



Universidade do Estado do Rio de Janeiro

Centro de Ciências Sociais

Faculdade de Ciências Econômicas

Gabriel de Almeida Arruda

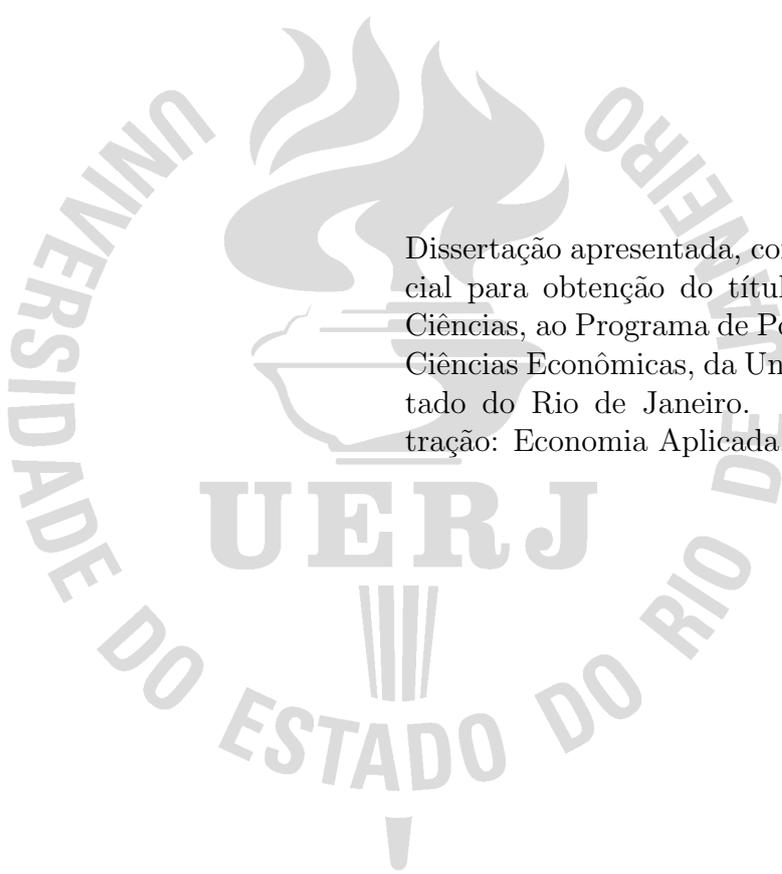
**Avaliação da estrutura de dependência entre Fundos de
Investimentos e variáveis macroeconômicas: Uma aplicação da
teoria de cópulas**

Rio de Janeiro

2024

Gabriel de Almeida Arruda

**Avaliação da estrutura de dependência entre Fundos de Investimentos e
variáveis macroeconômicas: Uma aplicação da teoria de cópulas**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Ciências, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de concentração: Economia Aplicada.

Orientador: Profa. Dra. Daiane Rodrigues dos Santos

Rio de Janeiro

2024

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CCS/B

A779 Arruda, Gabriel de Almeida
Avaliação da estrutura de dependência entre Fundos de Investimentos e variáveis macroeconômicas: uma aplicação da teoria de cópula / Gabriel de Almeida Arruda .– 2024.
87 f.

Orientadora: Daiane Rodrigues dos Santos.
Dissertação (Mestrado em Ciências Econômicas) – Universidade do Estado do Rio de Janeiro, Faculdade de em Ciências Econômicas.

1. Fundos de investimento – Teses. 2. Cópulas (Estatística matemática) - Teses. I. Santos, Daiane Rodrigues dos. II. Universidade do Estado do Rio de Janeiro. Faculdade de Ciências Econômicas. III. Título.

CDU 336.76

Bibliotecária: Lucia Andrade - CRB7/5272

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Gabriel de Almeida Arruda

**Avaliação da estrutura de dependência entre Fundos de Investimentos e
variáveis macroeconômicas: Uma aplicação da teoria de cópulas**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Ciências, ao Programa de Pós-Graduação em Ciências Econômicas, da Universidade do Estado do Rio de Janeiro. Área de concentração: Economia Aplicada.

Aprovada em 23 de Fevereiro de 2024.

Banca Examinadora:

Profa. Dra. Daiane Rodrigues dos Santos (Orientador)
Faculdade de Ciências Econômicas - UERJ

Prof. Ph.D. Elcyon Caiado Rocha Lima
Faculdade de Ciências Econômicas - UERJ

Prof. Dr. Marco Aurélio dos Santos Sanfins
Universidade Federal Fluminense

Rio de Janeiro

2024

RESUMO

DE ALMEIDA ARRUDA, Gabriel. *Avaliação da estrutura de dependência entre Fundos de Investimentos e variáveis macroeconômicas*: Uma aplicação da teoria de cópulas. 2024. 87 f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

A compreensão e o monitoramento das variáveis macroeconômicas e seus impactos nos fundos de investimento configuram-se como ferramentas de suma importância para a gestão eficaz desses instrumentos financeiros. Essa assertiva se fundamenta na capacidade de tais variáveis em influenciar, de maneira significativa, o desempenho dos fundos. A presente dissertação aborda essa relação complexa, utilizando a teoria de cópulas como instrumento analítico, realçando sua importância em um contexto marcado por volatilidades e incertezas econômicas. A partir de uma revisão sistemática da literatura e uma análise bibliométrica, identificou-se um crescimento no interesse pelo uso de cópulas em finanças, particularmente pós-crise financeira. No desenvolvimento teórico, o estudo delinea os conceitos fundamentais das cópulas, abordando suas principais famílias, medidas de dependência e concordância, e métodos de estimação e testes. Essa base teórica permite uma análise mais detalhada e flexível da dependência entre variáveis, ultrapassando as limitações impostas pelas distribuições marginais. O presente trabalho utilizou modelos de cópula para examinar a relação entre os retornos de cinco fundos de investimento brasileiros selecionados (BTG Discovery, JGP Hedge, Vokin Everest, Opportunity e Gávea Macro) e cinco variáveis macroeconômicas (Ibovespa, IMA-B, CDS, IBC-Br e Câmbio), ao longo do período de 2008 a início de 2023. Através de testes de adequação e simulações, avaliou-se a eficácia dos modelos em capturar a estrutura de dependência entre essas variáveis. Os resultados apontaram para uma relação entre os retornos dos fundos e o desempenho das variáveis macroeconômicas, com uma tendência clara de melhores desempenhos em cenários econômicos favoráveis. A sensibilidade dos fundos às condições de mercado e à atividade econômica é evidente, reforçando a importância de considerar essas variáveis na tomada de decisões estratégicas de investimento. Os resultados obtidos nesta pesquisa podem subsidiar gestores de fundos e investidores na avaliação do impacto de cenários macroeconômicos adversos, permitindo aprimorar estratégias de alocação de ativos e gerenciamento de riscos.

Palavras-chave: Cópulas; Fundos de investimento; Variáveis macroeconômicas; Estrutura de dependência.

ABSTRACT

DE ALMEIDA ARRUDA, Gabriel. *Assessment of the dependency structure between Investment Funds and macroeconomic variables: An application of copula theory*. 2024. 87 f. Dissertação (Mestrado em Ciências Econômicas) – Faculdade de Ciências Econômicas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

The understanding and monitoring of macroeconomic variables and their impacts on investment funds are recognized as tools of utmost importance for the effective management of these financial instruments. This assertion is based on the ability of such variables to significantly influence the performance of funds. This dissertation addresses this complex relationship, utilizing copula theory as an analytical tool, highlighting its importance in a context marked by economic volatilities and uncertainties. Through a systematic literature review and a bibliometric analysis, an increase in interest in the use of copulas in finance, particularly post-financial crisis, was identified. In the theoretical development, the study outlines the fundamental concepts of copulas, discussing their main families, measures of dependence and concordance, and estimation methods and tests. This theoretical framework allows for a more detailed and flexible analysis of dependence between variables, overcoming the limitations imposed by marginal distributions. This work used copula models to examine the relationship between the returns of five selected Brazilian investment funds (BTG Discovery, JGP Hedge, Vokin Everest, Opportunity, and Gávea Macro) and five macroeconomic variables (Ibovespa, IMA-B, CDS, IBC-Br, and Exchange Rate) over the period from 2008 to early 2023. Through adequacy tests and simulations, the effectiveness of the models in capturing the dependence structure among these variables was assessed. The results pointed to a relationship between the returns of the funds and the performance of macroeconomic variables, with a clear trend of better performances in favorable economic scenarios. The sensitivity of the funds to market conditions and economic activity is evident, reinforcing the importance of considering these variables in strategic investment decision-making. The findings of this research can provide fund managers and investors with insights into assessing the impact of adverse macroeconomic scenarios, allowing for the improvement of asset allocation strategies and risk management.

Keywords: Copulas; Investment funds; Macroeconomic variables; Dependence structure.

LISTA DE ILUSTRAÇÕES

Figura 1 - Evolução da produção anual.	17
Figura 2 - Construção do mapa temático	19
Figura 3 - Rede de co-citações	22
Figura 4 - Evolução das palavras-chaves por período	27
Figura 5 - Mapa temático para o período 2004-2016.	27
Figura 6 - Mapa temático para o período 2017-2020.	28
Figura 7 - Mapa temático para o período 2021-2024.	29
Figura 8 - Evolução do retorno das variáveis.	52
Figura 9 - Evolução do retorno dos Fundos.	53
Figura 10 - Distribuição do retorno do fundo Vokin Everest e IBC-Br.	55
Figura 11 - BTG Discovery - Densidade da copulas.	56
Figura 12 - JGP Hedge - Densidade da copulas.	56
Figura 13 - Vokin Everest - Densidade da copulas.	57
Figura 14 - Opportunity - Densidade da copulas.	57
Figura 15 - Gávea Macro - Densidade da copulas.	58
Figura 16 - BTG Discovery - Densidade das variáveis observadas e simuladas.	61
Figura 17 - JGP Hedge - Densidade das variáveis observadas e simuladas.	62
Figura 18 - Vokin Everest - Densidade das variáveis observadas e simuladas.	63
Figura 19 - Opportunity - Densidade das variáveis observadas e simuladas.	64
Figura 20 - Gávea Macro II - Densidade das variáveis observadas e simuladas.	65

LISTA DE TABELAS

Tabela 1 - Tipos de Publicações	16
Tabela 2 - Distribuição de revistas científicas.	17
Tabela 3 - Os 10 autores com mais citações.	20
Tabela 4 - Estatística descritiva dos retornos.	54
Tabela 5 - Parâmetros das copulas.	59
Tabela 6 - Probabilidades de ocorrência e retorno esperado do fundo BTG Discovery.	67
Tabela 7 - Probabilidades de ocorrência e retorno esperado do fundo JGP Hedge.	68
Tabela 8 - Probabilidades de ocorrência e retorno esperado do fundo Vokin Everest.	70
Tabela 9 - Probabilidades de ocorrência e retorno esperado do fundo Opportunity.	71
Tabela 10 - Probabilidades de ocorrência e retorno esperado do fundo Gávea Macro.	73
Tabela 11 - Intervalo das variáveis.	86

SUMÁRIO

	INTRODUÇÃO	9
1	REVISÃO SISTEMÁTICA DE LITERATURA	14
1.1	Revisão Sistemática da Literatura	14
1.1.1	<u>Análise Bibliométria</u>	17
1.1.1.1	Análise temática	18
1.2	Resultado	19
1.2.1	<u>Análise de citação</u>	19
1.2.2	<u>Análise de co-citações</u>	22
1.2.3	<u>Análise temática</u>	26
2	CÓPULAS	30
2.1	Análise de Dependência	31
2.1.1	<u>Teorema de Sklar</u>	32
2.1.2	<u>Limites de Fréchet-Hoeffding</u>	34
2.1.3	<u>Medidas de dependência e concordância</u>	35
2.1.3.1	Coeficiente de Correlação Linear	36
2.1.4	<u>Medidas de concordância</u>	37
2.1.4.1	Tau de Kendall	37
2.1.4.2	Rho de Sperman	38
2.2	Famílias de Cópulas	39
2.2.1	<u>Cópulas Elípticas</u>	39
2.2.1.1	Cópula Gaussiana	40
2.2.1.2	Cópula t-Student	40
2.2.2	<u>Cópulas Arquimedianas</u>	41
2.2.2.1	Cópula Gumbel-Hougaard	42
2.2.2.2	Cópula Frank	42
2.2.2.3	Cópula Clayton	43
2.3	Estimação	43
2.3.1	<u>Método de Máxima Verossimilhança (EMV)</u>	43
2.3.2	<u>Cópulas Empíricas</u>	44
2.4	Copula de sobrevivência	45
2.5	Copulas rotacionadas	45
2.6	Métricas de desempenho	46
2.7	Testes	46
2.7.1	<u>Kolmogorov-Smirnov</u>	46
2.7.2	<u>Teste de bondade do ajuste</u>	47

2.8	Cópu las - Avaliação da estrutura de dependência entre fundos de investimentos e variáveis macroeconômicas	47
3	RESULTADO	51
3.1	Análise dos resultados	51
3.1.1	<u>BTG Discovery</u>	66
3.1.2	<u>JGP Hedge</u>	67
3.1.3	<u>Vokin Everest</u>	68
3.1.4	<u>Opportunity</u>	70
3.1.5	<u>Gávea Macro</u>	71
	CONCLUSÃO	74
	REFERÊNCIAS	77
	ANEXO A – Pseudo-código	83
	ANEXO B – Tabela dos intervalos	86

INTRODUÇÃO

O estudo da relação entre fundos de investimento e variáveis macroeconômicas é de suma importância para a compreensão do comportamento dos mercados financeiros e para a tomada de decisões de investimento. Essa relação é complexa e envolve uma série de fatores que podem influenciar o desempenho dos fundos de investimento, incluindo taxas de juros, inflação, crescimento econômico (CHEN; ROLL; ROSS, 1986).

A década de 60 foi um período significativo na teoria de finanças, pois marcou o início dos estudos que buscavam modelar teoricamente o retorno esperado dos ativos. Foram desenvolvidos os modelos que formaram a base do *Capital asset pricing models* - CAPM (TREYNOR, 1961; SHARPE, 1964; LINTNER, 1965), que se tornou uma das principais contribuições. O CAPM consolidou a hipótese de que o retorno esperado de um ativo deve ser compensado pelo seu risco não diversificável, medida pelo beta em relação ao mercado. Embora tenha sido alvo de críticas, a relevância do CAPM é tamanha que, ainda nos dias atuais, é utilizado para calcular o custo de capital das empresas e mensurar a performance de gestores de fundos de investimentos. O modelo tornou-se uma referência na teoria financeira, e suas contribuições foram fundamentais para a consolidação de outras linhas de pesquisa, como a teoria das finanças comportamentais e a teoria dos mercados de capitais incompletos.

O CAPM foi desenvolvido a partir da teoria da carteira de Markowitz (1952). Esta teoria pressupõe que o investidor é avesso ao risco e seleciona sua carteira com base no risco e retorno, utilizando como medida estatística a média e a variância. Essa metodologia ficou conhecida como carteira de "média-variância", na qual cada investidor seleciona a carteira que oferece o melhor retorno dado uma variância mínima. Ainda que mantenha uma considerável relevância nos cursos de finanças, constata-se que o modelo de precificação de ativos financeiros (CAPM) é alvo de críticas substanciais, conforme sustentado por Fama e French (2004). Tais críticas decorrem do fato de que o modelo não se sustenta mediante evidências empíricas, bem como de hipóteses irrealistas, como aquela que pressupõe que os investidores se preocupam exclusivamente com o retorno médio e a variância em um único período de sua carteira, o que é extremamente limitado. Empiricamente, o CAPM não se mostrou capaz de explicar os retornos médios dos ativos, conforme constatado por Fama e French (1992). Além disso, a suposição de que apenas o beta do mercado constitui um bom fator de risco a ser considerado é falha, o que pode ser demonstrado mediante a proposta de Merton (1973) de uma extensão do CAPM, conhecida como *intertemporal capital asset pricing*, que demonstrou que um portfólio "ótimo" depende de várias variáveis distintas.

Embora o CAPM seja um modelo amplamente utilizado na teoria financeira, ele não é capaz de englobar todos os riscos aos quais um ativo pode estar sujeito. Com

efeito, a hipótese central do CAPM é de que o risco é medido apenas pelo beta em relação ao mercado, deixando de lado outros fatores que podem afetar o retorno de um ativo. Dessa forma, Ross (1976) desenvolveu o *Arbitrage pricing theory* (APT), que sugere que um ativo pode ter múltiplos fatores de risco que afetam seu preço de mercado. Esses fatores podem ser associados a variáveis macroeconômicas, como taxas de juros, inflação, mudanças no ambiente regulatório, entre outros. O APT pressupõe que, em um mercado eficiente, o preço de um ativo deve refletir seu valor intrínseco, ajustado pelos fatores de risco relevantes. Em outras palavras, o retorno de um ativo seria uma função linear dos seus fatores de risco, e os investidores seriam capazes de identificar e precificar esses fatores no mercado.

Um fundo de investimento é um veículo de investimento coletivo, no qual um gestor profissional seleciona e gerencia uma carteira de ativos em nome dos investidores que compram a cota do fundo. De acordo com Haddock, Reilly e Brown (2017), os fundos de investimentos podem oferecer aos investidores a oportunidade de obter exposição a uma ampla gama de ativos com gestão profissional mais assertiva e qualificada. Existem vários tipos de fundos de investimento disponíveis no mercado, cada um com características e objetivos distintos. Os fundos de renda fixa, por exemplo, investem em títulos de dívida, como títulos públicos e privados, com o objetivo de gerar renda para os investidores (JIANG; ZHOU, 2020). Por esse motivo são considerados investimentos de baixo risco e são indicados para investidores mais conservadores. Os Fundos de ações são outro exemplo, esses fundos investem em ações de empresas listadas em bolsa e são considerados investimentos de maior risco, mas com potencial de retorno mais elevado (CHEN et al., 2021). Os Fundos multimercado, por exemplo, investem em diversas classes de ativos, como renda fixa, ações, câmbio e *commodities*, com o objetivo de diversificar a carteira de investimentos (SHARPE, 1966; QAMAR; SINGH, 2016). Os fundos imobiliários investem em imóveis comerciais, como shoppings, escritórios e galpões logísticos, com o objetivo adicional, além da valorização da cota, de gerar renda por meio do aluguel desses imóveis. Outros Fundos são os cambiais, esses fundos investem em moedas estrangeiras, como dólar e euro, com o objetivo, por exemplo, de proteger o investidor contra a variação cambial ou aproveitar oportunidades de valorização dessas moedas (MALINDA; CHEN, 2022).

No campo da economia e das finanças, muito estudo e análise tem sido realizado sobre como as variáveis macroeconômicas afetam os fundos de investimentos. Para a tomada de decisões estratégicas, gestão de riscos e busca de retornos sólidos, é fundamental entender como essas variáveis afetam o desempenho e a rentabilidade dos fundos de investimentos. Ao direcionar investimentos para vários setores da economia e promover o crescimento econômico, os fundos de investimentos desempenham um papel importante na distribuição de recursos financeiros. Em tal situação, é fundamental ter uma compreensão das variáveis macroeconômicas, como taxas de juros, inflação, taxas de câmbio, crescimento do PIB e políticas governamentais, para analisar e antecipar os efeitos desses

elementos nos fundos de investimentos.

Os fundos de investimentos desempenham um papel substancial no sistema financeiro e no sistema econômico global, pois canaliza recursos para diversos setores da economia, rentabiliza os detentores dos recursos financeiro e contribui para a diversificação dos riscos com alocação heterogênea. Nesse contexto, a compreensão e análise aprofundada das variáveis macroeconômicas que afetam esses fundos são de suma importância para a tomada de decisões estratégicas de investimentos.

É possível verificar na prática o desempenho dos fundos de investimentos no Brasil. Segundo a AMBIMA¹, os fundos de investimento encerraram 2022 com alta de 7,1% no patrimônio líquido, passando de R\$ 6,9 trilhões, em 2021, para R\$ 7,4 trilhões. Dentre os fundos com maiores representatividades no patrimônio líquido da indústria, os multimercados do tipo macro (realizam operações em diversas classes de ativos e têm estratégias de médio e longo prazos) registraram a melhor rentabilidade média de 2022, com 17%. Quase todos os fundos da classe de multimercados tiveram retornos positivos: outros destaques são os tipos livre (aqueles sem compromisso de concentração em nenhuma estratégia específica) e *long and short* neutro (fazem operações de ativos e derivativos ligados ao mercado de renda variável, montando posições compradas e vendidas, com o objetivo de manter a exposição financeira líquida limitada a 5%), com retornos de 9,4% e 29,6%, respectivamente.

Ainda de acordo com a AMBIMA, na renda fixa, a maioria dos fundos também teve rentabilidade positiva no acumulado do ano. Nos fundos de ações, os destaques são de queda na rentabilidade: os do tipo investimento no exterior (podem aplicar mais de 40% em ativos internacionais) encerraram o ano com recuo de 13,9%, seguidos pelo valor/crescimento (buscam retorno por meio da seleção de empresas cujo valor das ações negociadas esteja abaixo do “preço justo”), com baixa de 10,8%, small caps (compostos por pelo menos 85% de ações de empresas que não estão incluídas entre as 25 maiores participações do IBr-X), com queda de 9,1%, e pelos livres (podem adotar diversos tipos de estratégias), com baixa de 7%.

Algumas variáveis econômicas podem influenciar diretamente fundos de investimentos. Inúmeros estudos foram publicados ao longo das últimas décadas versando sobre essa relação. Elton, Gruber e Blake (1995) publicaram o artigo *Fundamental economic variables, expected returns, and bond fund performance*, no artigo em questão procurou estabelecer o efeito de variáveis macroeconômicas sobre o desempenho financeiro de fundos mútuos selecionados na Índia. Teker, Tasseven e Tukul (2009), examinaram os indicadores macroeconômicos que influenciam os preços dos fundos mútuos. Khoo et al. (2020)

¹ https://www.ambima.com.br/pt_br/noticias/fundos-de-investimento-encerram-2022-com-avanco-no-patrimonio-liquido-e-captacao-negativa.htm

encontraram relação significativa entre taxa de câmbio, taxa de inflação, produto interno bruto e oferta monetária com desempenho seletivo de fundos mútuos e taxa de juros é a única variável independente com relação insignificante com desempenho seletivo de fundos mútuos na Malásia.

Ha uma relação entre o rendimento dos fundos e a evolução de variáveis macroeconômicas. Sendo seu desempenho influenciado por variáveis como taxas de juros, inflação e PIB. As taxas de juros, por exemplo, tem uma forte influencia sobre os retornos dos fundos no curto prazo e no longo prazo (PANIGRAHI; KARWA; JOSHI, 2020; KARIUKI, 2014; CHU, 2011; GYIMAH; ADDAI; ASAMOAH, 2021). Lobão e Levi (2016) usam o caso de Portugal para mostram os fundos de investimentos carregam informação sobre o crescimento industrial.

Estudar a rentabilidade dos fundos de investimentos e a relação com variáveis macroeconômicas pode gerar benefícios e minimização de risco para os investidores e potenciais investidores. A avaliação e previsão de rentabilidade permite que os investidores possam tomar decisões mais informadas e estratégicas em relação aos seus investimentos. Com base na previsão de rentabilidade, os investidores podem decidir quais fundos investir ou resgatar os investimentos, de acordo com seus objetivos financeiros e perfil de risco. Além disso, o estudo e previsão da rentabilidade dos fundos pode ajudar os gestores de fundos de investimentos a ajustar suas estratégias de investimento de acordo com as condições do mercado e da economia em geral, de forma a maximizar os retornos dos seus fundos.

Utilizar variáveis macroeconômicas para ajudar a entender e prever a rentabilidade dos fundos de investimento pode trazer vantagens para os agentes econômicos envolvidos nas negociações. As variáveis macroeconômicas podem influenciar significativamente o desempenho dos mercados financeiros em geral e, conseqüentemente, dos fundos de investimentos. Asad e Siddiqui (2019) analisaram os determinantes do desempenho dos fundos mútuos no Paquistão de 2010 a 2017, usando dados trimestrais. Os autores utilizam variáveis macroeconômicas e de mercado para explicar a variação do desempenho dos fundos mútuos, dentre as variáveis utilizadas pelos autores estão: o índice de preços ao consumidor (CPI), a taxa de juros interbancária (KIBOR), o índice de preços de ações (KSE-100) e o câmbio PKR/USD.

Estudar a rentabilidade dos fundos de investimentos e a relação com variáveis macroeconômicas pode gerar benefícios e minimização de risco para os investidores e potenciais investidores. A avaliação e previsão de rentabilidade permite que os investidores possam tomar decisões mais informadas e estratégicas em relação aos seus investimentos. Com base na previsão de rentabilidade, os investidores podem decidir quais fundos investir ou resgatar os investimentos, de acordo com seus objetivos financeiros e perfil de risco. Isso lhes permite ajustar suas estratégias de investimento de acordo com as condições do mercado e da economia em geral, buscando maximizar os retornos para os investidores.

Apesar da relação entre as variáveis econômicas e os fundos de investimento ser um tema amplamente estudado, há uma lacuna na literatura, que busca entender a estrutura de dependência dos fundos com a macroeconomia, principalmente aplicada ao contexto brasileiro. Nesse sentido, as Copulas têm se destacado como uma ferramenta poderosa nesse campo, permitindo analisar essas estruturas independentemente das distribuições marginais. Tal abordagem possibilita a avaliação do impacto de movimentos conjuntos de mercado nos preços dos ativos e no risco, bem como a precificação consistente de produtos multivariados, a avaliação da diversificação de portfólio e a análise de sensibilidade das medidas de risco a mudanças na estrutura de dependência (CHERUBINI; LUCIANO; VECCHIATO, 2004).

Após essas considerações, torna-se evidente a importância de compreender a estrutura de dependência entre fundos de investimento e as variáveis macroeconômicas. Sendo assim, o presente estudo foi motivado pela necessidade de aprofundar a compreensão das relações complexas entre os fundos e as variáveis macroeconômicas, diante da escassez de estudos abrangentes e metodologias que examinem essas interações complexas.

Portanto, este trabalho foi cuidadosamente planejado para proporcionar uma compreensão aprofundada do tema. Após esta introdução, o primeiro capítulo consiste em uma revisão sistemática da literatura, com o objetivo de estabelecer uma base sólida que sustente a pesquisa, fornecendo um contexto metodológico para as relações estudadas. O segundo capítulo inicia com uma revisão teórica e apresenta a metodologia utilizada no estudo. O terceiro capítulo apresenta os resultados, detalhando a análise dos dados, a estrutura das cópulas e, por fim, a análise da probabilidade de ocorrência e do retorno esperado. Finalmente, a dissertação é concluída, resumindo os principais achados.

1 REVISÃO SISTEMÁTICA DE LITERATURA

O objetivo principal deste capítulo é conduzir uma revisão sistemática de literatura, detalhada e abrangente sobre a aplicação das Cópulas em finanças, com um foco na sua utilização em fundos de investimento e variáveis macroeconômicas utilizando a metodologia apresentada por Thomé, Scavarda e Scavarda (2016). Destaca-se que a presente pesquisa não se concentra na discussão da teoria fundamental das Cópulas, mas sim na exploração de suas práticas aplicadas ao setor financeiro.

Apesar de a bibliometria e a revisão sistemática de literatura começarem a ser discutidas desde a década de 50, no campo das finanças, economia e ciências sociais, é relativamente recente. Ao invés de ir à literatura e buscar evidências ou lacunas na literatura, as perguntas nas pesquisas são guiadas a partir de narrativas, e a inclusão ou exclusão de artigos é escolhida a dedo, sem nenhum critério rigoroso, (DONTU et al., 2021; LINNENLUECKE; MARRONE; SINGH, 2020). Com a disponibilidade de grandes bases de dados e softwares especializados, fez com que esse tipo de análise se tornasse mais prática e popular. O motivo pelo qual esse tipo de análise deve ser feito se dá pelo fato de que o conhecimento acadêmico se expande muito rapidamente, com milhares de artigos sendo publicados diariamente. Com isso, há a necessidade de se fazer revisões rigorosas na literatura a fim de identificar os trabalhos mais relevantes e para que os pesquisadores possam estabelecer contexto para suas pesquisas, buscando suporte teórico e lacunas na literatura (LINNENLUECKE; MARRONE; SINGH, 2020).

1.1 Revisão Sistemática da Literatura

A Revisão Sistemática da Literatura (RSL) é uma abordagem rigorosa para examinar a literatura existente em um campo específico. Seu objetivo é proporcionar um relato do estado da literatura no momento da revisão (XIAO; WATSON, 2019). De acordo com Thomé, Scavarda e Scavarda (2016), foram desenvolvidos oito passos para conduzir uma RSL: (1) Planejamento e formulação do problema, (2) Busca na literatura, (3) Coleta de dados, (4) Avaliação da qualidade, (5) Análise e síntese dos dados, (6) Interpretação, (7) Apresentação dos resultados e (8) Atualização da revisão.

O primeiro passo consiste em estabelecer o objetivo do trabalho, formular as perguntas que se pretende responder ao longo do artigo, definindo, dessa forma, o escopo e a contextualização do tópico. Isso implica em estabelecer as seguintes perguntas:

1. Quais são os autores mais influentes do uso de cópulas com variáveis macroeconômicas?
2. Quais são os artigos de maior impacto que utiliza cópulas com variáveis macroe-

conômicas?

3. quais são os principais temas relacionados ao uso de cópulas em Macroeconomia?

O segundo passo diz respeito à escolha do banco de dados utilizado, dos termos e critérios de inclusão e exclusão. Foram utilizados os bancos de dados *Web of Science* e *Scopus* em conjunto. A escolha de usar ambas as bases de dados se deu pelo fato de permitir a inclusão da maior quantidade possível de artigos sobre o tema. A seleção das palavras-chave² foi feita com critérios rigorosos e amplos, com o objetivo de abranger todos os artigos relacionados ao tema, sem restringir nenhum trabalho possível. O critério de exclusão se deu pelo fato de a palavra "copula" ser usada no campo da biologia e a exclusão do termo "zero-inflation" se dá por se confundir com o termo desejado "inflation".

Após a busca, encontrou-se um total de 544 artigos referentes ao tema. A Tabela 1 mostra os diversos tipos de publicações relacionadas ao tema, sendo a maioria composta por artigos, com 473 publicações. Os documentos abrangem o período de 2004 a 2024³.

O primeiro trabalho feito relacionado ao tema foi o Hull, White et al. (2004), o artigo propõe uma alternativa ao uso das transformações rápidas de Fourier utilizando um modelo de copula de fatores. O artigo em questão propõe duas metodologias para precificação de CDOs e CDS sem simulação de Monte Carlo, ambas baseadas em um modelo de copula fator de tempo de default. A primeira abordagem utiliza uma equação de recorrência para calcular a probabilidade exata dos k defaults ocorrerem até um determinado tempo. Esta abordagem é particularmente eficiente para precificar n-ésimo default e segmentos do CDOs onde todas as dívidas possuem pesos iguais e as taxas de recuperação são constantes. A segunda abordagem emprega um procedimento de "bucketing" probabilístico para construir a distribuição de perdas do CDO. Este método divide o espaço de perdas em buckets distintos e estima a probabilidade de cada bucket ser atingido. A soma ponderada do valor esperado das perdas em cada bucket fornece o valor do CDO. A copula determinou a estrutura de dependência entre os tempos de default através da distribuição conjunta dos termos de erro. Para os autores, a aplicação de copulas na precificação de CDOs e n-ésimo default oferece uma alternativa poderosa e flexível à simulação de Monte Carlo.

O trabalho mais recente foi: *A changepoint analysis of exchange rate and commodity price risks for Latin American stock markets* de Manner, Rodríguez e Stöckler (2024). O trabalho investigou a vulnerabilidade dos mercados de ações da América La-

² Em ambos os casos, foi utilizada a seguinte consulta: copula AND ("Macroeconomic Variable" OR "Macroeconomic indicator" OR "Economic variable" OR "Macroeconomic determinant" OR "business cycle" OR "inflation" OR "treasure bill" OR "exchange rate" OR "money supply") NOT ("zero-inflation" OR "female" OR "male")

³ O trabalho referente a 2024 é o Manner, Rodríguez e Stöckler (2024), que já está disponível, mas será oficialmente publicado apenas em 2024

tina aos choques de preços de commodities e taxas de câmbio. Os autores argumentaram que a dependência entre esses fatores e os retornos das ações não é estática, podendo sofrer mudanças ao longo do tempo. Assim, o estudo buscou identificar potenciais "pontos de mudança" (change points) onde a relação entre as variáveis sofreram alterações. O artigo Utilizou uma combinação de copulas e modelos GARCH para capturar a dependência dinâmica entre retornos de ações, preços de commodities e taxas de câmbio. As copulas permitem modelar a dependência conjunta das variáveis, enquanto o GARCH captura a heterocedasticidade (volatilidade variável) dos retornos. O estudo utiliza dados mensais de preços de commodities, taxas de câmbio e retornos de ações para diferentes países da América Latina, cobrindo um período que vai de janeiro de 1990 a dezembro de 2020.

O estudo de Manner, Rodríguez e Stöckler (2024) encontrou evidência de vários pontos de mudança na dependência entre preços de commodities, taxas de câmbio e retornos de ações para diferentes países da América Latina. De acordo com os autores os pontos de mudança tendem a coincidir com eventos econômicos e financeiros globais ou regionais, sugerindo que fatores externos influenciam a dinâmica de dependência. Em geral, o estudo observou um aumento na dependência entre os fatores de risco após a crise financeira global de 2008, indicando que os mercados latinos se tornaram mais sensíveis a choques externos.

observa-se na Figura 1, nota-se um aumento na publicação de artigos nos anos que sucedem crises, como em 2008 com a crise do *sub-prime*, em 2013 e 2014 com a crise da dívida soberana europeia e 2023 com a crise da COVID-19.

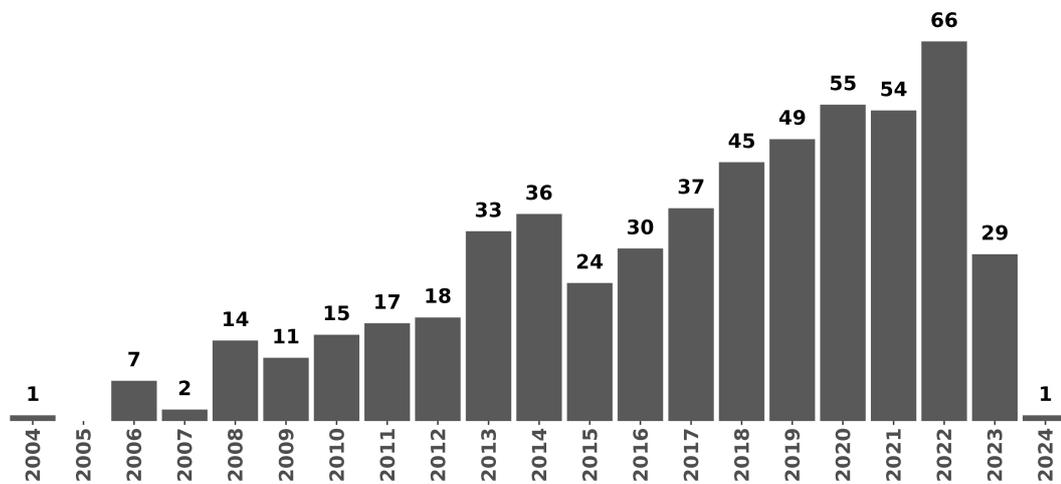
Tabela 1 - Tipos de Publicações

Tipo	Quantidade
Artigo	473
Artigo de Conferencia	36
Capitulo de Livro	18
Livro	2
Outros	15
Total	544

Fonte: O autor, 2024. Dados da
Web Of Science e Scopus

Ao analisar a distribuição de artigos por periódicos, foi encontrado um total de 303 revistas científicas diferentes. A Tabela 2 sintetiza essa distribuição, mostrando as 6 revistas com maior número de citações. Nota-se que a influência entre as revistas é bastante heterogênea, onde as três com maior número de citações concentram aproximadamente apenas 30% do total. Sendo o *International Economic Review* com o maior número de citações, vale ressaltar que com apenas um trabalho, sendo este o Patton (2006b), que

Figura 1 - Evolução da produção anual.



Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

utilizou um modelo de copula para mostrar que diferentes taxas de câmbio estão mais correlacionadas quando as mesmas estão se desvalorizando em relação ao dólar do que apreciando.

Tabela 2 - Distribuição de revistas científicas.

revista	n artigos	citações
International Economic Review	1	1190
Energy Economics	18	1080
Journal Of Banking And Finance	4	517
Journal Of International Money And Finance	8	462
Journal Of Econometrics	5	384
Journal Of Derivatives	2	292

Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

1.1.1 Análise Bibliométria

Para responder às perguntas, será realizada uma análise bibliométrica. Este tipo de análise é útil para mensurar e quantificar a performance e o impacto de pesquisas científicas (COBO et al., 2011; THOMÉ; SCAVARDA; SCAVARDA, 2016).

A primeira análise consiste nas citações. Esta análise busca entender os autores mais influentes, medindo quantos artigos foram publicados, quantas citações os autores

têm, suas instituições e país de origem. A segunda análise feita foi a de co-citações, que é de extrema importância, pois ela nos dirá quais são os trabalhos mais influentes. Em seguida, será feita a análise temática para poder analisar e acompanhar as mudanças dos principais tópicos ou conceitos tratados por um campo de pesquisa ao longo do tempo (COBO et al., 2011). O software utilizado para a construção da análise foi o Bibliometrix (ARIA; CUCCURULLO, 2017).

1.1.1.1 Análise temática

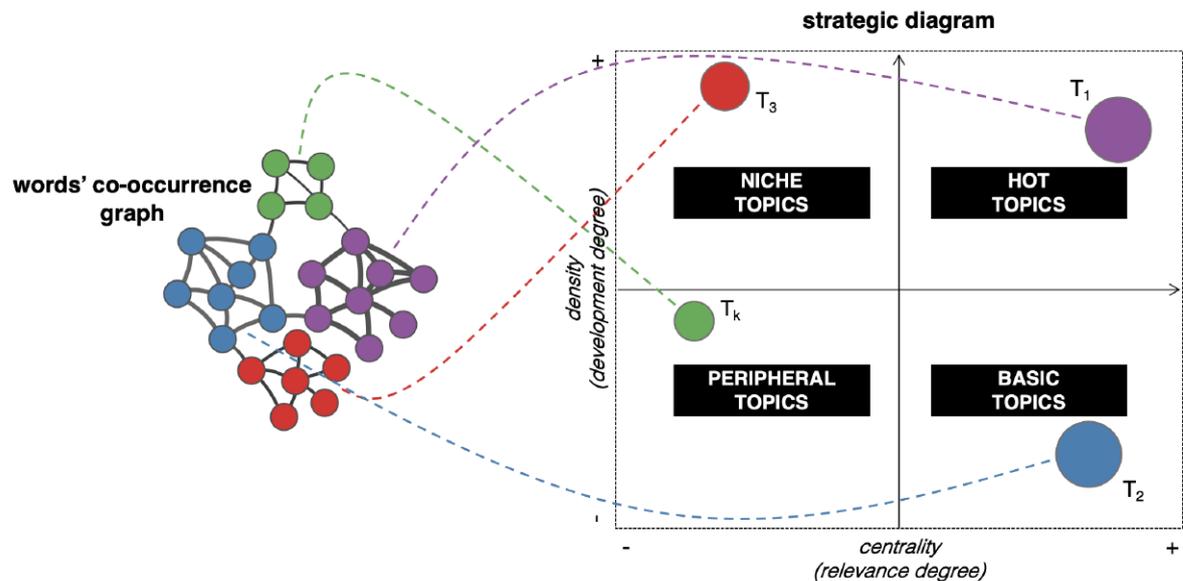
Um método tradicional para realizar análise temática é a análise de co-palavras, que considera a co-ocorrência de termos-chave para descrever o conteúdo dos documentos em um arquivo. Ao comparar os perfis baseados em palavras-chave, é possível medir a similaridade entre um par de publicações e identificar os tópicos ou temas que são frequentemente associados a um conjunto de palavras ou frases (COBO et al., 2011; ARIA et al., 2022).

Esses tópicos identificados são mapeados e mostrados no Diagrama de Callon, fornecendo uma representação visual dos tópicos e sua relevância e desenvolvimento dentro do domínio de pesquisa. O Diagrama de Callon usa duas medidas, a centralidade de Callon e a densidade de Callon, para determinar a relevância e o desenvolvimento de um tópico, respectivamente (Figura 2). Os tópicos são então mapeados em quatro quadrantes, cada um representando um tipo diferente de tópico: tópicos quentes ou motor (altamente relevantes e bem desenvolvidos), tópicos básicos (significativos para o domínio e interdisciplinares), tópicos periféricos (não totalmente desenvolvidos ou marginalmente interessantes para o domínio) e tópicos de nicho (fortemente desenvolvidos, mas ainda marginais para o domínio) (ARIA et al., 2022).

A centralidade de Callon é uma medida de quão central é um tema em um campo de pesquisa. Sendo definida como: $c = 10 \times \sum e_{kh}$ onde, k são palavras-chaves relacionada a um tema e h palavras-chaves relacionada a outro. Um tema com uma alta centralidade de Callon é um que está fortemente associado a muitos outros temas no campo. (COBO et al., 2011)

A densidade de Callon é uma medida de quão densamente um tema está conectado a outros temas em um campo de pesquisa. Sendo definida como: $d = 100 \times \sum e_{ij}/w$, onde, i e j são palavras-chave pertencente ao tema e w numero de palavras-chave do tema. Um tema com alta densidade de Callon é um que está fortemente associado a outros temas dentro do mesmo cluster temático. (COBO et al., 2011)

Figura 2 - Construção do mapa temático



Fonte: Aria et al. (2022)

1.2 Resultado

1.2.1 Análise de citação

Esta subseção está destinada a responder à primeira pergunta: "Quais são os autores mais influentes do uso de cópulas com variáveis macroeconômicas?". A Tabela 3 mostra os autores com maior número de citações. O autor com o maior número de citações é Andrew Patton, com seus trabalhos mais influentes Patton (2006b) e Patton (2006a), com respectivamente 1190 e 268 citações.

No artigo "Estimation of multivariate models for time series of possibly different lengths" de 2006 Patton propôs uma forma de estimar modelos multivariados lineares para séries temporais com comprimentos de tempo diferentes. Patton apresentou um estimador denominado Multi-Stage Maximum Likelihood Estimator (MSMLE). O MSMLE utiliza todas as observações disponíveis de cada série temporal, mesmo que elas não coincidam temporalmente. Embora o artigo não se concentre explicitamente em modelos de cópulas, o autor propõe um método adaptado e utilizado para a modelagem de dependências entre séries temporais através de copulas. A integração de cópulas no MSMLE oferece uma ferramenta para modelar dependências complexas entre séries temporais. Essa abordagem permite flexibilidade, precisão e interpretabilidade nas análises.

O segundo trabalho discute a estimativa de modelos multivariados de séries temporais com diferentes comprimentos, propondo o uso de um estimador de máxima veros-

semelhança em múltiplos estágios. A cópula é usada para descrever a dependência entre as variáveis e encontra evidências significativas de variação no tempo na cópula condicional de diferentes taxas de câmbio. O autor especificou uma copula Vine bidimensional para capturar a dependência assimétrica. Essa copula permitiu modelar uma dependência maior durante depreciações do dólar e uma dependência menor durante apreciações. Patton sugeriu que essa assimetria poderia ser explicada por motivos políticos e competitivos. Durante depreciações do dólar, os governos da Alemanha e do Japão podem intervir para impedir uma apreciação excessiva de suas moedas, levando a uma maior correlação entre as taxas. Em contraste, durante apreciações do dólar, a intervenção governamental é menos provável, resultando em uma correlação mais fraca. O estudo em questão, traves da teoria de copulas, encontrou evidências substanciais de dependência assimétrica entre as taxas de câmbio. As correlações entre o Marco e o Iene em relação ao Dólar foram significativamente maiores durante depreciações do dólar do que durante apreciações.

Tabela 3 - Os 10 autores com mais citações.

Autor	Artigos	Citações	Instituição	País
Patton A	4	1563	Duke University	EUA
Reboredo J	6	781	University of Santiago de Compostela	Espanha
Czado C	7	424	University of Munich	Alemanha
Aloui R	3	423	University of Tunis El Manar	Tunísia
Fan Y	3	364	University of Washington	China
Hull J	3	293	University of Toronto	Canada
White A	3	293	University of Toronto	Canada
Chen X	5	276	Yale University	China
Ben A M	2	254	University of Tunis El Manar	Tunisia
Hammoudeh S	2	219	University of Economics Ho Chi Minh City	Palestina

Legenda: O número de citações diz respeito apenas aos artigos relacionados ao tema.

Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

O segundo autor com maior quantidade de citações é Juan C. Reboredo, com 781 citações. Seu trabalho mais citado é Reboredo (2013), onde ele usa cópulas para avaliar o papel do ouro como forma de proteção contra a desvalorização do dólar americano. A pesquisa revelou uma dependência média positiva e significativa entre o ouro e a desvalorização do dólar, indicando que o ouro pode atuar como um *hedge* contra movimentos da taxa do dólar.

Além disso, Reboredo tem o sexto artigo mais citado, com 246 citações, sendo o trabalho Reboredo (2012). No artigo em questão o autor investiga a relação entre o preço do petróleo e as taxas de câmbio (dados mensais sobre o preço do petróleo bruto (West Te-

xas Intermediate, WTI) e diversas taxas de câmbio em relação ao dólar americano (USD) de janeiro de 1990 a dezembro de 2010), focando especificamente na "co-movimentação", ou seja, como eles se movem juntos ao longo do tempo.

Neste estudo, o autor explora como os preços do petróleo e as taxas de câmbio se movem juntos, utilizando duas medidas de dependência: correlações e cópulas. O estudo revela que a dependência entre os preços do petróleo e as taxas de câmbio é geralmente fraca (Os coeficientes de correlação foram geralmente baixos, indicando que mesmo fortes movimentos no preço do petróleo podem não se traduzir diretamente em grandes mudanças nas taxas de câmbio, embora tenha aumentado substancialmente após a crise financeira global de 2008). Por outro lado, as copulas revelaram uma relação não linear, onde as taxas de câmbio respondem de maneira diferente a aumentos e reduções do preço do petróleo. Em média, as taxas de câmbio tendem a depreciar (enfraquecer) mais em relação ao USD quando o preço do petróleo sobe, mas apreciar (fortalecer) menos quando o preço do petróleo cai.

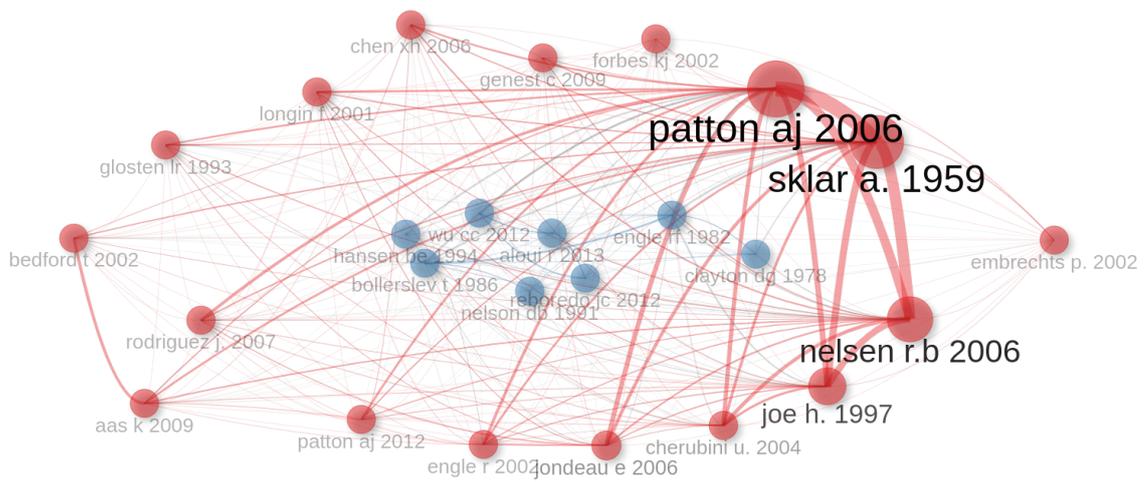
Adicionalmente, o estudo identificou um aumento na dependência entre o preço do petróleo e as taxas de câmbio no período posterior à crise financeira global. Essa intensificação da relação sugere que choques no preço do petróleo podem ter um impacto mais amplificado nas flutuações cambiais após a crise, elevando os riscos para a estabilidade financeira. Cabe destacar que compreender essa relação não linear é importante para formuladores de políticas, investidores e analistas que estejam preocupado com o impacto das flutuações do preço do petróleo nas economias e nos mercados financeiros.

Czado (2019) no seu livro intitulado "Analyzing dependent data with vine copulas" oferece uma introdução abrangente e acadêmica à classe de copulas Vine, explorando sua eficiência e flexibilidade em face de desafios tradicionais. O foco do livro recai nas copulas Vine, uma construção específica que permite a representação de dependências complexas através de pares ordenados de copulas bivariadas ("vinhas"). São abordados métodos de estimação baseados em máxima verossimilhança e inferência bayesiana, juntamente com procedimentos de seleção de copulas bivariadas e especificação da estrutura Vine. O livro apresenta alguns algoritmos computacionais e técnicas de diagnóstico para avaliar a adequação do modelo. O autor apresentou exemplos empíricos em áreas como finanças, seguros, hidrologia e ciências ambientais. O livro analisa modelos para dados financeiros, dependência entre retornos e volatilidade, e predição de riscos de carteira.

1.2.2 Análise de co-citações

A análise de co-citação⁴ é uma técnica de mapeamento que pressupõe que publicações que são citadas juntas frequentemente são semelhantes tematicamente. Essa análise pode ser usada para revelar a estrutura intelectual de um campo de pesquisa, como seus temas subjacentes (DONTU et al., 2021). Esta subseção está destinada a responder à pergunta "Quais são os trabalhos mais influentes que utilizam cópulas com variáveis macroeconômicas?". A Figura 3 mostra a rede de co-citações. Esta figura apresenta os 25 trabalhos mais influentes no tema. O tamanho da circunferência do círculo é proporcional ao número de citações que o artigo recebeu, e a largura da linha é proporcional ao número de coocorrências

Figura 3 - Rede de co-citações



Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

O modelo de cópula pode ser descrito como uma ligação de dependência entre variáveis aleatórias, e esta concepção teórica foi inicialmente apresentada por Sklar (1959). O foco principal da pesquisa de Sklar era a análise da distribuição tridimensional conjunta, e, assim, ele começou sua abordagem ao introduzir funções auxiliares que foram definidas no intervalo unitário, estabelecendo uma conexão entre a função de distribuição e suas distribuições marginais.

Embora não tenha sido descrito pelo autor como sendo um tipo de cópula, o trabalho de Clayton (1978) introduziu uma nova família de cópulas, sendo conhecida por

⁴ Não será comentada nesta sub-seção os trabalhos de Patton (2006b) e Reboredo (2012), pois já foram comentados anteriormente.

sua capacidade de modelar a ocorrência de eventos extremos negativos. Em seu trabalho original, o autor propõe um modelo de associação em distribuições bivariadas, do tempo de sobrevivência para a análise da tendência familiar na incidência de doenças. O modelo foi desenhado para estimar associação em tabelas de mortalidade multivariadas, considerando a distribuição da idade de incidência da doença.

Para lidar com problemas de heterocedasticidade, Engle (1982) introduz uma nova classe de processo estocástico chamada de *autoregressive conditional heteroskedasticity* (ARCH), onde não se assume que a variância dos erros ao longo do tempo seja constante. Bollerslev (1986) introduz o modelo *Generalized autoregressive conditional heteroskedasticity* (GARCH), sendo este modelo uma extensão do modelo ARCH. Este modelo inclui valores defasados da própria variância da previsão, além dos termos de erro quadrático do modelo ARCH, podendo ser utilizado em um modelo ARMA.

Nelson (1991) apresenta mais um modelo do tipo ARCH, o *Exponencial ARCH* (EARCH). O modelo EARCH tem como vantagem permitir respostas assimétricas a choques positivos e negativos. Isto significa que o impacto de uma grande mudança positiva numa série temporal pode ser diferente de uma grande mudança negativa. O modelo de Glosten-Jagannathan-Runkle, conhecido por GJR-GARCH, introduzido por Glosten, Jagannathan e Runkle (1993), também leva em consideração esta assimetria nos retornos.

O livro mais citado foi o *"Multivariate Models and Multivariate Dependence Concepts"* do Joe (1997), o livro foca em apresentar conceitos sobre dependência multivariada e distribuições multivariadas não normais para modelagem de dados de resposta binária, contagem, ordinal e de valor extremo. Longin e Solnik (2001) utilizaram da teoria do valor extremo para modelar as caudas da distribuição multivariada e testar a hipótese de que a correlação do mercado acionário internacional aumenta em tempos de alta volatilidade, mostrando que a correlação aumenta nos mercados em baixa, mas não nos mercados em alta.

Ainda sobre a importância da análise de dependência, Embrechts, McNeil e Straumann (2002) discutem a importância de compreender a dependência estocástica além da simples correlação linear. Os autores enfatizam a representação cópula da dependência para um vetor aleatório e argumentam que, embora a correlação linear seja uma medida de dependência natural para riscos multivariados normalmente e elipticamente distribuídos, outros conceitos de dependência, como comonotonicidade e correlação de classificação, também devem ser compreendidos pelos profissionais de gestão de risco. Os autores argumentam que a correlação, embora seja um conceito central em finanças e seguros, é frequentemente mal compreendida e mal utilizada, levando a falácias quando suposições ingênuas sobre propriedades de dependência do mundo elíptico são aplicadas ao mundo não elíptico.

Engle (2002) introduz uma nova classe de modelos GARCH, o *Dynamic Conditional Correlation* (DCC). Os modelos DCC têm a flexibilidade dos modelos GARCH

univariados combinados com modelos paramétricos parcimoniosos para correlações. Esses modelos não são lineares, mas muitas vezes podem ser estimados simplesmente com métodos univariados ou de duas etapas baseados na função de verossimilhança. Os modelos DCC são utilizados para estimar correlações que variam ao longo do tempo.

No artigo ” *Vines — A New Graphical Model for Dependent Random Variables*” de Bedford e Cooke (2002), um novo modelo gráfico chamado vine é introduzido para variáveis aleatórias dependentes, sendo uma generalizam árvores de Markov. Vines podem ser usadas para especificar distribuições multivariadas especificando distribuições marginais e como elas estão acopladas. Tendo utilidade na análise de incerteza, permitindo determinar a sensibilidade de um modelo a parâmetros incertos, e o artigo explica como criar e amostrar rapidamente uma distribuição de vine de informação mínima com base em informações *a priori*, marginais dadas e correlações de *rank* especificadas. O uso da *vine copula* ou *pair-copula* foi explorada posteriormente por Aas et al. (2009). Este tipo de cópula, são construídas a partir de uma aplicação sequencial de cópulas bivariadas para construir uma cópula de dimensão superior.

Forbes e Rigobon (2002) desafiam o entendimento convencional de contágio no contexto de comovimentos do mercado de ações. Os autores argumentam que estudos anteriores, que sugeriam que o contágio ocorreu durante crises como a crise asiática de 1997, a desvalorização mexicana de 1994 e o crash do mercado dos EUA em 1987, foram tendenciosos devido à heteroscedasticidade. Eles propõem que os coeficientes de correlação, usados para medir o contágio, são condicionais à volatilidade do mercado. Após ajustar para este viés, os autores encontraram praticamente nenhum aumento nos coeficientes de correlação incondicionais durante essas crises, sugerindo que nenhum contágio ocorreu. Em vez disso, eles observaram um alto nível de comovimento do mercado em todos os períodos, que eles denominam como interdependência.

O segundo livro mais influente no tópico, é o Cherubini, Luciano e Vecchiato (2004). O livro explica copulas através de sua aplicação em tópicos significativos como precificação de derivativos e análise de risco de crédito. O livro é estruturado para primeiro introduzir o leitor ao caso da copula bivariada, apresentando o contexto matemático e probabilístico no qual as aplicações são construídas. Em seguida, discute as falhas da correlação linear e destaca como as funções de copula, juntamente com medidas de associação não paramétricas, podem fornecer uma maneira muito mais flexível de representar comovimentos de mercado. O livro também aborda a inferência estatística para copulas, incluindo tanto aspectos metodológicos quanto aplicações a partir de dados de mercado, como a calibração de comovimentos de fatores de risco reais e a medição do VaR (Valor em Risco). Mais tarde Nelsen (2006) publica um livro introdutório, começando com as propriedades básicas das copulas e, em seguida, apresenta métodos para construir copulas e discutir o papel desempenhado pelas copulas na modelagem e no estudo da dependência. O foco é nas copulas bivariadas, embora a maioria dos capítulos termine com uma dis-

cussão do caso multivariado.

Chen, Fan e Tsyrennikov (2006), propõem um novo método que possibilita contornar a "maldição da dimensionalidade" associada a distribuições multivariadas puramente não paramétricas. O procedimento de estimação de máxima verossimilhança de peneira para uma ampla classe de distribuições multivariadas semi-paramétricas. Essas distribuições são caracterizadas por uma função de cópula paramétrica avaliada em distribuições marginais não paramétricas. Os autores demonstram que seus estimadores de máxima verossimilhança de peneira *plug-in* (MLEs) de todas as funções suaves, incluindo os parâmetros de cópula de dimensão finita e as distribuições marginais desconhecidas, são semi-parametricamente eficientes. O artigo também mostra que suas variâncias assintóticas podem ser estimadas de forma consistente.

Jondeau e Rockinger (2006) os autores argumentam que os retornos de ativos financeiros são frequentemente não normais, impulsionados por distribuições assimétricas e/ou de cauda pesada, levando a questionar a adequação do critério de média-variância proposto por Markowitz para a alocação de riqueza. Os autores usam uma expansão em série de Taylor da utilidade esperada para se concentrar em certos momentos e calcular a alocação ótima de portfólio numericamente. Eles descobrem que, enquanto o critério de média-variância fornece uma boa aproximação da maximização da utilidade esperada sob moderada não normalidade, ele pode ser ineficaz sob grandes desvios da normalidade. Nesses casos, as estratégias de otimização de três momentos ou quatro momentos podem fornecer uma boa aproximação da utilidade esperada.

Rodriguez (2007) utiliza copulas com mudança de parâmetro para modelar a dependência e entender o contágio de uma crise financeira de um país no mercado acionário em outro. Os resultados sugerem evidências da mudança de dependência durante períodos de alta volatilidade. Mostrando que as quebras estruturais na dependência de cauda são uma dimensão do fenômeno de contágio. Portanto, a rejeição da hipótese de quebra de correlação não deve ser considerada.

Genest, Rémillard e Beaudoin (2009) focam em discutir os testes de adequação dos modelos de cópula, especificamente "testes abrangentes" que não requerem categorização arbitrária de dados ou escolhas estratégicas de parâmetros de suavização, funções de peso, kernels ou janelas. Os autores apresentam uma revisão crítica dos procedimentos existentes, sugerem novos e descrevem os resultados de um grande experimento de Monte Carlo projetado para avaliar o efeito do tamanho da amostra e da força da dependência no nível e na potência dos testes abrangentes para várias combinações de modelos de cópula sob a hipótese nula e a alternativa.

Patton (2012) faz revisão de literatura sobre modelos baseados em cópulas para séries temporais econômicas e financeiras. O autor analisa métodos de estimativa e inferência e testes de adequação do ajuste para tais modelos, bem como aplicações empíricas dessas cópulas para séries temporais econômicas e financeiras. O artigo também discute

o uso de modelos baseados em cópulas para processos de séries temporais univariadas e multivariadas.

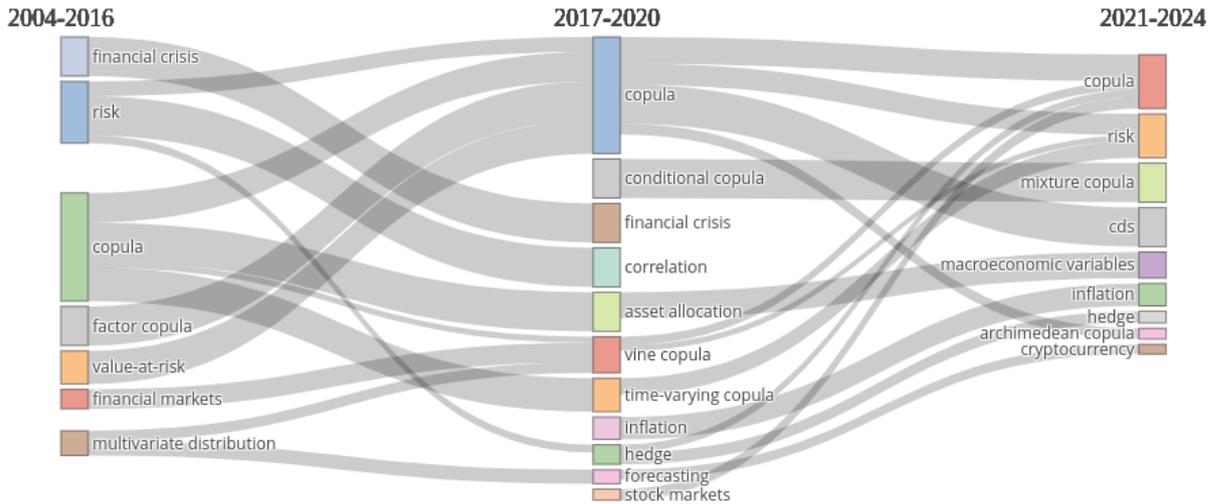
Aloui, Aïssa e Nguyen (2013) e Wu, Chung e Chang (2012) utilizam do modelo copula-GARCH (CGARCH) para investigar a relação entre os preços petróleo bruto e as taxas de câmbio do dólar americano. Os trabalhos mostraram que uma estratégia dinâmica baseada no modelo CGARCH com a copula t-Student oferece maiores benefícios econômicos do que estratégias estáticas e outras estratégias dinâmicas. Aloui, Aïssa e Nguyen (2013) também mostram que as copulas t-Student capturam melhor a dependência extrema, e levar em conta o comovimento extremo leva a uma melhoria na precisão das previsões de Valor em Risco (VaR).

1.2.3 Análise temática

Para responder à pergunta "quais são os principais temas relacionados ao uso de cópulas em Macroeconomia", foi feita uma análise a partir das palavras-chave. Existem duas possibilidades de fazer esta análise: uma é utilizando as palavras-chave escolhidas pelos autores e outra é utilizar as palavras-chave que foram indexadas pela própria *Web Of Science* e *Scopus*.

As palavras-chave do autor são escolhidas pelos próprios autores dos artigos e, portanto, refletem as perspectivas dos autores sobre os principais tópicos de seu trabalho. Por outro lado, as palavras-chave geradas pelos bancos de dados são geradas automaticamente com base nos títulos das referências citadas em um artigo. Para se manter o mais fiel possível à intenção dos autores, foram utilizadas as palavras-chave escolhidas pelos próprios autores.

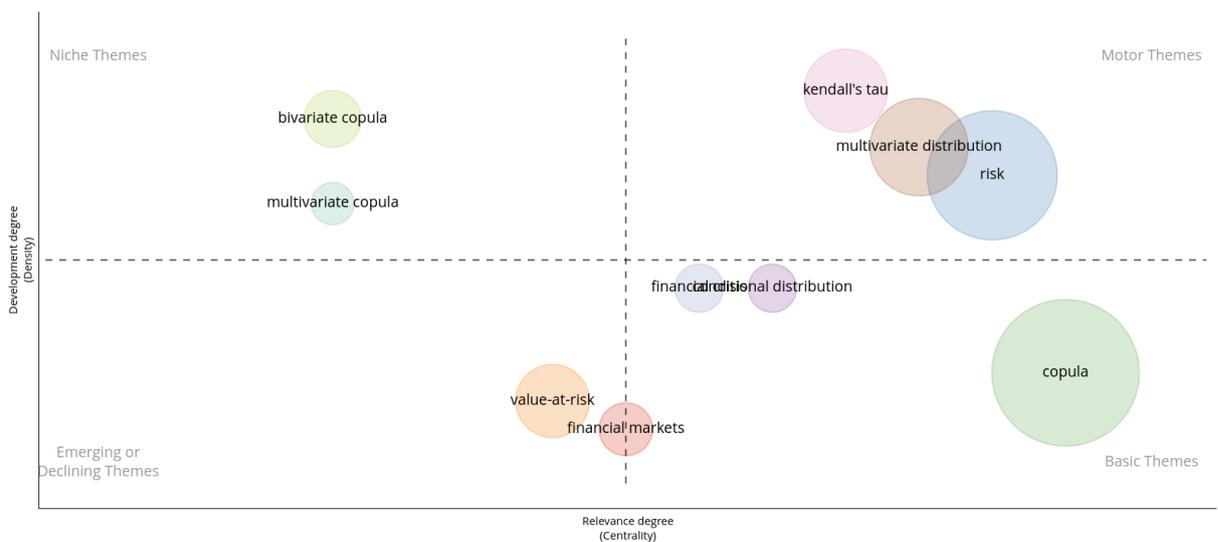
Figura 4 - Evolução das palavras-chaves por período



Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

A Figura 4 mostra a evolução das palavras-chave ao longo do tempo, onde a escolha da duração dos períodos foi feita a partir da equalização dos números de artigos dos mais recentes para os mais antigos. A associação entre os temas de um período para o outro se dá através do índice de inclusão (coeficiente de Simpson) ponderado pela ocorrência da palavra (ARIA; CUCCURULLO, 2017). Esse índice é uma métrica de sobreposição que mede o quanto um conjunto está incluído em outro. A força dessa associação é representada pela largura da linha que associa dois termos.

Figura 5 - Mapa temático para o período 2004-2016.



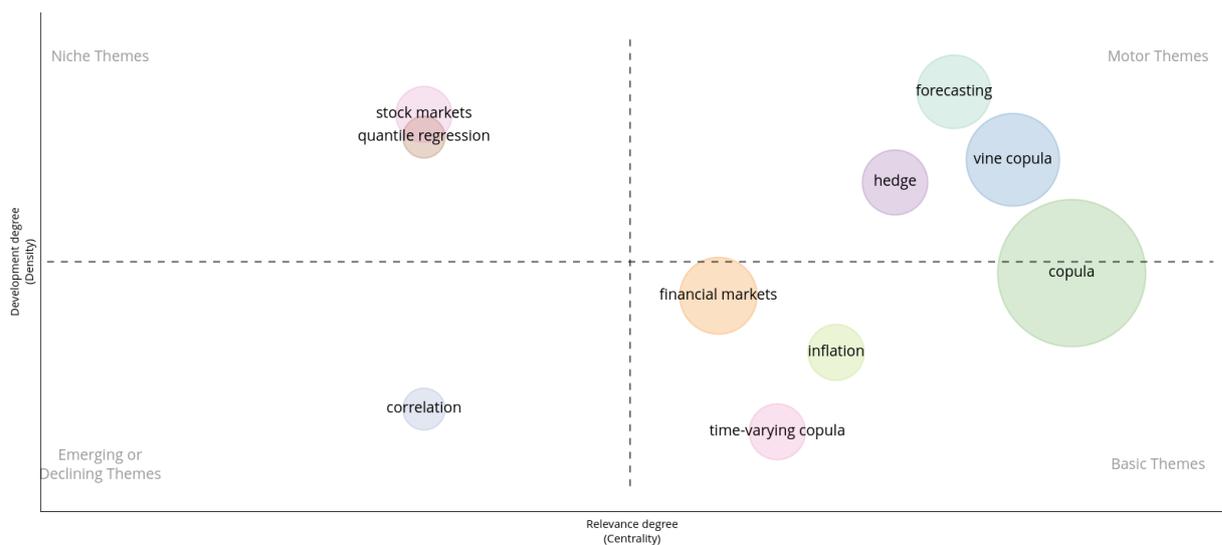
Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

A Figura 5 exibe o primeiro mapa temático, o qual apresenta os temas mais relevantes no início da literatura. É possível observar uma concentração nos tópicos motores e básicos, localizados no primeiro e segundo quadrante. O tema "copula" se apresenta, conforme esperado, como um tema básico, juntamente com "financial markets", "financial crisis" e "conditional distribution". O tema "risk" se posiciona no tópico motor, devido ao uso da cópula para avaliar o risco de um investimento, em conjunto com "kendall's tau" e "multivariate distribution". Os tópicos de nicho incluem "bivariate copula" e "multivariate copula", sendo que o primeiro apresenta uma densidade maior. Um tema que se destaca como emergente no período é "value-at-risk".

A Figura 6 apresenta o mapa temático do período de 2017 a 2020. Temas básicos como "copula" e "financial markets" se mantiveram, enquanto temas como "time-varying copula" e "inflation" foram introduzidos. Observa-se uma mudança dos tópicos quentes em relação ao período anterior. Neste período, as cópulas passaram a ser mais utilizadas para "hedge" e previsão ("forecasting"). O tema "vine copula" também se estabeleceu fortemente durante este período.

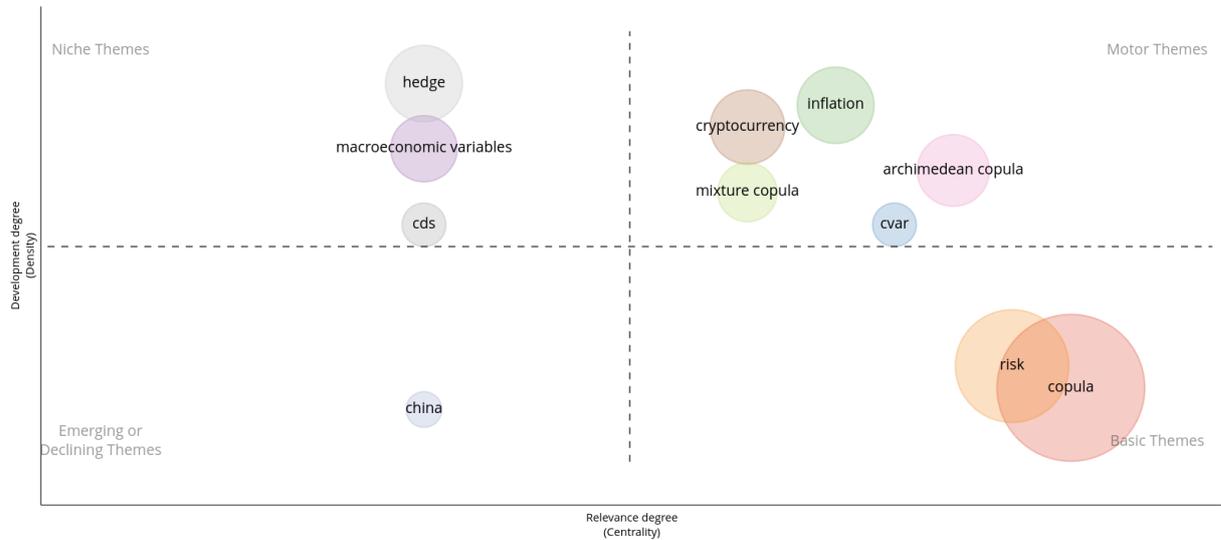
O mapa temático referente ao período de 2021 a 2024 é mostrado na Figura 7. É interessante observar que o tema "risk" passou de tópico motor no primeiro período para tema básico no último, e "inflation" evoluiu de tópico básico para motor. Outros temas como "cryptocurrency", "cvar", "mixture copula" e "archimedean copula" foram introduzidos como tópicos motores. O tema "hedge" passou de tópico motor para tópico de nicho, juntamente com "cbs" e "macroeconomic variables".

Figura 6 - Mapa temático para o período 2017-2020.



Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

Figura 7 - Mapa temático para o período 2021-2024.



Fonte: O autor, 2024. Dados da *Web Of Science* e *Scopus*

A análise temática dos resultados mostra uma mudança significativa em seus interesses e usos. A literatura se concentrou inicialmente nos fundamentos e aplicações diretas das cópulas nos mercados financeiros, com ênfase em distribuições condicionais e multivariadas e avaliações de risco, conforme ilustrado na Figura 5. O conceito de "cópula" emergiu como um conceito fundamental e foi principalmente usado para compreender crises e mercados financeiros. Este é o momento em que o foco estava em estabelecer a importância das cópulas na modelagem de dependências no contexto financeiro.

O uso de cópulas evoluiu com o tempo, como demonstrado nas Figuras 6 e 7. Entre 2017 e 2020, os temas que antes eram considerados essenciais continuam sendo importantes, mas novas aplicações surgem, como a "vine copula", que reflete um aprofundamento no uso de cópulas para modelar dependências mais complexas. Além disso, a inclusão de termos como "hedge" e "forecasting" indica uma mudança para estratégias de previsão e proteção financeira.

A presente dissertação não só se insere nesse contexto de pesquisa avançada, como também impulsiona o tema, ao aplicar a teoria de cópulas para examinar a estrutura de dependência entre fundos de investimento e variáveis macroeconômicas. A escolha deste enfoque permite aos investidores, formuladores de políticas e acadêmicos uma compreensão mais matizada das inter-relações do mercado, essenciais para a tomada de decisões informadas e para a construção de estratégias econômicas resilientes. A aplicação da teoria de cópulas, neste sentido, emerge como uma contribuição teórica e uma ferramenta auxiliar na tomada de decisão.

2 CÓPULAS

A análise da estrutura de dependência entre variáveis financeiras desempenha um papel crucial na gestão de riscos, na precificação de ativos e na alocação de portfólio. Nesse contexto, as Copulas surgem como uma ferramenta estatística de extrema importância, capaz de construir distribuições multivariadas e investigar a estrutura de dependência entre variáveis aleatórias. No campo das finanças, as Copulas têm sido amplamente empregadas para modelar a dependência entre retornos de ativos, taxas de juros, taxas de câmbio e outras variáveis financeiras relevantes. (BOUYÉ et al., 2000; CHERUBINI; LUCIANO; VECCHIATO, 2004; PATTON, 2009)

Modelos baseados em copulas oferecem uma estrutura flexível para modelar séries temporais econômicas e financeiras. As copulas são funções que permitem a separação das distribuições marginais da estrutura de dependência de uma distribuição multivariada específica, isto é, cópulas são funções que ligam funções de distribuição multivariadas com suas funções de distribuição marginais univariadas (NELSEN, 2006; ANJOS et al., 2004). Esses modelos têm despertado considerável interesse devido à sua capacidade de capturar relacionamentos complexos entre variáveis, conferindo uma maior flexibilidade ao modelar dependências, mesmo quando as variáveis não seguem distribuições normais ou suas relações são complexas. Podendo representar várias formas de dependência, incluindo positiva, negativa, linear, não linear e extrema (PATTON, 2012).

A relevância das copulas na finanças é evidente em diversas aplicações. Tendo sido utilizadas para monitoramento de risco de mercado de produtos em cesta, mensuração de risco de crédito para carteiras de empréstimos, cálculo de requisito de capital para risco operacional e modelagem da dependência entre variáveis na análise de risco financeiro. Elas têm se mostrado valiosas em análise de portfólio, gestão de risco, precificação de derivativos. Sendo especialmente utilizados em bancos e seguradoras (ZHANG; JIANG, 2019; BOUYÉ et al., 2000).

Não obstante, as copulas demonstram grande utilidade na modelagem de risco de carteira, permitindo a quantificação do impacto de choques em diferentes ativos sobre o valor total da carteira. Através da estimulação de cenários de perdas e da avaliação da probabilidade de eventos extremos, é possível otimizar a diversificação e a alocação de ativos, minimizando o risco sistêmico e maximizando o retorno esperado. O estudo de Bilbao-Terol, Cañal-Fernández e Mar (2016) propõe um modelo inovador para a construção de carteiras de investimentos sustentáveis e socialmente responsáveis (SR). O modelo se baseia na Teoria de Cópulas para capturar a dependência entre os retornos de fundos SR e convencionais, permitindo uma análise mais precisa do risco e da diversificação. O modelo foi aplicado a um conjunto de 38 fundos convencionais e 12 fundos SR na Espanha. O estudo utilizou dados de retornos mensais de fundos de investimento e cobriu o período

de janeiro de 2008 a dezembro de 2014, totalizando 84 meses. Os autores demonstram que o modelo é capaz de construir carteiras SR com melhor desempenho em termos de retorno e risco em comparação com carteiras diversificadas tradicionais. O estudo em questão demonstrou que o modelo baseado na Teoria de Cópulas para a construção de carteiras de investimentos sustentáveis e socialmente responsáveis (SR) é capaz de gerar carteiras eficientes com melhor desempenho em termos de retorno e risco em comparação com carteiras diversificadas tradicionais.

As copulas também podem ser utilizadas na precificação de derivativos, como opções e swaps, que dependem do comportamento de múltiplos ativos subjacentes. As copulas permitem modelar a dependência entre os ativos e incorporar essa informação no cálculo do preço do derivativo, resultando em uma precificação mais precisa e eficiente. Em um estudo de Brigo e Mercurio (2006), as copulas foram utilizadas para precificar swaps de taxa de juros. Os autores demonstram que a utilização de copulas para capturar a dependência entre as taxas de juros resulta em uma precificação mais precisa do swap em comparação com métodos tradicionais que assumem normalidade e independência.

2.1 Análise de Dependência

O modelo de cópula, inicialmente proposto por Sklar (1959), é uma ferramenta que estabelece uma ligação de dependência entre variáveis aleatórias. Sklar focou na análise da distribuição tridimensional conjunta, introduzindo funções auxiliares definidas no intervalo unitário para estabelecer uma conexão entre a função de distribuição e suas distribuições marginais (SKLAR, 1959). Essas funções, conhecidas como cópulas, permitem a separação das distribuições marginais da estrutura de dependência de uma distribuição multivariada específica. Em outras palavras, cópulas são funções que ligam funções de distribuição multivariadas com suas funções de distribuição marginais univariadas, ou, alternativamente, são funções de distribuição multivariadas cujas marginais unidimensionais são uniformes em $[0, 1]$ (NELSEN, 2006; ANJOS et al., 2004).

O Teorema de Sklar afirma que, para variáveis aleatórias X_1, \dots, X_d com dependência completamente descrita pela função de distribuição conjunta H e distribuições marginais uniformes no intervalo $[0, 1]$, existe uma função C , denominada cópula d -dimensional, que relaciona as funções de distribuição marginais com a função de distribuição conjunta.

Anjos et al. (2004) mostra duas maneiras de se caracterizar uma cópula. A primeira delas como uma função distribuição conjunta e outra como uma função que precisa cumprir determinados requisitos.

Definição 2.1.1. Uma cópula é uma função de distribuição conjunta

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d), \quad \forall (u_1, \dots, u_d) \in [0, 1]^d \quad (1)$$

cujas marginais são todas uniformes no intervalo $[0, 1]$.

Portanto a condição que caracteriza uma função distribuição conjunta como cópula é que as marginais devem ser uniformes. Isto implica que toda cópula é uma função distribuição conjunta mas nem toda função distribuição conjunta é uma cópula.

A segunda maneira é:

Definição 2.1.2. Uma cópula é qualquer função $C : [0, 1]^n \rightarrow [0, 1]$ que tem as seguintes propriedades:

- (i) $C(x_1, \dots, x_n)$ é crescente em cada componente x_i ;
- (ii) $C(1, \dots, 1, x_i, 1, \dots, 1) = x_i$ para todo $i = 1, \dots, n$;
- (iii) Para todos $(a_1, \dots, a_n), (b_1, \dots, b_n) \in [0, 1]^n$ com $a_i \leq b_i$ tem-se que

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1+\dots+i_n} C(x_{1i_1}, \dots, x_{ni_n}) \geq 0, \quad (2)$$

com $x_{j1} = a_j$ e $x_{j2} = b_j$, para $j = 1, \dots, n$.

2.1.1 Teorema de Sklar

A teoria de cópulas é embasada no teorema de Sklar (1959). Conforme esse teorema, é possível analisar uma distribuição multivariada por meio de sua estrutura de dependência, conhecida como cópula, juntamente com suas distribuições marginais.

Teorema 2.1.1 (Teorema de Sklar). *Suponha X_1, \dots, X_d sejam variáveis aleatórias com funções de distribuição contínuas F_1, \dots, F_d e função de distribuição conjunta F , então existe uma única cópula C (uma função definida no cubo unitário $[0, 1]^d$ com marginais uniformes) tal que para todo $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$:*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (3)$$

Inversamente, dadas quaisquer funções de distribuição F_1, \dots, F_d e uma cópula C , pode-se definir uma função de distribuição d -variada F com funções de distribuição marginais F_1, \dots, F_d .

Também se pode reescrever (3) para $u = (u_1, \dots, u_d)^T \in [0, 1]^d$ como:

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)). \quad (4)$$

Sendo, F_i^{-1} , uma função inversa generalizada, ou seja, $F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)$, são as funções inversas de $F_1(x_1), \dots, F_n(x_n)$.

Portanto, $C(u_1, \dots, u_n)$ é conhecida, ela é a distribuição multivariada das variáveis transformadas (ANJOS et al., 2004). Em que, a densidade associada a uma cópula é definida através das derivadas parciais em relação a \mathbf{u} da função $C(u_1, \dots, u_n)$:

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} \quad (5)$$

Assim, conforme mostrado por Luciano e Cherubini (2001), a maior vantagem do uso da Teoria de Cópuas se dá pela forma como elas expressam uma distribuição conjunta de probabilidades. A adoção de cópuas permite que a seleção das distribuições marginais seja feita de forma autônoma da modelagem da estrutura de ligação entre as variáveis em estudo.

Nesse sentido, ao construir uma distribuição conjunta envolvendo duas ou mais variáveis aleatórias, é possível modelar cada uma delas individualmente com diferentes distribuições marginais, tais como a Normal, a T-Student, a Exponencial, a Quiquadrado, a Weibull, entre outras. Assim, ao mesmo tempo, a dependência entre essas variáveis pode assumir diversas estruturas, incluindo aquelas que são não-lineares, dependendo do tipo de cópula utilizado (BARCELLOS, 2022).

Exemplo 2.1.1. (GRUGEL, 2015 apud BARCELLOS, 2022) Vamos considerar um vetor aleatório bi-dimensional (X, Y) , com função de distribuição conjunta dada por:

$$F_{X,Y}(x, y) = \begin{cases} 1 - e^{-\lambda y} - e^{-\lambda x} + e^{-\lambda(x+y)} & , \text{ se } x > 0, y > 0 \\ 0 & , \text{ caso contrário.} \end{cases}$$

Posto isto, o primeiro passo é encontrar as distribuições marginais, o que pode ser feito da seguinte maneira:

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = \lim_{y \rightarrow \infty} (1 - e^{-\lambda y} - e^{-\lambda x} + e^{-\lambda(x+y)}) = 1 - e^{-\lambda x}.$$

e

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = \lim_{x \rightarrow \infty} (1 - e^{-\lambda y} - e^{-\lambda x} + e^{-\lambda(x+y)}) = 1 - e^{-\lambda y}.$$

Após a aplicação da teoria supramencionada, obteve-se as distribuições marginais

de X e Y , sendo elas distribuições Exponenciais com parâmetro λ .

$$X \sim \text{Exp}(\lambda) \text{ e } Y \sim \text{Exp}(\lambda)$$

Encontrando $x = h(u)$ (x em função de u e $y = g(v)$ (y em função de v , tem-se que:

$$\begin{aligned} u &= 1 - e^{-\lambda x} \\ u - 1 &= -e^{-\lambda x} \\ 1 - u &= e^{-\lambda x} \\ \log(1 - u) &= -\lambda x \\ x &= \frac{-\log(1 - u)}{\lambda} \end{aligned}$$

$$\text{Analogamente, } y = \frac{-\log(1-v)}{\lambda}$$

Desta maneira, pode-se encontrar a cópula associada ao vetor aleatório (X, Y) :

$$\begin{aligned} C(u, v) &= F\left(\frac{-\log(1-u)}{\lambda}, \frac{-\log(1-v)}{\lambda}\right) \\ &= 1 - e^{-\frac{\lambda \log(1-v)}{\lambda}} - e^{-\frac{\lambda \log(1-u)}{\lambda}} + e^{-\lambda\left(\frac{\log(1-u)}{\lambda} + \frac{\log(1-v)}{\lambda}\right)} \\ &= 1 - (1-v) - (1-u) + (1-u)(1-v) \\ &= 1 - 1 + v - 1 + u + 1 - v - u + uv \\ &= uv \end{aligned}$$

Com base no Teorema de Sklar, como $C(u, v) = uv$, pode-se dizer que X e Y são independentes.

2.1.2 Limites de Fréchet-Hoeffding

O Teorema 2.1.2 afirma que qualquer cópula C é limitada pontualmente por baixo pelo limite inferior de Fréchet-Hoeffding, W , e por cima pelo limite superior de Fréchet-Hoeffding, M , onde W e M são definidos por (HOFERT et al., 2018)

$$W(u) = \max\left\{\sum_{j=1}^d u_j - d + 1, 0\right\} \quad \text{e} \quad M(u) = \min\{u_j\}, \quad u \in [0, 1]^d. \quad (6)$$

É importante notar que W é uma cópula somente se $d = 2$ enquanto M é uma cópula para todo $d \geq 2$.

Teorema 2.1.2 (Limites de Fréchet–Hoeffding). *Para qualquer cópula C de di-*

mensão d ,

$$W(u) \leq C(u) \leq M(u), \quad u \in [0, 1]^d. \quad (7)$$

Se $U \sim U(0, 1)$, então

$$(U, 1 - U) \sim W \quad \text{e} \quad (U, \dots, U) \sim M. \quad (8)$$

2.1.3 Medidas de dependência e concordância

A relação entre variáveis aleatórias é estabelecida pela função de distribuição conjunta. Essa função contém todas as informações sobre a relação entre as variáveis. Para entender a natureza dessa dependência e como ela é medida, são utilizadas medidas de dependência. Essas medidas quantificam a relação entre as variáveis, podendo manifestar-se de diferentes maneiras, como dependência linear, caudal, entre outras. Cada medida de dependência captura certos aspectos da dependência, priorizando alguns em detrimento de outros. Isso é um fator crucial ao medir e, principalmente, avaliar o que a medida de dependência informa, pois cada medida de dependência é adequada para medir um aspecto da dependência, e se for usada para medir um aspecto diferente, pode levar a conclusões e interpretações incorreta (ANJOS et al., 2004).

A análise da dependência entre variáveis aleatórias é essencial em várias áreas do conhecimento, como finanças, seguros, economia e engenharia. Uma melhor avaliação de riscos, modelagem de cenários e tomada de decisões são possíveis com uma compreensão das relações entre as variáveis. O estudo dos vários tipos de dependência é importante pois um dado modelo de cópula pode ser mais adequado para um tipo de dependência pretendido do que para outro (VIOLA, 2009). De acordo com o autor, a teoria de cópula, que surgiu recentemente, é um dos primeiros tópicos a ser levados em consideração quando se trata de modelagem de dependência hoje em dia. Esta teoria é atraente porque as cópulas abrangem uma ampla gama de estruturas de dependência e podem modelar a estrutura de dependência dos dados de forma completa. A modelagem por distribuição normal, segundo Viola, pode modelar uma grande variedade de estruturas, bem como sua simplicidade analítica e a facilidade de estimar a matriz de correlação, seu único parâmetro de dependência. Por outro lado, sua curtose e simetria, por exemplo, limitam sua utilização. Devido à sua maior diversidade de estruturas de dependência, a modelagem por cópulas torna-se atraente neste contexto.

Em modelagem estatística e econométrica de séries de dados multivariados as medidas de dependência se tornam ferramentas essenciais para a análise, pois permitem não apenas quantificar as relações entre variáveis, mas também caracterizá-las e diferenciá-las

de acordo com características de dependência como caudalidade ou linearidade. No entanto, a validade e precisão das inferências feitas são determinadas pela seleção criteriosa da medida de dependência adequada para capturar o aspecto relevante da interação.

2.1.3.1 Coeficiente de Correlação Linear

Como bem define Anjos et al. (2004), seja (X, Y) um vetor de variáveis aleatórias com variâncias positivas finitas. O coeficiente de correlação linear de (X, Y) é definido como

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (9)$$

onde $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$ é a covariância entre X e Y e $Var(X)$, $Var(Y)$ são as variâncias de X e Y respectivamente.

As principais propriedades do coeficiente de correlação linear são:

1. $-1 \leq \rho(X, Y) \leq 1$;
2. Se X e Y são independentes, então $\rho(X, Y) = 0$;
3. $\rho(aX + b, cY + d) = \text{sign}(ac)\rho(X, Y)$ para todos os reais $a \neq 0$ e $c \neq 0$, b e d .

O coeficiente $\rho(X, Y)$ atua como um parâmetro que quantifica a dependência linear entre duas variáveis, mantendo-se constante sob transformações lineares estritamente ascendentes. Além disso, $\rho(X, Y)$ serve como um indicador de dependência inerente em distribuições normais multivariadas e, de maneira mais abrangente, em famílias de distribuições elípticas e copulas.

A correlação linear é uma medida de dependência que só é não-invariante sob transformações não lineares estritamente crescentes $T : \mathbb{R} \rightarrow \mathbb{R}$, isto é

$$\rho(X, T(Y)) \neq \rho(X, Y). \quad (10)$$

Tem-se então, que esta trata-se de uma medida de dependência, muito utilizada devido à seu método simplificado de estimação, porém a mesma apresenta algumas limitações como:

- O coeficiente de correlação linear não é invariante para transformações não lineares das variáveis aleatórias;
- Não é uma medida de associação robusta, pois uma única observação pode ter uma influência significativa no valor de ρ , logo, não se pode utilizá-lo quando a amostra apresenta *outliers*.

Outras falhas da correlação linear discutidas por Embrechts, McNeil e Straumann (2002 apud ANJOS et al., 2004) mostram que a correlação pode facilmente ser mal interpretada. Alguns casos típicos de uso incorreto são:

1. $\rho(X, Y) = 0$ não implica que X e Y são independentes.
2. $\rho(X, Y) = 0$ não implica que não existe dependência entre X e Y .
3. Para distribuições marginais de X e Y dadas, o intervalo de variação de $\rho(X, Y)$ é $[-1, 1]$.

2.1.4 Medidas de concordância

Para definir a noção de concordância (e também para entender melhor a definição do tau de Kendall e Rho de Sperman), considere dois pontos em \mathbb{R}^2 denotados por (x_1, x_2) e (x'_1, x'_2) . Diz-se que esses pontos são concordantes se $(x_1 - x'_1)(x_2 - x'_2) > 0$ (ou seja, se a inclinação da linha que passa pelos dois pontos for positiva) e discordantes se $(x_1 - x'_1)(x_2 - x'_2) < 0$ (indicando uma inclinação negativa) (HOFERT et al., 2018). Não é necessário se preocupar com empates, já que o caso limite $(x_1 - x'_1)(x_2 - x'_2) = 0$ ocorre com probabilidade zero sob a suposição de que x e x' são contínuos (GENEST; FAVRE, 2007).

2.1.4.1 Tau de Kendall

A medida de concordância Tau de Kendall é definida como a diferença entre a probabilidade de concordância e a probabilidade de discordância entre duas variáveis aleatórias. Seja τ a medida de associação de Kendall, definida por:

$$\tau = P((X_1 - X'_1)(X_2 - X'_2) > 0) - P((X_1 - X'_1)(X_2 - X'_2) < 0), \quad (11)$$

isto é, o tau de Kendall é a probabilidade de concordância menos a probabilidade de discordância dos pares aleatórios (X_1, X_2) e (X'_1, X'_2) . Se X_2 tende a aumentar (respectivamente, diminuir) com X_1 , então a probabilidade de concordância (relativamente à probabilidade de discordância) é alta (respectivamente, baixa). Esta é a intuição por trás da definição do tau de Kendall. Observando que, sob a suposição de marginais contínuas, $(X_1 - X'_1)(X_2 - X'_2) \neq 0$ com probabilidade um, obtém-se imediatamente que

$$\tau = 2P((X_1 - X'_1)(X_2 - X'_2) > 0) - 1. \quad (12)$$

Para calcular o estimador do tau de Kendall. Seja uma amostra bivariada aleatória $(X_{11}, X_{12}), \dots, (X_{n1}, X_{n2})$ de (X_1, X_2) :

$$\tau_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{signal}((X_{i1} - X_{j1})(X_{i2} - X_{j2})). \quad (13)$$

Esta expressão pode ser reescrita como uma versão amostral da Equação 12:

$$\tau_n = 2 \frac{p_n}{\binom{n}{2}} - 1 = \frac{4p_n}{n(n-1)} - 1, \quad (14)$$

onde p_n é o número de pares concordantes na amostra.

Tendo isso em consideração, sejam X e Y variáveis aleatórias contínuas com distribuição conjunta definida pela cópula C , podemos afirmar que a versão do coeficiente tau de Kendall, definida como uma função da cópula (τC), é expressa na seguinte equação.

$$\tau = \tau C = 4E(C(U, V)) - 1 \Leftrightarrow \iint_{I^2} C(u, v) dC(u, v) - 1 \quad (15)$$

Portanto, a estimativa do coeficiente Tau de Kendall pode ser empregada na dedução da cópula que estabelece a relação entre as variáveis aleatórias X e Y (BARCELLOS, 2022).

2.1.4.2 Rho de Spearman

O coeficiente de correlação de Spearman, também conhecido como correlação ordinal, é definido como o coeficiente de correlação linear entre as ordens dos valores de X e Y .

Definição 2.1.3. Seja (X, Y) vetores aleatórios reais. É chamado de coeficiente de correlação de Spearman entre X e Y o número real.

$$\rho_S(X, Y) = \rho_L(F_x(X), F_y(Y)) \quad (16)$$

Proposição 2.1.1. *Seja, (X, Y) um vetor aleatório real com cópula C , é possível escrever ρ_S em função da cópula pertinente:*

$$\rho_S(X, Y) = 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} dudv \quad (17)$$

2.2 Famílias de Cópulas

Conforme o Teorema de Sklar, é possível abordar uma distribuição multivariada considerando sua estrutura de dependência, cópula, e suas distribuições marginais. Esse enfoque confere uma notável flexibilidade à modelagem por meio de funções cópulas (CHERUBINI; LUCIANO; VECCHIATO, 2004). Portanto, existe uma ampla variedade de funções cópulas propostas na literatura. Abaixo, apresentaremos algumas das famílias de cópulas mais reconhecidas e mencionadas, ressaltando que há inúmeras famílias distintas disponíveis. Aqui, listaremos apenas as principais para destacar suas características distintivas e suas estruturas, que podem ser compreendidas por meio das curvas de nível associadas a cada uma.

Para realizar uma modelagem precisa da dependência entre variáveis aleatórias, é necessário selecionar a família de cópula correta. Cada família tem características e atributos distintos que a tornam mais adequada para lidar com vários tipos de dependência. Para auxiliar na escolha da família de cópula mais adequada considera-se as propriedades da dependência (Linear, monotônica, assimétrica, por exemplo), o objetivo da análise (Medição de risco e modelagem de séries temporais), a flexibilidade da família, ou seja, a possibilidade de ajustar a modelagem à dependência real entre as variáveis e a interpretabilidade da família, o que se traduz na Facilidade de compreensão dos resultados da modelagem (CHERUBINI; LUCIANO; VECCHIATO, 2004). De acordo com os autores, a escolha da família de cópula é um passo crucial na modelagem da dependência. A família selecionada deve ser capaz de capturar as características da dependência real entre as variáveis, como linearidade, monotonicidade, assimetria e cauda pesada.

2.2.1 Cópulas Elípticas

A classe de distribuições elípticas oferece uma variedade de distribuições multivariadas, e essa família de cópulas apresenta várias propriedades em comum com a distribuição normal multivariada, o que facilita a geração de dados a partir dessa distribuição (EMBRECHTS; LINDSKOG; MCNEIL, 2001). As principais distribuições elípticas incluem as distribuições multivariadas Gaussiana e a distribuição t-Student, utilizando o coeficiente de correlação linear como medida de dependência.

Definição 2.2.1. (HOFERT et al., 2018) Um vetor aleatório X de dimensão d tem uma distribuição elíptica com vetor de média $\mu \in \mathbb{R}^d$ e matriz de covariância $\Sigma = AA'$ com $\text{posto}(\Sigma) = k \leq d$ para uma matriz $A \in \mathbb{R}^{d \times k}$ se

$$X \stackrel{d}{=} \mu + RS, \tag{18}$$

onde $S \sim U(\{x \in \mathbb{R}^k : \|x\| = 1\})$, $\|\cdot\|$ denota a norma Euclidiana (isto é, S é uniformemente distribuído na esfera unitária em \mathbb{R}^k) e R e S são independentes.

2.2.1.1 Cópula Gaussiana

A cópula Gaussiana é estabelecida inicialmente ao utilizar a função de distribuição multivariada Normal como ponto de partida. Posteriormente, essa cópula é obtida ao substituir as variáveis aleatórias por suas inversas generalizadas.

Definição 2.2.2. Caso Bivariado A cópula gaussiana é definida da seguinte forma:

$$C^{Ga}(v, z) = \Phi_{\rho_{XY}}(\Phi^{-1}(v), \Phi^{-1}(z)) \quad (19)$$

onde $\Phi_{\rho_{XY}}$ é a função de distribuição conjunta de um vetor normal padrão bidimensional com coeficiente de correlação linear ρ_{XY} , e Φ é a função de distribuição normal padrão. Portanto

$$\Phi_{\rho_{XY}}(\Phi^{-1}(v), \Phi^{-1}(z)) = \int_{-\infty}^{\Phi^{-1}(v)} \int_{-\infty}^{\Phi^{-1}(z)} \frac{1}{2\pi\sqrt{1-\rho_{XY}^2}} \exp\left(\frac{2\rho_{XY}st - s^2 - t^2}{2(1-\rho_{XY}^2)}\right) ds dt \quad (20)$$

Uma vez que é parametrizada pelo coeficiente de correlação linear, também podemos escrever C^{Ga} . A representação a seguir é equivalente a (CHERUBINI; LUCIANO; VECCHIATO, 2004)

$$C^{Ga}(v, z) = \int_0^v \Phi\left(\frac{\Phi^{-1}(z) - \rho_{XY}\Phi^{-1}(t)}{\sqrt{1-\rho_{XY}^2}}\right) dt \quad (21)$$

A característica notável da cópula Gaussiana é que ela possui um parâmetro de correlação no intervalo de $-1 \leq \rho \leq 1$ e é simétrica. A intensidade da dependência aumenta à medida que o valor de ρ se aproxima de 1, e ela não exibe dependência na cauda, sugerindo que as variáveis aleatórias não estão correlacionadas em eventos extremos (BARCELLOS, 2022).

2.2.1.2 Cópula t-Student

De maneira similar à cópula Gaussiana, cópula t-Student também simétrica. Sua vantagem reside na capacidade de modelar com um grau mais elevado de dependência nas caudas da distribuição.

Definição 2.2.3. Caso Bivariado A cópula bivariado t-Student, $T_{\rho,\nu}(v, z)$, é definida como:

$$\begin{aligned} T_{\rho,\nu}(v, z) &= t_{\rho,\nu}(t_{\nu}^{-1}(v), t_{\nu}^{-1}(z)) \\ &= \int_{-\infty}^{t_{\nu}^{-1}(v)} \int_{-\infty}^{t_{\nu}^{-1}(z)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt \end{aligned} \quad (22)$$

Onde $t_{\rho,\nu}$ corresponde à distribuição bivariada t-Student com ν graus de liberdade, e t_{ν}^{-1} é a inversa da distribuição t-Student.

2.2.2 Cópuas Arquimedianas

As cópuas Arquimedianas apresenta diversas subfamílias, como a Gumbel, Joe, Frank, Clayton, entre outras. Esta cópula tem como vantagem, que sua construção permite simplificar a análise de cópuas multivariadas para o de funções univariadas.

Cherubini, Luciano e Vecchiato (2004) mostram que as Copulas arquimedianas podem ser construídas usando uma função $\phi : I \rightarrow \mathbb{R}^{++}$, contínua, decrescente, convexa e tal que $\phi(1) = 0$. Tal função ϕ é chamada de *geradora*. É chamada de geradora estrita sempre que $\phi(0) = +\infty$. A pseudo-inversa de ϕ também deve ser definido, da seguinte forma:

$$\phi^{[-1]}(v) = \begin{cases} \phi^{-1}(v) & 0 \leq v \leq \phi(0) \\ 0 & \phi(0) \leq v \leq +\infty \end{cases} \quad (23)$$

Essa pseudo-inversa é tal que, pela composição com a geradora, ela fornece a identidade, assim como inversos ordinários fazem para funções com domínio e imagem \mathbb{R} :

$$\phi^{[-1]}(\phi(v)) = v \quad \forall v \in I \quad (24)$$

Além disso, ele coincide com o inverso usual se ϕ é um gerador estrito.

Definição 2.2.4. Dado uma função geradora e sua pseudo-inversa, uma cópula arquimediana C^A é gerada da seguinte forma:

$$C^A(v, z) = \phi^{[-1]}(\phi(v) + \phi(z)) \quad (25)$$

Se a geradora é estrita, a cópula é dita ser uma cópula arquimediana estrita.

Definição 2.2.5. Dado que C^A é uma Cópula Arquimediana contínua, então a sua den-

sidade é dada por:

$$C^A(v, z) = \frac{-\phi''(C(v, z))\phi'(v)\phi'(z)}{(\phi'(C(v, z)))^3} \quad (26)$$

2.2.2.1 Cópula Gumbel-Hougaard

A Função geradora da cópula de Gumbel-Hougaard é dada por (CHERUBINI; LUCIANO; VECCHIATO, 2004):

$$\phi_\alpha(t) = (-\ln t)^\alpha, \quad \forall \theta \geq 1 \quad (27)$$

Logo, sua cópula é dada pela equação:

$$C(v, z) = \exp\{-[(-\ln v) + (-\ln z)^\theta]^{1/\theta}\} \quad (28)$$

onde $\alpha \in [-1, +\infty)$. O parâmetro α é o indica o nível de dependência entre as variáveis. Quando $\alpha = 1$, a cópula de marginais independentes é obtida. Por outro lado, quando $\alpha \rightarrow \infty$, a cópula que representa a dependência perfeita positiva é alcançada. Assim, a cópula de Gumbel-Hougaard é capaz de modelar situações que vão desde a independência até a dependência perfeita positiva. Esta cópula é a única cópula Arquimediana de valores extremos e exibe dependência na cauda superior. Nelsen (2006 apud BARCELLOS, 2022).

2.2.2.2 Cópula Frank

A função geradora da cópula de Frank dada por:

$$\phi_\alpha(t) = -\ln \left(\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1} \right) \quad (29)$$

em que sua cópula é dada por

$$C(v, z) = -\frac{1}{\alpha} \ln \left(1 + \frac{(\exp(-\alpha v) - 1)(\exp(-\alpha z) - 1)}{\exp(-\alpha) - 1} \right) \quad (30)$$

onde $\alpha \in (-\infty, \infty)$. A cópula de Frank se torna uma escolha apropriada quando as variáveis têm igual probabilidade de exibir valores simultaneamente baixos ou elevados.

2.2.2.3 Cópula Clayton

A cópula Clayton, dentre as Arquimedianas, é apresentada uma dependência mais pronunciada na cauda negativa em comparação com a cauda positiva e apresenta a seguinte função geradora

$$\phi_\alpha(t) = \frac{1}{\alpha}(t^\alpha - 1) \quad (31)$$

em que sua cópula é dada por

$$C(v, z) = \max [(v^\alpha + z^\alpha - 1)^{1/\alpha}, 0] \quad (32)$$

onde, $\alpha \in [-1, \infty)$

2.3 Estimação

Nesta seção, será mostrado duas formas de estimação de uma cópula bivariada. O primeiro é o método da Máxima Verossimilhança (caso paramétrico), enquanto o segundo é através da cópula empírica (caso não paramétrico).

2.3.1 Método de Máxima Verossimilhança (EMV)

Anjos et al. (2004) mostra o Método de Máxima Verossimilhança para uma cópula n-variada, aqui será mostrado para o caso da cópula bivariada. Seja X_1, X_2 o conjunto de variáveis aleatórias de interesse com funções distribuição marginais $F_1(x_1)$ e $F_2(x_2)$ respectivamente e $H(x_1, x_2)$ sua função distribuição conjunta. Se $H(x_1, x_2)$ for absolutamente contínua então,

$$h(x_1, x_2) = \frac{\partial^2 H(x_1, x_2)}{\partial x_1 \partial x_2} \quad (33)$$

aplicando o teorema de Sklar (1959) em Equação 33 temos que,

$$h(x_1, x_2) = \frac{\partial^2 C(F_1(x_1), F_2(x_2))}{\partial F_1(x_1) \partial F_2(x_2)} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \quad (34)$$

em que,

$$\frac{\partial^2 C(F_1(x_1), F_2(x_2))}{\partial F_1(x_1) \partial F_2(x_2)} = c(F_1(x_1), F_2(x_2)) \quad (35)$$

é a densidade da cópula. A Equação 35, também é chamada de função dependência. Deste modo, a densidade conjunta $h(\cdot)$, em termos de cópula, é dada por

$$h(x_1, x_2; \Theta) = c(F_1(x_1; \theta_1), F_2(x_2; \theta_2); \theta) \prod_{i=1}^2 f_i(x_i; \theta_i) \quad (36)$$

onde $\Theta = (\theta_1, \theta_2, \theta)$ é o vetor de parâmetros.

Seja \mathbf{X} uma amostra de tamanho m

$$L(\mathbf{X}; \Theta) = \prod_{j=1}^m c(F_1(x_{1j}), F_2(x_{2j}); \theta) \prod_{j=1}^m \prod_{i=1}^2 f_i(x_{i,j}; \theta_i) \quad (37)$$

A log-verossimilhança é então:

$$\ell(\mathbf{X}; \Theta) = \sum_{j=1}^m \ln c(F_1(x_{1j}), F_2(x_{2j}); \theta) + \sum_{j=1}^m \sum_{i=1}^2 \ln f_i(x_{i,j}; \theta_i) \quad (38)$$

Os estimadores de máxima verossimilhança para os parâmetros são obtidos maximizando a log-verossimilhança em relação a θ :

$$\hat{\theta} = \arg \max_{\theta} \sum_{j=1}^m \frac{\partial \ln c(F_1(x_{1j}), F_2(x_{2j}); \theta)}{\partial \theta} \quad (39)$$

e

$$\hat{\theta}_i = \arg \max_{\theta_i} \left(\sum_{j=1}^m \frac{\partial \ln c(F_1(x_{1j}), F_2(x_{2j}); \theta)}{\partial \theta_i} + \sum_{j=1}^m \frac{\partial \ln f_i(x_{i,j}; \theta)}{\partial \theta_i} \right) \quad (40)$$

2.3.2 Cópulas Empíricas

O estimador não paramétrico de uma cópula bivariada⁵ é dada por:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(U_{i,1} \leq u, U_{i,2} \leq v), \quad (u, v) \in [0, 1]^2 \quad (41)$$

onde $U_{i,1}$ e $U_{i,2}$ são as pseudo-observações para as duas dimensões. As desigualdades são entendidas componente a componente, o que significa que ambas as condições $U_{i,1} \leq u$ e $U_{i,2} \leq v$ devem ser satisfeitas para que a função indicadora seja igual a 1.

⁵ Uma definição mais detalhada pode ser encontrada em Hofert et al. (2018)

2.4 Copula de sobrevivência

Considere a distribuição de probabilidade conjunta $F(x_1, x_2) \equiv \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$. E vamos ter uma probabilidade conjunta $\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2)$. Em que a relação entre $F(x_1, x_2)$ e $\bar{F}(\mathbf{x})$ é dada por (CHERUBINI et al., 2011):

$$\bar{F}(x_1, x_2) = \mathbb{P}(X_1 > x_1) + \mathbb{P}(X_2 > x_2) - 1 + \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \quad (42)$$

A probabilidade $\bar{F}(x_1, x_2)$ é chamada de *probabilidade de sobrevivência*. Aplicando o teorema de Sklar (Teorema 2.1.1), podemos obter uma função de cópula $\hat{C}(u_1, u_2)$ tal que

$$\bar{F}(x_1, x_2) = \hat{C}(1 - F_1(x_1), 1 - F_2(x_2)) \quad (43)$$

$\hat{C}(u_1, u_2)$ é chamado de *cópula de sobrevivência* correspondente a $C(u_1, u_2)$.

2.5 Copulas rotacionadas

Considere C como uma cópula com dimensão d . Conforme Hofert et al. (2018, pg. 42 - Proposição 2.5.2) mostra, que se temos $U \sim C$ ⁶, então $1 - U$ terá distribuição segundo a cópula de sobrevivência \hat{C} , que é associada a C . Especificamente em duas dimensões, podemos obter uma amostra de \hat{C} refletindo ortogonalmente uma amostra de C primeiramente na linha $u_1 = 1/2$ e em seguida na $u_2 = 1/2$. É importante notar que a sequência das reflexões ortogonais não altera o resultado final, o que também pode ser interpretado como um tipo específico de rotação (HOFERT et al., 2018).

Hofert et al. (2018) argumenta que a ideia de cópulas rotacionadas expande a ideia sobre cópulas de sobrevivência. Dada uma cópula C com dimensão d e uma variável aleatória U distribuída conforme C , e dado um vetor r em $\{0, 1\}^d$, a cópula rotacionada $\text{rot}_r(C)$ baseada em r é definida de tal forma que $U \sim C$ é equivalente a

$$((1 - r_1)U_1 + r_1(1 - U_1), \dots, (1 - r_d)U_d + r_d(1 - U_d)) \sim \text{rot}_r(C). \quad (44)$$

É fácil notar que a cópula de sobrevivência \hat{C} de uma cópula C é simplesmente $\text{rot}_1(C)$. De maneira simplificada, para obter uma amostra de $\text{rot}_r(C)$ a partir de uma amostra de C , invertamos as dimensões de r que são iguais a 1.

⁶ Lembrando que $U_i = F_i(x_i)$

2.6 Métricas de desempenho

O critério de seleção dos modelos se da por meio dos critérios de informação de Akaike (AIC) ou o critério de informação Schwartz (*Bayesian information Criterion* - BIC). A ideia por trás do critério de informação é tentar mensurar a qualidade de um modelo de acordo com a informação que ele carrega e também eles presam pela simplicidade, quer dizer que quanto mais parâmetros e mais defasagens for adicionado, mais penalizado será o modelo de acordo com o critério de informação prezando sempre por um modelo mais parcimonioso.

O ideal é que o AIC e BIC sejam menor possível. Quanto melhor for o ajuste do modelo, mais AIC e BIC irá se aproximar de $-\infty$. São calculados da seguinte maneira

$$\text{AIC} = \frac{-2\ln(L)}{T} + \frac{2n}{T} \quad (45)$$

$$\text{BIC} = \frac{-2\ln(L)}{T} + \frac{n\ln(T)}{T} \quad (46)$$

onde $n = p + q + 1$ que é soma dos parâmetros do modelo; T é o tamanho da amostra; L é o valor máximo da função de máxima verossimilhança.

Note que a diferença entre eles é que no BIC vamos ter $\ln(T)$ ao invés do 2 multiplicando o n e como $\ln(T)$ sempre será maior que 2, quer dizer que o BIC é mais parcimonioso que o AIC; o custo marginal de se adicionar regressores é maior no BIC do que no AIC.

2.7 Testes

2.7.1 Kolmogorov-Smirnov

DeGroot e Schervish (2021) demonstram que o teste de Kolmogorov-Smirnov visa avaliar se uma função de distribuição acumulada (f.d.a.) desconhecida, $F(x)$, é, na verdade, uma f.d.a. contínua $F^*(x)$. Em outras palavras, o teste de hipótese é formulado da seguinte maneira:

$$H_0 : F(x) = F^*(x) \text{ para } -\infty < x < \infty, \quad (47)$$

$$H_1 : F(x) \neq F^*(x) \quad (48)$$

Este problema é não paramétrico, uma vez que a distribuição desconhecida, da qual a amostra aleatória é retirada, pode ser qualquer distribuição contínua. A estatística de teste associada é definida como:

$$D_n^* = \sup_{-\infty < x < \infty} |F_n(x) - F^*(x)|. \quad (49)$$

Em outras palavras, D_n^* representa a diferença máxima entre a f.d.a. da amostra $F_n(x)$ e a f.d.a. hipotetizada $F^*(x)$. Quando a hipótese nula H_0 é verdadeira, a distribuição de probabilidade de D_n^* segue uma certa distribuição, que é a mesma para qualquer possível f.d.a. contínua $F^*(x)$ (DEGROOT; SCHERVISH, 2021).

2.7.2 Teste de bondade do ajuste

Segundo Hofert et al. (2018), o teste de bondade de ajuste busca verificar se dada um copula C ela realmente vai pertencer a uma família de copula \mathcal{C} , o teste de hipótese é dado como:

$$H_0 : C \in \mathcal{C} = \{C_\theta : \theta \in \Theta\} \quad (50)$$

$$H_1 : C \notin \mathcal{C} \quad (51)$$

2.8 **Cópulas - Avaliação da estrutura de dependência entre fundos de investimentos e variáveis macroeconômicas**

O objetivo deste trabalho é investigar a estrutura de dependência do retorno fundos multimercados brasileiros com variáveis macroeconômicas e financeiras. Para tal, foi utilizada uma métrica específica a fim de minimizar o viés de seleção dos fundos. A amostra incluiu fundos com mais de 15 anos de existência e que ainda estavam em atividade até 01/03/2023, sendo filtrados aqueles com os maiores retornos acumulados e maior patrimônio líquido.

A relação funcional entre o retorno dos fundos e as variáveis selecionadas está descrita na Equação 52:

$$Retorno = f(IBov, IMA.Geral, CDS.5y, IBC.Br, Cambio) \quad (52)$$

A decisão de representar o retorno dos fundos como uma função dessas variáveis baseia-se no fato de que o índice IBovespa é composto pelas ações mais negociadas na bolsa, o índice IMA-Geral reflete a política monetária por meio dos títulos públicos atrelados a inflação negociados pelos fundos, o *Credit Default Swap* de 5 anos (CDS-5y) serve como uma medida do risco do país, o IBC-Br é uma aproximação mensal do Produto Interno Bruto (PIB) brasileiro, e a taxa de câmbio.

Foi escolhido o software R para fazer toda a análise e modelagem dos dados, o pseudo-código pode ser visto no Anexo A. O método de estimação se deu em duas partes, a primeira foi tentar estimar as cópulas de forma paramétrica, onde primeiro foi testar qual é a melhor distribuição que se ajusta aos dados e para isso, foi utilizada a função

`fitdistrplus::fitdist()`, esta função ajusta um modelo univariado a um conjunto de dados, para que então possamos aplicar um teste de *Kolmogorov-Smirnov*.

Após identificar a distribuição que melhor se ajusta a cada série, foram calculadas suas respectivas funções de distribuição acumulada. O próximo passo foi identificar a família de cópulas que melhor se ajusta ao conjunto de dados. Para isso, foi utilizada a função `VineCopula::BiCopSelect()` para selecionar a função de cópulas que melhor se ajusta aos dados. Como critério de seleção, utilizou-se o valor da verossimilhança.

Após selecionar a família, utilizou-se a função `VineCopula::BiCopGofTest()` para realizar o teste de bondade do ajuste e garantir que esta família é adequada à respectiva cópula bivariada. Os parâmetros utilizados para realizar o teste de cada cópula foram os estimados pela função `VineCopula::BiCopSelect()`. O teste pode ser aplicado por dois métodos: pela matriz de informação, de acordo com Huang e Prokhorov (2014), ou baseado no processo de Kendall, conforme Wang e Wells (2000). Para realizar o teste da t-cópula foi utilizado o método da matriz de informação, quanto ao resto, foi utilizado o método baseado no processo de Kendall. Conforme demonstrado por Genest, Rémillard e Beaudoin (2009), o número de simulações (bootstrap) deve ser significativamente maior que o tamanho da amostra para aumentar o poder do teste, onde foram usadas 500 simulações.

Após a função selecionada passar no teste de bondade do ajuste, foi feita a estimação da cópula através do método de máxima verossimilhança. Um segundo teste de adequação foi feito, nele foi plotado gráficos de densidade entre as séries simuladas e observada, onde foi constatada que as copula não representavam os dados adequadamente.

Essa metodologia, de adotar um modelo paramétrico para a densidade da cópula c , cujos parâmetros são estimados através do método de máxima verossimilhança. Apesar da ampla gama de modelos paramétricos de cópula disponíveis, eles são conhecidos por sua falta de flexibilidade e risco de má especificação. A solução para essas questões pode ser encontrada nos estimadores de densidade não paramétricos. No entanto, dado que as cópulas existem em um suporte limitado - o hipercubo unitário - é necessário adaptar cuidadosamente os estimadores a essa situação (NAGLER, 2018).

A partir disso, ouve a decisão de se estimar as copula pelo estimador não paramétrico de kernel. O pacote `kdecopula` (NAGLER, 2018) implementa vários estimadores de densidade de cópula de kernel bivariados. O pacote fornece, métodos de estimação, seleção de largura de banda, simulação e visualização.

é comum na literatura, focamos no caso bivariado. Esta seção revisará diferentes abordagens para a estimativa do núcleo da densidade da cópula. Como

Um estimador de kernel da densidade da copula pode ser definido como: Suponha que temos observações iid $(U_i, V_i), i = 1, \dots, n$, de uma cópula bivariada C , onde queremos

a estimativa da densidade correspondente a $c(u, v)$. Pode-se aplicar o estimador de kernel:

$$\hat{c}_n(u, v) = \frac{1}{n} \sum_{i=1}^n K_{b_n}(u - U_i) K_{b_n}(v - V_i), \quad (u, v) \in [0, 1]^2, \quad (53)$$

onde $K_{b_n}(\cdot) = K(\cdot/b_n)/b_n$. A função kernel K é assumida como uma função de densidade de probabilidade simétrica e limitada em \mathbb{R}^2 e $b_n > 0$ é o parâmetro de suavização ou largura de banda. Contudo, surge uma dificuldade com o uso deste estimador. Uma parcela significativa da massa de probabilidade será atribuída para além dos limites do hipercubo unitário. Consequentemente, \hat{c}_n não pode ser considerada uma função de densidade válida sobre o domínio $[0, 1]^2$, visto que a sua integral não resulta em um. Ademais, o estimador é suscetível a um viés acentuado nas bordas do domínio (NAGLER, 2018).

Para contornar este problema, existe três métodos: *mirror-reflexion*, *beta kernel* e *transformation*. O presente trabalho fez a utilização do método *transformation*. Este método transforma os dados de modo com que estejam em todo \mathbb{R}^2 . Neste domínio técnicas padrão de kernel podem ser usadas para estimar a densidade. Uma retro transformação adequada então fornece uma estimativa da densidade da cópula. Para a transformação, o inverso da f.d.a. normal padrão é utilizada, dado que os estimadores kernel apresentam bom funcionamento com variáveis aleatórias Gaussianas (NAGLER, 2018).

Seja Φ uma f.d.a. normal padrão e ϕ como sua primeira derivada. Então $(X_i, Y_i) = (\Phi^{-1}(U_i), \Phi^{-1}(V_i))$ é um vetor aleatório com marginais gaussianas e cópula C . Pelo teorema de Sklar 2.1.1, a densidade correspondente f pode ser escrita como

$$f(x, y) = c(\Phi(x), \Phi(y))\phi(x)\phi(y). \quad (54)$$

A Equação 54 pode ser estimada por kernel. A partir de um estimador \hat{f}_n , pode-se derivar um estimador para a densidade da cópula c

$$\hat{c}_n^{(T)}(u, v) = \frac{\hat{f}_n(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \quad (u, v) \in [0, 1]^2. \quad (55)$$

Para a função \hat{f}_n , o pacote emprega uma função de kernel gaussiano. A estimação dessa função é realizada através do método de verossimilhança local em dois estágios com vizinhos próximos⁷ que envolve duas etapas. Na primeira etapa, as distribuições marginais dos dados são estimadas por kernel de maneira independente. Na segunda etapa, a cópula é estimada utilizando uma abordagem de vizinhança mais próxima, onde a log-densidade da cópula é ajustada levando em consideração as marginais já estimadas e a proximidade

⁷ Se refere ao algoritmo de *K-Vizinhos Mais Próximos*.

dos pontos de dados. Isso permite uma adaptação mais localizada e potencialmente mais precisa da densidade da cópula em regiões onde os dados são esparsos.

Após a estimação das cópulas, elaborou-se um gráfico de densidade entre as variáveis observadas e as simulações, de maneira que pudéssemos avaliar o quão bem ajustada está a cópula estimada.

Em seguida, analisou-se a probabilidade de ocorrência e o retorno esperado médio, observando-se a probabilidade de o retorno do fundo ser positivo ou negativo, dado que o retorno da variável está em um determinado intervalo.

3 RESULTADO

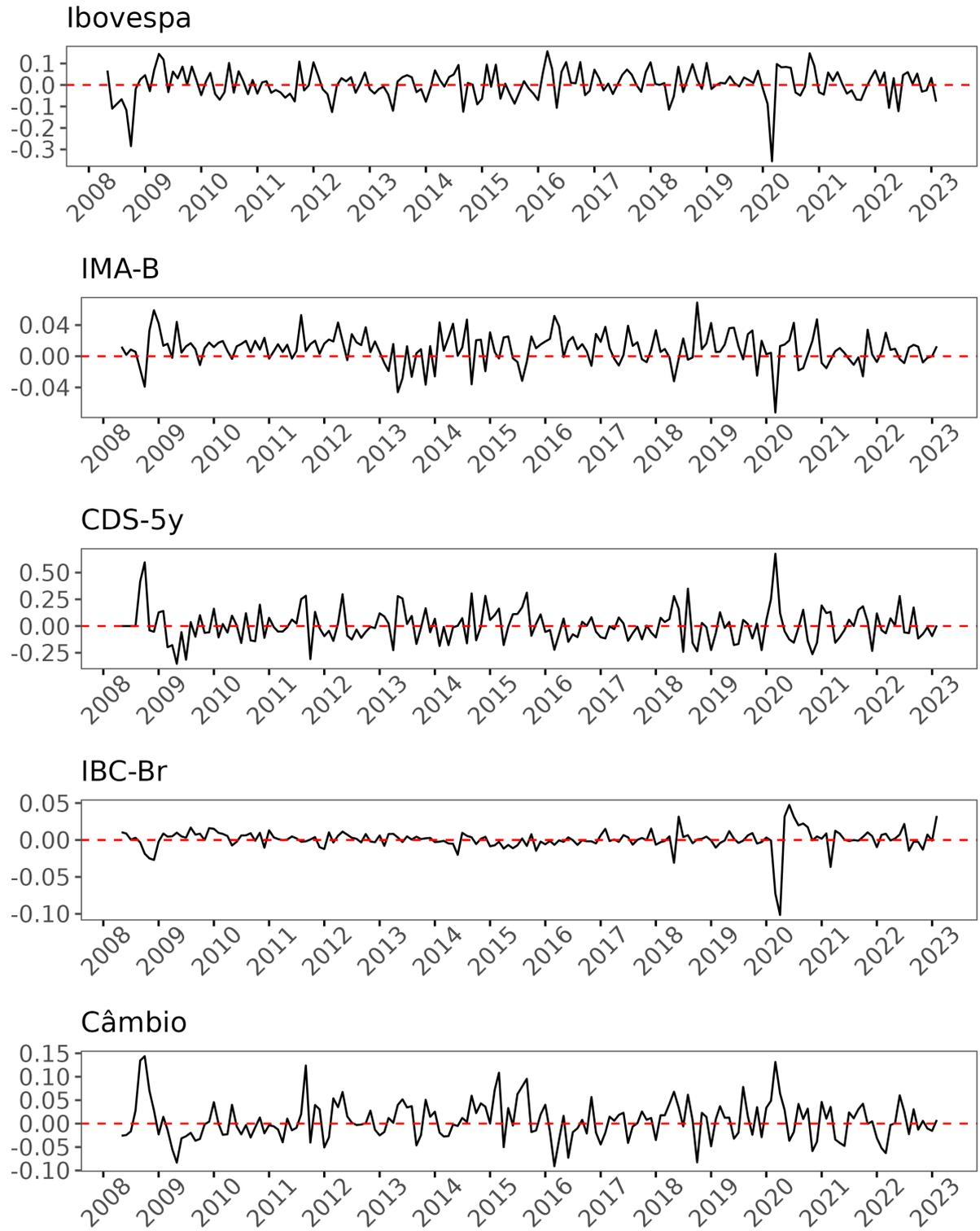
Esta seção está dividida em três partes, primeira será apresentar as estatísticas descritivas das variáveis, a segunda será mostrar os testes realizados durante a estimação das cópulas, assim como as funções cópulas que melhor se ajustaram junto a seus parâmetros. Em seguida, será feita a análise de probabilidade de ocorrência do retorno dos fundos dado que as variáveis macroeconômicas estão em um intervalo. Os intervalos foram escolhidos da seguinte forma: [Mínimo, 1Q), [1Q, Média), [Média, 3Q) e [3Q, Máximo). Os valores exatos destes intervalos estão apresentados no Anexo B.

3.1 Análise dos resultados

Antes de apresentar os resultados do modelo, é importante examinar os dados para obter uma melhor compreensão de sua distribuição. A Figura 8 mostra a evolução dos retornos das variáveis macroeconômicas e financeiras utilizadas. Vale ressaltar que os dados utilizados no presente trabalho são o log-retorno de maio de 2008 a fevereiro de 2023. Na figura, é possível notar a alta volatilidade das variáveis, em especial do CDS e do câmbio.

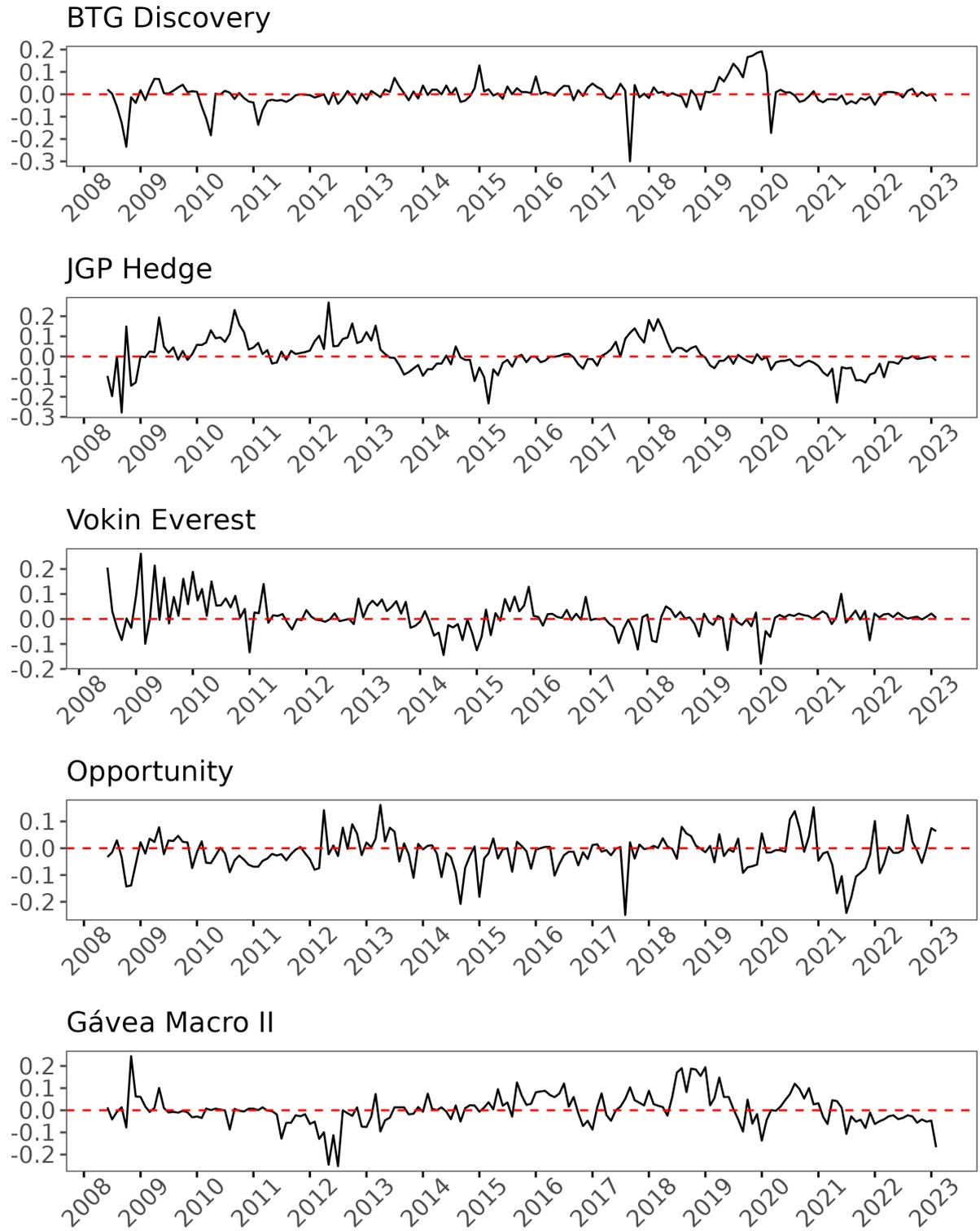
A Figura 8 ilustra a evolução de diversas variáveis econômicas ao longo do tempo, destacando a volatilidade como um indicador significativo de alterações abruptas nos preços dos ativos financeiros. Observa-se que os picos de alta volatilidade estão frequentemente associados a eventos de grande impacto econômico e político. É possível identificar quatro períodos notáveis de volatilidade: o primeiro ocorreu em 2008, desencadeado pela crise financeira global, que afetou todas as variáveis de maneira similar. Posteriormente, em meados da década de 2010, o país enfrentou uma crise política e econômica que também se refletiu nas variáveis analisadas. Em 2018, a greve dos caminhoneiros provocou um pico de volatilidade, especialmente visível no IBC-Br e no CDS. Por fim, a pandemia de COVID-19 em 2020 gerou uma onda de incertezas e impactou significativamente os mercados financeiros, como evidenciado pelos movimentos acentuados nas séries. Esses eventos geram incertezas e alteram as expectativas dos investidores, o que se traduz em movimentos bruscos nos mercados financeiros. Isto pode ser observado na Figura 9 que mostra a série de retorno dos fundos.

Figura 8 - Evolução do retorno das variáveis.



Fonte: O autor, 2024.

Figura 9 - Evolução do retorno dos Fundos.



Fonte: O autor, 2024.

A Tabela 4 apresenta as estatísticas descritivas dos dados. Observa-se que, em todos os casos, a média e a mediana são muito próximas uma da outra e também próximas de zero. É possível observar ainda que a distância entre o mínimo e o máximo dos retornos dos fundos JGP Hedge, Vokin Everest, Gávea Macro e da variável CDS são mais acentuadas, indicando que possuem caudas mais alongadas que as demais.

Tabela 4 - Estatística descritiva dos retornos.

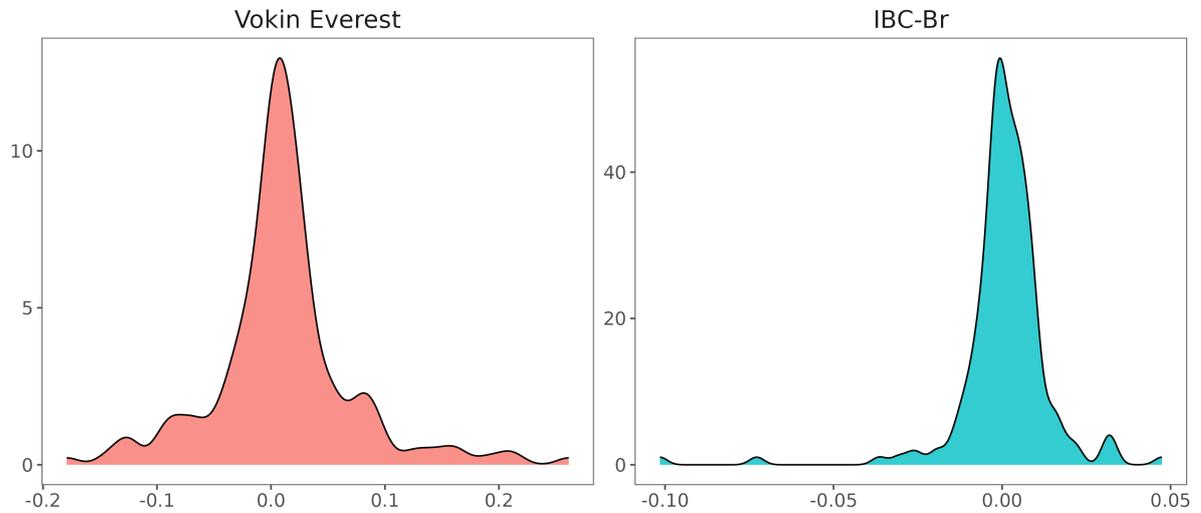
	Média	Dp	Mediana	Mínimo	Máximo	Intervalo
BTG	0,001	0,057	0,003	-0,300	0,192	0,492
JGP	0,000	0,081	-0,004	-0,279	0,268	0,547
Vokin	0,011	0,062	0,008	-0,179	0,261	0,440
Opportunity	-0,016	0,063	-0,013	-0,249	0,162	0,411
Gávea	0,004	0,070	0,001	-0,253	0,244	0,497
Ibovespa	0,002	0,069	0,006	-0,355	0,157	0,512
IMA-B	0,009	0,020	0,010	-0,072	0,069	0,141
CDS-5y	0,004	0,155	-0,008	-0,353	0,676	1,029
IBC-Br	0,001	0,014	0,001	-0,102	0,047	0,149
Câmbio	0,006	0,040	0,003	-0,091	0,144	0,235

Legenda: Dp significa desvio padrão. A coluna "intervalo" é referente a diferença do retorno mínimo e máximo.

Fonte: O autor, 2024.

A análise das séries de fundos de investimentos e das variáveis macroeconômicas e financeiras revelou a existência de padrões de densidade distintivamente marcados, os quais se caracterizam pela complexidade na sua captura através de métodos convencionais de estimação de densidade paramétrica. Esta complexidade é particularmente evidente na presença frequente de distribuições bimodais, que desafiam a adequação de abordagens simplistas. Diante deste cenário, a adoção do método de densidade de Kernel emergiu como uma solução metodologicamente robusta, permitindo uma estimação de densidade mais fidedigna e ajustada às peculiaridades das distribuições em análise. A eficácia deste método é demonstrada pela precisão com que conseguiu mapear as características intrínsecas das séries, conforme ilustrado na Figura 10. Estas representações gráficas evidenciam não apenas a especificidade das densidades observadas, mas também a superioridade da abordagem de densidade de Kernel na captura das nuances das distribuições em estudo. Assim, a escolha dessa metodologia não se deu arbitrariamente, mas como resposta à necessidade de um modelo que se ajustasse de maneira ótima aos dados observados, garantindo assim uma representação mais acurada e informativa das distribuições subjacentes.

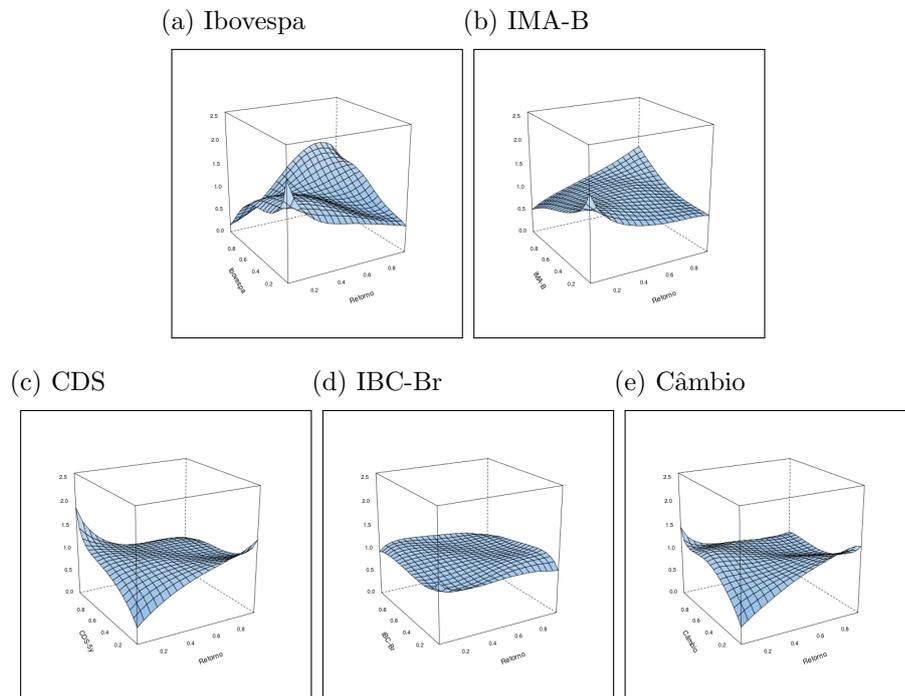
Figura 10 - Distribuição do retorno do fundo Vokin Everest e IBC-Br.



Fonte: O autor, 2024.

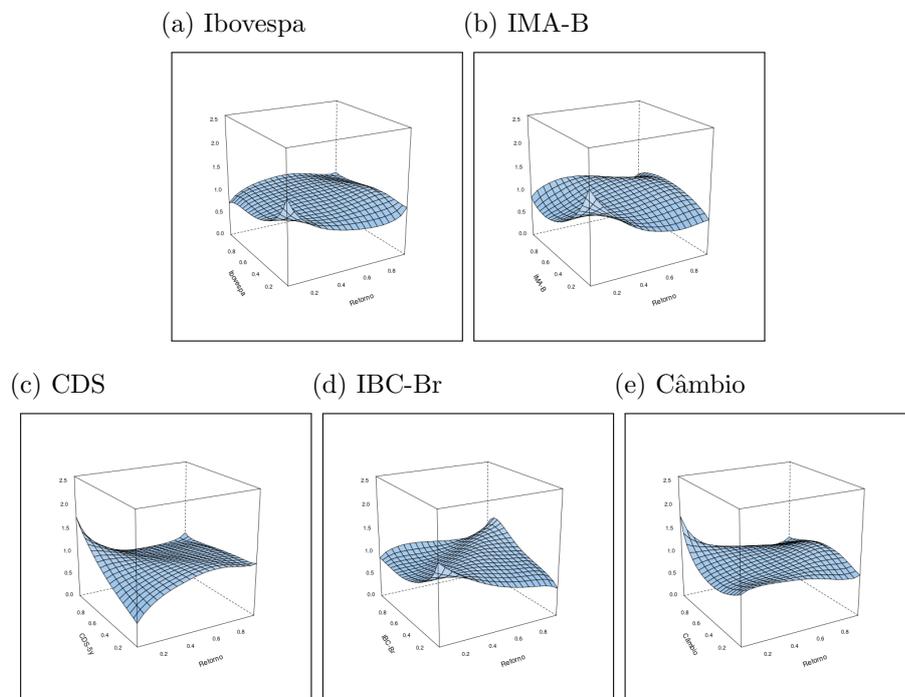
A Figura 11 expõe a representação gráfica da densidade da cópula, uma ferramenta analítica crucial que esboça a intrincada estrutura de dependência bivariada entre os fundos de investimento e as variáveis macroeconômicas e financeiras. Este método, baseado na aplicação da metodologia de Kernel juntamente com a Teoria de Copulas, permite a caracterização detalhada da relação entre cada fundo e variável. A densidade da cópula, exemplificada na Figura 11, ilustra vividamente a complexidade da interdependência, destacando nuances e padrões que emergem da conjunção dos referidos ativos financeiros. Este intrincado mapeamento de dependência é resultado direto das densidades previamente identificadas nas séries temporais observadas. Notavelmente, a densidade da estrutura de dependência entre o Ibovespa e o fundo BTG Discovery, conforme apresentado, oferece uma visão aprofundada da natureza das relações subjacentes. Cada cópula bivariada, assim gerada, não só captura a dependência entre os pares específicos de fundo e variável, mas também fornece uma representação visual tangível das características distintivas das relações financeiras em análise. Este enfoque analítico, fundamentado na cópula, emerge como uma abordagem robusta para desvelar a estrutura de dependência bivariada, enriquecendo a compreensão das complexas interações presentes nas dinâmicas financeiras.

Figura 11 - BTG Discovery - Densidade da copulas.



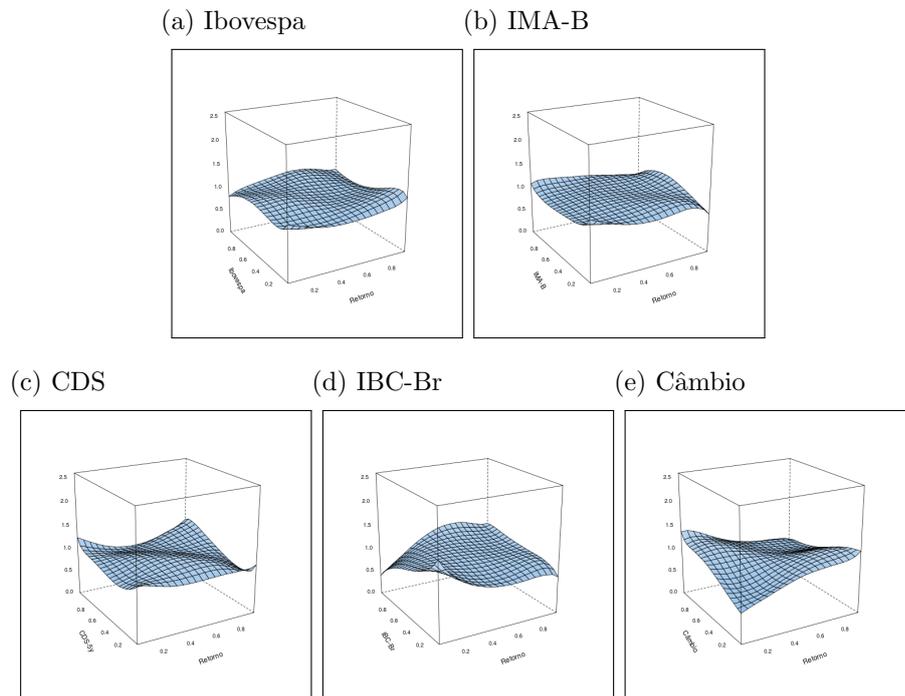
Fonte: O autor, 2024.

Figura 12 - JGP Hedge - Densidade da copulas.



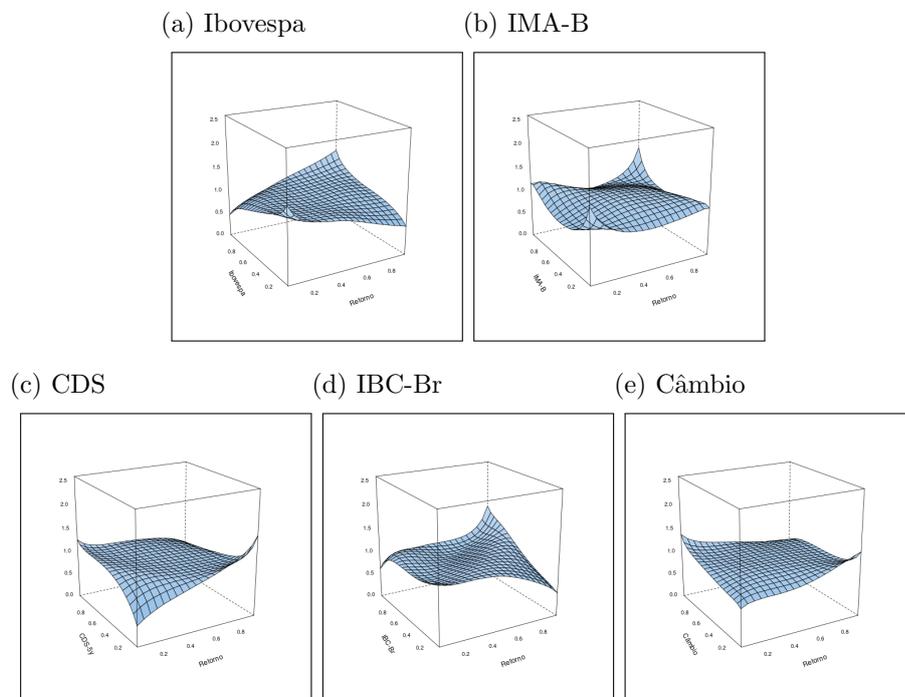
Fonte: O autor, 2024.

Figura 13 - Vokin Everest - Densidade da copulas.



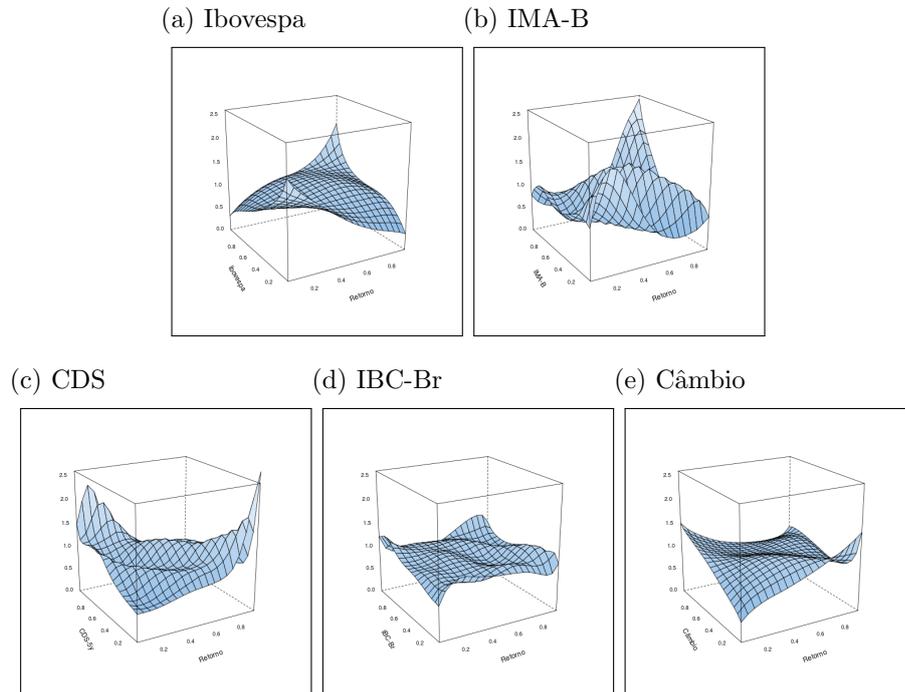
Fonte: O autor, 2024.

Figura 14 - Opportunity - Densidade da copulas.



Fonte: O autor, 2024.

Figura 15 - Gávea Macro - Densidade da copulas.



Fonte: O autor, 2024.

Os parâmetros apresentados na Tabela 5 são o parâmetro de dependência (α) e o coeficiente de correlação de Kendall (Tau). O parâmetro de dependência mede a força da dependência entre as variáveis, enquanto o coeficiente de correlação de Kendall mede a concordância entre as classificações das variáveis.

Observando a Tabela 5, podemos observar que o parâmetro de dependência é bastante semelhante entre os diferentes fundos e variáveis, variando principalmente entre 0,6 e 0,8. Isso sugere que a dependência entre o retorno dos fundos e as variáveis macroeconômicas é de moderada a alta, com exceção do fundo Gávea Macro em relação ao IMA-B que apresenta o menor valor do α (0,301).

O coeficiente de correlação de Kendall (Tau), por outro lado, mostra uma maior variação. Valores positivos de Tau indicam uma correlação positiva entre o retorno do fundo e a variável macroeconômica, enquanto valores negativos indicam uma correlação negativa. Por exemplo, para o fundo BTG Discovery, a correlação com o Ibovespa é positiva (0,184), enquanto a correlação com o CDS é negativa (-0,119). Isso sugere que um aumento no Ibovespa está associado a um aumento no retorno do fundo, enquanto um aumento no CDS está associado a uma diminuição no retorno do fundo. Indo de acordo com o que é observado nas figuras 11a e 11c.

Tabela 5 - Parâmetros das copulas.

	Ibovespa	IMA-B	CDS-5y	IBC-Br	Câmbio
BTG Discovery					
α	0,621	0,787	0,787	0,787	0,787
Tau	0,184	0,111	-0,119	-0,019	-0,097
JGP Hedge					
α	0,787	0,715	0,787	0,787	0,787
Tau	0,034	0,066	-0,087	0,086	-0,040
Vokin Everest					
α	0,789	0,789	0,789	0,789	0,789
Tau	-0,005	0,003	0,005	0,083	-0,104
Opportunity					
α	0,787	0,715	0,787	0,787	0,787
Tau	0,104	0,015	-0,085	0,081	-0,063
Gávea Macro II					
α	0,726	0,301	0,342	0,415	0,726
Tau	0,171	0,128	-0,159	-0,028	-0,087

Legenda: A primeira coluna indica a variável utilizada para criar a cópula bivariada juntamente com o retorno do respectivo Fundo.

Fonte: O autor, 2024.

Em resumo, a análise da tabela sugere que existe uma dependência de moderada para alta entre o retorno dos fundos de investimento e as variáveis macroeconômicas, com a direção e a força da correlação variando dependendo do fundo e da variável específica.

A partir da conclusão da etapa de estimação das cópulas, procedeu-se à simulação de um total de 20.000 observações, empregando as cópulas previamente estimadas como base. Este processo foi meticulosamente desenhado para preservar a estrutura de dependência entre as variáveis identificada na fase de estimação, garantindo que as simulações refletissem de maneira fiel as relações estabelecidas no conjunto de dados original. Posteriormente, as observações simuladas foram submetidas a um procedimento de transformação inversa, com o objetivo de reverter os dados ao seu espaço original, conforme definido pelas distribuições marginais das variáveis em análise.

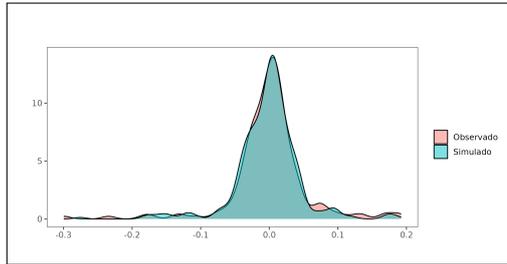
A relevância deste procedimento de simulação reside na sua capacidade de gerar um conjunto de dados que, embora sintético, captura de forma estatisticamente significativa e assertiva as características e dependências observadas nos dados reais. Tal abordagem é fundamental para a validação de modelos estatísticos e para a realização de inferências robustas sobre o comportamento das variáveis sob estudo.

Para avaliar a adequação e a precisão das simulações realizadas, foi implementado um teste de ajuste baseado em critérios gráficos. Este teste envolve a comparação visual entre as distribuições de densidade dos dados simulados e dos dados observados originalmente. A aplicação deste método gráfico permite uma análise intuitiva e imediata da congruência entre os conjuntos de dados, facilitando a identificação de possíveis discrepâncias ou desvios significativos.

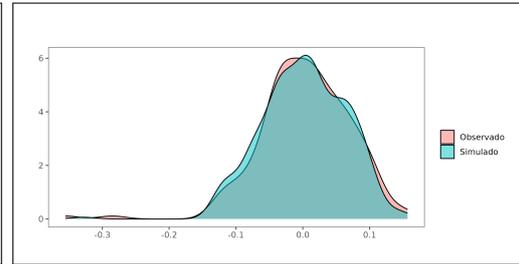
A análise detalhada dos gráficos (abaixo) gerados revelou que os dados simulados apresentam um elevado grau de aderência em relação aos dados observados, indicando que o processo de simulação conseguiu replicar de forma eficaz as características e dependências inerentes ao conjunto de dados original. Tal constatação reforça a validade das técnicas de estimação e simulação empregadas, corroborando a sua aplicabilidade como ferramentas valiosas na modelagem estatística e na análise de dados complexos.

Figura 16 - BTG Discovery - Densidade das variáveis observadas e simuladas.

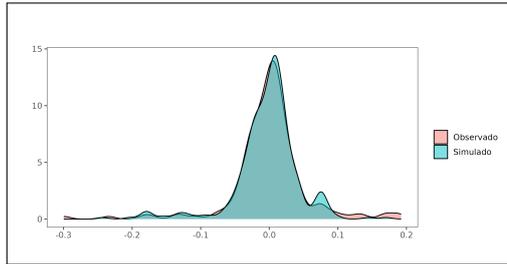
(a) Retorno - Ibovespa



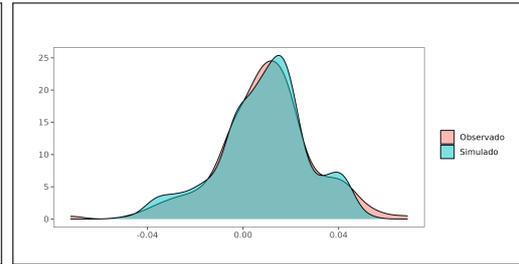
(b) Ibovespa



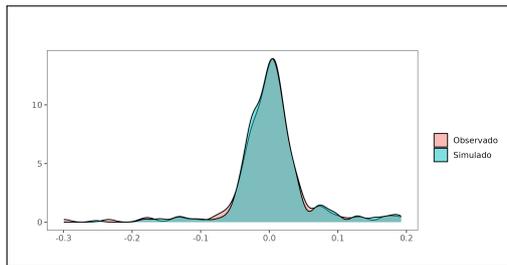
(c) Retorno - IMA-B



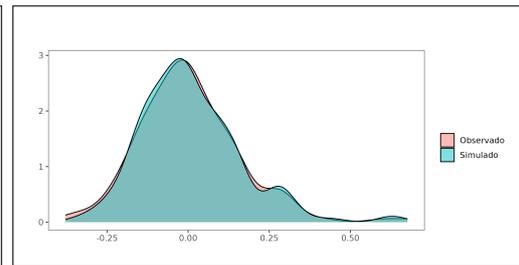
(d) IMA-B



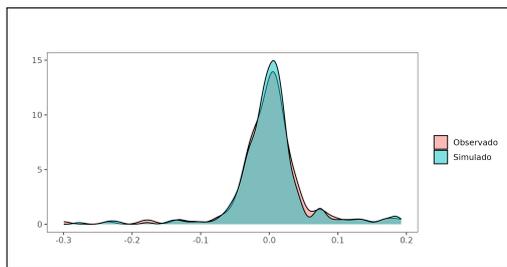
(e) Retorno - CDS



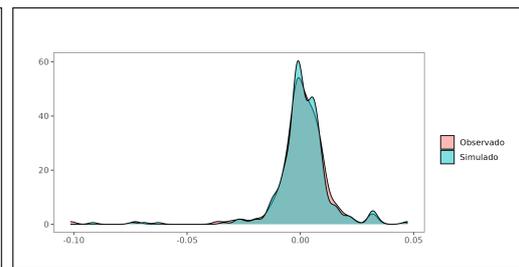
(f) CDS



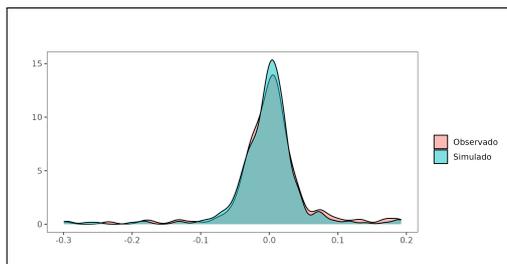
(g) Retorno - IBC - Br



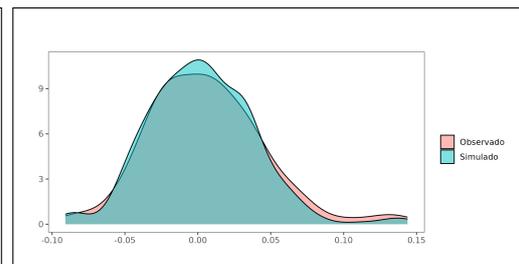
(h) IBC-Br



(i) Retorno - Câmbio



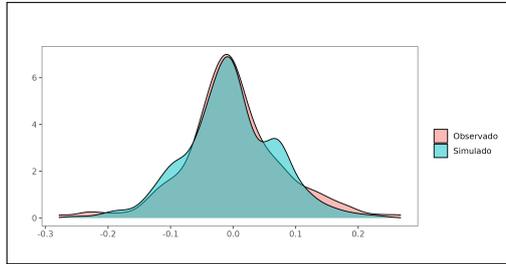
(j) Câmbio



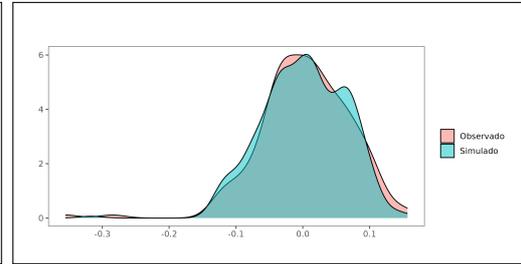
Fonte: O autor, 2024.

Figura 17 - JGP Hedge - Densidade das variáveis observadas e simuladas.

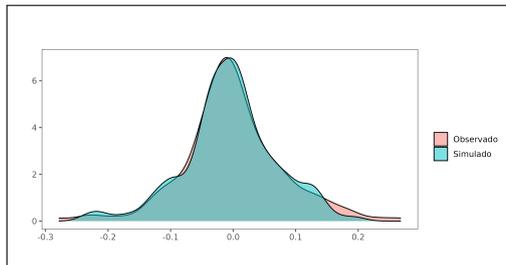
(a) Retorno - Ibovespa



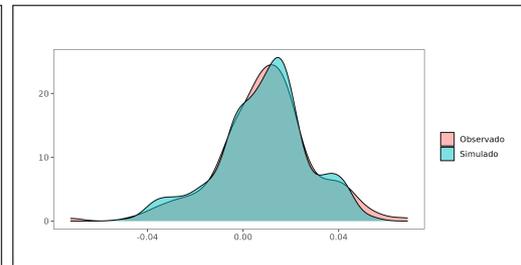
(b) Ibovespa



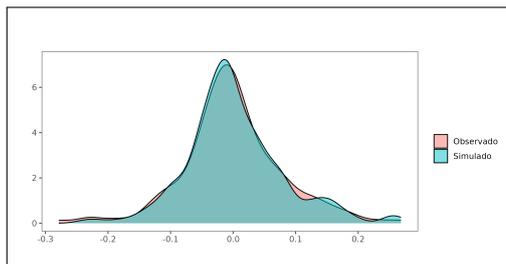
(c) Retorno - IMA-B



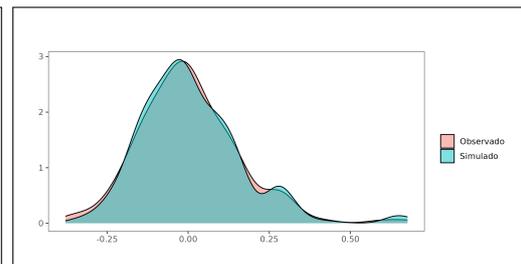
(d) IMA-B



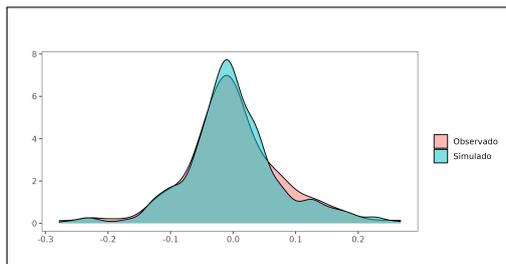
(e) Retorno - CDS



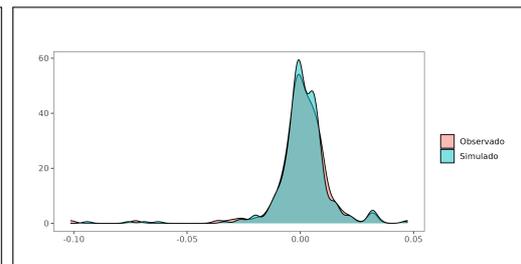
(f) CDS



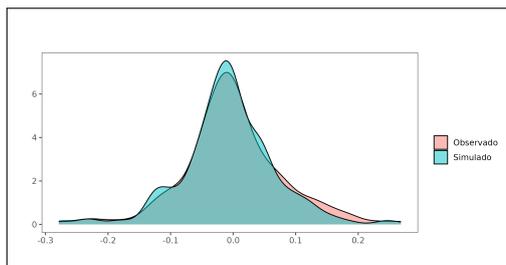
(g) Retorno - IBC - Br



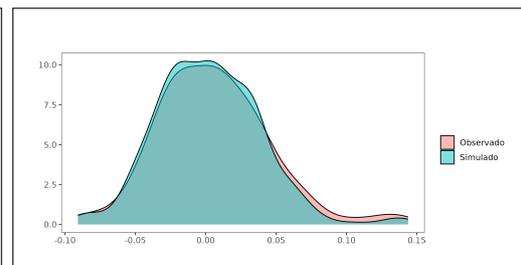
(h) IBC-Br



(i) Retorno - Câmbio



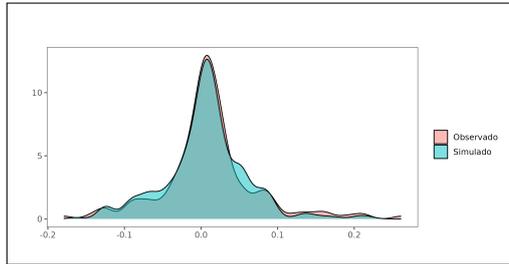
(j) Câmbio



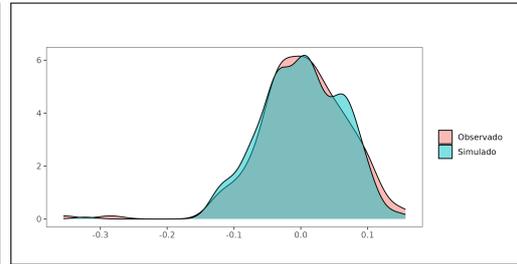
Fonte: O autor, 2024.

Figura 18 - Vokin Everest - Densidade das variáveis observadas e simuladas.

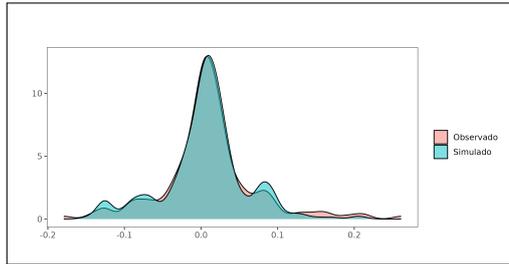
(a) Retorno - Ibovespa



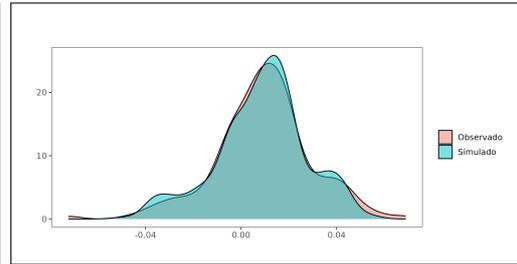
(b) Ibovespa



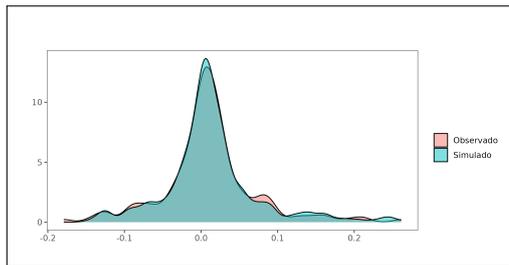
(c) Retorno - IMA-B



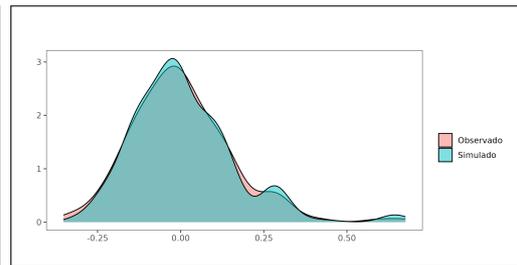
(d) IMA-B



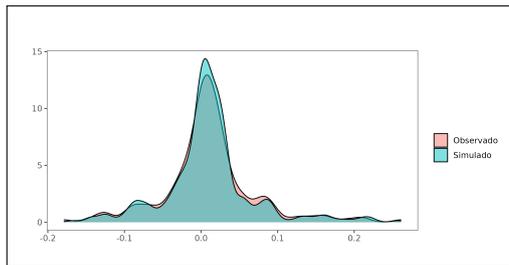
(e) Retorno - CDS



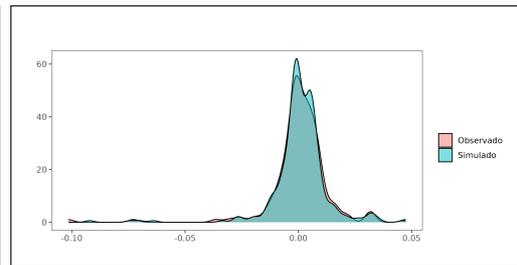
(f) CDS



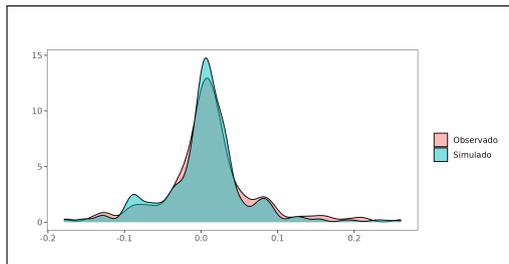
(g) Retorno - IBC - Br



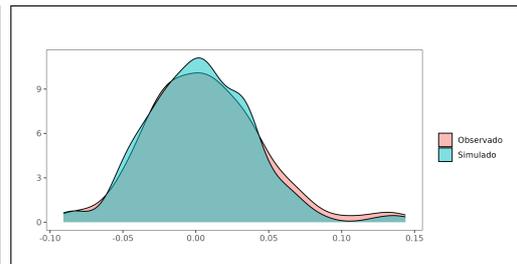
(h) IBC-Br



(i) Retorno - Câmbio



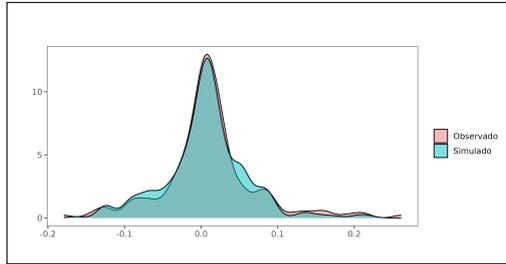
(j) Câmbio



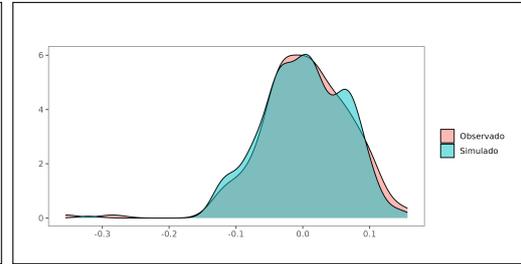
Fonte: O autor, 2024.

Figura 19 - Opportunity - Densidade das variáveis observadas e simuladas.

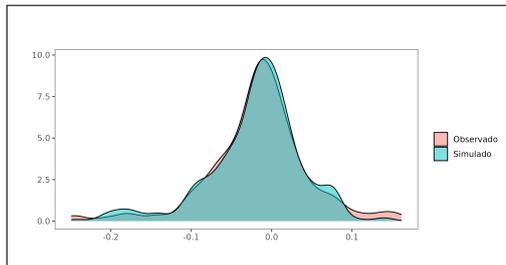
(a) Retorno - Ibovespa



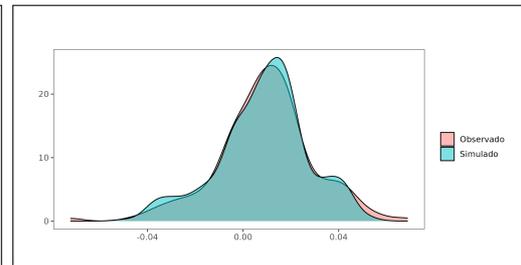
(b) Ibovespa



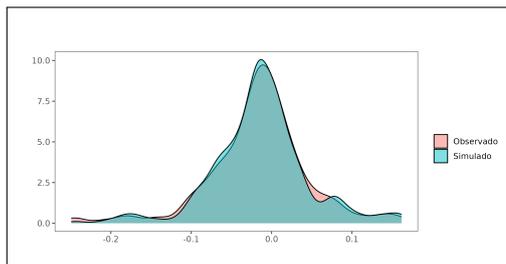
(c) Retorno - IMA-B



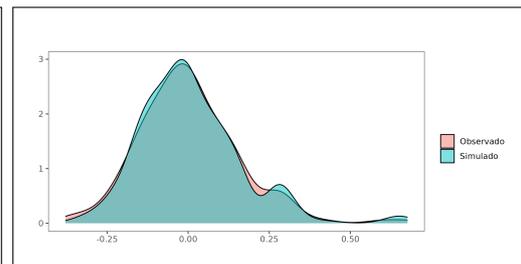
(d) IMA-B



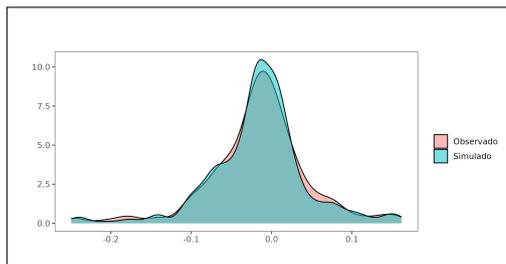
(e) Retorno - CDS



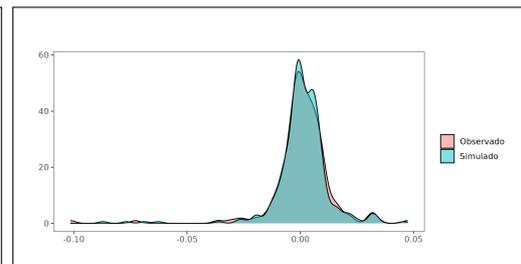
(f) CDS



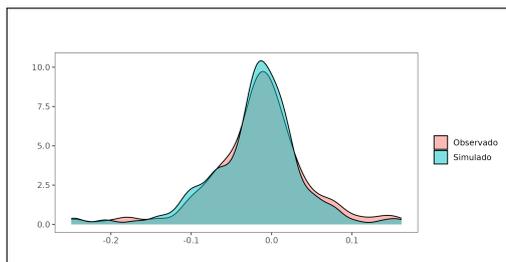
(g) Retorno - IBC - Br



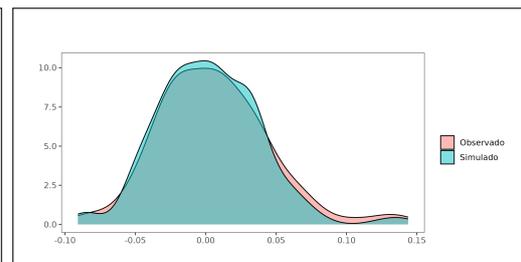
(h) IBC-Br



(i) Retorno - Câmbio

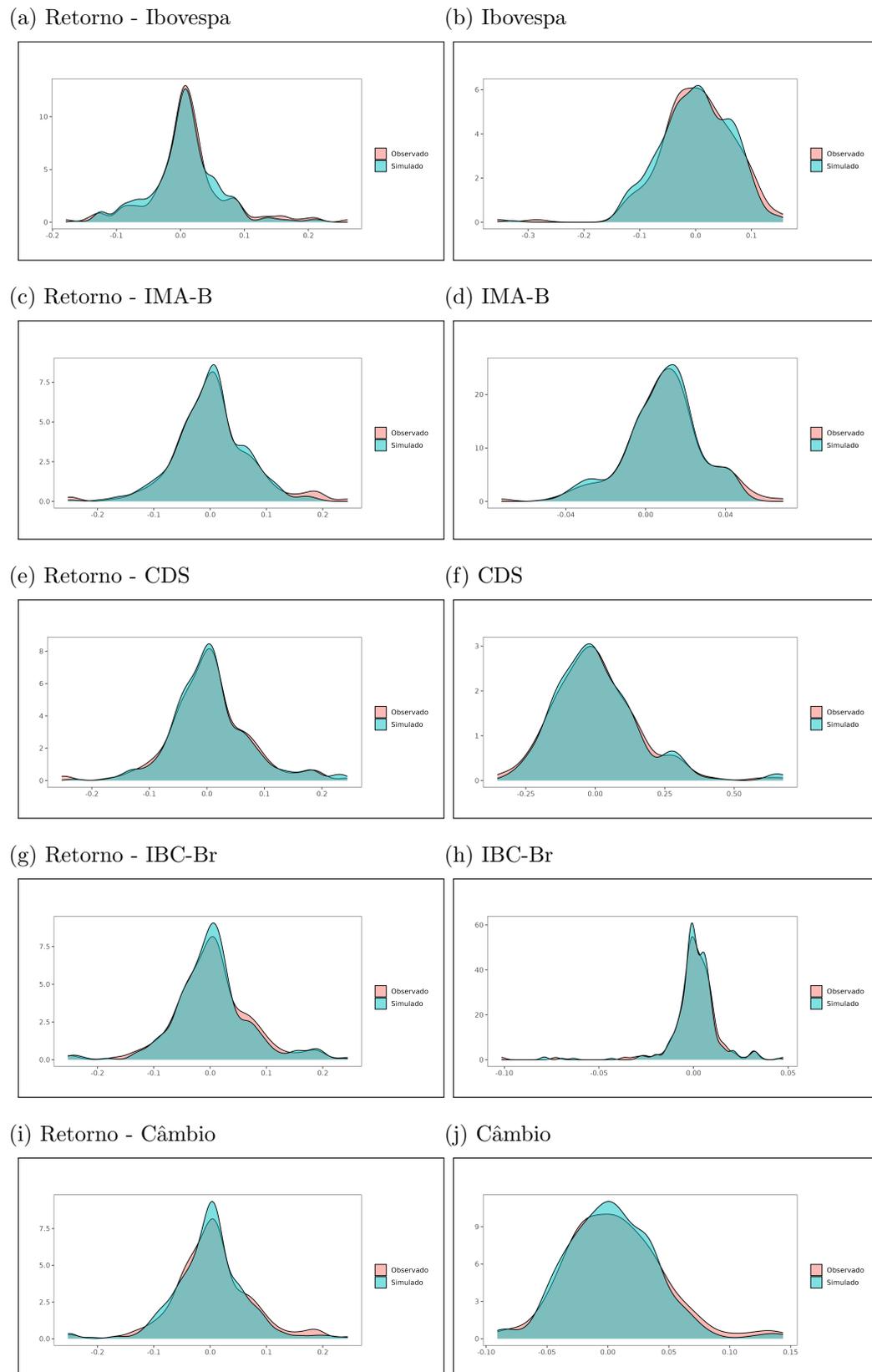


(j) Câmbio



Fonte: O autor, 2024.

Figura 20 - Gávea Macro II - Densidade das variáveis observadas e simuladas.



Fonte: O autor, 2024.

3.1.1 BTG Discovery

A Tabela 6, mostra que a probabilidade de um retorno positivo⁸ do fundo de investimento, condicionada à variável estar no intervalo mais baixo (entre o mínimo e o primeiro quartil), é nula para todas as variáveis. Mostrando que quando as variáveis econômicas estão em seus níveis mais baixos, o fundo de investimento não tende a apresentar retornos positivos.

À medida que as variáveis macroeconômicas (Ibovespa, IMA-B CDS Câmbio e IBC-Br) se movem para o intervalo entre o primeiro quartil e a média, a probabilidade de retorno positivo do fundo aumenta, embora de forma modesta para a maioria das variáveis, com exceção do CDS, que apresenta um aumento mais significativo para 9,115%. Isso pode refletir uma sensibilidade maior do fundo a mudanças no risco de crédito do país.

Para o Ibovespa, observamos que a probabilidade de um retorno positivo do fundo aumenta à medida que o retorno do Ibovespa aumenta, passando de 0% no primeiro intervalo para 26,830% no terceiro intervalo. Isso sugere uma relação positiva entre o desempenho do Ibovespa e a probabilidade de um retorno positivo do fundo. No entanto, quando o retorno do fundo é negativo, a probabilidade é alta (25%) quando o Ibovespa está no primeiro intervalo e diminui para 0% no terceiro e quarto intervalo.

Um padrão semelhante é observado para o IMA-B, CDS-5y, IBC-Br e Câmbio. Em todos esses casos, a probabilidade de um retorno positivo do fundo aumenta à medida que o log-retorno da variável aumenta. Isso sugere que o fundo tende a ter um desempenho melhor quando essas variáveis econômicas também estão se saindo bem.

Além disso, o retorno esperado do fundo também varia dependendo do intervalo de retorno da variável. Por exemplo, para o Ibovespa, o retorno esperado do fundo é de -5,494% quando o retorno do Ibovespa está no primeiro intervalo, mas aumenta para 5,785% quando o retorno do Ibovespa está no quarto intervalo. Isso sugere que o fundo tende a ter um retorno esperado mais alto quando o Ibovespa está se saindo bem.

Quanto ao retorno esperado do fundo, os resultados mostram que, uma progressão do negativo para o positivo à medida que as variáveis econômicas se movem de seus níveis mais baixos para os mais altos. Quando o retorno das variáveis está entre o mínimo e o primeiro quartil, o retorno esperado é negativo, variando de -5,510% para IMA-B a -5,695% para CDS. No entanto, quando o retorno das variáveis está entre o primeiro

⁸ Vale ressaltar que probabilidade de ocorrência refere-se à chance de um determinado evento acontecer. Por outro lado, o retorno esperado é uma medida do ganho ou perda provável. Mesmo que a probabilidade de ocorrência seja 0%, isso não significa necessariamente que o retorno esperado também seja 0%. Portanto, mesmo que a probabilidade de ocorrência seja 0% para alguns intervalos, o retorno esperado não necessariamente será 0%, pois ele é influenciado por todos os possíveis resultados, não apenas por aqueles com probabilidade de ocorrência de 0%.

quartil e a média, o retorno esperado ainda é negativo, mas menos severo. Quando o retorno das variáveis está entre a média e o terceiro quartil, o retorno esperado torna-se positivo, e aumenta significativamente quando o retorno das variáveis está entre o terceiro quartil e o máximo. Isso indica que o fundo tende a ter um desempenho melhor em períodos de alto retorno das variáveis.

Tabela 6 - Probabilidades de ocorrência e retorno esperado do fundo BTG Discovery.

Intervalo	Ibovespa	IMA-B	CDS-5y	Câmbio	IBC-Br
P(Fundo > 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	0 %	0 %	0 %	0 %	0 %
[1Q, Média)	1,865 %	3,310 %	9,115 %	5,780 %	4,005 %
[Média, 3Q)	26,830 %	25,705 %	19,400 %	22,620 %	24,370 %
[3Q, Máximo)	24,995 %	24,985 %	24,995 %	25,000 %	24,990 %
P(Fundo < 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	25,000 %	24,995 %	25,000 %	25,000 %	25,000 %
[1Q, Média)	21,305 %	20,990 %	21,485 %	21,595 %	21,625 %
[Média, 3Q)	0 %	0 %	0 %	0 %	0 %
[3Q, Máximo)	0 %	0 %	0 %	0 %	0 %
E[Fundo Variável ∈ [x, y)]					
[Mínimo, 1Q)	-5,494 %	-5,510 %	-5,695 %	-5,596 %	-5,469 %
[1Q, Média)	-0,804 %	-0,721 %	-0,557 %	-0,670 %	-0,725 %
[Média, 3Q)	0,968 %	0,993 %	1,161 %	1,056 %	1,007 %
[3Q, Máximo)	5,784 %	5,924 %	5,773 %	5,710 %	5,926 %

Fonte: O autor, 2024.

3.1.2 JGP Hedge

Observando as probabilidades de ocorrência ilustrado na A Tabela 7, é evidente que a probabilidade do fundo ter um retorno positivo aumenta à medida que as variáveis econômicas se movem do mínimo para o máximo. Por exemplo, a probabilidade de um retorno positivo do fundo quando o Ibovespa está entre o mínimo e o primeiro quartil é de 0%, mas aumenta para quase 25% quando o Ibovespa está entre o terceiro quartil e o máximo. Por outro lado, a probabilidade do fundo ter um retorno negativo diminui à medida que as variáveis econômicas se movem do mínimo para o máximo. Este padrão é mantido para as demais variáveis.

Em relação ao retorno esperado do fundo, os dados mostram que o retorno esperado aumenta à medida que as variáveis econômicas se movem do mínimo para o máximo. Por

exemplo, o retorno esperado do fundo quando o Ibovespa está entre o mínimo e o primeiro quartil é de -9,163%, mas aumenta para 9,972% quando o Ibovespa está entre o terceiro quartil e o máximo. Mais uma vez, um padrão semelhante é observado para as outras variáveis.

Ao observar a distribuição da probabilidade de ocorrência, nota-se que o fundo tende a performar pior quando as variáveis econômicas estão abaixo de sua média histórica. Isso mostra que o fundo tende a ser mais sensível a variações negativas. A Tabela 7 mostra ainda, que o fundo é mais sensível aos extremos. Isso é evidenciado pelo fato de que as probabilidades de ocorrência e os retornos esperados mudam mais drasticamente quando as variáveis econômicas estão em seus extremos (ou seja, entre o mínimo e o primeiro quartil ou entre o terceiro quartil e o máximo).

Tabela 7 - Probabilidades de ocorrência e retorno esperado do fundo JGP Hedge.

Intervalo	Ibovespa	IMA-B	CDS-5y	Câmbio	IBC-Br
P(Fundo > 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	0 %	0 %	0 %	0 %	0 %
[1Q, Média)	0 %	0 %	0 %	0 %	0 %
[Média, 3Q)	18,915 %	19,205 %	18,620 %	18,535 %	18,600 %
[3Q, Máximo)	25,000 %	24,990 %	25,000 %	24,995 %	24,990 %
P(Fundo < 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	25,000 %	25,000 %	25,000 %	25,000 %	25,000 %
[1Q, Média)	23,280 %	24,345 %	30,610 %	27,285 %	25,815 %
[Média, 3Q)	7,800 %	6,450 %	0,765 %	4,180 %	5,585 %
[3Q, Máximo)	0 %	0 %	0 %	0 %	0 %
E[Fundo Variável ∈ [x, y)]					
[Mínimo, 1Q)	-9,163 %	-9,127 %	-9,367 %	-9,281 %	-9,113 %
[1Q, Média)	-2,272 %	-2,135 %	-1,828 %	-2,046 %	-2,123 %
[Média, 3Q)	1,209 %	1,235 %	1,639 %	1,368 %	1,279 %
[3Q, Máximo)	9,972 %	10,128 %	9,878 %	9,846 %	10,090 %

Fonte: O autor, 2024.

3.1.3 Vokin Everest

Observando a Tabela 8, nota-se que a probabilidade de o fundo ter um retorno positivo aumenta significativamente à medida que nos movemos dos quartis inferiores para os superiores. Por exemplo, para todas as variáveis, a probabilidade de um retorno positivo do fundo é de 0% quando as variáveis estão entre seus valores mínimos e o primeiro

quartil, mas aumenta para aproximadamente 10% e 17% quando as variáveis estão entre o primeiro quartil e a média, e chegando a aproximadamente 26% entre a média e o terceiro quartil. Este padrão sugere que retornos mais altos das variáveis estão associados a uma maior probabilidade de retornos positivos do fundo. Indicando que o fundo apresenta uma leve dependência negativa quando a variável esta performando abaixo da média, mas não próxima ao mínimo, ou seja, é muito improvável o fundo ter ganhos com os ativos estão próximo do seu mínimo, mas provável que o fundo tenha ganha com as variáveis abaixo de sua média

A probabilidade de o fundo ter um retorno negativo é inversamente proporcional. Há uma alta probabilidade (25%) de retorno negativo do fundo quando as variáveis estão entre seus valores mínimos e o primeiro quartil, mas essa probabilidade cai drasticamente tornando-se nula para os intervalos subsequentes. Isso reforça a noção de que retornos mais altos das variáveis estão associados a menores chances de retornos negativos do fundo.

A Tabela 8 ainda mostra o retorno esperado do fundo, condicional aos intervalos de retorno das variáveis. Os resultados são consistentes com a probabilidade de ocorrência, onde o retorno esperado é negativo quando as variáveis estão entre seus valores mínimos e o primeiro quartil, mas torna-se positivo e cresce à medida que as variáveis se movem para intervalos superiores. Notavelmente, o retorno esperado é substancialmente mais alto no intervalo entre o terceiro quartil e o valor máximo das variáveis, indicando que retornos excepcionalmente altos das variáveis estão associados a retornos esperados significativamente mais altos do fundo.

Ao olhar para a distribuição da probabilidade de ocorrência e o retorno esperado, nota-se que o fundo, se beneficia mais dos períodos econômicos favoráveis e dos ativos de referência do que sofrendo com os períodos em que a economia está em baixa. Isso é evidenciado pelo fato de que a probabilidade de um retorno positivo do fundo é maior quando o retorno das variáveis está nos intervalos superiores, enquanto a probabilidade de um retorno negativo do fundo é maior quando o retorno das variáveis está nos intervalos inferiores.

Tabela 8 - Probabilidades de ocorrência e retorno esperado do fundo Vokin Everest.

Intervalo	Ibovespa	IMA-B	CDS-5y	Câmbio	IBC-Br
P(Fundo > 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	0 %	0 %	0 %	0 %	0 %
[1Q, Média)	10,175 %	11,820 %	16,825 %	13,995 %	12,500 %
[Média, 3Q)	26,690 %	25,490 %	19,640 %	22,670 %	24,140 %
[3Q, Máximo)	25,000 %	24,990 %	24,995 %	24,995 %	24,995 %
P(Fundo < 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	25,000 %	25,000 %	25,000 %	24,990 %	25,000 %
[1Q, Média)	13,130 %	12,690 %	13,535 %	13,345 %	13,360 %
[Média, 3Q)	0 %	0 %	0 %	0 %	0 %
[3Q, Máximo)	0 %	0 %	0 %	0 %	0 %
E[Fundo Variável ∈ [x, y)]					
[Mínimo, 1Q)	-5,661 %	-5,655 %	-5,807 %	-5,743 %	-5,603 %
[1Q, Média)	-0,090 %	-0,017 %	0,093 %	0,011 %	-0,020 %
[Média, 3Q)	1,701 %	1,734 %	1,938 %	1,804 %	1,755 %
[3Q, Máximo)	8,252 %	8,432 %	8,192 %	8,140 %	8,369 %

Fonte: O autor, 2024.

3.1.4 Opportunity

Observa-se na Tabela 9 que, quando as variáveis estão entre seus valores mínimos e primeiro quartil, a probabilidade do fundo ter retorno positivo é nula. Por outro lado, a probabilidade do retorno ser negativo é de 25%. O retorno esperado do fundo nesse cenário é bastante negativo, variando entre -9,021% e -9,240%, dependendo da variável.

À medida que as variáveis assumem valores mais altos, aumenta a probabilidade de retorno positivo do fundo e diminui a de retorno negativo. Quando as variáveis estão no intervalo entre a média e o terceiro quartil, a probabilidade de retorno positivo do fundo varia entre 11% e 12%. Já no intervalo entre o terceiro quartil e o valor máximo, tal probabilidade fica aproximadamente 25% para todas as variáveis

Analogamente, o retorno esperado do fundo torna-se menos negativo conforme as variáveis assumem valores mais elevados, em torno de 0% no intervalo entre a média e terceiro quartil. No último intervalo, o retorno esperado é positivo, por volta de 5%.

A Tabela 9 sugere uma relação positiva entre o desempenho das variáveis macroeconômicas e financeiras e o retorno do fundo de investimento analisado. A probabilidade de ocorrência de retornos negativos do fundo Opportunity é maior do que a de retornos positivos em determinados intervalos destaca uma maior sensibilidade do fundo a condições de mercado desfavoráveis. Isso sugere que o fundo possui um perfil de risco

que tende a ser mais vulnerável a perdas em períodos de baixa performance das variáveis econômicas e financeiras consideradas. Além disso, o impacto negativo no retorno do fundo é geralmente mais acentuado do que os ganhos observados em cenários favoráveis, especialmente quando as variáveis estudadas estão abaixo de suas médias históricas.

Tabela 9 - Probabilidades de ocorrência e retorno esperado do fundo Opportunity.

Intervalo	Ibovespa	IMA-B	CDS-5y	Câmbio	IBC-Br
P(Fundo > 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	0 %	0 %	0 %	0 %	0 %
[1Q, Média)	0 %	0 %	0 %	0 %	0 %
[Média, 3Q)	11,885 %	11,810 %	11,600 %	11,470 %	11,330 %
[3Q, Máximo)	24,995 %	24,995 %	24,995 %	24,995 %	24,995 %
P(Fundo < 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	25,000 %	24,990 %	25,000 %	25,000 %	25,000 %
[1Q, Média)	23,295 %	24,090 %	30,520 %	27,280 %	25,690 %
[Média, 3Q)	14,820 %	14,110 %	7,880 %	11,250 %	12,980 %
[3Q, Máximo)	0 %	0 %	0 %	0 %	0 %
E[Fundo Variável ∈ [x, y)]					
[Mínimo, 1Q)	-9,056 %	-9,047 %	-9,240 %	-9,158 %	-9,021 %
[1Q, Média)	-2,605 %	-2,537 %	-2,325 %	-2,473 %	-2,525 %
[Média, 3Q)	-0,148 %	-0,107 %	0,192 %	0,015 %	-0,071 %
[3Q, Máximo)	5,406 %	5,546 %	5,357 %	5,314 %	5,498 %

Fonte: O autor, 2024.

3.1.5 Gávea Macro

A análise das informações contidas na Tabela 10 permite compreender a relação entre o desempenho do fundo Gávea Macro e as flutuações das variáveis macroeconômicas selecionadas. Primeiramente, é importante notar na Tabela 10 que a probabilidade do fundo ter um retorno positivo é nula quando as variáveis estão no intervalo mínimo até o primeiro quartil. No entanto, essa probabilidade aumenta significativamente à medida que as variáveis se movem para intervalos mais altos. No intervalo entre o primeiro quartil e a média a maior probabilidade de ocorrência é com a variável CDS, com 6,860%. Isso mostra que o fundo tem uma maior sensibilidade a variações no risco do país.

Em contraste, a probabilidade do fundo ter um retorno negativo é alta quando as variáveis estão no intervalo mínimo até o primeiro quartil, mas diminui à medida que as variáveis se movem para intervalos mais altos. De fato, a probabilidade do fundo ter um

retorno negativo é nula quando as variáveis estão no intervalo entre a média e o terceiro quartil, bem como no intervalo entre o terceiro quartil e o máximo, exceto para o Ibovespa e IMA-B.

Como pode ser visto na Tabela 10, o retorno esperado do fundo Gávea Macro é negativo quando as variáveis macroeconômicas selecionadas estão nos intervalos mais baixos, mas torna-se positivo à medida que as variáveis se movem para intervalos mais altos. Quando as variáveis macroeconômicas apresentam variações entre o mínimo e o primeiro quartil, em todos os casos, a probabilidade do fundo superar 0% é de 0%, indicando que, quando as variáveis estão em seus valores mais baixos, o fundo Gávea Macro tende a não ter desempenho positivo.

Quando as variáveis macroeconômicas apresentam variações entre a média e o terceiro quartil, há um aumento significativo na probabilidade, variando de 19,115% a 26,700%, indicando que, à medida que as variáveis alcançam valores intermediários, a chance do fundo ter um retorno positivo aumenta consideravelmente. Quando as variáveis macroeconômicas apresentam variações entre o terceiro quartil e o máximo, as probabilidades são muito similares, em torno de 24,995% a 25,000%, mostrando que, nos valores mais altos das variáveis macroeconômicas selecionadas, o fundo tende a manter uma alta chance de desempenho positivo.

Ou seja, o retorno esperado é particularmente alto quando as variáveis estão no intervalo entre o terceiro quartil e o máximo. Isso sugere que o retorno do fundo é mais sensível aos extremos, particularmente aos extremos superiores.

A Tabela 10 sugere que há uma dependência positiva entre o retorno do fundo Gávea Macro e as variáveis econômicas consideradas. Em outras palavras, o fundo tende a ter um desempenho melhor quando as condições econômicas, são favoráveis. Além disso, o benefício de um cenário econômico favorável parece ser maior do que o desvantagem de um cenário econômico desfavorável, uma vez que a probabilidade de ocorrência do fundo performar positivamente é levemente superior.

Tabela 10 - Probabilidades de ocorrência e retorno esperado do fundo Gávea Macro.

Intervalo	Ibovespa	IMA-B	CDS-5y	Câmbio	IBC-Br
P(Fundo > 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	0 %	0 %	0 %	0 %	0 %
[1Q, Média)	0,195 %	2,080 %	7,780 %	3,995 %	2,270 %
[Média, 3Q)	26,700 %	25,650 %	19,115 %	22,690 %	24,635 %
[3Q, Máximo)	24,995 %	24,990 %	24,995 %	25,000 %	24,995 %
P(Fundo < 0 Variável ∈ [x, y))					
[Mínimo, 1Q)	25,000 %	25,000 %	24,995 %	25,000 %	25,000 %
[1Q, Média)	23,105 %	22,270 %	23,110 %	23,310 %	23,095 %
[Média, 3Q)	0 %	0 %	0 %	0 %	0 %
[3Q, Máximo)	0 %	0 %	0 %	0 %	0 %
E[Fundo Variável ∈ [x, y)]					
[Mínimo, 1Q)	-7,359 %	-7,392 %	-7,564 %	-7,467 %	-7,374 %
[1Q, Média)	-1,476 %	-1,362 %	-1,058 %	-1,286 %	-1,401 %
[Média, 3Q)	1,445 %	1,500 %	1,764 %	1,585 %	1,490 %
[3Q, Máximo)	9,039 %	9,211 %	8,990 %	8,938 %	9,157 %

Fonte: O autor, 2024.

CONCLUSÃO

Compreender e monitorar as variáveis macroeconômicas e seus efeitos nos fundos de investimentos pode ser de extrema valia para auxiliar na tomada de decisões sobre a alocação de recursos, a fim de garantir uma gestão eficaz desses veículos de investimento. Os investidores, mediante conhecimento prévio de variáveis macroeconômicas podem anteciper tendências, encontrar oportunidades de investimento, reduzir riscos e maximizar os retornos usando uma análise criteriosa.

Este trabalho realizou uma revisão sistemática da literatura sobre a aplicação de cópulas em finanças, com foco na sua utilização em fundos de investimento e variáveis macroeconômicas. Através da análise bibliométrica, foi possível identificar os principais autores, trabalhos influentes e temas de pesquisa relacionados ao uso de cópulas nesse contexto.

Com base na revisão sistemática realizada, é possível concluir que as Copulas têm desempenhado um papel significativo na análise da dependência entre variáveis financeiras, sendo amplamente empregadas para modelar a dependência entre retornos de ativos, taxas de juros, taxas de câmbio e outras variáveis econômicas e financeiras relevantes.

A análise bibliométrica realizada revelou que a produção acadêmica sobre o uso de Copulas em finanças tem aumentado, especialmente em períodos que sucedem crises financeiras. Além disso, a análise temática mostrou que os tópicos de pesquisa têm evoluído ao longo do tempo, com novas aplicações e abordagens sendo exploradas. Reforça a relevância das Copulas como uma ferramenta estatística poderosa para a análise de dependência entre variáveis econômicas e financeiras.

O referencial teórico apresentou os principais conceitos, teoremas e famílias de cópulas, discutindo sua relevância para análise de dependência e aplicações em finanças. As cópulas permitem estudar a estrutura de dependência entre variáveis independentemente das distribuições marginais, oferecendo maior flexibilidade.

A presente pesquisa aprofundou-se na teoria de cópulas, explorando em detalhes as medidas de dependência e concordância entre variáveis. O estudo se fundamentou em princípios estatísticos robustos para realizar uma avaliação minuciosa da dependência entre os retornos de cinco fundos de investimento brasileiros de destaque (BTG Discovery, JGP Hedge, Vokin Everest, Opportunity e Gávea Macro) e cinco variáveis macroeconômicas de relevância fundamental (Ibovespa, IMA-B, CDS, IBC-Br e Câmbio). A análise abrange um período de 2008 ao início de 2023. A metodologia descreveu os procedimentos adotados para avaliar a relação entre fundos de investimento brasileiros e variáveis macroeconômicas via cópulas. Foram utilizados testes de adequação e simulações para validar os modelos e analisar a estrutura de dependência identificada.

O fundo Vokin Everest, como supramencionado, demonstrou uma tendência de au-

mento na probabilidade de retornos positivos à medida que as condições macroeconômicas se aprimoram. Observamos que, quando as variáveis macroeconômicas se encontram em seus níveis mais baixos (entre os valores mínimos e o primeiro quartil), a probabilidade de o fundo gerar retornos positivos é nula (0%). No entanto, essa probabilidade aumenta gradativamente à medida que as variáveis macroeconômicas apresentam variações positivas e maiores. Esse comportamento sugere uma associação direta entre a melhora das variáveis macroeconômicas e a maior probabilidade de retornos positivos para o fundo Vokin Everest. O fundo BTG, por sua vez, apresentou dificuldade em gerar retornos positivos em cenários macroeconômicos desfavoráveis, mas sua probabilidade de retorno positivo aumenta significativamente com a melhora das variáveis macroeconômicas, especialmente o Ibovespa, sugerindo forte relação com o índice.

O fundo Opportunity apresentou uma probabilidade nula de alcançar retornos positivos em condições econômicas desfavoráveis (valores entre mínimos e o primeiro quartil), com um retorno esperado fortemente negativo que varia entre -9,021% e -9,240%. À medida que as condições econômicas e financeiras melhoram, observa-se um aumento na probabilidade de retornos positivos, que alcança cerca de 25% nos cenários mais otimistas (entre o terceiro quartil e o valor máximo), enquanto o retorno esperado se torna positivo, aproximadamente 5%. Esta tendência sugere uma relação positiva entre o desempenho das variáveis macroeconômicas e o retorno do fundo, indicando uma maior sensibilidade a condições de mercado desfavoráveis e um perfil de risco inclinado a perdas mais significativas em períodos de baixa performance econômica. O fundo em questão apresentou vulnerabilidade a variações negativas das variáveis macroeconômicas, com impactos negativos no retorno sendo mais acentuados do que os ganhos em cenários favoráveis.

A análise do fundo Gávea Macro, conforme supramencionado, revela uma forte relação entre seu desempenho e as flutuações das variáveis macroeconômicas. Observa-se que a probabilidade de retorno positivo é nula em cenários macroeconômicos desfavoráveis (correspondentes aos valores mínimos até o primeiro quartil). Essa probabilidade aumenta significativamente à medida que as condições macroeconômicas melhoram, evidenciando a sensibilidade particular do fundo ao risco do país. Em cenários macroeconômicos mais otimistas (entre a média e o máximo das variáveis), a probabilidade de retornos negativos diminui até se tornar nula. Esse padrão sugere que o fundo Gávea Macro apresenta uma maior chance de desempenho positivo em condições econômicas favoráveis. A dependência positiva entre o desempenho do fundo e as variáveis macroeconômicas destaca a capacidade do fundo de se beneficiar mais em cenários de crescimento e estabilidade econômica, indicando uma sensibilidade maior aos extremos superiores das variáveis macroeconômicas.

Essa análise sugere que o retorno do fundo de investimento está fortemente atrelado ao desempenho das variáveis econômicas, com uma tendência clara de melhores retornos em ambientes econômicos mais robustos. A sensibilidade do fundo às condições

de mercado, risco de crédito, câmbio e atividade econômica, conforme representado pelas variáveis analisadas, é evidente e deve ser considerada por investidores e gestores na tomada de decisões estratégicas.

Percebe-se, portanto, uma relação positiva entre o aumento dos retornos das variáveis analisadas e a probabilidade e magnitude dos retornos positivos do fundo. Isso indica que o fundo se beneficia de cenários de valorização no mercado, apresentando maiores retornos esperados quando as variáveis macroeconômicas têm performance positiva. O retorno do fundo de investimento é geralmente menor quando o retorno das variáveis está em intervalos mais restritos, independentemente do sinal do retorno do fundo. Isso indica que o fundo de investimento tende a ter um retorno mais modesto quando as variáveis têm retornos mais concentrados.

Destaca-se ainda que os resultados mostram que o tau de Kendall indica uma associação negativa entre o CDS e o retorno dos fundos, mas as probabilidades de retornos positivos são influenciadas por vários fatores além do risco de crédito do país. Os fundos demonstram capacidade de gerar retornos positivos independentemente dos níveis de CDS, indicando uma gestão eficiente do risco por parte dos gestores. A correlação negativa fornece uma visão geral, mas as decisões de investimento são baseadas em uma avaliação mais ampla que inclui análise de risco, diversificação de portfólio e projeções econômicas.

Este estudo contribui para o conhecimento sobre a aplicação de cópulas no âmbito das finanças, particularmente no que tange à análise de fundos de investimento e variáveis macroeconômicas. Os resultados obtidos fornecem subsídios para gestores de fundos e investidores na avaliação do impacto de cenários macroeconômicos adversos, possibilitando o aprimoramento de estratégias de alocação de ativos e gerenciamento de riscos.

REFERÊNCIAS

- AAS, K. et al. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, Elsevier, v. 44, n. 2, p. 182–198, 2009.
- ALOUÏ, R.; AÏSSA, M. S. B.; NGUYEN, D. K. Conditional dependence structure between oil prices and exchange rates: A copula-garch approach. *Journal of International Money and Finance*, Elsevier, v. 32, p. 719–738, 2013.
- ANJOS, U. U. dos et al. *Modelando Dependências via Cópulas*. 2004. Minicurso do 16o SINAPE, Caxambu, MG.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, Elsevier, v. 11, n. 4, p. 959–975, 2017.
- ARIA, M. et al. Thematic analysis as a new culturomic tool: the social media coverage on covid-19 pandemic in italy. *Sustainability*, MDPI, v. 14, n. 6, p. 3643, 2022.
- ASAD, M.; SIDDIQUI, D. A. Determinants of mutual funds performance in pakistan. *International Journal of Social and Administrative Sciences*, v. 4, n. 2, p. 85–107, 2019.
- BARCELLOS, T. *Modelagem da relação de dependência entre as variáveis de velocidade do vento e a geração de energia eólica: Uma aplicação da teoria de cópulas*. Dissertação (Engenharia de Produção) — Pontifícia Universidade Católica do Rio de Janeiro, set 2022.
- BEDFORD, T.; COOKE, R. M. Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 30, n. 4, p. 1031–1068, 2002.
- BILBAO-TEROL, A.; CAÑAL-FERNÁNDEZ, V.; MAR, A.-P. A model based on copula theory for sustainable and social responsible investments: Un modelo basado en la teoría de cópulas para la inversión sostenible y socialmente responsable. *Revista de Contabilidad-Spanish Accounting Review*, v. 19, n. 1, p. 55–76, 2016.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, Elsevier, v. 31, n. 3, p. 307–327, 1986.
- BOUYÉ, E. et al. Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*, 2000.
- BRIGO, D.; MERCURIO, F. Interest rate models-theory and practice springer-verlag. *Berlin, Heidelberg, New York*, 2006.
- CHEN, H. et al. Predicting fund performance using machine learning: Evidence from equity mutual funds. *Pacific-Basin Finance Journal*, v. 69, p. 101506, 2021.
- CHEN, N.-F.; ROLL, R.; ROSS, S. A. Economic forces and the stock market. *The Journal of Business*, University of Chicago Press, v. 59, n. 3, p. 383–403, 1986. ISSN 00219398, 15375374.

- CHEN, X.; FAN, Y.; TSYRENNIKOV, V. Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, Taylor & Francis, v. 101, n. 475, p. 1228–1240, 2006.
- CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. *Copula methods in finance*. [S.l.]: John Wiley & Sons, 2004.
- CHERUBINI, U. et al. *Dynamic copula methods in finance*. [S.l.]: John Wiley & Sons, 2011.
- CHU, P. K.-K. Relationship between macroeconomic variables and net asset values (nav) of equity funds: Cointegration evidence and vector error correction model of the hong kong mandatory provident funds (mpfs). *Journal of International Financial Markets, Institutions and Money*, Elsevier, v. 21, n. 5, p. 792–810, 2011.
- CLAYTON, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 65, n. 1, p. 141–151, 1978. ISSN 00063444.
- COBO, M. J. et al. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, Elsevier, v. 5, n. 1, p. 146–166, 2011.
- CZADO, C. Analyzing dependent data with vine copulas. *Lecture Notes in Statistics*, Springer, Springer, v. 222, 2019.
- DEGROOT, M. H.; SCHERVISH, M. J. *Probability and Statistics*. 4th. ed. [S.l.]: Addison-Wesley, 2021.
- DONTHU, N. et al. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, v. 133, p. 285–296, 2021. ISSN 0148-2963.
- ELTON, E. J.; GRUBER, M. J.; BLAKE, C. R. Fundamental economic variables, expected returns, and bond fund performance. *The Journal of Finance*, [American Finance Association, Wiley], v. 50, n. 4, p. 1229–1256, 1995. ISSN 00221082, 15406261.
- EMBRECHTS, P.; LINDSKOG, F.; MCNEIL, A. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, Citeseer, v. 14, 2001.
- EMBRECHTS, P.; MCNEIL, A.; STRAUMANN, D. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, New York, v. 1, p. 176–223, 2002.
- ENGLE, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 20, n. 3, p. 339–350, 2002.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, [Wiley, Econometric Society], v. 50, n. 4, p. 987–1007, 1982. ISSN 00129682, 14680262.

FAMA, E. F.; FRENCH, K. R. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, v. 47, n. 2, p. 427–465, 1992. ISSN 0022-1082. Publisher: [American Finance Association, Wiley].

FAMA, E. F.; FRENCH, K. R. The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, American Economic Association, v. 18, n. 3, p. 25–46, 2004.

FORBES, K. J.; RIGOBON, R. No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, Wiley Online Library, v. 57, n. 5, p. 2223–2261, 2002.

GENEST, C.; FAVRE, A.-C. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, American Society of Civil Engineers, v. 12, n. 4, p. 347–368, 2007.

GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, Elsevier, v. 44, n. 2, p. 199–213, 2009.

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, Wiley Online Library, v. 48, n. 5, p. 1779–1801, 1993.

GRUGEL, D. M. *Estudo de dependência de extremos via teoria de cópulas*. 62 p. Monografia (Bacharel em Estatística) — Instituto de Matemática e Estatística, Universidade Federal Fluminense, Niterói, 2015.

GYIMAH, A. G.; ADDAI, B.; ASAMOAH, G. K. Macroeconomic determinants of mutual funds performance in Ghana. *Cogent Economics & Finance*, v. 9, n. 1, p. 1913876, jan. 2021. ISSN null. Publisher: Cogent OA.

HADDOCK, D. D.; REILLY, F. K.; BROWN, K. C. *Investment Funds and Portfolio Management*. [S.l.]: Pearson, 2017.

HOFERT, M. et al. *Elements of copula modeling with R*. [S.l.]: Springer, 2018.

HUANG, W.; PROKHOROV, A. A goodness-of-fit test for copulas. *Econometric Reviews*, Taylor & Francis, v. 33, n. 7, p. 751–771, 2014.

HULL, J.; WHITE, A. et al. Valuation of a cdo and an nth to default cds without monte carlo simulation. *Journal of Derivatives*, v. 12, n. 2, p. 8–23, 2004.

JIANG, X.; ZHOU, Z. A neural network based deep learning model for forecasting fixed income mutual fund returns. *Journal of Forecasting*, v. 39, n. 2, p. 191–208, 2020.

JOE, H. *Multivariate models and multivariate dependence concepts*. [S.l.]: CRC press, 1997.

JONDEAU, E.; ROCKINGER, M. Optimal portfolio allocation under higher moments. *European Financial Management*, Wiley Online Library, v. 12, n. 1, p. 29–55, 2006.

KARIUKI, E. C. *Effect of Macro economic variables on financial performance of mutual funds industry in Kenya*. Tese (Doutorado) — University of Nairobi, 2014.

- KHOO, Y. C. et al. *Effect of macroeconomic factors on selective mutual fund performance in Malaysia*. Tese (Doutorado) — UNIVERSITI TUNKU ABDUL RAHMAN, 2020.
- LINNENLUECKE, M. K.; MARRONE, M.; SINGH, A. K. Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, v. 45, n. 2, p. 175–194, 2020.
- LINTNER, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, v. 47, n. 1, p. 13–37, 1965. ISSN 0034-6535.
- LOBÃO, J.; LEVI, A. The relation between mutual fund flows, stock returns and macroeconomic variables: evidence from Portugal. *Portuguese Journal of Finance, Management and eounting*, 2016. ISSN 2183-3826.
- LONGIN, F.; SOLNIK, B. Extreme correlation of international equity markets. *The journal of finance*, Wiley Online Library, v. 56, n. 2, p. 649–676, 2001.
- LUCIANO, E.; CHERUBINI, U. Multivariate option pricing with copulas. *Available at SSRN 269868*, 2001.
- MALINDA, M.; CHEN, J. H. The forecasting of consumer exchange-traded funds (etfs) via grey relational analysis (gra) and artificial neural network (ann). *Empirical Economics*, v. 62, n. 2, p. 779–823, 2022.
- MANNER, H.; RODRÍGUEZ, G.; STÖCKLER, F. A changepoint analysis of exchange rate and commodity price risks for latin american stock markets. *International Review of Economics & Finance*, Elsevier, v. 89, p. 1385–1403, 2024.
- MARKOWITZ, H. Portfolio selection. *The Journal of Finance*, [American Finance Association, Wiley], v. 7, n. 1, p. 77–91, 1952. ISSN 00221082, 15406261.
- MERTON, R. C. An Intertemporal Capital Asset Pricing Model. *Econometrica*, v. 41, n. 5, p. 867–887, 1973. ISSN 0012-9682. Publisher: [Wiley, Econometric Society].
- NAGLER, T. kdecopula: An r package for the kernel estimation of bivariate copula densities. *Journal of Statistical Software*, v. 84, n. 7, p. 1–22, 2018.
- NELSEN, R. B. *An Introduction to Copulas*. New York, NY: Springer, 2006. (Springer Series in Statistics). ISBN 978-0-387-28659-4.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, JSTOR, p. 347–370, 1991.
- PANIGRAHI, C. . A.; KARWA, P.; JOSHI, P. SSRN Scholarly Paper, *Impact of Macroeconomic Variables on the Performance of Mutual Funds: A Selective Study*. Rochester, NY: [s.n.], 2020.
- PATTON, A. J. Estimation of multivariate models for time series of possibly different lengths. *Journal of applied econometrics*, Wiley Online Library, v. 21, n. 2, p. 147–173, 2006.
- PATTON, A. J. Modelling asymmetric exchange rate dependence. *International economic review*, Wiley Online Library, v. 47, n. 2, p. 527–556, 2006.

- PATTON, A. J. Copula-based models for financial time series. In: ANDERSEN, T. et al. (Ed.). *Handbook of financial time series*. [S.l.]: Springer, 2009. p. 767–785.
- PATTON, A. J. A review of copula models for economic time series. *Journal of Multivariate Analysis*, v. 110, p. 4–18, 2012. ISSN 0047-259X. Special Issue on Copula Modeling and Dependence.
- QAMAR, H.; SINGH, S. Mutual fund performance prediction. In: IEEE. *2016 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*. [S.l.], 2016. p. 185–189.
- REBOREDO, J. C. Modelling oil price and exchange rate co-movements. *Journal of Policy Modeling*, Elsevier, v. 34, n. 3, p. 419–440, 2012.
- REBOREDO, J. C. Is gold a safe haven or a hedge for the us dollar? implications for risk management. *Journal of Banking & Finance*, Elsevier, v. 37, n. 8, p. 2665–2676, 2013.
- RODRIGUEZ, J. C. Measuring financial contagion: A copula approach. *Journal of empirical finance*, Elsevier, v. 14, n. 3, p. 401–423, 2007.
- ROSS, S. A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, v. 13, n. 3, p. 341–360, 1976. ISSN 0022-0531.
- SHARPE, W. F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. *The Journal of Finance*, v. 19, n. 3, p. 425–442, 1964. ISSN 1540-6261.
- SHARPE, W. F. Mutual Fund Performance. *The Journal of Business*, v. 39, n. 1, p. 119–138, 1966. ISSN 0021-9398. Publisher: University of Chicago Press.
- SKLAR, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, v. 8, p. 229–231, 1959.
- TEKER, D.; TASSEVEN, O.; TUKEL, A. Determinants of capital structure for turkish firms: A panel data analysis. *International Research Journal of Finance and Economics*, v. 1, p. 179–187, 07 2009.
- THOMÉ, A. M. T.; SCAVARDA, L. F.; SCAVARDA, A. J. Conducting systematic literature review in operations management. *Production Planning & Control*, Taylor & Francis, v. 27, n. 5, p. 408–420, 2016.
- TREYNOR, J. L. SSRN Scholarly Paper, *Market Value, Time, and Risk*. Rochester, NY: [s.n.], 1961.
- VIOLA, M. L. L. Tipos de dependência entre variáveis aleatórias e teoria de cópulas. Instituto de Matemática, Estatística e Computação Científica, p. 34, 2009. Disponível em: <<https://www.ime.unicamp.br/~veronica/dependence/book.pdf>>. Acesso em: 06 de abril de 2024.
- WANG, W.; WELLS, M. T. Model selection and semiparametric inference for bivariate failure-time data. *Journal of the american statistical association*, JSTOR, p. 62–72, 2000.

WU, C.-C.; CHUNG, H.; CHANG, Y.-H. The economic value of co-movement between oil price and exchange rate using copula-based garch models. *Energy Economics*, Elsevier, v. 34, n. 1, p. 270–282, 2012.

XIAO, Y.; WATSON, M. Guidance on conducting a systematic literature review. *Journal of planning education and research*, SAGE Publications Sage CA: Los Angeles, CA, v. 39, n. 1, p. 93–112, 2019.

ZHANG, X.; JIANG, H. Application of copula function in financial risk analysis. *Computers & Electrical Engineering*, Elsevier, v. 77, p. 376–388, 2019.

ANEXO A – Pseudo-código

Algoritmo 1: Parte 1

```

Função read_data(caminho)
/* Lê o arquivo CSV e processa os dados */
raw_data ← read_csv(caminho)
funds_name ← extrair_nomes_únicos(raw_data)
data_list ← dividir_dados_por_fundo(raw_data, funds_name)
Retorna data_list

Função melhor_distribuicao(df, distribuicoes)
/* Função utilizada para encontrar a melhor distribuição
estatística */
Para cada coluna em df faça
/* Testa diferentes distribuições para cada série e retorna a
mais adequada. */
Retorna resultados_lista

Função create_cdf(df, dist)
/* Cria uma função de distribuição cumulativa com base nos
resultados da função melhor_distribuicao(). */
Retorna cdf_list

Função encontrando_melhor_familia(data, B, criterio)
/* Aplica o teste de bondade na família de cópula escolhida pela
função VineCopula::BiCopSelect() com base no valor da
verossimilhança. Caso p-valor < 0.05, refaz o processo. */
Retorna resultados

Função estimando_copula_bivariada(data, familia)
/* Estima a cópula de acordo com as famílias encontradas na função
encontrando_melhor_familia() */
Retorna resultado

```

Algoritmo 1: Parte 2

```

Função simulando_dados_copula_bivar(copulas, orig_data, simulations)
/* Simula dados das cópulas bivariadas */
Retorna teste_df
Função estimando_kernel(data)
/* Estima a copula por kernel. */
Retorna copulas

Função simulando_dados_kernel(data, copulas, simulacoes)
/* Simula os dados e faz a transformação para a distribuição
original. */
Retorna sim

Função pegando_os_parametros_kernel(copula)
/* Extrai parâmetros das copulas do kernel */
Retorna x

Função calculate_probabilities(df_sim)
/* Calcula probabilidades de ocorrência e o retorno médio */
Retorna list_result

Início
/* Execução das funções em sequência */
dados ← read_data("caminho_para_arquivo.csv")
distribuicoes ← c("Vetor das distribuicoes que sera testada")
melhor_dist ← melhor_distribuicao(dados, distribuicoes)
cdf_dados ← create_cdf(dados, melhor_dist)
familia ← encontrando_melhor_familia(cdf_dados, 100, "logLik")
copulas_bivar ← estimando_copula_bivariada(cdf_dados, familia)
simulacao_bivar ← simulando_dados_copula_bivar(copulas_bivar, dados, 2000)
/* O código responsável pelos dados é a partir daqui. */
pseudo_obs ← VineCopula::pobs(dados)
copula_kernel ← estimando_kernel(pseudo_obs)
simulacao_kernel ← simulando_dados_kernel(dados, copula_kernel)
probabilidade ← calculate_probabilities(simulacao_kernel)

Fim

```

O pseudo-código apresentado descreve de forma abrangente um procedimento computacional destinado a analisar a interação entre fundos de investimento brasileiros e variáveis macroeconômicas, utilizando para tal a teoria de cópulas. Este procedimento

inicia com a função *read_data*, , que processa dados de um arquivo CSV, extraindo nomes únicos dos fundos e dividindo os dados por fundo, fundamentando a base para análises subsequentes. Em seguida, o script aborda a seleção da melhor distribuição que melhor se ajusta aos dados através da função *melhor_distribuicao*, com isso, foi possível a criação de funções de distribuição cumulativa com *create_cdf*, e a determinação da família de cópulas mais adequada via *encontrando_melhor_familia*. Estes passos são cruciais para a estimação de cópulas paramétrica *estimando_copula_bivariada* que permitem uma análise mais detalhada da dependência entre as variáveis em estudo, em seguida, foi simulado os dados com a função *simulando_dados_copula_bivar*. Dado que a estimação por cópulas paramétricas não se mostrou adequada, houve a necessidade da estimação das copulas de maneira não paramétrica. Para isso foi utilizada a função *estimando_kernel* que transforma os dados em pseudo-observações antes de estimar as cópulas. Após a estimação por kernel, foi simulado os dados com *simulando_dados_kernel*, permitindo a avaliação da performance dos modelos em condições variadas. A função *calculate_probabilities* finaliza o processo ao calcular as probabilidades de ocorrências e retornos médios.

ANEXO B – Tabela dos intervalos

Na condução da análise envolvendo copulas bivariadas entre retornos dos fundos de investimento e variáveis macroeconômicas, objetiva-se elucidar a interação e a relação existentes entre um fundo selecionado e uma variável macroeconômica específica. Essencial para este escopo é a determinação da probabilidade de incidência de valores dentro de segmentos pré-definidos, em função das oscilações da variável macroeconômica em questão. Embora a precisão do cálculo da probabilidade de ocorrência de valores exatos (pontuais) esteja ao alcance teórico, a escolha metodológica recai sobre a análise de intervalos, conforme delineado na Tabela 11.

Além disso, a adoção de intervalos propicia uma análise mais abrangente da dinâmica intervável, facilitando a identificação de padrões e tendências significativos. Ao segmentar as variáveis em intervalos tais como [Mínimo, 1Q), [1Q, Média), [Média, 3Q) e [3Q, Máximo), obtêm-se percepções mais abrangentes sobre a distribuição e a probabilidade de eventos.

Por intermédio de exemplos, como a segmentação da variável Ibovespa em intervalos que variam de [-35,493, -3,523) a [4,801, 15,658), e o tratamento similar aplicado às variáveis IMA-B, CDS, Câmbio e IBC-Br, procede-se à análise das probabilidades de resultados específicos dos fundos, considerando-se que a variável macroeconômica esteja situada dentro de um segmento particular.

Tabela 11 - Intervalo das variáveis.

Variável	[Mínimo, 1Q)	[1Q, Média)	[Média, 3Q)	[3Q, Máximo)
Ibovespa	[-35,493, -3,523)	[-3,523, 0,319)	[0,319, 4,801)	[4,801, 15,658)
IMA-B	[-7,191, -0,192)	[-0,192, 0,970)	[0,970, 1,974)	[1,974, 6,891)
CDS-5y	[-35,228, -10,396)	[-10,396, 0,105)	[0,105, 8,048)	[8,048, 67,597)
IBC-Br	[-10,142, -0,280)	[-0,280, 0,100)	[0,100, 0,635)	[0,635, 4,703)
Câmbio	[-9,077, -2,122)	[-2,122, 0,576)	[0,576, 3,045)	[3,045, 14,379)

Fonte: Elaboração própria do Autor.

A funcionalidade específica desta componente da modelagem é elucidada da seguinte maneira: tomemos, por exemplo, a circunstância em que o valor do CDS esteja enquadrado no intervalo [-35,228, -10,3963). Questiona-se, então, qual seria a probabilidade de o fundo sob análise registrar um desempenho no intervalo mais baixo, isto é, do mínimo até o primeiro quartil, com base no histórico da série temporal até o presente? Ou ainda, qual a probabilidade de o fundo demonstrar um desempenho que varie entre o primeiro quartil e a média, considerando que o valor do CDS no período em análise esteja próximo do seu valor mínimo histórico? Estes são exemplos de indagações pertinentes

que emergem ao se optar pela divisão dos dados em intervalos.