000
cópula de Clayton	$\tau = 0,25$	0,0962	0,0916	0,0878	0,0853	0,0836	0,0828	0,0819	0,0810	0,0802	0,0798	0,0796	0,0790	0,0789
	$\tau = 0,5$	1,7451	1,7395	1,7338	1,7337	1,7323	1,7327	1,7322	1,7321	1,7316	1,7324	1,7319	1,7317	1,7321
	$\tau = 0,75$	26,1553	26,1528	26,1562	26,1522	26,1545	26,1531	26,1537	26,1537	26,1541	26,1567	26,1546	26,1560	26,1540
Cópula de Gumbel	$\tau = 0,25$	0,8230	0,8495	0,8640	0,8735	0,8768	0,8811	0,8818	0,8882	0,8892	0,8889	0,8909	0,8926	0,8936
	$\tau = 0,5$	1,6749	1,6804	1,6814	1,6849	1,6859	1,6869	1,6868	1,6873	1,6878	1,6878	1,6879	1,6884	1,6878
	$\tau = 0,75$	9,5034	9,5058	9,5056	9,5053	9,5050	9,5048	9,5061	9,5047	9,5050	9,5051	9,5052	9,5048	9,5046
Cópula de Frank	$\tau = 0,25$	4,3024	4,3026	4,2873	4,2893	4,2852	4,2832	4,2838	4,2789	4,2797	4,2776	4,2758	4,2727	4,2752
	$\tau = 0,5$	26,6495	26,6136	26,6002	26,5873	26,5811	26,5776	26,5694	26,5630	26,5599	26,5600	26,5617	26,5599	26,5569
	$\tau = 0,75$	177,3761	177,3280	177,3283	177,3190	177,3145	177,3069	177,3125	177,3081	177,3110	177,3143	177,3108	177,3111	177,3108
Cópula de Joe	$\tau = 0,25$	1,5158	1,5025	1,4931	1,4879	1,4842	1,4774	1,4751	1,4684	1,4662	1,4645	1,4636	1,4603	1,4592
	$\tau = 0,5$	4,7920	4,7667	4,7543	4,7512	4,7455	4,7375	4,7353	4,7341	4,7345	4,7329	4,7312	4,7306	4,7330
	$\tau = 0,75$	34,8186	34,8032	34,7940	34,7873	34,8820	34,7868	34,7858	34,7835	34,7827	34,7841	34,7852	34,7835	34,7840

2.5 Considerações finais

Nota-se que, tanto o critério de seleção de Akaike (AIC) quanto o Bayesiano (BIC) se mostraram satisfatório para selecionar modelos de cópulas dentre os analisados. Inclusive para amostras de tamanho próximo de 200 já se notou uma taxa de acerto bastante alta para quase todos os modelos analisados, principalmente quando as medidas de dependência são maiores ($\tau = 0,50$ e $0,75$).

No entanto, foram identificados problemas na utilização dos critérios, sobretudo na seleção de cópulas para amostras pequenas ($n \leq 100$), combinadas com medida de dependência também pequena ($\tau = 0,25$), além de vários modelos apresentarem valores elevados para o EQM referente aos parâmetros estimados das cópulas. Para as séries que apresentam alta dependência ($\tau = 0,75$) notou-se que amostras de tamanho inferior a 100 já foram suficientes, com exceção da cópula t de Student que apresentou a menor taxa de acerto dentre as comparadas, sendo necessário amostras de tamanhos maiores para atingir a taxa de acerto definida, de 90%.

É mais frequente observar a utilização da metodologia de cópulas em problemas com dados registrados com frequências diários e semanais. Contudo, certos problemas das áreas econômicas e agrárias, por exemplo, possuem dados de frequências menores e com poucos registros. Diversas séries históricas não possuem tamanho amostral maiores que 200, por exemplo. As séries de cotação de algumas commodities que, geralmente, apresentam registros históricos mensais. Outras séries têm observações trimestrais (PIB de um país, resultados contábeis de algumas empresas listadas na bolsa) e mensal (inflação, taxa de desemprego e empregos gerados). Desta forma, os resultados deste estudo permitem que o pesquisador se valha de modelos de cópulas também quando se tem séries com número de observações pequenas, desde que a seleção seja feita de forma cuidadosa, no sentido de evitar má identificação, em

especial para séries com menos de ($n \leq 200$) observações.

Nesse sentido, este estudo abre espaço para que novos critérios sejam analisados e empregados para a solução de problemas de escolhas de cópulas em amostras de tamanho pequeno combinado com medidas de dependência baixas.

Agradecimentos

A pesquisa do autor correspondente foi financiada pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Brasil, à qual o autor agradece.

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6):716–723.
- Beaudoin, D. e Lakhali-Chaieb, L. (2008). Archimedean copula model selection under dependent truncation. Statistics in medicine, 27(22):4440–4454.
- Cherubini, U., Luciano, E. e Vecchiato, W. (2004). Copula methods in finance. John Wiley & Sons.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(1):141–151.
- Dissmann, J., Brechmann, E. C., Czado, C. e Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59:52–69.
- dos Santos Silva, R. e Lopes, H. F. (2008). Copula, marginal distributions and model selection: a bayesian note. Statistics and Computing, 18(3):313–320.
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R. e Jermain, L. S. (2020). Sensitivity and specificity of information criteria. Briefings in bioinformatics, 21(2):553–565.
- Embrechts, P., McNeil, A. e Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond, 1:176–223.
- Emiliano, P. C., Vivanco, M. J. e De Menezes, F. S. (2014). Information criteria: How do they behave in different models? Computational Statistics & Data Analysis, 69:141–153.
- Frank, M. J. (1979). On the simultaneous associativity of (x, y) and $x+y - f(x, y)$. Aequationes mathematicae, 19(1):194–226.
- Grønneberg, S. e Hjort, N. L. (2014). The copula information criteria. Scandinavian Journal of Statistics, 41(2):436–459.
- Gülöksüz, Ç. T. (2016). Comparison of some selection criteria for selecting bivariate archimedean copulas. Afyon kocatepe üniversitesi fen ve mühendislik bilimleri dergisi, 16(2):250–255.
- Gumbel, E. J. (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55(292):698–707.

- Huard, D., Evin, G. e Favre, A.-C. (2006). Bayesian copula selection. Computational Statistics & Data Analysis, 51(2):809–822.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. Journal of multivariate analysis, 46(2):262–282.
- Joe, H. (1997). Multivariate models and multivariate dependence concepts. CRC Press.
- Jordanger, L. A. e Tjøstheim, D. (2014). Model selection of copulas: Aic versus a cross validation copula information criterion. Statistics & Probability Letters, 92:249–255.
- Kazianka, H. e Pilz, J. (2011). Bayesian spatial modeling and interpolation using copulas. Computers & Geosciences, 37(3):310–319.
- Ko, V. e Hjort, N. L. (2019). Copula information criterion for model selection with two-stage maximum likelihood estimation. Econometrics and Statistics, 12:167–180.
- Ko, V., Hjort, N. L. e Hobæk Haff, I. (2019). Focused information criteria for copulas. Scandinavian Journal of Statistics, 46(4):1117–1140.
- Kullback, S. e Leibler, R. A. (1951). On information and sufficiency. The annals of mathematical statistics, 22(1):79–86.
- Morettin, P. A. e Bussab, W. O. (2017). Estatística básica. Saraiva Educação SA.
- Morettin, P. A. e Toloi, C. (2006). Análise de séries temporais. Em Análise de séries temporais, página 538.
- Nabaei, S., Sharafati, A., Yaseen, Z. M. e Shahid, S. (2019). Copula based assessment of meteorological drought characteristics: regional investigation of iran. Agricultural and Forest Meteorology, 276:107611.
- Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M. et al. (2019). Vinecopula: statistical inference of vine copulas. R package version, 2(0).
- Nelsen, R. B. (2006). An introduction to copulas. Springer, New York.
- Nelsen, R. B. (2007). An introduction to copulas. Springer Science & Business Media.
- Perez-Rodríguez, J. V., Ledesma-Rodríguez, F. e Santana-Gallego, M. (2015). Testing dependence between gdp and tourism's growth rates. Tourism Management, 48:268–282.
- Reboredo, J. C. (2012). Do food and oil prices co-move? Energy policy, 49:456–467.
- Resende, A. G. e Candido, O. (2015). Dependence analysis of ethanol, sugar, oil, brl/usd exchange rate and bovespa: A vine copula approach. Economia Aplicada, 19:455–480.
- Rossi, J. L., Ehlers, R. S., Andrade Filho, M. G. d. et al. (2012). Copula-garch model selection: a bayesian approach.

- Sato, M., Ichiki, K. e Takeuchi, T. T. (2010). Precise estimation of cosmological parameters using a more accurate likelihood function. Physical Review Letters, 105(25):251301.
- Schoelzel, C. e Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research—introduction to the copula approach. Nonlinear Processes in Geophysics, 15(5):761–772.
- Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, páginas 461–464.
- Serra, T. e Gil, J. M. (2013). Price volatility in food markets: can stock building mitigate price fluctuations? European Review of Agricultural Economics, 40(3):507–528.
- Sklar, A. (1959). Fonctions de répartition à n dimension et leurs marges. Université Paris, 8.
- Swift, A., Heale, R. e Twycross, A. (2020). What are sensitivity and specificity? Evidence-Based Nursing, 23(1):2–4.
- Tang, J., Ramos, V., Cang, S. e Sriboonchitta, S. (2017). An empirical study of inbound tourism demand in china: a copula-garch approach. Journal of Travel & Tourism Marketing, 34(9):1235–1246.
- Tovar Cuevas, J. R., Portilla Yela, J. e Achcar, J. A. (2019). A method to select bivariate copula functions. Revista Colombiana de Estadística, 42(1):61–80.

Conclusão geral

Objetivou-se com este trabalho fazer uma vasta revisão acerca de cópulas, que se mostrou um método mais adequado para calcular a dependência entre duas ou mais séries de tempo. A princípio, o interesse consistia apenas em mensurar a dependência entre as séries históricas de algumas commodities de uma determinada região do Brasil. No entanto, durante o desenvolvimento da pesquisa detectou-se um ponto de grandes dúvidas no processo de seleção de uma cópula.

Desta forma, o segundo trabalho foi planejado e posteriormente desenvolvido, objetivando trazer mais discussão e suprir algumas lacunas ao processo de seleção de um modelo de cópula, via AIC e BIC, para diversos cenários possíveis. Ambos os critérios se mostraram interessantes no processo de seleção das cópulas, não sendo claro a vantagem de um se comparado ao outro.

O bom ajustes dos modelos aplicados no Capítulo 1 deste trabalho indicaram uma dependência fraca, que é bastante comum em diversas séries históricas. Com isso, o tamanho amostral e a medida de dependência baixa foi o principal foco do Capítulo 2 desta tese, o que não tinha sido encontrado na literatura até então com tamanha exploração.

Com este trabalho, conclui-se que as cópulas se mostram uma ferramenta bastante interessante quando o objetivo de um estudo é mensurar a medida de dependência que há entre duas ou mais séries históricas, sobretudo em alguns cenários que foram melhores esclarecidos no Capítulo 2. Assim, acredita-se encorajar mais pesquisadores a se valer de tais métodos para investigações futuras.

Referências Bibliográficas

- Abitante, K. G. (2008). Co-integração entre os mercados spot e futuro: evidências dos mercados de boi gordo e soja. Revista de Economia e Sociologia Rural, 46:75–96.
- Adami, M., Rudorff, B. F. T., Freitas, R. M., Aguiar, D. A., Sugawara, L. M. e Mello, M. P. (2012). Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in brazil. Sustainability, 4(4):574–585.
- Aguiar, D. A., Rudorff, B. F. T., Silva, W. F., Adami, M. e Mello, M. P. (2011). Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in são paulo state, brazil. Remote Sensing, 3(12):2682–2703.
- Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6):716–723.
- Aloui, R., Aïssa, M. S. B. e Nguyen, D. K. (2011). Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? Journal of Banking & Finance, 35(1):130–141.
- Beaudoin, D. e Lakhali-Chaieb, L. (2008). Archimedean copula model selection under dependent truncation. Statistics in medicine, 27(22):4440–4454.
- Belik, W. (2015). A heterogeneidade e suas implicações para as políticas públicas no rural brasileiro. Revista de Economia e Sociologia Rural, 53(1):9–30.
- Bernardo, R., Lourenzani, W. L., Satolo, E. G. e Caldas, M. M. (2019). Analysis of the agricultural productivity of the sugarcane crop in regions of new agricultural expansions of sugarcane. Gestão & Produção, 26.
- Box, G. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
- Brasil (2009). Decreto nº 6.961, de 17 de setembro de 2009. dispõe sobre aprovação do zoneamento agroecológico para o cultivo da cana-de-açúcar no território brasileiro, além de orientar financiamento da produção. Relatório técnico.
- Brasil (2019). Decreto nº 10.084, de 5 de novembro de 2019. revogou o decreto nº 6.961. Relatório técnico.
- Breymann, W., Dias, A. e Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance.
- Caldas, M. M., Granco, G., Bishop, C., Kastens, J. e Brown, J. C. (2017). Effects of sugarcane ethanol expansion in the brazilian cerrado: Land use response in the new frontier. Bioenergy and Land Use Change, 231:39.

- Capaz, R. S., Carvalho, V. S. B. e Nogueira, L. A. H. (2013). Impact of mechanization and previous burning reduction on ghg emissions of sugarcane harvesting operations in brazil. Applied Energy, 102:220–228.
- Casella, G. e Berger, R. L. (2010). Inferência estatística. São Paulo: Cengage Learning.
- Chen, S.-T., Kuo, H.-I. e Chen, C.-C. (2010). Modeling the relationship between the oil price and global food prices. Applied Energy, 87(8):2517–2525.
- Cherubin, M. R., Franco, A. L. C., Cerri, C. E. P., da Silva Oliveira, D. M., Davies, C. A. e Cerri, C. C. (2015). Sugarcane expansion in brazilian tropical soils—effects of land use change on soil chemical attributes. Agriculture, Ecosystems & Environment, 211:173–184.
- Cherubini, U., Luciano, E. e Vecchiato, W. (2004). Copula methods in finance. John Wiley & Sons.
- Chollete, L., de la Pena, V. e Lu, C.-C. (2011). International diversification: A copula approach. Journal of banking & finance, 35(2):403–417.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(1):141–151.
- Coelho, S. T., Goldemberg, J., Lucon, O. e Guardabassi, P. (2006). Brazilian sugarcane ethanol: lessons learned. Energy for Sustainable Development, 10(2):26–39.
- Cohn, A. S., Mosnier, A., Havlík, P., Valin, H., Herrero, M., Schmid, E., O'Hare, M. e Obersteiner, M. (2014). Cattle ranching intensification in brazil can reduce global greenhouse gas emissions by sparing land from deforestation. Proceedings of the National Academy of Sciences, 111(20):7236–7241.
- Conab - Companhia Nacional de Abastecimento (2019). Acompanhamento da safra brasileira: Cana-de-açúcar. 6(3). Safra 19/20. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em 15 de agosto de 2020. Relatório técnico.
- Conover, W. J. (1999). Practical nonparametric statistics, volume 350. john wiley & sons.
- Crago, C. L., Khanna, M., Barton, J., Giuliani, E. e Amaral, W. (2010). Competitiveness of brazilian sugarcane ethanol compared to us corn ethanol. Energy Policy, 38(11):7404–7415.
- de Araujo, G. J. F. e de Oliveira, S. V. W. B. (2021). Energy and environmental analysis of vinasse processing using internal circulation biodigesters and concentrators for different production ranges of ethanol and its potential impact on brazil. Environment, Development and Sustainability, 23(3):3130–3163.
- Defante, L. R., Vilpoux, O. F. e Sauer, L. (2020). Importance of the sugarcane industry in the formal employment in the state of mato grosso do sul during the period of 2008 to 2014. Revista de Economia e Sociologia Rural, 58.

- Dhillon, R. e von Wuehlisch, G. (2013). Mitigation of global warming through renewable biomass. Biomass and bioenergy, 48:75–89.
- Dickey, D. A. e Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a):427–431.
- Dissmann, J., Brechmann, E. C., Czado, C. e Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59:52–69.
- dos Santos Silva, R. e Lopes, H. F. (2008). Copula, marginal distributions and model selection: a bayesian note. Statistics and Computing, 18(3):313–320.
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R. e Jermin, L. S. (2020). Sensitivity and specificity of information criteria. Briefings in bioinformatics, 21(2):553–565.
- Embrechts, P., McNeil, A. e Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond, 1:176–223.
- Emiliano, P. C., Vivanco, M. J. e De Menezes, F. S. (2014). Information criteria: How do they behave in different models? Computational Statistics & Data Analysis, 69:141–153.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the econometric society, páginas 987–1007.
- Ferreira, M. E., Ferreira Jr, L. G., Latrubesse, E. M. e Miziara, F. (2016). Considerations about the land use and conversion trends in the savanna environments of central brazil under a geomorphological perspective. Journal of Land Use Science, 11(1):33–47.
- Ferreira, M. E., Ferreira Jr, L. G., Miziara, F. e Soares-Filho, B. S. (2013). Modeling landscape dynamics in the central brazilian savanna biome: future scenarios and perspectives for conservation. Journal of Land Use Science, 8(4):403–421.
- Figueira, S. R. F. (2020). Impactos dos preços e do crédito rural sobre a produção de cana-de-açúcar no estado de são paulo. Revista de Economia e Sociologia Rural, 58.
- Frank, M. J. (1979). On the simultaneous associativity off (x, y) and $x+y- f(x, y)$. Aequationes mathematicae, 19(1):194–226.
- Gandolfi, M. R. C., Jesus, C. M. e Gandolfi, P. E. (2019). Qualidade do emprego nas culturas de café, cana-de-açúcar e milho: estudo comparativo 2000/2010 entre mesorregiões de minas gerais. Revista de economia e sociologia rural, 57:63–80.
- Genest, C., Ghoudi, K. e Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3):543–552.

- Genest, C. e Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Em Annales de l'IHP Probabilités et statistiques, volume 44, páginas 1096–1127.
- Genest, C., Rémillard, B. e Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and economics, 44(2):199–213.
- Gilio, L. e de Moraes, M. A. F. D. (2016). Sugarcane industry's socioeconomic impact in são paulo, brazil: A spatial dynamic panel approach. Energy Economics, 58:27–37.
- Goldemberg, J. (2006). The ethanol program in brazil. Environmental Research Letters, 1(1):014008.
- Goldemberg, J., Mello, F. F., Cerri, C. E., Davies, C. A. e Cerri, C. C. (2014). Meeting the global demand for biofuels in 2021 through sustainable land use change policy. Energy Policy, 69:14–18.
- Graesser, J., Aide, T. M., Grau, H. R. e Ramankutty, N. (2015). Cropland/pastureland dynamics and the slowdown of deforestation in latin america. Environmental Research Letters, 10(3):034017.
- Granco, G., Caldas, M., Featherstone, A., Sant'Anna, A. C. e Bergtold, J. (2018a). Farmers' acreage responses to the expansion of the sugarcane ethanol industry: The case of goiás and mato grosso do sul, brazil. Em Land Allocation for Biomass Crops, páginas 103–123. Springer.
- Granco, G., Caldas, M. M., Bergtold, J. S. e Sant'Anna, A. C. (2017). Exploring the policy and social factors fueling the expansion and shift of sugarcane production in the brazilian cerrado. GeoJournal, 82(1):63–80.
- Granco, G., Sant'Anna, A. C., Bergtold, J. S. e Caldas, M. M. (2018b). Factors influencing ethanol mill location in a new sugarcane producing region in brazil. Biomass and bioenergy, 111:125–133.
- Grønneberg, S. e Hjort, N. L. (2014). The copula information criteria. Scandinavian Journal of Statistics, 41(2):436–459.
- Gülöksüz, Ç. T. (2016). Comparison of some selection criteria for selecting bivariate archimedean copulas. Afyon kocatepe üniversitesi fen ve mühendislik bilimleri dergisi, 16(2):250–255.
- Gumbel, E. J. (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55(292):698–707.
- Hausman, C. (2012). Biofuels and land use change: sugarcane and soybean acreage response in brazil. Environmental and Resource Economics, 51(2):163–187.
- Hofert, M., Kojadinovic, I., Maechler, M. e Yan, J. (2020). copula: Multivariate Dependence with Copulas. R package version 1.0-1.
- Huard, D., Evin, G. e Favre, A.-C. (2006). Bayesian copula selection. Computational Statistics & Data Analysis, 51(2):809–822.

- IBGE - Instituto Brasileiro de Geografia e Estatística (2011). Sistema IBGE de Recuperação Automática. Disponível em: <http://www.sidra.ibge.gov.br>. Acesso em 15 de junho de 2020. Relatório técnico.
- IPEA - Instituto de Pesquisa Econômica Aplicada - Ipeadata (2019). Dados macroeconômicos e regionais. Disponível em: <http://www.ipeadata.gov.br>. Acesso em 20 de outubro de 2019. Relatório técnico.
- Jarque, C. M. e Bera, A. K. (1987). A test for normality of observations and regression residuals. International Statistical Review/Revue Internationale de Statistique, páginas 163–172.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. Journal of multivariate analysis, 46(2):262–282.
- Joe, H. (1997). Multivariate models and multivariate dependence concepts. CRC Press.
- Jordanger, L. A. e Tjøstheim, D. (2014). Model selection of copulas: Aic versus a cross validation copula information criterion. Statistics & Probability Letters, 92:249–255.
- Kaltalioglu, M. e Soytas, U. (2009). Price transmission between world food, agricultural raw material, and oil prices. Em GBATA International Conference Proceedings, páginas 596–603.
- Katchova, A. L. e Sant’Anna, A. C. (2019). Impact of ethanol plant location on corn revenues for us farmers. Sustainability, 11(22):6512.
- Kazianka, H. e Pilz, J. (2011). Bayesian spatial modeling and interpolation using copulas. Computers & Geosciences, 37(3):310–319.
- Khanna, M., Louviere, J. e Yang, X. (2017). Motivations to grow energy crops: the role of crop and contract attributes. Agricultural Economics, 48(3):263–277.
- Ko, V. e Hjort, N. L. (2019). Copula information criterion for model selection with two-stage maximum likelihood estimation. Econometrics and Statistics, 12:167–180.
- Ko, V., Hjort, N. L. e Hobæk Haff, I. (2019). Focused information criteria for copulas. Scandinavian Journal of Statistics, 46(4):1117–1140.
- Kullback, S. e Leibler, R. A. (1951). On information and sufficiency. The annals of mathematical statistics, 22(1):79–86.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P. e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of econometrics, 54(1-3):159–178.
- Ljung, G. M. e Box, G. E. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2):297–303.
- Lourenzani, W. L. e Caldas, M. M. (2014). Mudanças no uso da terra decorrentes da expansão da cultura da cana-de-açúcar na região oeste do estado de são paulo. Ciência Rural, 44:1980–1987.

- Macedo, L. O. B. (2005). Modernização da pecuária de corte bovina no brasil e a importância do crédito rural. *Agroanalysis*, Rio de Janeiro, 25(6):35–36.
- Mann, M. L., Kaufmann, R. K., Bauer, D. M., Gopal, S., Nomack, M., Womack, J. Y., Sullivan, K. e Soares-Filho, B. S. (2014). Pasture conversion and competitive cattle rents in the amazon. *Ecological economics*, 97:182–190.
- Manoel, A. A. S., Moraes, M. B. d. C., Santos, D. F. L. e Neves, M. F. (2018). Determinants of corporate cash holdings in times of crisis: insights from brazilian sugarcane industry private firms. *International Food and Agribusiness Management Review*, páginas 201–218.
- Manzatto, C. V., Assad, E. D., BACA, J. F. M., Zaroni, M. J. e Pereira, S. E. M. (2009). Zoneamento agroecológico da cana-de-açúcar: expandir a produção, preservar a vida, garantir o futuro. *Embrapa Solos-Documents (INFOTECA-E)*.
- Marin, F. R. (2016). Understanding sugarcane production, biofuels, and market volatility in brazil—a research perspective.
- Martinelli, L. A. e Filoso, S. (2008). Expansion of sugarcane ethanol production in brazil: environmental and social challenges. *Ecological applications*, 18(4):885–898.
- McManus, C., Barcellos, J. O. J., Formenton, B. K., Hermuche, P. M., Carvalho Jr, O. A. d., Guimarães, R., Gianezini, M., Dias, E. A., Lampert, V. d. N., Zago, D. et al. (2016). Dynamics of cattle production in brazil. *PloS one*, 11(1):e0147138.
- Moraes, M. A. F. D. e Zilberman, D. (2014). *Production of Ethanol from Sugarcane in Brazil: from state intervention to a free market*, volume 43. Springer Science & Business Media.
- Morettin, P. A. e Bussab, W. O. (2017). *Estatística básica*. Saraiva Educação SA.
- Morettin, P. A. e Tolo, C. (2006). Análise de séries temporais. Em *Análise de séries temporais*, página 538.
- Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Arai, E., del Bon Espirito-Santo, F., Freitas, R. e Morisette, J. (2006). Cropland expansion changes deforestation dynamics in the southern brazilian amazon. *Proceedings of the National Academy of Sciences*, 103(39):14637–14641.
- Mueller, S. A., Anderson, J. E. e Wallington, T. J. (2011). Impact of biofuel production and other supply and demand factors on food price increases in 2008. *Biomass and bioenergy*, 35(5):1623–1632.
- Mutuc, M., Pan, S. e Hudson, D. (2011). Response of cotton to oil price shocks. *Agricultural Economics Review*, 12(389-2016-23468).
- Nabaei, S., Sharafati, A., Yaseen, Z. M. e Shahid, S. (2019). Copula based assessment of meteorological drought characteristics: regional investigation of iran. *Agricultural and Forest Meteorology*, 276:107611.
- Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M. et al. (2019). Vinecopula: statistical inference of vine copulas. *R package version*, 2(0).

- Nassar, A. M. (2009). Brazil as an agricultural and agroenergy superpower. Brazil as an economic superpower, páginas 55–80.
- Nassar, A. M., Rudorff, B. F., Antoniazzi, L. B., Aguiar, D. d., Bacchi, M. R., Adami, M. et al. (2008). Prospects of the sugarcane expansion in brazil: impacts on direct and indirect land use changes. Sugarcane ethanol: contributions to climate change mitigation and the environment, páginas 63–93.
- Nazlioglu, S. e Soytas, U. (2011). World oil prices and agricultural commodity prices: Evidence from an emerging market. Energy Economics, 33(3):488–496.
- Nelsen, R. B. (2006). An introduction to copulas. Springer, New York.
- Nelsen, R. B. (2007). An introduction to copulas. Springer Science & Business Media.
- Neves, M., Trombin, V. e Consoli, M. (2011). The sugar energy map of brazil in ethanol and bioelectricity: sugarcane in the future of the energy matrix. So Paulo: Unica. Nicholson, FA, Smith, SR, Alloway, BJ, Carlton-Smith, C.
- Neves, M. F., Waack, R. S. e Marino, M. K. (1998). Sistema agroindustrial da cana-de-açúcar: caracterização das transações entre empresas de insumos, produtores de cana e usinas. Em Congresso Brasileiro de Economia e Sociologia Rural, volume 36, páginas 559–572.
- Ning, C. (2010). Dependence structure between the equity market and the foreign exchange market—a copula approach. Journal of International Money and Finance, 29(5):743–759.
- Nogueira, L. A. H., Moreira, J. R., Schuchardt, U. e Goldemberg, J. (2013). The rationality of biofuels. Energy Policy, 61:595–598.
- Nuñez, H. M., Önal, H. e Khanna, M. (2013). Land use and economic effects of alternative biofuel policies in brazil and the united states. Agricultural Economics, 44(4-5):487–499.
- Nyko, D., Valente, M. S., Milanez, A. Y., Tanaka, A. K. R. e Rodrigues, A. V. P. (2013). A evolução das tecnologias agrícolas do setor sucroenergético: estagnação passageira ou crise estrutural? BNDES Setorial, n. 37, mar. 2013, p 399-442.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International economic review, 47(2):527–556.
- Pereira, C. N. e Silveira, J. M. F. J. (2016). Análise exploratória da eficiência produtiva das usinas de cana-de-açúcar na região centro-sul do brasil. Revista de Economia e Sociologia Rural, 54:147–166.
- Perez-Rodríguez, J. V., Ledesma-Rodríguez, F. e Santana-Gallego, M. (2015). Testing dependence between gdp and tourism's growth rates. Tourism Management, 48:268–282.
- Phillips, P. C. e Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2):335–346.

- Rada, N. (2013). Assessing brazil's cerrado agricultural miracle. Food Policy, 38:146–155.
- Reboredo, J. C. (2012). Do food and oil prices co-move? Energy policy, 49:456–467.
- Resende, A. G. e Candido, O. (2015). Dependence analysis of ethanol, sugar, oil, brl/usd exchange rate and bovespa: A vine copula approach. Economia Aplicada, 19:455–480.
- Rodrigues, D. M. T. e Miziara, F. (2008). Expansão da fronteira agrícola: a intensificação da pecuária bovina no estado de goiás.
- Rossi, J. L., Ehlers, R. S., Andrade Filho, M. G. d. et al. (2012). Copula-garch model selection: a bayesian approach.
- Saiani, C. C. S. e Perosa, B. B. (2016). Saúde respiratória e mecanização da colheita da cana-de-açúcar nos municípios paulistas: a importância do protocolo agroambiental. Revista de Economia e Sociologia Rural, 54:29–50.
- Sano, E. E., Rosa, R., Brito, J. L. e Ferreira, L. G. (2010). Land cover mapping of the tropical savanna region in brazil. Environmental monitoring and assessment, 166(1):113–124.
- Sant'Anna, A. C., Granco, G., Bergtold, J. e Caldas, M. M. (2015). Os desafios da expansão da cana-de-açúcar: como pensam e agem arrendatários e produtores?
- Sant'Anna, A. C., Shanoyan, A., Bergtold, J. S., Caldas, M. M. e Granco, G. (2016). Ethanol and sugarcane expansion in brazil: what is fueling the ethanol industry? International Food and Agribusiness Management Review, 19(1030-2017-2132):163–182.
- Sato, M., Ichiki, K. e Takeuchi, T. T. (2010). Precise estimation of cosmological parameters using a more accurate likelihood function. Physical Review Letters, 105(25):251301.
- Schoelzel, C. e Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research—introduction to the copula approach. Nonlinear Processes in Geophysics, 15(5):761–772.
- Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, páginas 461–464.
- Serra, T. e Gil, J. M. (2013). Price volatility in food markets: can stock building mitigate price fluctuations? European Review of Agricultural Economics, 40(3):507–528.
- Silva, A. A. e Miziara, F. (2011). Sucroalcohol sector and agricultural frontier expansion in the goiás state, brazil. Pesquisa Agropecuária Tropical, 41(3).
- Sklar, A. (1959). Fonctions de répartition à n dimension et leurs marges. Université Paris, 8.

- Strassburg, B. B., Latawiec, A. E., Barioni, L. G., Nobre, C. A., Da Silva, V. P., Valentim, J. F., Vianna, M. e Assad, E. D. (2014). When enough should be enough: Improving the use of current agricultural lands could meet production demands and spare natural habitats in brazil. Global Environmental Change, 28:84–97.
- Swift, A., Heale, R. e Twycross, A. (2020). What are sensitivity and specificity? Evidence-Based Nursing, 23(1):2–4.
- Tang, J., Ramos, V., Cang, S. e Sriboonchitta, S. (2017). An empirical study of inbound tourism demand in china: a copula-garch approach. Journal of Travel & Tourism Marketing, 34(9):1235–1246.
- Team, R. C. et al. (2022). R: A language and environment for statistical computing.
- Tovar Cuevas, J. R., Portilla Yela, J. e Achcar, J. A. (2019). A method to select bivariate copula functions. Revista Colombiana de Estadística, 42(1):61–80.
- Trombeta, N. d. C. e Caixeta, J. V. (2017). Potencial e disponibilidade de biomassa de cana-de-açúcar na região centro-sul do brasil: indicadores agroindustriais. Revista de Economia e Sociologia Rural, 55:479–496.
- UDOP - União Nacional dos Produtores de Bioenergia (2019). Valores de ATR e o preço da tonelada de cana-de-açúcar - Consecana do Estado de São Paulo. Disponível em: https://udop.com.br/cana/tabela_consecana_saopaulo.pdf. Acesso em 04 de novembro de 2019. Relatório técnico.
- Valdes, C., Hjort, K. e Seeley, R. (2016). Brazil's agricultural land use and trade: effects of changes in oil prices and ethanol demand. Relatório técnico.
- Van den Wall Bake, J., Junginger, M., Faaij, A., Poot, T. e Walter, A. (2009). Explaining the experience curve: Cost reductions of brazilian ethanol from sugarcane. Biomass and bioenergy, 33(4):644–658.
- Walker, R., Moore, N. J., Arima, E., Perz, S., Simmons, C., Caldas, M., Vergara, D. e Bohrer, C. (2009). Protecting the amazon with protected areas. Proceedings of the National Academy of Sciences, 106(26):10582–10586.