



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

**Avaliando o impacto de medidas de confiabilidade
do tipo *stress-strength* com marginais extremais:
aplicações em dados financeiros e hidrológicos**

por

Rebeca Klamerick Lima

Brasília, Julho de 2024

Avaliando o impacto de medidas de confiabilidade do tipo *stress-strength* com marginais extremas: aplicações em dados financeiros e hidrológicos

por

Rebeca Klamerick Lima¹

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Pushpa N. Rathie

Co-orientador: Prof. Dr. Felipe S. Quintino

Brasília, Julho de 2024

¹A autora contou com apoio financeiro CAPES durante a realização deste trabalho.

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Prof. Dr. Pushpa N. Rathie

Orientador, EST/UnB

Prof. Dr. Felipe S. Quintino

Co-orientador:, EST/UnB

Prof. Dr. Cira Etheowalda Guevara Otiniano

Membro interno, EST/UnB

Prof. Dr. José Aílton Alencar Andrade

Membro externo, DEMA/UFC

Prof. Dr. Frederico Machado Almeida

Suplente, EST/UnB

Agradecimentos

Primeiramente, agradeço a Deus pela minha vida e saúde, por ter me concedido mais esta conquista. A conclusão desta dissertação marca um momento importante na minha jornada acadêmica.

Gostaria de expressar minha profunda gratidão à Universidade de Brasília pelo apoio e recursos fornecidos durante a realização deste estudo. Agradeço imensamente aos meus orientadores, Prof. Dr. Pushpa N. Rathie e Prof. Dr. Felipe S. Quintino, pela orientação, paciência e valiosas contribuições ao longo deste processo. Também sou grata aos colegas de pesquisa, Prof. Dr. Tiago da Fonseca e Dr. Luan Ozelim, que compartilharam suas ideias e conhecimentos, enriquecendo significativamente este trabalho. Agradeço especialmente ao Prof. Dr. Helton Saulo por contribuir com revisões neste trabalho.

Não posso deixar de mencionar o apoio inabalável da minha família, especialmente meu marido, Melquisadec de Souza Oliveira, pelo apoio constante, pela paciência e compreensão durante os desafios enfrentados na elaboração desta dissertação.

Agradeço a todos os professores do Programa de Pós-Graduação em Estatística da Universidade de Brasília e aos colegas de curso, especialmente a Estevão Sérgio Zeco, João Victor Monteiro de Andrade, Leonardo Santos da Cruz, Mariana Fehr Nicácio, Melquisadec de Souza Oliveira, Moisés Lima e Tiago Chandiona Ernesto, que tornaram esta caminhada mais leve. Obrigado pela troca de conhecimentos.

Agradeço também à Prof. Dr. Cira Etheowalda Guevara Otiniano e ao Prof. Dr. Frederico Machado Almeida pelas contribuições e sugestões feitas durante a qualificação desta dissertação.

E desde já, agradeço à Prof. Dr. Cira Etheowalda Guevara Otiniano e ao Prof. José Aílton Alencar Andrade por aceitarem participar da banca de defesa deste trabalho.

Por fim, dedico este trabalho aos meus pais e ao meu marido, que são minha inspiração e motivação constantes. Agradeço a todos que, de alguma forma, contribuíram para a conclusão deste trabalho e para o meu crescimento acadêmico e pessoal. Esta conquista é também de vocês. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Em estudos de confiabilidade, estamos interessados no comportamento de um sistema quando ele interage com seu ambiente circundante. Para avaliar o comportamento do sistema em termos de confiabilidade, podemos considerar a qualidade analisadas do sistema como sua resistência e o resultado das interações como tensão. A falha é observada sempre que a tensão excede a resistência. Tomando Y como uma variável aleatória que representa a tensão (*stress*) que o sistema experimenta e a variável aleatória X como sua resistência (*strength*), a probabilidade de não falha pode ser tomada como um indicador da confiabilidade do componente e expressa como $P(Y < X) = 1 - P(X < Y)$. Dessa forma, neste trabalho, consideramos que X e Y seguem distribuições generalizadas de valores extremos (GEV), que representam uma família de distribuições de probabilidade contínuas amplamente aplicadas em contextos de Engenharia e Economia. Nossa contribuição lida com um cenário mais geral em que tensão e resistência possuem dependência, e cópulas são usadas para modelar a dependência entre as variáveis aleatórias envolvidas. As cópulas de Gumbel-Hougaard, Frank e Clayton foram usadas para modelar conjuntos de dados bivariados. Em cada caso, critérios de informação foram considerados para comparar as capacidades de modelagem de cada cópula. Duas aplicações económicas em conjuntos de dados reais são discutidos, bem como uma aplicação em Engenharia. Outro tema abordada neste trabalho foi a confiabilidade em um sistema multicomponente. A confiabilidade pode ser estendida para um sistema multicomponente que consiste em k componentes de resistência com uma tensão comum, ou seja, $R_{s,k} = P(\text{pelo menos } s \text{ das } (X_1, \dots, X_k) \text{ excedem } Y)$, sendo X_1, X_2, \dots, X_k variáveis aleatórias independentes e idênticamente distribuídas que representam

a resistência de k componentes em um sistema e seguirem uma função de distribuição acumulada comum $G_X(\cdot)$, onde cada componente está sujeito a uma tensão aleatória independente Y , com a FDP $f_Y(\cdot)$. Dessa forma foram consideradas as leis estáveis l-máximas e as leis p-max estáveis como as distribuição para a aplicação deste estudo. Uma aplicação em conjuntos de dados reais é discutida. No geral, é descrito um quadro metodológico fácil de usar, permitindo que os profissionais o apliquem em seus próprios projetos de pesquisa.

Palavras-chaves: Confiabilidade *stress-strength*; Distribuições dos valores extremos; Cópu-
las bivariadas; Multicomponente.

Abstract

In reliability studies, we are interested in the behaviour of a system when it interacts with its surrounding environment. To assess the system's behavior in a reliability sense, we can take the system's intrinsic quality as strength, and the outcome of interactions as stress. Failure is observed whenever the stress exceeds the strength. Taking Y as a random variable representing the stress the system experiences and the random variable X as its strength, the probability of not failing can be taken as a proxy for the reliability of the component and given as $P(Y < X) = 1 - P(X < Y)$. This way, in the present paper it is considered that X and Y follow generalized extreme value distributions, which represent a family of continuous probability distributions that have been extensively applied in Engineering and Economy contexts. Our contribution deals with a more general situation when stress and strength are not independent, and copulas are used to model the dependence between the involved random variables. Gumbel-Hougaard, Frank and Clayton copulas were used for modelling bivariate data sets. In each case, information criteria were considered to compare the modelling capabilities of each copula. Two economic applications on real data sets are discussed, as well as an Engineering one. Another topic addressed in this work was reliability in a multicomponent system. Reliability can be extended to a multicomponent system consisting of k resistance components with a common voltage, i.e. $R_{s,k} = P(\text{at least } s \text{ of } (X_1, \dots, X_k) \text{ exceed } Y)$, with X_1, X_2, \dots, X_k independent and identically distributed random variables that represent the resistance of k components in a system and follow a common cumulative distribution function $G_X(\cdot)$, where each component is subject to an independent random voltage Y , with the PDF $f_Y(\cdot)$. Thus, the l-maximum stable laws and

the stable p-max laws were considered as the distributions for the application of this study. An application to real data sets is discussed. Overall, an easy-to-use methodological framework is described, allowing practitioners to apply it to their own research projects.

Keywords: Stress–strength reliability; Distributions of extreme values; Bivariate copulas; Multicomponent.

Sumário

1	Introdução	1
2	Preliminaries	7
2.1	Introdução	7
2.2	Cópuas	8
2.3	Distribuições dos valores extremos	11
2.3.1	Distribuição de Fréchet	12
2.3.2	Distribuição Weibull reversa	13
2.3.3	Distribuição de Gumbel	14
2.3.4	Distribuição GEV	15
2.3.5	Leis p-max estáveis	17
2.4	Seleção de Modelos	22
2.4.1	Critério de Informação de Akaike	23
2.4.2	Critério de Informação Bayesiano	23
2.4.3	Critério de Desvio Esperado	23
2.5	Problemas Abordados	24
3	Avaliando o impacto da seleção de copulas nas medidas de confiabilidade do tipo <i>stress-strength</i> com marginais GEV.	25
3.1	Introdução	25

3.2	Estimação da probabilidade de <i>stress-strength</i>	26
3.3	Estudo de Simulações	28
3.4	Aplicações	33
3.4.1	Seleção de ativos financeiros	33
3.4.2	Dados de renda e consumo	49
3.4.3	Fluxos mínimos mensais de água.	54
4	MSSR com Marginais de Distribuições de Valores Extremos: Teoria e Aplicação a	
	Dados Hidrológicos	58
4.1	Introdução	58
4.2	Confiabilidade do sistema multicomponente	59
4.2.1	Uma expressão alternativa para MSSR	60
4.2.2	Leis estáveis l-máximas	61
4.2.3	Leis p-max estáveis	64
4.3	Aplicações a dados hidrológicos	67
4.3.1	Aplicação aos dados hidrológicos	68
5	Conclusões	73
A	Artigo: Lima et al. (2024)	75
	Referências Bibliográficas	77

Capítulo 1

Introdução

A probabilidade de falha de um sistema ou componente pode ser calculada comparando estatisticamente a tensão (*stress*) aplicado à sua resistência (*strength*). Sejam X (*strength*) e Y (*stress*) variáveis aleatórias contínuas com função de densidade de probabilidade (FDP) conjunta $f_{X,Y}(\cdot, \cdot)$. A confiabilidade (ou probabilidade) de *stress-strength* (SSR) é definida como:

$$R = P(X < Y) = \int_{-\infty}^{\infty} \int_{-\infty}^y f_{X,Y}(x, y) dx dy. \quad (1.1)$$

Há várias aplicações dessa estrutura, como na Seleção de Ativos Financeiros (Quintino et al., 2024), Fragilidade Financeira Doméstica (Domma e Giordano, 2012) e Engenharia (Surles e Padgett, 1998), entre outros. Pode consultar Kotz, Lumelskii e Pensky (2003) para mais detalhes.

A Equação (1.1) pode ser avaliada quando a representação analítica da FDP conjunta é conhecida (ou qualquer outra fórmula estatística equivalente que possa ser transformada em $f_{X,Y}$). Portanto, avaliar adequadamente essa formulação conjunta é de extrema importância para aplicações de SSR do tipo $P(X < Y)$. Encontrar uma representação precisa para $f_{X,Y}$ envolve dois passos: entender como são as distribuições marginais e como é a estrutura de dependência dessas variáveis aleatórias.

Assim, determinar as melhores distribuições marginais para X e Y é um dos problemas

de interesse. Os experimentos envolvendo dados financeiros no século XX apontaram para a adequação de distribuições de caudas pesadas na modelagem, seja por meio de processos α -estáveis (alternativa de cauda pesada ao movimento Browniano de cauda leve (Bachelier, 1900)) ou por meio de modelos de séries temporais de caudas pesadas (Embrechts, Klüppelberg e Mikosch, 2013; Taleb, 2020). Embora a literatura proponha a hipótese geral de que os retornos logarítmicos em dados financeiros seguem um processo α -estável com parâmetro $0 < \alpha < 2$ (Mandelbrot, 1967) (e variância indefinida), sem perda de generalidade, os resultados da Teoria dos Valores Extremos (TVE) em relação à distribuição de valores extremos generalizada (GEV) (Jenkinson, 1955) apresentam essa distribuição como uma alternativa às distribuições α -estáveis. Essa abordagem pode ser considerada válida, uma vez que a distribuição GEV apresenta comportamento de cauda pesada. Do ponto de vista econômico, é bem conhecido que retornos extremos de ações nos mercados de ações podem ter importantes implicações para o gerenciamento de risco financeiro, e vários estudos aplicaram com sucesso a GEV para modelar dados financeiros (Cirillo e Taleb, 2016; Gettinby et al., 2004; Goncu et al., 2012; Hussain e Li, 2015; Quintino et al., 2024).

Além disso, distribuições de cauda pesada como Log-Normal e Pareto são conhecidas por sua adequação em abranger tanto as caudas quanto o modo das funções de densidade de renda empírica (Domma e Giordano, 2012). No entanto, algumas limitações dessas distribuições deixam a oportunidade para a aplicação de outras distribuições que se assemelham ao comportamento de cauda pesada, mas trazem mais flexibilidade no conjunto de parâmetros (Dagum, 1977). Dessa forma, explorar como a distribuição GEV se comporta ao modelar dados de renda e consumo em um quadro de avaliação de fragilidade é um dos interesses desse trabalho.

Uma discussão das características da distribuição GEV foi feita em Kotz e Nadarajah (2000), destacando sua ampla relevância em vários domínios, como testes de vida acelerada, desastres naturais, corridas de cavalos, chuvas, filas de supermercado, correntes marítimas, velocidades do vento, registros de corridas, e mais.

Após selecionar as marginais para X e Y , sua estrutura de dependência precisa ser consi-

derada. A confiabilidade SSR quando X e Y são variáveis aleatórias independentes seguindo distribuições dos valores extremos tem sido amplamente estudada na literatura. Nadarajah (2003) considerou a classe de distribuições de valores extremos e derivou as formas correspondentes para a confiabilidade $R = P(X < Y)$ em termos de funções especiais. Vários autores trabalharam na estimativa e aplicação da probabilidade de *stress-strength* para as distribuições de valores extremos como por exemplo, Abbas e Tang (2014), Jia, Nadarajah e Guo (2017), Krishnamoorthy e Lin (2010) e Kundu e Raqab (2009). A confiabilidade SSR para distribuições GEV independentes foi estudada por Quintino et al. (2024), que derivaram fórmulas de confiabilidade SSR e investigaram a aplicação da medida de confiabilidade $P(X < Y)$ na seleção de ativos financeiros.

O uso de cópulas para modelar a estrutura de dependência de variáveis aleatórias é uma consequência direta do Teorema de Sklar, que possibilita a criação de várias famílias de cópulas, cada uma das quais captura melhor situações de dependência específicas. Entre as cópulas mais exploradas, a cópula de Gumbel-Hougaard é especialmente eficaz em lidar com a dependência da cauda superior entre variáveis. Por outro lado, a cópula de Clayton é adequada para delinear a dependência da cauda inferior entre variáveis, enquanto a cópula de Frank se destaca na captura da dependência simétrica da cauda entre variáveis (Frees e Valdez, 1998). Para um estudo mais detalhado sobre a teoria de cópulas, o leitor pode consultar o livro de Nelsen (2006).

Trabalhos recentes exploraram cópulas em diferentes cenários, como o uso da Cópula de Frank para modelar ativos financeiros com marginais de Dagum para Gestão de Portfólio de Ativos Dependentes (Rathie, Ozelim e Andrade, 2021), a aplicação de cópulas em dados de empresas de seguros sobre perdas e despesas (Frees e Valdez, 1998), a derivação da distribuição bivariada de chuvas de monção em subdivisões meteorológicas vizinhas (Ghosh, 2010) e uma simulação de vazão mensal usando uma cópula de máxima entropia-Gumbel-Hougaard (Kong et al., 2015). Propriedades teóricas de cópulas de valores extremos também podem ser encontradas em Gudendorf e Segers (2010) e Kasper, Fuchs e Trutschnig (2021).

Considerando seu uso comum tanto em aplicações financeiras quanto de engenharia, as

cópuas de Gumbel-Hougaard, Frank e Clayton foram consideradas no presente trabalho. Além dos aspectos teóricos, exploramos os procedimentos de estimação, propondo uma estrutura geral que pode ser aplicado por profissionais ao considerar marginais GEV com as três famílias de cópuas mencionadas. Nesse caso, pode-se observar que o valor de R dado em (1.1) depende dos parâmetros das distribuições marginais (para a distribuição GEV, como será visto, localização μ , escala σ , e forma γ) e um parâmetro de dependência θ (introduzido pelo modelo de cópuas). Motivados pela melhoria no tempo computacional, optamos por uma estimação em duas etapas, ou seja, as marginais são modeladas e então o parâmetro de dependência é estimado, através da estimação de máxima verossimilhança, inferência marginais em 2 passos (Genest e Rivest, 1993).

Outra temática abordada nesse trabalho são os casos de confiabilidade em um sistema multicomponente. A teoria SSR pode ser estendida para um sistema multicomponente que consiste em k componentes de resistência com uma tensão comum, e o sistema funciona quando pelo menos s ($1 \leq s \leq k$) componentes sobrevivem simultaneamente.

Sejam X_1, X_2, \dots, X_k variáveis aleatórias (v.a.'s) independentes e idênticamente distribuídas (i.i.d.) que representam a resistência de k componentes em um sistema e seguirem uma função de distribuição acumulada (FDA) comum $G_X(\cdot)$. Cada componente está sujeito a uma tensão aleatória independente Y , com a FDP $f_Y(\cdot)$. O modelo de confiabilidade *stress-strength* multicomponente (MSSR) é dado por:

$$\begin{aligned} R_{s,k} &= P(\text{pelo menos } s \text{ das } (X_1, \dots, X_k) \text{ excedem } Y) \\ &= \sum_{j=s}^k \binom{k}{j} \int_{-\infty}^{\infty} (1 - G_X(u))^j (G_X(u))^{k-j} f_Y(u) du. \end{aligned} \quad (1.2)$$

A estimativa de confiabilidade em um modelo de MSSR foi realizada por vários autores. Procedimentos de inferência ótima para a função de confiabilidade sob o modelo paramétrico de distribuições exponenciais foram obtidos por Bhattacharyya e Johnson (1974). De igual forma

Wang et al. (2021) estudaram a estimativa de confiabilidade SSR para sistemas multicomponentes com distribuição Rayleigh. Estimativas Bayesianas de confiabilidade para o modelo de MSSR baseada na distribuição Topp-Leone foi estudada por Akram e Yousif (2022). Os autores Lio et al. (2022) desenvolveram uma metodologia inferencial para MSSR quando as variáveis de resistência e tensão seguem as distribuições Burr XII sob censura do Tipo-II nos dados de força.

O objetivo desse trabalho é dividido em duas partes: um estudo detalhado de Lima et al. (2024a), onde estendemos a abordagem de Quintino et al. (2024) estudando (1.1) quando X e Y seguem distribuições GEV dependentes com uma FDP conjunta dada por cópulas; obter novas expressões fechadas para (1.2) quando as componentes do sistema multicomponente seguem distribuições de valores extremos. Tais resultados obtidos foram aplicados em estudo de simulações e à dados reais.

A presente dissertação está organizada da seguinte forma: no Capítulo 2 são apresentadas definições e resultados gerais necessários para realizar o estudo proposto. São apresentadas a FDA e a FDP das cópulas de Gumbel-Hougaard, Frank e Clayton. Além disso, a FDA, a FDP e algumas propriedades das distribuições de valores extremos também são apresentadas.

No Capítulo 3, um estudo de simulações é apresentado para validar os estimadores de $R = P(X < Y)$ propostos, quando X e Y seguem distribuições GEV dependentes. O modelo metodológico aqui proposto é aplicado para estimar R em três situações reais. Primeiramente, consideramos uma situação de seleção de ativos onde existe correlação entre os retornos de pares de diferentes ações. Em resumo, quando X e Y representam variáveis aleatórias de retorno financeiro e $R < 1/2$, é aconselhável que o investidor escolha o ativo correspondente à variável X . Se $R > 1/2$, ocorre o oposto. O caso $R = 1/2$ é inconclusivo. A medição da fragilidade financeira domiciliar é considerada como uma segunda situação de modelagem de dados reais. Usando dados da Pesquisa sobre Renda e Consumo das Famílias do Banco de dados da Itália de 2008 (Domma e Giordano, 2012), investigamos com que frequência as famílias têm seu consumo anual maior que sua renda. Finalmente, um terceiro banco de dados é modelado, permitindo comparar a vazão mínima mensal para o rio Piracicaba no Brasil. Essa comparação é útil para

definir medidas de contingência para complementar a matriz elétrica, que é em grande parte dependente de fontes hídricas. Por fim, destacamos que os resultados apresentados neste capítulo foram publicados em Lima et al. (2024a).

No Capítulo 4, estudamos estimativas para (1.2) quando X_1, \dots, X_k e Y seguem distribuições de valores extremos. Novas expressões analíticas são obtidas para várias classes de distribuições de valores extremos. Uma aplicação de conjunto de dados reais é realizada para mostrar o desempenho de tais expressões na modelagem de dados. Os dados representam a capacidade mensal de água do Reservatório Shasta na Califórnia (EUA) durante agosto, setembro e dezembro de 1980 a 2015.

Finalmente, no Capítulo 5 as considerações finais do trabalho são apresentadas.

Capítulo 2

Preliminaries

2.1 Introdução

Neste capítulo, apresentamos os conceitos que serão utilizados ao longo desta dissertação. Iniciamos a Seção 2.2 apresentando um breve contexto histórico sobre cópulas, as cópulas escolhidas para este trabalho e como estimar os parâmetros das mesmas.

A Seção 2.3 é dedicada a apresentação das distribuições de valores extremos utilizadas nos capítulos seguintes.

Na Seção 2.5 apresentamos a função dos valores extremos \mathbb{H} dado por

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^{\infty} y^{a_6} \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy.$$

Esta função foi introduzida por Rathie et al. (2023). Em nosso estudo, utilizamos a função \mathbb{H} para reescrever a forma de estimar a confiabilidade sem impondo mínimo de restrições aos parâmetros das distribuições envolvidas.

Finalmente, destacamos que um leitor familiarizado com os tópicos abordados nesse capítulo pode omitir a leitura do capítulo e ir diretamente ao Capítulo 3, retornando a este capítulo apenas para citações ocasionais.

As principais referências para esse capítulo são: Domma e Giordano (2012), Genest e Rivest (1993), Kolev, Anjos e Mendes (2006), Mohan e Ravi (1993), Nadarajah (2003), Nelsen (2006) e Quintino et al. (2024).

2.2 Cópulas

O estudo das cópulas e suas aplicações em estatística tem ganhado destaque nas últimas décadas. O interesse por cópulas cresceu significativamente nas últimas décadas, como evidenciado pelo aumento no número de publicações e conferências dedicadas ao tema. A primeira utilização formal do termo “copula” em um contexto matemático foi por Sklar (1959), ao descrever funções que conectam distribuições unidimensionais para formar distribuições multivariadas. Os trabalhos de Fréchet, Hoeffding, e outros pesquisadores também contribuíram significativamente para o desenvolvimento das cópulas (Nelsen, 2006).

Definição 2.2.1 (Nelsen, 2006). Uma cópula bi-dimensional é uma função $C : [0, 1]^2 \rightarrow [0, 1]$ satisfazendo as seguintes propriedades:

1. Para cada u, v em $[0, 1]$,

$$C(u, 0) = 0 = C(0, v) \quad (2.1)$$

$$C(u, 1) = u \quad \text{e} \quad C(1, v) = v; \quad (2.2)$$

2. Para cada u_1, u_2, v_1, v_2 em $[0, 1]$ tal que $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (2.3)$$

O Teorema de Sklar apareceu pela primeira vez em Sklar (1959) e este teorema fundamenta o uso de cópulas. Ele também estabelece a possibilidade de expressar qualquer distribuição

conjunta multivariada em termos de funções de distribuição marginal univariadas e uma cópula que descreve a estrutura de dependência entre as variáveis.

Teorema 2.2.1 (Teorema de Sklar). Seja H uma função de distribuição conjunta com marginais F e G . Então existe uma cópula C tal que, para todos $x, y \in \mathbb{R}$,

$$H(x, y) = C(F(x), G(y)). \quad (2.4)$$

Se F e G são contínuas, então C é única; caso contrário, C é unicamente determinada em $\text{Im } F \times \text{Im } G$. Por outro lado, se C é uma cópula e F e G são funções de distribuição, então a função H definida por (2.4) é uma função de distribuição conjunta com marginais F e G .

Note que (2.4) está associada a vetor aleatório, isto é, $(X, Y) \sim H$ e $(F(X), G(Y)) = (U, V) \sim C$.

Para a demonstração do Teorema de Sklar ver Teorema 2.3.3 em Nelsen (2006).

As cópulas são especialmente úteis em aplicações estatísticas, permitindo a modelagem e a estimação da distribuição de vetores aleatórios de forma separada para marginais e cópulas (Cherubini, Luciano e Vecchiato, 2004). Existem diversas famílias de cópulas paramétricas disponíveis, cada uma com parâmetros que controlam a intensidade da dependência entre as variáveis. Alguns exemplos populares incluem a Cópula Gaussiana, Cópula de Frank, Cópula de Clayton, Cópula de Gumbel e Cópula de Student-t (Cherubini, Mulinacci e Gobbi, 2009). Porém no Capítulo 3 abordaremos três tipos de cópulas que são as cópulas de Gumbel-Hougaard, Frank e Clayton, apresentadas a seguir:

A cópula de Gumbel-Hougaard é dada por:

$$C^{GH}(u, v; \theta) = \exp \left\{ - \left[(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}} \right]^{\theta} \right\}, \quad \theta \in (0, 1], \quad (2.5)$$

onde θ é o valor que determina o grau e a estrutura de dependência entre as variáveis aleatórias. Observe que se $\theta = 1$, então $C^{GH}(u, v; 1) = uv$, ou seja, as marginais são independentes.

As cópulas de Frank e Clayton são dadas, respectivamente, por:

$$C^F(u, v; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(\exp\{-\theta u\} - 1)(\exp\{-\theta v\} - 1)}{\exp\{-\theta\} - 1} \right), \quad \theta \in \mathbb{R} - \{0\}, \quad (2.6)$$

e

$$C^C(u, v; \theta) = \max \left\{ (u^{-\theta} + v^{-\theta} - 1); 0 \right\}^{-\frac{1}{\theta}}, \quad \theta \in [-1, \infty) - \{0\}. \quad (2.7)$$

Para obter a FDP associada a cada FDA, basta considerar que:

$$c(u, v; \theta) = \frac{\partial^2 C(u, v; \theta)}{\partial u \partial v},$$

onde $\frac{\partial C(u, v; \theta)}{\partial u}$ simboliza a derivada parcial da função com respeito u . A Tabela 2.1 apresenta as FDPs das cópulas de Gumbel-Hougaard, Frank e Clayton.

Tabela 2.1: Cópulas, FDP e seus espaços de parâmetros

Cópula	$c(u, v; \theta)$	espaços de parâmetros
Gumbel-Hougaard	$\frac{C(u, v; \theta)(\log u \log v)^{\frac{1}{\theta}-1}}{uv \left[(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}} \right]^{2-\theta}} \left\{ \left[(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}} \right]^{\theta} + \frac{1}{\theta} - 1 \right\}$	$\theta \in (0, 1]$
Frank	$\frac{-\theta \exp\{-\theta(u+v)\}(\exp\{-\theta\}-1)}{(\exp\{-\theta\}-\exp\{-\theta u\}-\exp\{-\theta v\}+\exp\{-\theta(u+v)\})^2}$	$\theta \in \mathbb{R} - \{0\}$
Clayton	$(1 + \theta)(uv)^{(-1-\theta)}(-1 + u^{-\theta} + v^{-\theta})^{-2-1/\theta}$	$\theta \in (0, \infty)$

Pode ser observado na Tabela 2.1 que, restringindo o espaço de parâmetros das cópulas de Clayton de $\theta \in [-1, \infty) - \{0\}$ para $\theta \in (0, \infty)$, o operador max pode ser desconsiderado.

A estimação de θ pode ser feita por meio de um Estimador de Máxima Verossimilhança (EMV). Neste caso, dado uma amostra aleatória $\{(x_i, y_i), i = 1, \dots, n\}$ de $(X, Y) \sim H$ tal que $(F(X), G(Y)) = (U, V) \sim C$. Seja c a densidade da cópula C , o EMV de $\hat{\theta}$ é dado por:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \log c(F(x_i), G(y_i); \theta). \quad (2.8)$$

As derivadas parciais (com respeito a v) das cópulas de Gumbel-Hougaard, Frank e Clayton

estão apresentadas abaixo.

$$\frac{\partial C^{GH}(u, v)}{\partial v} = \exp \left\{ -[(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}}]^{\theta} \right\} \times \left\{ [(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}}]^{\theta-1} \right\} (-\log v)^{\frac{1}{\theta}-1} \frac{1}{v}; \quad (2.9)$$

$$\frac{\partial C^F(u, v)}{\partial v} = \left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right)^{-1} \times \frac{(\exp(-\theta u) - 1)(\exp(-\theta v))}{\exp(-\theta) - 1} \quad (2.10)$$

$$\frac{\partial C^C(u, v)}{\partial v} = \frac{v^{-\theta-1}}{(u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}+1}} \quad (2.11)$$

As derivadas parciais das cópulas de Gumbel-Hougaard, Frank e Clayton apresentadas anteriormente, serão utilizadas para geração de amostra via algoritmo de transformada inversa, no Capítulo 3.

2.3 Distribuições dos valores extremos

A TVE propõe FDAs assintóticas para o máximo parcial $M_n = \max\{X_j; j = 1, \dots, n\}$ de variáveis aleatórias i.i.d. X_1, X_2, \dots com uma FDA comum $F(\cdot)$. Mais especificamente, se existem $a_n > 0$ e $b_n \in \mathbb{R}$ tais que

$$\lim_{n \rightarrow \infty} P \left(\frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (2.12)$$

onde $G(\cdot)$ é uma função de distribuição não-degenerada.

Existem três classes de FDA $G(\cdot)$ que satisfazem (2.12) (ver, por exemplo, Resnick (2008)).

São elas Fréchet, Weibull reversa e Gumbel. Essas distribuições são conhecidas como distribuições de valores extremos sob normalização linear ou leis estáveis l -max. As condições sobre a função de distribuição $G(\cdot)$ para (2.12) ser verdadeira são bem documentadas e podem ser encontradas em Resnick (2008), Galambos (1987), e Haan e Ferreira (2006). Aplicações da TVE são encontradas em finanças, catástrofes naturais e falhas de equipamentos, entre outros (Embrechts, Klüppelberg e Mikosch, 2013).

Normalizações mais gerais do máximo parcial M_n também podem ser consideradas. Foi demonstrado em Pantcheva (1985) que sob normalização potência existem seis classes de FDAs $H(\cdot)$ tais que

$$\lim_{n \rightarrow \infty} F^n(\alpha_n |x|^{\beta_n} \text{sign}(x)) = H(x), \quad (2.13)$$

onde $\alpha_n, \beta_n > 0$ são sequências de números reais positivos. Essas distribuições são chamadas de distribuições de valores extremos sob normalização potência ou leis estáveis p -max e são conhecidas, respectivamente, como log-Fréchet, log-Weibull, inversa log-Fréchet, inversa log-Weibull, Fréchet padrão e Weibull padrão. A expansão das definições e diversas propriedades das distribuições estáveis l -max para distribuições estáveis p -max ocorre naturalmente. Para uma caracterização completa dessas distribuições, consulte Mohan e Ravi (1993).

2.3.1 Distribuição de Fréchet

Dizemos que uma variável aleatória (v.a.) X possui distribuição de Fréchet, e denotamos por $X \sim \text{Fréchet}(\mu, \sigma, \alpha)$, com parâmetro de locação $\mu \in \mathbb{R}$, parâmetro de escala $\sigma \in \mathbb{R}_+$ e parâmetro de forma $\alpha \in \mathbb{R}_+$, se X possui, respectivamente, uma FDA e uma FDP dadas por

$$F(x; \mu, \sigma, \alpha) = \begin{cases} 0, & x < \mu, \\ \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right)^{-\alpha} \right\}, & x \geq \mu, \end{cases}$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma} \left(\frac{x - \mu}{\sigma} \right)^{-\alpha-1} F(x; \mu, \sigma, \alpha) \mathbb{1}_{(\mu, \infty)}, \quad (2.14)$$

onde $\mathbb{1}_A$ denota a função indicadora do conjunto A . A Figura 2.1 mostra o comportamento da FDP de Fréchet quando os parâmetros variam.

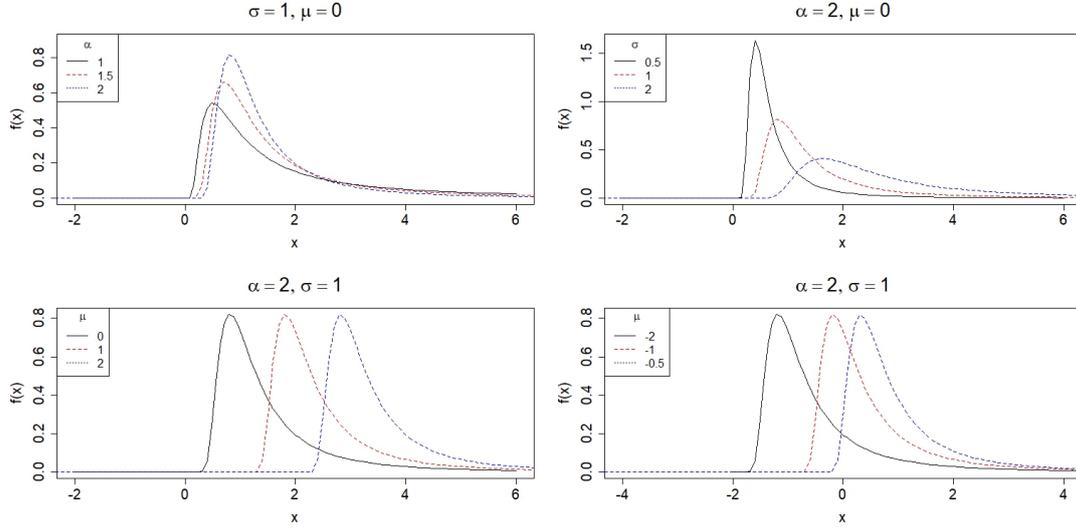


Figura 2.1: Formas da distribuição de Fréchet com parâmetros de localização μ , escala σ e forma α .

2.3.2 Distribuição Weibull reversa

Dizemos que uma v.a. X possui distribuição Weibull reversa e denotamos por $X \sim \text{RWeibull}(\mu, \sigma, \alpha)$, com parâmetro de localização $\mu \in \mathbb{R}$, parâmetro de escala $\sigma \in \mathbb{R}_+$ e parâmetro de forma $\alpha \in \mathbb{R}_+$, se X possui, respectivamente, uma FDA e uma FDP dadas por

$$F(x; \mu, \sigma, \alpha) = \begin{cases} \exp \left\{ - \left(\frac{\mu - x}{\sigma} \right)^\alpha \right\}, & x < \mu, \\ 1, & x \geq \mu, \end{cases}$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma} \left(\frac{\mu - x}{\sigma} \right)^{\alpha-1} F(x; \mu, \sigma, \alpha) \mathbb{1}_{(-\infty, \mu)}(x). \quad (2.15)$$

A Figura 2.2 mostra o comportamento da FDP da Weibull reversa quando os parâmetros variam.

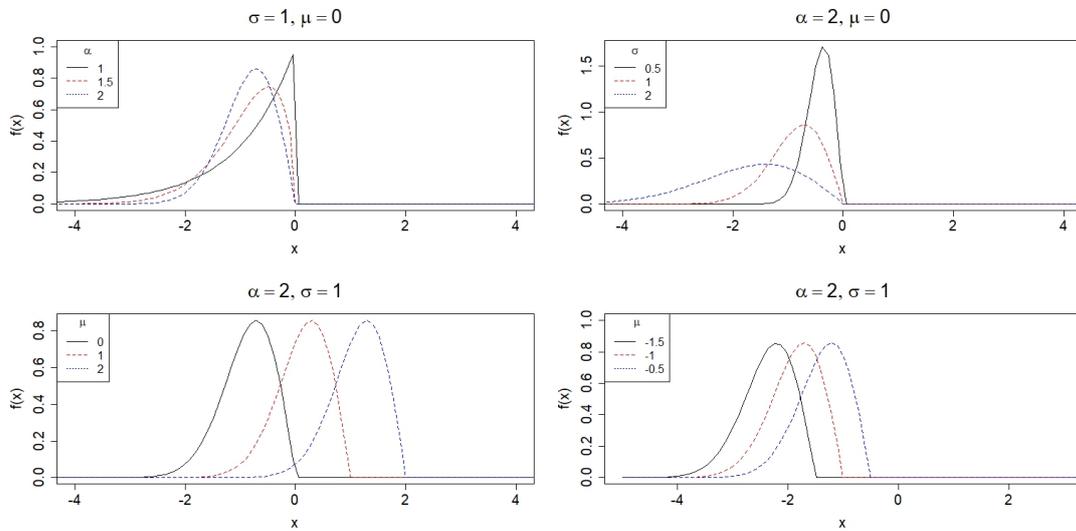


Figura 2.2: Formas da distribuição Weibull reversa com parâmetros de locação μ , escala σ e forma α .

2.3.3 Distribuição de Gumbel

Dizemos que uma v.a. X possui distribuição de Gumbel, e denotamos por $X \sim \text{Gumbel}(\mu, \sigma)$, com parâmetro de locação $\mu \in \mathbb{R}$ e parâmetro de escala $\sigma \in \mathbb{R}_+$, se X possui, respectivamente, uma FDA e uma FDP dadas por

$$F(x; \mu, \sigma, \alpha) = \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \right\}, \quad x \in \mathbb{R},$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{1}{\sigma} \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \right\}, \quad x \in \mathbb{R}.$$

A Figura 2.3 apresenta a FDP da Gumbel geradas utilizando diferentes conjuntos de parâmetros.

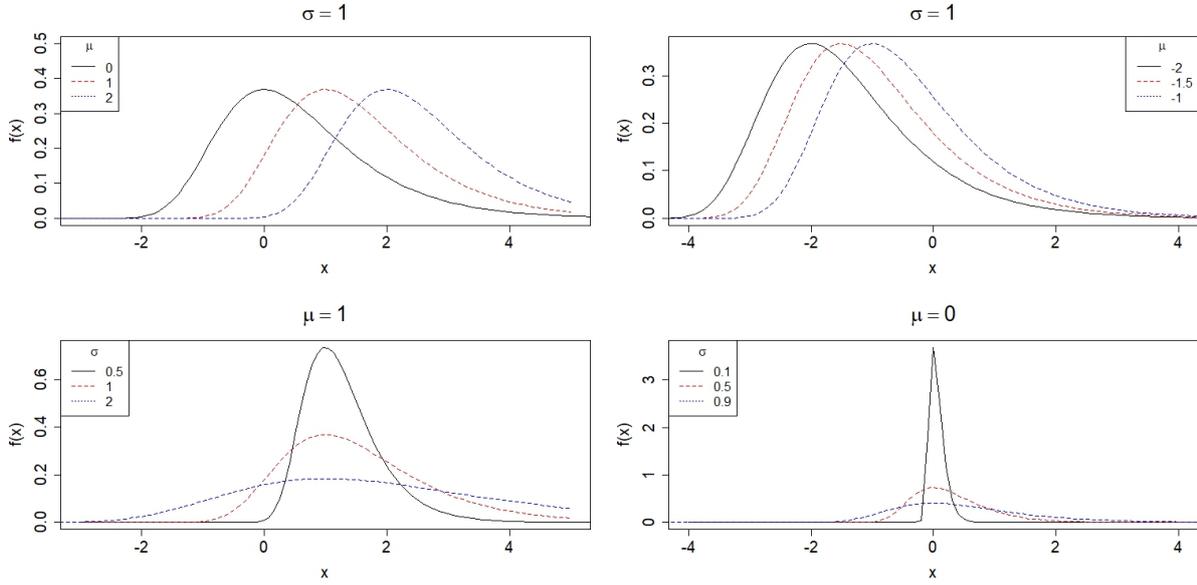


Figura 2.3: Formas da distribuição Gumbel com parâmetros de locação μ e escala σ .

2.3.4 Distribuição GEV

Para aplicações estatísticas, as distribuições l-max-estáveis podem ser resumidas em uma única função de distribuição chamada de distribuição de valores extremos generalizada (GEV, do inglês *Generalized Extreme Value Distribution*). Com a leitura de Embrechts, Klüppelberg e Mikosch (2013), é possível realizar um estudo detalhado sobre a distribuição GEV.

Dizemos que uma v.a. X possui distribuição GEV e denotamos por $X \sim GEV(\mu, \sigma, \gamma)$, $\mu, \gamma \in \mathbb{R}$ e $\sigma > 0$, se X possui, respectivamente, uma FDA e uma FDP dadas por

$$F(x; \mu, \sigma, \gamma) = \exp \left\{ - \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}} \right\}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0, \quad (2.16)$$

onde μ é o parâmetro de locação, σ é o parâmetro de escala e γ é o parâmetro de forma.

$$f(x; \mu, \sigma, \gamma) = F(x, \gamma, \mu, \sigma) \frac{1}{\sigma} \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}-1}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0. \quad (2.17)$$

A Figura 2.4 mostra o comportamento de (2.17) para algumas escolhas de parâmetros.

Observe que o parâmetro de locação desloca a curva, a escala controla a dispersão e a densidade muda de acordo com o sinal da forma.

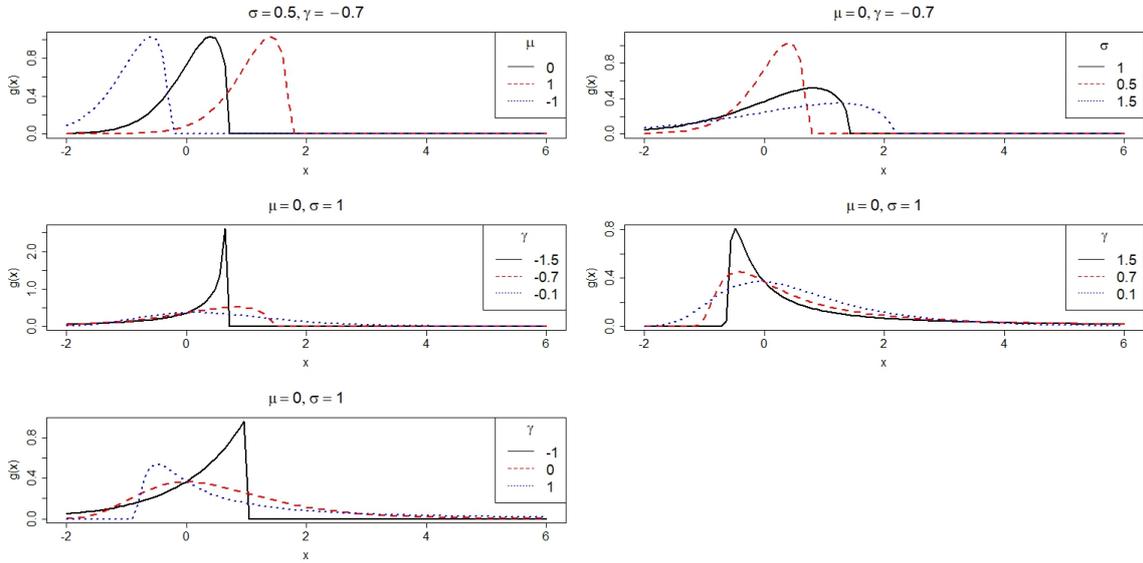


Figura 2.4: Gráficos para a FDP da GEV.

Considere uma amostra aleatória observada $\mathbf{x} = (x_1, x_2, \dots, x_n)$ de uma v.a. X com distribuição GEV. A função de verossimilhança é dada por:

$$L(\mu, \sigma, \gamma; \mathbf{x}) = \prod_{i=1}^n g(x_i; \mu, \sigma, \gamma) \mathbb{1}_{1+\gamma(x_i-\mu)/\sigma > 0}, \quad (2.18)$$

onde $\mathbb{1}_A$ denota a função indicadora do conjunto A . Observe que o suporte de G depende da escolha dos parâmetros (exceto para o caso $\gamma = 0$). Assim, as condições usuais de regularidade não são satisfeitas (ver o trabalho de Smith (1985) para ver condições que garantem a existência do EMV). No entanto, é possível realizar a estimação de máxima verossimilhança, que é formalmente escrita como:

$$(\hat{\mu}, \hat{\sigma}, \hat{\gamma}) = \arg \max_{\mu, \sigma, \gamma} L(\mu, \sigma, \gamma; \mathbf{x}). \quad (2.19)$$

Observe que para $\gamma = 0$ podemos usar o vetor gradiente da função de verossimilhança para

encontrar o estimador. Para o caso $\gamma \neq 0$, essa abordagem não funciona, então são necessários métodos numéricos. Para uma leitura detalhada sobre o assunto, recomendamos ao leitor que consulte Embrechts, Klüppelberg e Mikosch (2013).

2.3.5 Leis p-max estáveis

Caso 1: Distribuição log-Fréchet

Seja X uma v.a. com distribuição $H_1(\alpha_x, \beta_x, \gamma_x)$, onde $\alpha, \beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_1(x; \alpha, \beta, \gamma) = \begin{cases} 0, & \text{se } \gamma x^\beta < 1, \\ \exp\{-(\log(\gamma x^\beta))^{-\alpha}\}, & \text{se } \gamma x^\beta \geq 1, \end{cases}$$

e

$$h_1(x; \alpha, \beta, \gamma) = \exp\{-[\log \gamma x^\beta]^{-\alpha}\} \frac{\alpha\beta}{x} [\log(\gamma x^\beta)]^{-\alpha-1} \mathbb{1}_{(\gamma^{-1/\beta}, \infty)}(\gamma x^\beta),$$

A Figura 2.5 apresenta a FDP de H_1 geradas utilizando diferentes valores de parâmetros.

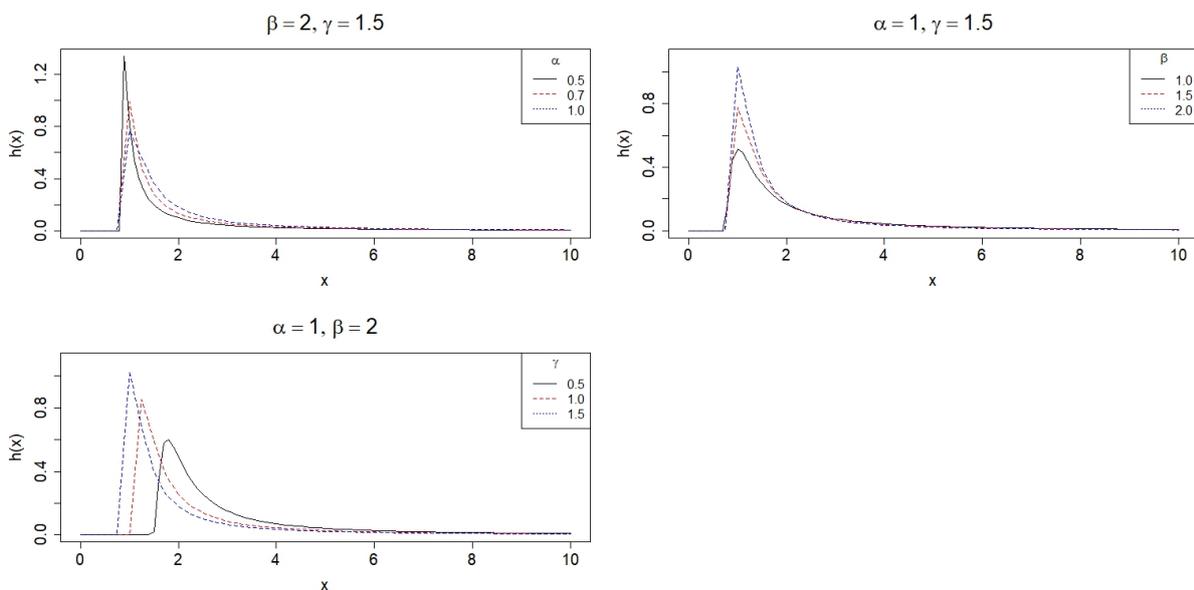


Figura 2.5: Formas de H_1 com parâmetros α, β, γ .

Caso 2: Distribuição log-Weibull

Seja X uma v.a. com distribuição $H_2(\alpha_x, \beta_x, \gamma_x)$, onde $\alpha, \beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_2(x; \alpha) = \begin{cases} 0, & \text{se } \gamma x^\beta < 0, \\ \exp\{-(-\log(\gamma x^\beta))^\alpha\}, & \text{se } 0 \leq \gamma x^\beta < 1, \\ 1, & \text{se } \gamma x^\beta \geq 1, \end{cases}$$

e

$$h_2(x; \alpha, \beta, \gamma) = \exp\{-[-\log(\gamma x^\beta)]^\alpha\} [-\log(\gamma x^\beta)]^{\alpha-1} \frac{\beta\alpha}{x} \mathbb{1}_{(0, \gamma^{-1/\beta})}(\gamma x^\beta),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$.

A Figura 2.6 apresenta a FDP de H_2 geradas utilizando diferentes valores de parâmetros.

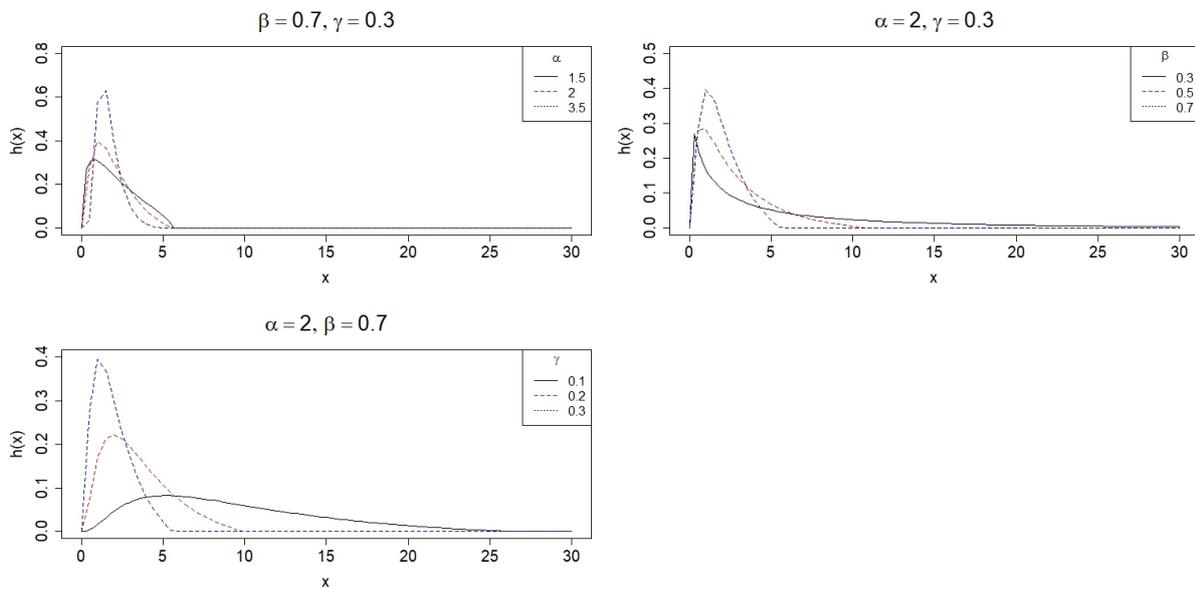


Figura 2.6: Formas de H_2 com parâmetros α, β, γ .

Caso 3: Distribuição log-Fréchet inversa

Seja X uma v.a. com distribuição $H_3(\alpha_x, \beta_x, \gamma_x)$, onde $\alpha, \beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_3(x; \alpha) = \begin{cases} 0, & \text{se } -\gamma(-x)^\beta < -1, \\ \exp\{-(-\log(\gamma(-x)^\beta))^{-\alpha}\}, & \text{se } -1 \leq -\gamma(-x)^\beta < 0, \\ 1, & \text{se } -\gamma(-x)^\beta \geq 0, \end{cases}$$

e

$$h_3(x; \alpha, \beta, \gamma) = \exp\left\{-[-\log(\gamma(-x)^\beta)]^{-\alpha}\right\} \frac{\alpha\beta}{-x} [-\log(\gamma(-x)^\beta)]^{-\alpha-1} \mathbb{1}_{(-1,0)}(-\gamma(-x)^\beta),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$.

A Figura 2.7 apresenta a FDP de H_3 geradas utilizando diferentes valores de parâmetros.

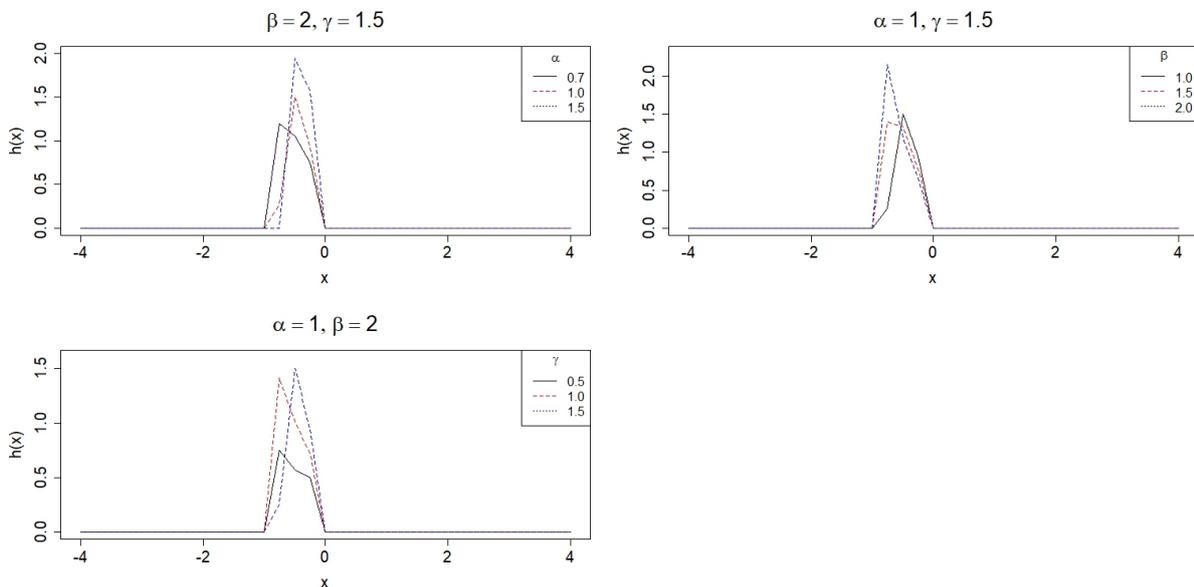


Figura 2.7: Formas de H_3 com parâmetros α, β, γ .

Caso 4: Distribuição log-Weibull inversa

Seja X uma v.a. com distribuição $H_4(\alpha_x, \beta_x, \gamma_x)$, onde $\alpha, \beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_4(x; \alpha) = \begin{cases} \exp\{-\log(\gamma(-x)^\beta)\}^\alpha, & \text{se } -\gamma(-x)^\beta < -1, \\ 1, & \text{se } -\gamma(-x)^\beta \geq -1, \end{cases}$$

e

$$h_4(x; \alpha, \beta, \gamma) = \exp\{-[\log(\gamma(-x)^\beta)]^\alpha\} [\log(\gamma(-x)^\beta)]^{\alpha-1} \frac{\alpha\beta}{-x} \mathbb{1}_{(-\infty, -1)}(-\gamma(-x)^\beta),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$.

A Figura 2.8 apresenta a FDP de H_4 geradas utilizando diferentes valores de parâmetros.

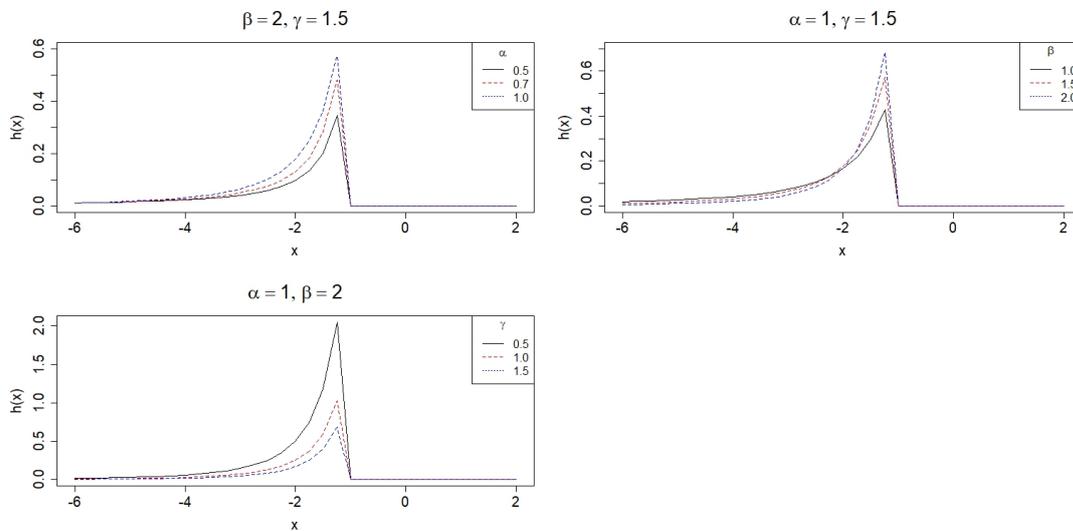


Figura 2.8: Formas de H_4 com parâmetros α, β, γ .

Caso 5: Distribuição Fréchet padrão

Seja X uma v.a. com distribuição $H_5(\beta_x, \gamma_x)$, onde $\beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_5(x) = \begin{cases} 0, & \text{se } \gamma x^\beta < 0, \\ \exp\{-(\gamma x^\beta)^{-1}\} & \text{se } \gamma x^\beta \geq 0, \end{cases}$$

e

$$h_5(x; \beta, \gamma) = \exp\{-(\gamma x^\beta)^{-1}\} \frac{\beta}{\gamma} x^{-\beta-1} \mathbb{1}_{[0, \infty)}(\gamma x^\beta),$$

onde $\beta, \gamma \in \mathbb{R}_+$.

A Figura 2.9 apresenta a FDP de H_5 geradas utilizando diferentes valores de parâmetros.

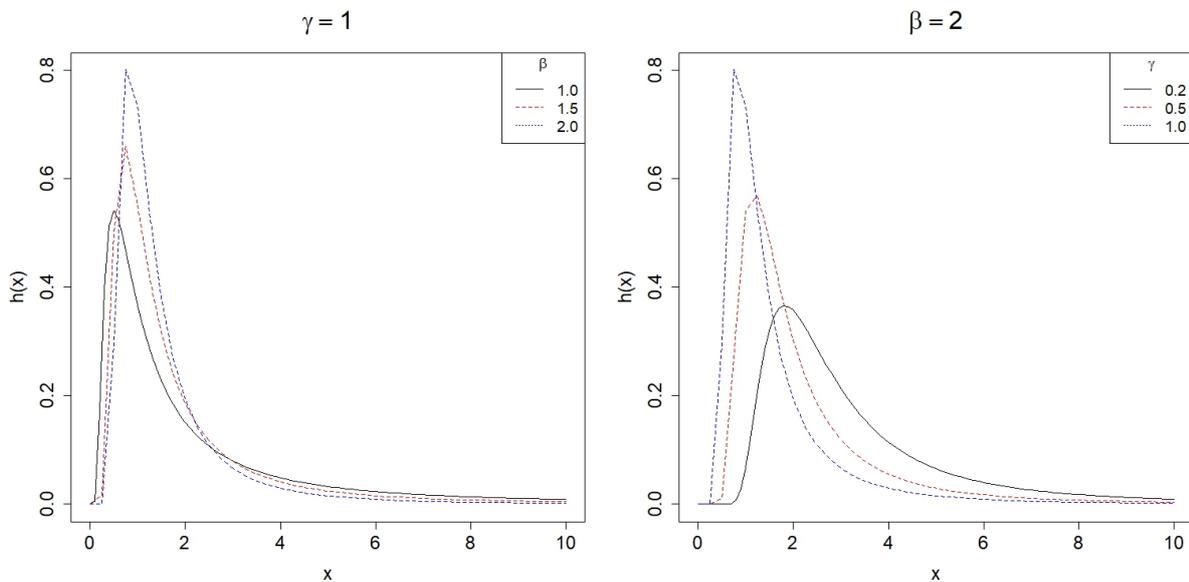


Figura 2.9: Formas de H_5 com parâmetros β, γ .

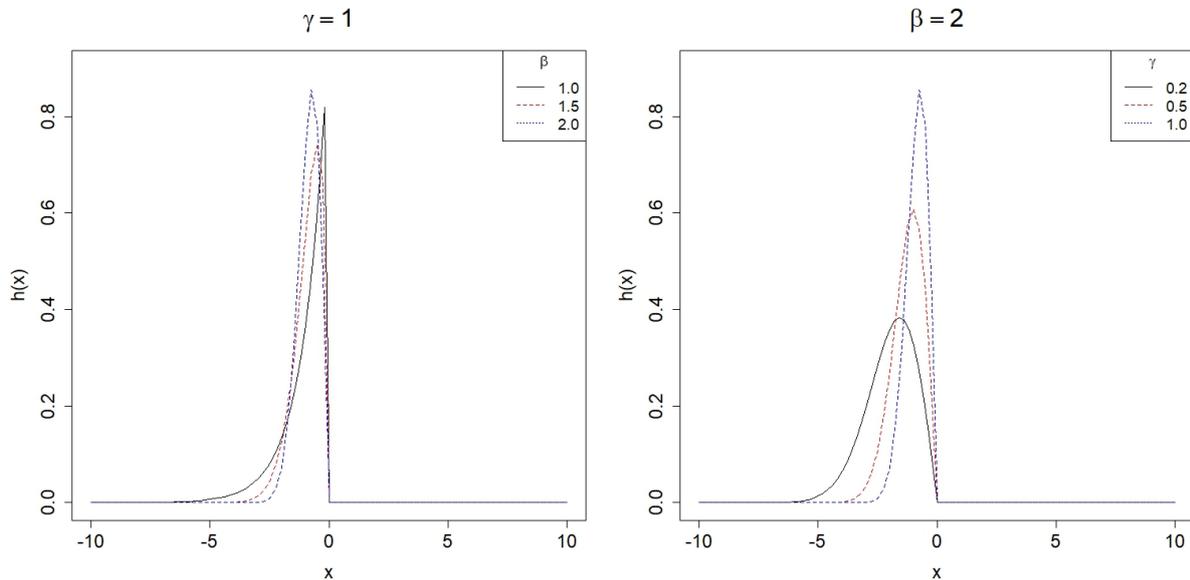
Caso 6: Distribuição Weibull padrão

Seja X uma v.a. com distribuição $H_6(\beta_x, \gamma_x)$, onde $\beta, \gamma \in (0, \infty)$. A FDA e a FDP são dadas, respectivamente, por

$$H_6(x) = \begin{cases} \exp\{-\gamma(-x)^\beta\}, & \text{se } -\gamma(-x)^\beta < 0, \\ 1 & \text{se } -\gamma(-x)^\beta \geq 0, \end{cases}$$

e

$$h_6(x; \beta, \gamma) = \exp\{-\gamma(-x)^\beta\} \gamma \beta (-x)^{\beta-1} \mathbb{1}_{(-\infty, 0)}(-\gamma(-x)^\beta),$$

onde $\beta, \gamma \in \mathbb{R}_+$.A Figura 2.10 apresenta a FDP de H_6 geradas utilizando diferentes valores de parâmetros.**Figura 2.10:** Formas de H_6 com parâmetros β, γ .

2.4 Seleção de Modelos

Nesta seção, apresentaremos o Critério de Informação de Akaike (AIC), o Critério de Informação Bayesiano (BIC) e o Critério de Desvio Esperado (EDC). Esses critérios foram utilizados na Seção 3.4 para a seleção da melhor cópula para representar o conjunto de dados e na Seção 4.3 para a escolha da melhor distribuição de TVE para os conjuntos de dados.

2.4.1 Critério de Informação de Akaike

De acordo com Akaike (1974), o AIC é um método utilizado para a seleção de modelos que minimiza a divergência de Kullback-Leibler. O modelo com o menor valor de AIC é considerado o melhor para representar os dados. A estimativa do AIC para um determinado modelo é dada por:

$$AIC = -2L + 2k \quad (2.20)$$

em que: L é o logaritmo da máxima verossimilhança e k o número de parâmetros.

2.4.2 Critério de Informação Bayesiano

O BIC é um método para a seleção de modelos que penaliza modelos mais complexos para evitar o sobreajuste (Schwarz, 1978). O modelo com o menor valor de BIC é considerado o melhor para representar os dados. A estimativa do BIC para um determinado modelo é dada por:

$$BIC = -2L + k \ln(n) \quad (2.21)$$

em que: L é o logaritmo da máxima verossimilhança, k é o número de parâmetros, e n é o número de observações.

2.4.3 Critério de Desvio Esperado

O EDC é um método utilizado para a seleção de modelos que busca minimizar a divergência esperada entre o modelo estimado e a verdadeira distribuição dos dados (Levy, 1969). O modelo com o menor valor de EDC é considerado o melhor para representar os dados. A estimativa do EDC para um determinado modelo é dada por:

$$EDC = -2L + 2 \log(\log(n))k \quad (2.22)$$

em que: L é o logaritmo da verossimilhança máxima, k é o número de parâmetros do modelo e n é o número de observações.

2.5 Problemas Abordados

A seguir, descrevemos os problemas de pesquisa abordados nessa dissertação:

1. Propor um procedimento de estimativa para $R = P(X < Y)$ quando X e Y seguem distribuição GEV e validar tal procedimento por meio de um estudo de simulação.
2. Aplicar a metodologia proposta em alguns conjuntos de dados reais.
3. Estudar o caso de SSR multicomponente.
4. Apresentar os códigos computacionais necessários para implementar a estrutura metodológica aqui desenvolvida.

Os tópicos 1 e 2 são abordados no Capítulo 3. O Capítulo 4 analisa o tópico 3.

O tópico 4 foi abordado e disponibilizado em Lima et al. (2024b).

Capítulo 3

Avaliando o impacto da seleção de copulas nas medidas de confiabilidade do tipo *stress-strength* com marginais GEV.

3.1 Introdução

Neste capítulo, estudamos a probabilidade de SSR $R = \mathbb{P}(X < Y)$ quando (X, Y) tem cópula associada C de Clayton, Frank ou Gumbel-Hougaard e as marginais seguem distribuição GEV. A FDA e a FDP da distribuição GEV são respectivamente como em (3.1) e (3.2), ou seja,

$$G(x; \mu, \sigma, \gamma) = \exp \left\{ - \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}} \right\}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0. \quad (3.1)$$

e

$$g(x; \mu, \sigma, \gamma) = G(x; \mu, \sigma, \gamma) \frac{1}{\sigma} \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}-1}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0. \quad (3.2)$$

As cópulas de Clayton, Frank e Gumbel-Hougaard são dadas respectivamente por

$$C^C(u, v; \theta) = \max \{ (u^{-\theta} + v^{-\theta} - 1); 0 \}^{-\frac{1}{\theta}}, \quad \theta \in [-1, \infty) - \{0\}. \quad (3.3)$$

$$C^F(u, v; \theta) = -\frac{1}{\theta} \log \left(1 + \frac{(\exp\{-\theta u\} - 1)(\exp\{-\theta v\} - 1)}{\exp\{-\theta\} - 1} \right), \quad \theta \in \mathbb{R} - \{0\}, \quad (3.4)$$

e

$$C^{GH}(u, v; \theta) = \exp \left\{ - \left[(-\log u)^{\frac{1}{\theta}} + (-\log v)^{\frac{1}{\theta}} \right]^{\theta} \right\}, \quad \theta \in (0, 1], \quad (3.5)$$

Na Seção 3.2 apresentamos uma forma de implementação da probabilidade SSR utilizando um método de Monte Carlo para aproximar numericamente as integrais duplas, além de apresentar estimadores para R. Na Seção 3.3, apresentamos a avaliação do estimador proposto por meio de um estudo de simulações. Na Seção 3.4, propomos e validamos a aplicação da metodologia R na modelagem de dados reais. No primeiro caso, avaliamos um indicador para comparar dois ativos financeiros, proporcionando um critério direto a ser utilizado por um investidor na seleção de ativos financeiros. Na segunda aplicação, avaliamos dados provenientes da Pesquisa do Banco de dados da Itália sobre Renda e Consumo das Famílias no ano de 2008 para modelagem de fragilidade econômica. Na terceira aplicação, analisamos os fluxos mínimos mensais de água.

Por fim, destacamos que os resultados apresentados nesse capítulo foram publicados em Lima et al. (2024a).

3.2 Estimação da probabilidade de stress-strength

Seja (X, Y) um vetor aleatório com distribuição H, que satisfaz a Equação (2.4) e seja $f_{X,Y}$ a FDP de H, ou seja,

$$f_{X,Y}(x, y) = c(F(x), G(y); \theta) f(x) g(y), \quad (3.6)$$

onde as FDP marginais são $f(x) = F'(x)$ e $g(y) = G'(y)$, respectivamente, e θ é o parâmetro de associação. Podemos escrever a probabilidade de SSR como em Nelsen (2006) e Rathie, Ozelim e Andrade (2021):

$$\begin{aligned}
 R &= P(X < Y) = \int_{-\infty}^{\infty} \int_{-\infty}^y c(F(x), G(y); \theta) f(x)g(y) dx dy \\
 &= \int_0^1 \int_0^{F(G^{-1}(v))} c(u, v; \theta) du dv.
 \end{aligned} \tag{3.7}$$

Assumindo que $X \sim GEV(\mu_x, \sigma_x, \gamma_x)$ e $Y \sim GEV(\mu_y, \sigma_y, \gamma_y)$, temos

$$R = R(\mu_x, \sigma_x, \gamma_x, \mu_y, \sigma_y, \gamma_y, \theta).$$

Como R é uma função integral (mensurável), podemos estima-la através da propriedade de invariância dos EMVs por meio de estimativas $(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\theta})$, isto é,

$$\hat{R} = R(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\theta}). \tag{3.8}$$

Como indicado no Capítulo 2, consideramos o método de estimação, onde os parâmetros das distribuições marginais são estimados separadamente dos parâmetros da cópula. Pode-se observar que (3.7) e (3.8) dependem de integrais da função densidade da cópula.

Nestes casos, algoritmos de integração numérica podem ser prontamente utilizados para avaliar adequadamente as integrais envolvidas. No presente trabalho, foi considerada uma abordagem de integração por Monte Carlo. Desta forma, primeiro geramos pontos aleatórios uniformes v_1, \dots, v_k em $[0, 1]$, e em seguida pontos aleatórios uniformes u_j são gerados em $[0, F(G^{-1}(v_j))]$, para cada $j = 1, \dots, k$. Finalmente, estimamos \hat{R} por:

$$\hat{R} \approx \frac{1}{k} \sum_{j=1}^k c(u_j, v_j; \hat{\theta}) \times [1 - 0] \times [F(G^{-1}(v_j); \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y); \hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x) - 0]. \tag{3.9}$$

Nas próximas seções, compararemos os resultados de \hat{R} com um estimador não paramétrico

denotado como \hat{R}_{NP} , que é definido como:

$$\hat{R}_{NP} = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j \leq y_j\}}, \quad (3.10)$$

onde $\mathbb{1}_A$ denota a função indicadora do conjunto A e n é o tamanho da amostra.

O Algoritmo 3.2.1 descreve a abordagem usada na Seção 3.4 para obter intervalos de confiança Bootstrap das estimativas de R .

Algoritmo 3.2.1. Seja (\mathbf{x}, \mathbf{y}) uma amostra de tamanho n de $(X, Y) \sim H$. Considere M um inteiro positivo que denota o número de repetições Bootstrap.

Passo 1 Gerar amostras Bootstrap $(\mathbf{x}, \mathbf{y})_0$

Passo 2 Calcular as estimativas $(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\theta})$ com base em $(\mathbf{x}, \mathbf{y})_0$. Neste caso, os parâmetros de cada distribuição marginal são estimados individualmente usando (2.19), e então os vetores transformados uniformes correspondentes são considerados em (2.8) para estimar $\hat{\theta}$.

Passo 3 Obter $\hat{R}_0 = \hat{R}(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\theta})$.

Passo 4 Repetir os passos 1 a 3 M vezes.

Passo 5 O intervalo de confiança (IC) aproximado de $100(1 - \alpha)\%$ de \hat{R} é dado por $[\hat{R}_M(\alpha/2), \hat{R}_M(1 - \alpha/2)]$, onde $\hat{R}_M(\alpha) \approx \hat{G}^{-1}(\alpha)$ e \hat{G} é a função de distribuição empírica de \hat{R} .

3.3 Estudo de Simulações

Para avaliar o desempenho das estimativas \hat{R} e \hat{R}_{NP} , simulamos amostras aleatórias das cópulas fornecidas em (3.3), (3.4) e (3.5) com marginais GEV. As amostras aleatórias foram

simuladas usando as distribuições condicionais dos vetores aleatórios $\mathbf{U} = \{U_1, U_2\}$. No caso bivariado, seguimos os passos descritos em Mai e Scherer (2017).

Dada uma cópula C , calculamos a função

$$F_{U_1|U_2=u_2}(u_1) := \frac{\partial}{\partial u_2} C(u_1, u_2)|_{U_2=u_2}, \quad u_1 \in [0, 1]. \quad (3.11)$$

As derivadas parciais de cada cópula foram apresentadas na Seção 2.2. Essa função representa $P(U_1 \leq u_1 | U_2 = u_2)$. Em seguida, calculamos a inversa generalizada

$$F_{U_1|U_2=u_2}^{-1}(v) := \inf\{u_1 > 0; F_{U_1|U_2=u_2}(u_1) \geq v\}, \quad v \in (0, 1). \quad (3.12)$$

A Equação (3.12) foi obtida numericamente. Por fim, o seguinte algoritmo de geração de valores aleatório foi utilizado:

- Algoritmo 3.3.1.**
1. Obter uma amostra u_2 de $U_2 \sim U[0, 1]$, ou seja, uniformemente distribuída em $[0, 1]$.
 2. Obter uma amostra v de $V \sim U[0, 1]$, independente de U_2 .
 3. Definir $u := F_{U_1|U_2=u_2}^{-1}(v)$ e considerar (u, u_2) como o vetor aleatório da cópula C .

Os Passos 1-3 descritos acima nos fornecem um vetor aleatório com distribuições marginais uniformes (Mai e Scherer, 2017). Podemos gerar um vetor aleatório com distribuição C e marginais $GEV(\mu_x, \sigma_x, \gamma_x)$ e $GEV(\mu_y, \sigma_y, \gamma_y)$ da seguinte forma:

$$(G^{-1}(u; \mu_x, \sigma_x, \gamma_x), G^{-1}(u_2; \mu_y, \sigma_y, \gamma_y)).$$

onde $G(\cdot)$ denota a FDA de variáveis aleatórias GEV dada em (3.1).

Os valores de $\mu_x, \sigma_x, \gamma_x, \mu_y, \sigma_y, \gamma_y, \theta$ e o tamanho da amostra n são pré-especificados. Simulações de Monte Carlo foram implementadas em R (R Core Team, 2023) com $M = 100$ replicas.

Para comparar as estimativas, calculamos a média \hat{R}_{MC} das 100 amostras Bootstrap, o Viés e o Raiz do Erro Quadrático Médio (REQM) em comparação com o verdadeiro R .

O único pacote R utilizado em nosso estudo foi o *extRemes* para usar a FDA e FDP da distribuição marginal GEV, através da função *fevd*. Após o ajuste das marginais, estimamos o parâmetro de dependência θ usando a Equação (2.8). Todos os outros algoritmos utilizados neste estudo foram programados com base nos procedimentos delineados neste trabalho. Para permitir que os leitores apliquem a metodologia aqui proposta, os códigos estão disponíveis em um repositório público (Lima et al., 2024b).

Nas Tabelas 3.1, 3.2 e 3.3 apresentam o desempenho dos estimadores \hat{R}_{MC} e \hat{R}_{NP} , para as cópulas de Gumbel-Hougaard, Frank e Clayton com marginais GEV, respectivamente. Em geral, os estimadores \hat{R}_{NP} parecem ser melhores do que \hat{R}_{MC} , apresentando Viés e REQM mais baixos.

Em resumo, as Tabelas 3.1, 3.2 e 3.3 fornecem uma análise detalhada da precisão e do viés das estimativas de R utilizando simulações de Monte Carlo e abordagens não paramétricas para diferentes parâmetros das cópulas de Gumbel-Hougaard, Frank e Clayton e marginais GEV. Isso ajuda a entender a robustez das estimativas em diferentes cenários de dependência e formas das distribuições marginais.

De modo geral as simulações para as cópulas de Gumbel-Hougaard, Frank e Clayton foram satisfatórias, ou seja, tanto \hat{R}_{MC} quanto \hat{R}_{NP} estão próximas do valor teórico, indicando um bom desempenho das simulações de Monte Carlo e da abordagem não paramétrica. O viés e o REQM para \hat{R}_{MC} e \hat{R}_{NP} mostram que a estimativa não paramétrica muitas vezes tem um viés menor e um REQM menor, indicando que pode ser uma abordagem mais robusta em alguns casos.

Tabela 3.1: Simulações de Monte Carlo para a cópula de Gumbel-Hougaard com marginais GEV e $n = 100$.

μ_1	σ_1	γ_1	μ_2	σ_2	γ_2	θ	R	\hat{R}_{MC}	Viés	REQM	\hat{R}_{NP}	Viés _{NP}	REQM _{NP}
0.0	0.7	-0.3	0.0	1.0	-1.0	0.3	0.3600	0.4047	0.0447	0.0692	0.3679	0.0079	0.0470
0.0	0.9	-0.3	0.0	1.5	-1.0	0.3	0.4306	0.4261	-0.0045	0.1393	0.4362	0.0056	0.0422
0.0	0.7	-1.0	0.0	1.0	-1.0	0.3	0.6361	0.6194	-0.0166	0.0719	0.6511	0.0150	0.0528
0.0	0.9	-1.0	0.0	1.5	-1.0	0.3	0.6431	0.5814	-0.0617	0.1616	0.6571	0.0140	0.0454
0.0	0.5	-1.5	0.0	1.5	-1.5	0.3	0.6587	0.6149	-0.0439	0.0874	0.6621	0.0034	0.0474
0.0	0.7	-1.5	0.0	1.0	-1.5	0.3	0.6596	0.5925	-0.0672	0.1306	0.6723	0.0127	0.0498
0.0	0.9	-1.5	0.0	1.5	-1.5	0.3	0.6649	0.6099	-0.0549	0.1118	0.6739	0.0090	0.0490
0.0	0.7	-0.3	0.0	1.0	-1.0	0.5	0.4030	0.4207	0.0178	0.0576	0.3953	-0.0077	0.0496
0.0	0.9	-0.3	0.0	1.5	-1.0	0.5	0.4450	0.4182	-0.0268	0.1303	0.4361	-0.0089	0.0578
0.0	0.7	-1.0	0.0	1.0	-1.0	0.5	0.6107	0.5984	-0.0123	0.0785	0.6314	0.0207	0.0551
0.0	0.9	-1.0	0.0	1.5	-1.0	0.5	0.6202	0.5666	-0.0537	0.1595	0.6386	0.0184	0.0511
0.0	0.5	-1.5	0.0	1.5	-1.5	0.5	0.6563	0.6135	-0.0428	0.1119	0.6588	0.0025	0.0414
0.0	0.7	-1.5	0.0	1.0	-1.5	0.5	0.6434	0.6013	-0.0421	0.0949	0.6486	0.0052	0.0490
0.0	0.9	-1.5	0.0	1.5	-1.5	0.5	0.6448	0.6062	-0.0386	0.0978	0.6567	0.0119	0.0453
0.0	0.7	-0.3	0.0	1.0	-1.0	0.7	0.4287	0.4492	0.0204	0.0463	0.4316	0.0029	0.0510
0.0	0.9	-0.3	0.0	1.5	-1.0	0.7	0.4647	0.4229	-0.0417	0.1646	0.4654	0.0007	0.0530
0.0	0.7	-1.0	0.0	1.0	-1.0	0.7	0.5896	0.5652	-0.0244	0.0921	0.5892	-0.0004	0.0545
0.0	0.9	-1.0	0.0	1.5	-1.0	0.7	0.6011	0.5448	-0.0563	0.1655	0.6129	0.0118	0.0498
0.0	0.5	-1.5	0.0	1.5	-1.5	0.7	0.6456	0.6072	-0.0384	0.1314	0.6568	0.0112	0.0467
0.0	0.7	-1.5	0.0	1.0	-1.5	0.7	0.6186	0.5836	-0.0350	0.1159	0.6244	0.0058	0.0472
0.0	0.9	-1.5	0.0	1.5	-1.5	0.7	0.6217	0.5930	-0.0287	0.0905	0.6255	0.0038	0.0410
0.0	0.7	-0.3	0.0	1.0	-1.0	0.9	0.4542	0.4577	0.0036	0.0575	0.4529	-0.0013	0.0515
0.0	0.9	-0.3	0.0	1.5	-1.0	0.9	0.4858	0.4275	-0.0583	0.1587	0.4819	-0.0039	0.0485
0.0	0.7	-1.0	0.0	1.0	-1.0	0.9	0.5701	0.5676	-0.0025	0.0703	0.5731	0.0030	0.0470
0.0	0.9	-1.0	0.0	1.5	-1.0	0.9	0.5851	0.5332	-0.0519	0.1605	0.5934	0.0083	0.0525
0.0	0.5	-1.5	0.0	1.5	-1.5	0.9	0.6336	0.5908	-0.0428	0.1371	0.6374	0.0038	0.0531
0.0	0.7	-1.5	0.0	1.0	-1.5	0.9	0.5965	0.5809	-0.0156	0.0938	0.6054	0.0089	0.0474
0.0	0.9	-1.5	0.0	1.5	-1.5	0.9	0.6028	0.5805	-0.0224	0.1436	0.6223	0.0195	0.0520

Tabela 3.2: Simulações de Monte Carlo para a cópula de Frank com marginais GEV e $n = 100$.

μ_1	σ_1	γ_1	μ_2	σ_2	γ_2	θ	R	\hat{R}_{MC}	Viés	REQM	\hat{R}_{NP}	Viés _{NP}	REQM _{NP}
1.0	0.7	-0.3	1.0	1.0	-1.0	-5.0	0.4960	0.5029	0.0069	0.0643	0.5041	0.0081	0.0579
1.0	0.9	-0.3	1.0	1.5	-1.0	-5.0	0.5259	0.4598	-0.0661	0.1574	0.5269	0.0010	0.0486
1.0	0.7	-1.0	1.0	1.0	-1.0	-5.0	0.5284	0.5288	0.0004	0.0530	0.5280	-0.0004	0.0532
1.0	0.9	-1.0	1.0	1.5	-1.0	-5.0	0.5458	0.4819	-0.0639	0.1721	0.5467	0.0009	0.0492
1.0	0.5	-1.5	1.0	1.5	-1.5	-5.0	0.5976	0.5562	-0.0415	0.1462	0.6017	0.0041	0.0573
1.0	0.7	-1.5	1.0	1.0	-1.5	-5.0	0.5493	0.5159	-0.0334	0.1199	0.5425	-0.0068	0.0492
1.0	0.9	-1.5	1.0	1.5	-1.5	-5.0	0.5628	0.5517	-0.0111	0.1386	0.5641	0.0013	0.0495
1.0	0.7	-0.3	1.0	1.0	-1.0	-0.9	0.4740	0.4688	-0.0052	0.0587	0.4729	-0.0011	0.0546
1.0	0.9	-0.3	1.0	1.5	-1.0	-0.9	0.5038	0.4440	-0.0598	0.1535	0.4897	-0.0141	0.0553
1.0	0.7	-1.0	1.0	1.0	-1.0	-0.9	0.5525	0.5575	0.0050	0.0608	0.5605	0.0080	0.0522
1.0	0.9	-1.0	1.0	1.5	-1.0	-0.9	0.5698	0.5059	-0.0640	0.1782	0.5734	0.0036	0.0537
1.0	0.5	-1.5	1.0	1.5	-1.5	-0.9	0.6211	0.5846	-0.0366	0.1337	0.6281	0.0070	0.0515
1.0	0.7	-1.5	1.0	1.0	-1.5	-0.9	0.5771	0.5534	-0.0237	0.1135	0.5721	-0.0050	0.0576
1.0	0.9	-1.5	1.0	1.5	-1.5	-0.9	0.5866	0.5498	-0.0368	0.1328	0.5864	-0.0002	0.0480
1.0	0.7	-0.3	1.0	1.0	-1.0	0.9	0.4569	0.4509	-0.0060	0.0580	0.4495	-0.0074	0.0493
1.0	0.9	-0.3	1.0	1.5	-1.0	0.9	0.4883	0.4260	-0.0623	0.1598	0.4774	-0.0109	0.0529
1.0	0.7	-1.0	1.0	1.0	-1.0	0.9	0.5690	0.5775	0.0085	0.0629	0.5711	0.0021	0.0512
1.0	0.9	-1.0	1.0	1.5	-1.0	0.9	0.5863	0.5207	-0.0656	0.1805	0.5876	0.0013	0.0463
1.0	0.5	-1.5	1.0	1.5	-1.5	0.9	0.6345	0.6002	-0.0343	0.1103	0.6277	-0.0068	0.0475
1.0	0.7	-1.5	1.0	1.0	-1.5	0.9	0.5966	0.5579	-0.0387	0.1443	0.6015	0.0049	0.0559
1.0	0.9	-1.5	1.0	1.5	-1.5	0.9	0.6036	0.5597	-0.0439	0.1574	0.6035	-0.0001	0.0522
1.0	0.7	-0.3	1.0	1.0	-1.0	5.0	0.4166	0.4119	-0.0047	0.0542	0.4117	-0.0049	0.0506
1.0	0.9	-0.3	1.0	1.5	-1.0	5.0	0.4537	0.4006	-0.0530	0.1665	0.4530	-0.0007	0.0500
1.0	0.7	-1.0	1.0	1.0	-1.0	5.0	0.6046	0.6087	0.0041	0.0715	0.6041	-0.0005	0.0488
1.0	0.9	-1.0	1.0	1.5	-1.0	5.0	0.6278	0.5706	-0.0572	0.1869	0.6300	0.0022	0.0504
1.0	0.5	-1.5	1.0	1.5	-1.5	5.0	0.6550	0.6224	-0.0326	0.1229	0.6584	0.0034	0.0499
1.0	0.7	-1.5	1.0	1.0	-1.5	5.0	0.6388	0.5944	-0.0445	0.1267	0.6333	-0.0055	0.0485
1.0	0.9	-1.5	1.0	1.5	-1.5	5.0	0.6430	0.6180	-0.0250	0.1399	0.6495	0.0065	0.0497

Tabela 3.3: Simulações de Monte Carlo para a cópula de Clayton com marginais GEV e $n = 100$.

μ_1	σ_1	γ_1	μ_2	σ_2	γ_2	θ	R	\hat{R}_{MC}	Viés	REQM	\hat{R}_{NP}	Viés _{NP}	REQM _{NP}
0.0	0.7	-0.3	0.0	1.0	-1.0	1.5	0.4048	0.4014	-0.0034	0.0462	0.4000	-0.0048	0.0512
0.0	0.9	-0.3	0.0	1.5	-1.0	1.5	0.4434	0.4152	-0.0282	0.1164	0.4405	-0.0029	0.0447
0.0	0.7	-1.0	0.0	1.0	-1.0	1.5	0.5662	0.5800	0.0138	0.0609	0.5641	-0.0021	0.0453
0.0	0.9	-1.0	0.0	1.5	-1.0	1.5	0.5836	0.5271	-0.0565	0.1881	0.5973	0.0137	0.0521
0.0	0.5	-1.5	0.0	1.5	-1.5	1.5	0.6265	0.5876	-0.0389	0.1395	0.6321	0.0056	0.0465
0.0	0.7	-1.5	0.0	1.0	-1.5	1.5	0.6029	0.5729	-0.0301	0.1322	0.6084	0.0055	0.0525
0.0	0.9	-1.5	0.0	1.5	-1.5	1.5	0.6125	0.5716	-0.0409	0.1332	0.6068	-0.0057	0.0482
0.0	0.7	-0.3	0.0	1.0	-1.0	2.5	0.3794	0.3747	-0.0048	0.0575	0.3754	-0.0040	0.0438
0.0	0.9	-0.3	0.0	1.5	-1.0	2.5	0.4244	0.3782	-0.0462	0.1665	0.4301	0.0057	0.0519
0.0	0.7	-1.0	0.0	1.0	-1.0	2.5	0.5642	0.5892	0.0250	0.0768	0.5682	0.0040	0.0489
0.0	0.9	-1.0	0.0	1.5	-1.0	2.5	0.5832	0.5486	-0.0346	0.1673	0.5933	0.0101	0.0409
0.0	0.5	-1.5	0.0	1.5	-1.5	2.5	0.6234	0.5658	-0.0576	0.1758	0.6281	0.0047	0.0493
0.0	0.7	-1.5	0.0	1.0	-1.5	2.5	0.6042	0.5658	-0.0384	0.1222	0.6004	-0.0038	0.0463
0.0	0.9	-1.5	0.0	1.5	-1.5	2.5	0.6076	0.5885	-0.0191	0.1359	0.6230	0.0154	0.0512
0.0	0.7	-0.3	0.0	1.0	-1.0	2.8	0.3734	0.3695	-0.0039	0.0574	0.3700	-0.0034	0.0444
0.0	0.9	-0.3	0.0	1.5	-1.0	2.8	0.4214	0.3682	-0.0531	0.1485	0.4079	-0.0135	0.0504
0.0	0.7	-1.0	0.0	1.0	-1.0	2.8	0.5639	0.5838	0.0199	0.1003	0.5747	0.0108	0.0484
0.0	0.9	-1.0	0.0	1.5	-1.0	2.8	0.5835	0.5418	-0.0416	0.1724	0.5839	0.0004	0.0503
0.0	0.5	-1.5	0.0	1.5	-1.5	2.8	0.6228	0.5723	-0.0505	0.1544	0.6174	-0.0054	0.0437
0.0	0.7	-1.5	0.0	1.0	-1.5	2.8	0.6046	0.5488	-0.0558	0.1691	0.5975	-0.0071	0.0539
0.0	0.9	-1.5	0.0	1.5	-1.5	2.8	0.6077	0.5685	-0.0392	0.1564	0.6056	-0.0021	0.0583
0.0	0.7	-0.3	0.0	1.0	-1.0	3.0	0.3697	0.3664	-0.0033	0.0574	0.3653	-0.0044	0.0443
0.0	0.9	-0.3	0.0	1.5	-1.0	3.0	0.4197	0.3980	-0.0217	0.1506	0.4130	-0.0067	0.0487
0.0	0.7	-1.0	0.0	1.0	-1.0	3.0	0.5637	0.5983	0.0346	0.0835	0.5782	0.0145	0.0538
0.0	0.9	-1.0	0.0	1.5	-1.0	3.0	0.5838	0.5442	-0.0396	0.1734	0.5896	0.0058	0.0514
0.0	0.5	-1.5	0.0	1.5	-1.5	3.0	0.6225	0.5798	-0.0427	0.1566	0.6219	-0.0006	0.0484
0.0	0.7	-1.5	0.0	1.0	-1.5	3.0	0.6049	0.5769	-0.0281	0.0969	0.6046	-0.0003	0.0464
0.0	0.9	-1.5	0.0	1.5	-1.5	3.0	0.6079	0.5788	-0.0291	0.1174	0.6092	0.0013	0.0438

3.4 Aplicações

Nesta seção discutimos três aplicações usando dados reais. Estudamos a validade do modelo para todos os casos e mostramos que a cópula de Gumbel-Hougaard, Frank ou Clayton com marginais GEV se ajustam adequadamente aos conjuntos de dados considerados.

3.4.1 Seleção de ativos financeiros

O uso de métricas do tipo $P(X < Y)$ pode servir como um guia para selecionar ativos financeiros ao gerenciar uma carteira. Em vez de depender do método tradicional de comparar

os valores esperados de X e Y (de acordo com a teoria moderna da carteira), exploramos o uso de uma medida de confiabilidade $P(X < Y)$ como um parâmetro alternativo para avaliar log-retornos.

Os log-retornos são usados para medir a variação percentual dos preços das ações de um dia para o outro, são definidos como a diferença logarítmica entre o preço de fechamento de um dia e o preço de fechamento do dia anterior. Matematicamente, o log-retorno R_t de um ativo no tempo t é dado por:

$$R_t = \log X_t - \log X_{t-1},$$

onde X_t e X_{t-1} são os preços das ações nos tempos t e $t - 1$, respectivamente.

Comparamos os log-retornos dos preços das ações¹ Banco do Brasil SA (BBAS3), Itaú Unibanco (ITUB4), Grupo Ultra (UGPA3), Petroleo Brasileiro SA Petrobras (PETR3), Vale S.A. (VALE3), VIA S.A (VIAA3) e Magazine Luiza (MGLU3). Essas ações são de empresas que representam uma ampla variedade de setores econômicos, a saber: bancário (BBAS3, ITUB4), extração e comercialização de gás e petróleo (UGPA3, PETR3), mineração (VALE3) e varejo (VIAA3 e MGLU3). Consideramos os preços de fechamento diários entre 01-01-2022 e 30-04-2023 obtidos do site Yahoo! Finance. Estatísticas descritivas para os conjuntos de dados são apresentadas na Tabela 3.4.

A Figura 3.1 apresenta os boxplots dos log-retornos, mostrando certa simetria dos log-retornos em torno de zero e maior variabilidade para MGLU3 e VIAA3. A estrutura de dependência entre os log-retornos dos conjuntos de dados é mostrada na Figura 3.2.

Tabela 3.4: Estatísticas resumidas para os log-retornos dos preços das ações ($n = 331$).

Variáveis	Min.	Q_1	Mediana	Média	Q_3	Max.	Desvio Padrão	Assimetria	Curtose
BBAS3	-0.1057	-0.0097	0.0019	0.0012	0.0136	0.0736	0.0204	-0.3452	5.7413
ITUB4	-0.0492	-0.0105	0.0004	0.0006	0.0109	0.0794	0.0172	0.3809	4.4864
UGPA3	-0.0802	-0.0169	-0.0023	0.0001	0.0158	0.0771	0.0252	0.0306	3.0732
PETR3	-0.1270	-0.0136	0.0007	-0.0005	0.0159	0.0849	0.0280	-1.0420	6.3563
VALE3	-0.0689	-0.0140	0.0001	-0.0002	0.0128	0.0989	0.0231	0.4092	4.5967
VIAA3	-0.1075	-0.0344	-0.0059	-0.0030	0.0231	0.1504	0.0447	0.6144	3.6044
MGLU3	-0.1435	-0.0329	-0.0043	-0.0021	0.0284	0.1635	0.0502	0.1849	3.1167

¹Negociados na BOVESPA, Bolsa de Valores de São Paulo.

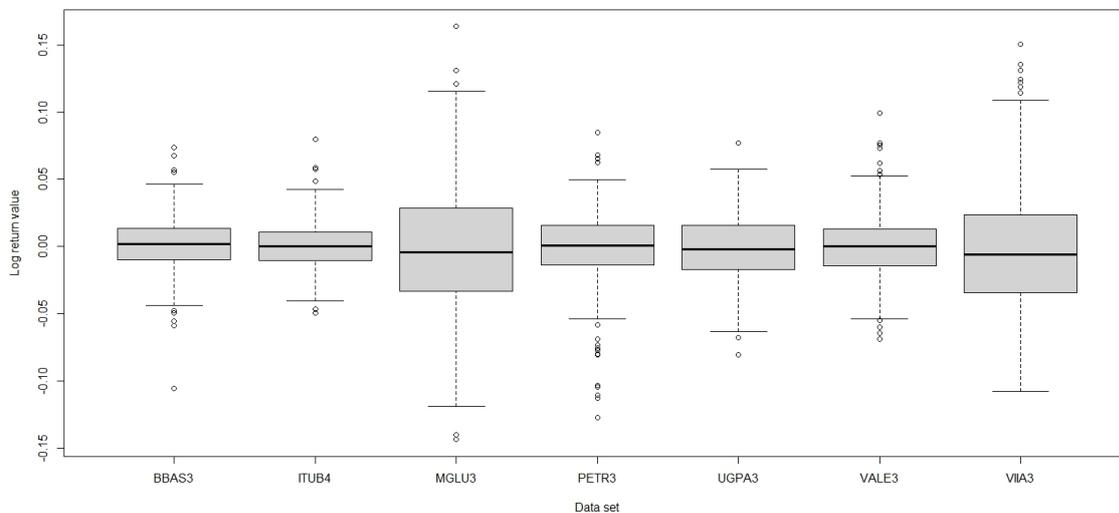


Figura 3.1: Boxplot para log-retornos.

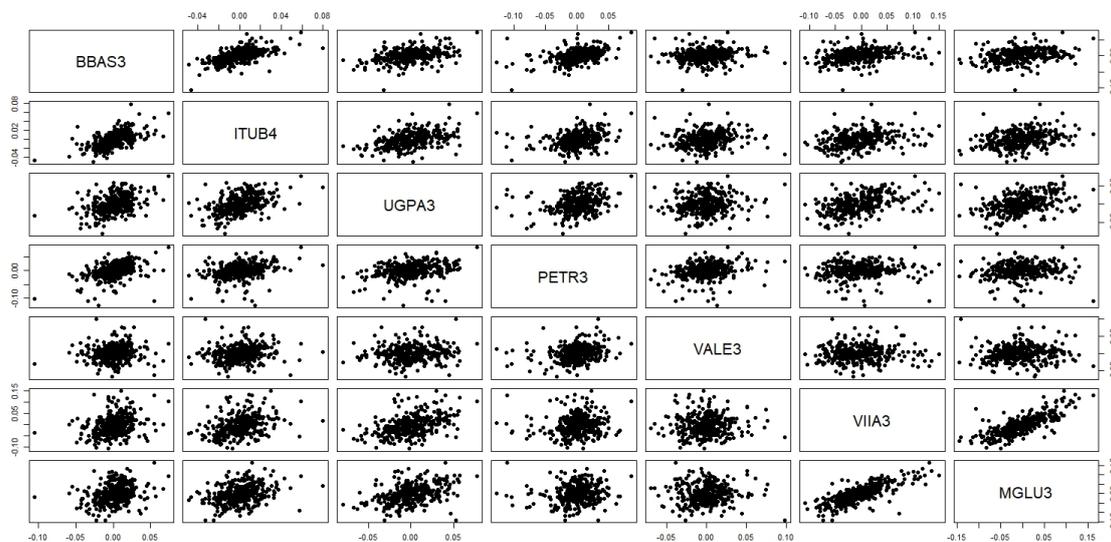


Figura 3.2: Grafico de dispersão dos log-retornos.

Os parâmetros estimados para a distribuição GEV são apresentados na Tabela 3.5, assim como o p-valor do teste de Kolmogorov-Smirnov (KS). As Figuras 3.3 a 3.6 mostram os ajustes GEV para os conjuntos de dados. Note que, apesar do teste KS rejeitar o ajuste GEV para os dados de BBAS3 e PETR3, uma análise gráfica do histograma e da função de distribuição

empírica (ECDF) não necessariamente descartaria a adequação da distribuição GEV para esses conjuntos de dados. Também é importante observar que o teste KS se torna muito sensível até mesmo a pequenas discrepâncias para tamanhos de amostra médios a grandes, o que explica essa questão.

Tabela 3.5: Parâmetros estimados da distribuição GEV e o p-valor do teste de Kolmogorov-Smirnov.

Dados	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	p-valor KS
BBAS3	-0.01	0.02	-0.25	0.01
ITUB4	-0.01	0.02	-0.15	0.43
UGPA3	-0.01	0.03	-0.26	0.73
PETR3	-0.01	0.03	-0.32	0.00
VALE3	-0.01	0.02	-0.16	0.23
VIIA3	-0.02	0.04	-0.12	0.70
MGLU3	-0.02	0.05	-0.22	0.64

A decisão sobre a melhor cópula para modelar conjuntamente os dados bivariados foi orientada pelos critérios AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiano) e EDC (Critério de Determinação Eficiente). Na Tabela 3.6, esses critérios sugerem a cópula de Gumbel-Hougaard para (MGLU3, VIIA3), e a cópula de Frank para (BBAS3, ITUB4), (UGPA3, MGLU3), (VIIA3, PETR3), (MGLU3, PETR3) e (VALE3, BBAS3).

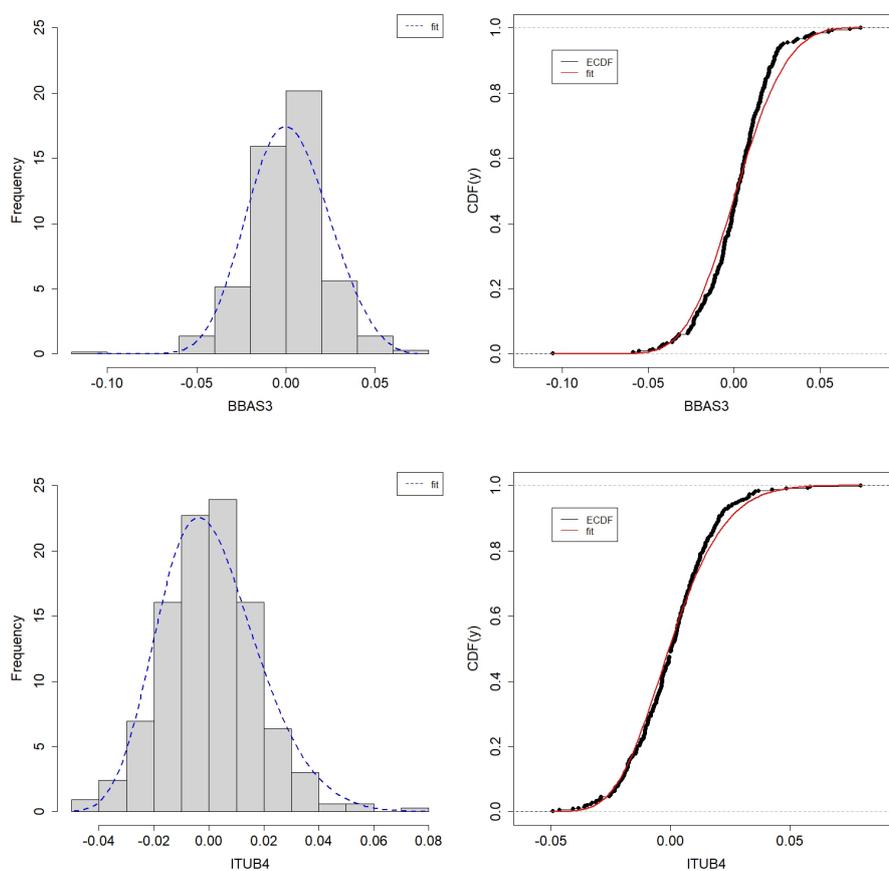


Figura 3.3: Histograma (esquerda) e ECDF (direita) dos log-retornos ajustados das instituições bancárias BBAS3 e ITUB4.

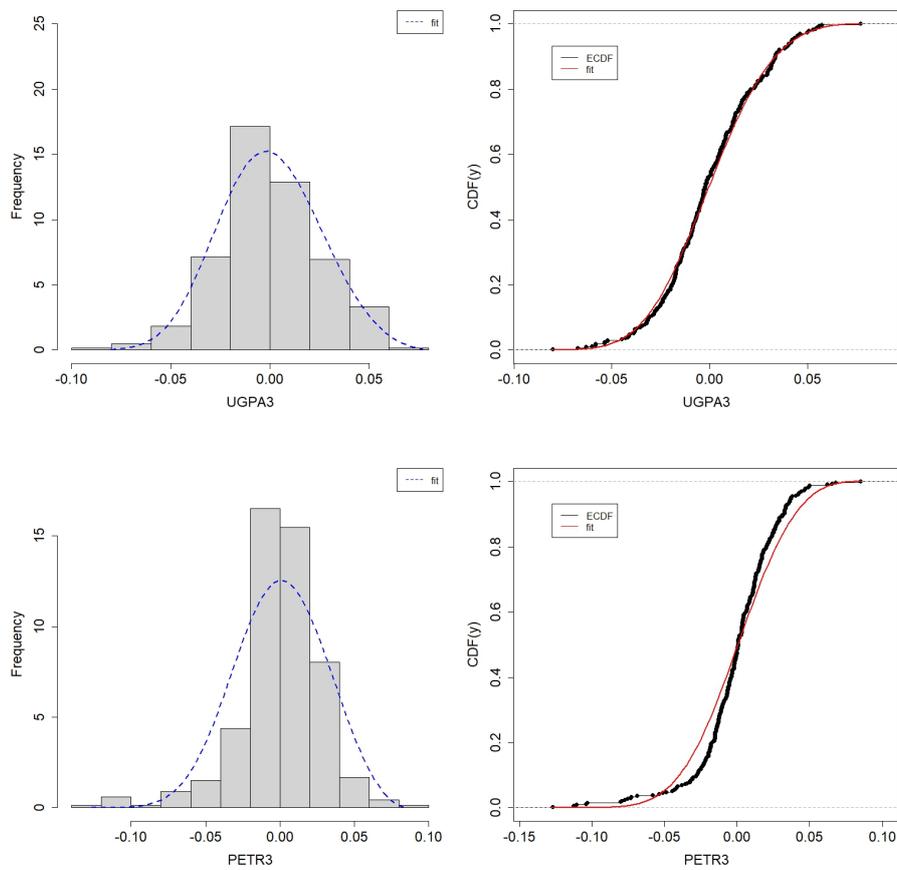


Figura 3.4: Histograma (esquerda) e ECDF (direita) dos log-retornos ajustados das empresas de petróleo e gás UGPA3 e PETR3.

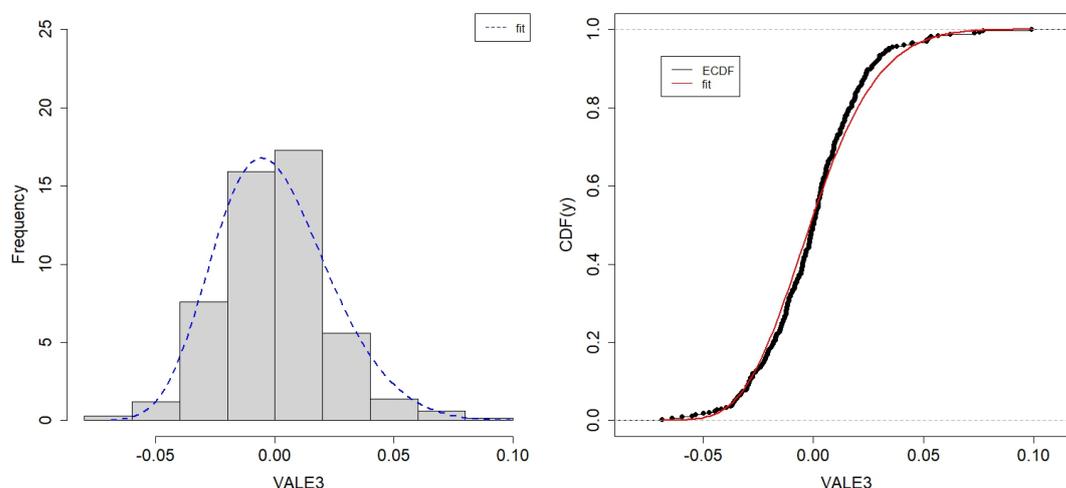


Figura 3.5: Histograma (esquerda) e ECDF (direita) dos log-retornos ajustados da empresa de mineração VALE3.

Tabela 3.6: Estimativas do parâmetro de dependência das cópulas e seleção de modelos com AIC, BIC e EDC.

Dados bivariados	cópula	$\hat{\theta}$	AIC	BIC	EDC
(BBAS3, ITUB4)	Gumbel-Hougaard	0.58	-121.74	-54.55	-111.13
	Frank	5.77	-159.13	-91.94	-148.52
	Clayton	0.78	-80.07	-12.88	-69.46
(UGPA3, MGLU3)	Gumbel-Hougaard	0.72	-51.81	15.38	-41.20
	Frank	3.22	-61.22	5.97	-50.61
	Clayton	0.43	-22.73	44.46	-12.12
(VIA3, PETR3)	Gumbel-Hougaard	0.94	11.24	78.42	21.85
	Frank	0.89	9.08	76.27	19.69
	Clayton	0.05	13.12	80.31	23.73
(MGLU3, VIA3)	Gumbel-Hougaard	0.43	-301.61	-234.43	-291.00
	Frank	7.98	-297.10	-229.92	-286.50
	Clayton	1.30	-152.77	-85.59	-142.17
(MGLU3, PETR3)	Gumbel-Hougaard	0.93	11.10	78.29	21.71
	Frank	1.03	7.59	74.77	18.19
	Clayton	0.06	12.25	79.44	22.86
(VALE3, BBAS3)	Gumbel-Hougaard	0.95	12.79	79.98	23.40
	Frank	1.08	7.46	74.65	18.07
	Clayton	0.04	12.83	80.02	23.44

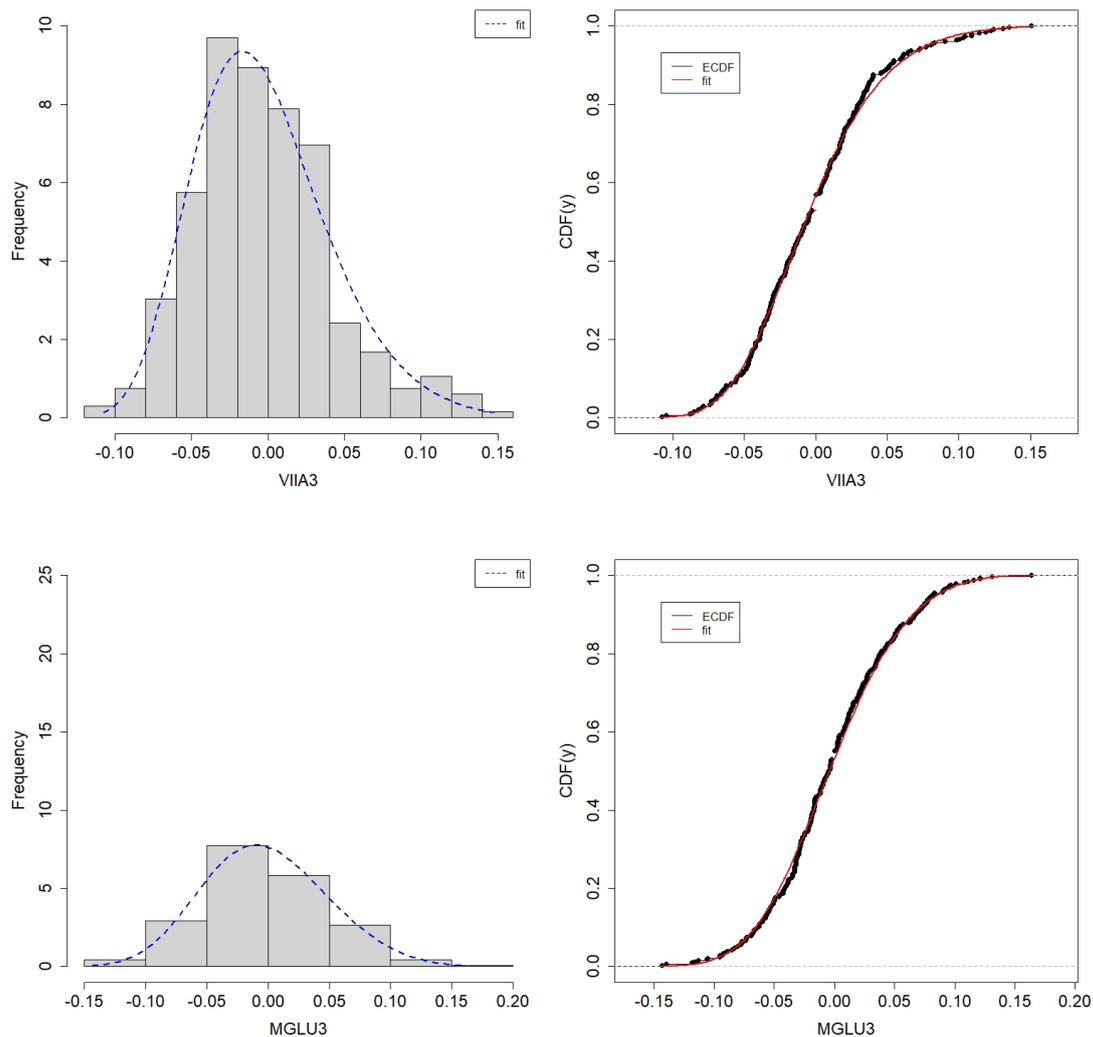


Figura 3.6: Histograma (esquerda) e ECDF (direita) dos log-retornos ajustados das empresas de varejo VAA3 e MGLU3.

Na Tabela 3.7, os ICs foram obtidos pelo método Bootstrap, conforme o Algoritmo 3.2.1. Em um cenário de seleção de ativos, a única situação conclusiva ocorre ao comparar UGPA3 e MGLU3, pois, este é o caso em que $R = 1/2$ não está dentro do intervalo de confiança $(0, 4035; 0, 4949)$, indicando que os retornos de UGPA3 são maiores do que os de MGLU3 no período analisado. Para todos os outros casos, a análise é inconclusiva.

Tabela 3.7: Estimativas de R e intervalo de confiança.

Dados bivariados	\hat{R}	\hat{R}_{NP}	95% IC
(BBAS3, ITUB4)	0.4749	0.4909	(0.4459; 0.5224)
(UGPA3, MGLU3)	0.4393	0.4788	(0.4035; 0.4949)
(VIA3, PETR3)	0.5482	0.5576	(0.5053; 0.5899)
(MGLU3, VIA3)	0.4696	0.4667	(0.4477; 0.533)
(MGLU3, PETR3)	0.5245	0.5455	(0.4843; 0.566)
(VALE3, BBAS3)	0.5296	0.5242	(0.4874; 0.5689)

A Figura 3.7 foi construída usando dados reais e simulados segundo o modelo ajustado. Das Figuras 3.10 as Figuras 3.13 mostram os gráficos QQ normais para os ajustes univariados. Podemos concluir que os modelos de cópula se ajustam bem aos dados. Além disso, é importante comparar diferentes distribuições como modelos candidatos para a modelagem dos log-retornos. Comparamos o desempenho da distribuição GEV e normal como modelos para os retornos diários, conforme apresentado na Tabela 3.8. É possível observar que tanto a GEV quanto a normal forneceram capacidades de modelagem bastante similares (valores de critérios de informação aproximadamente iguais).

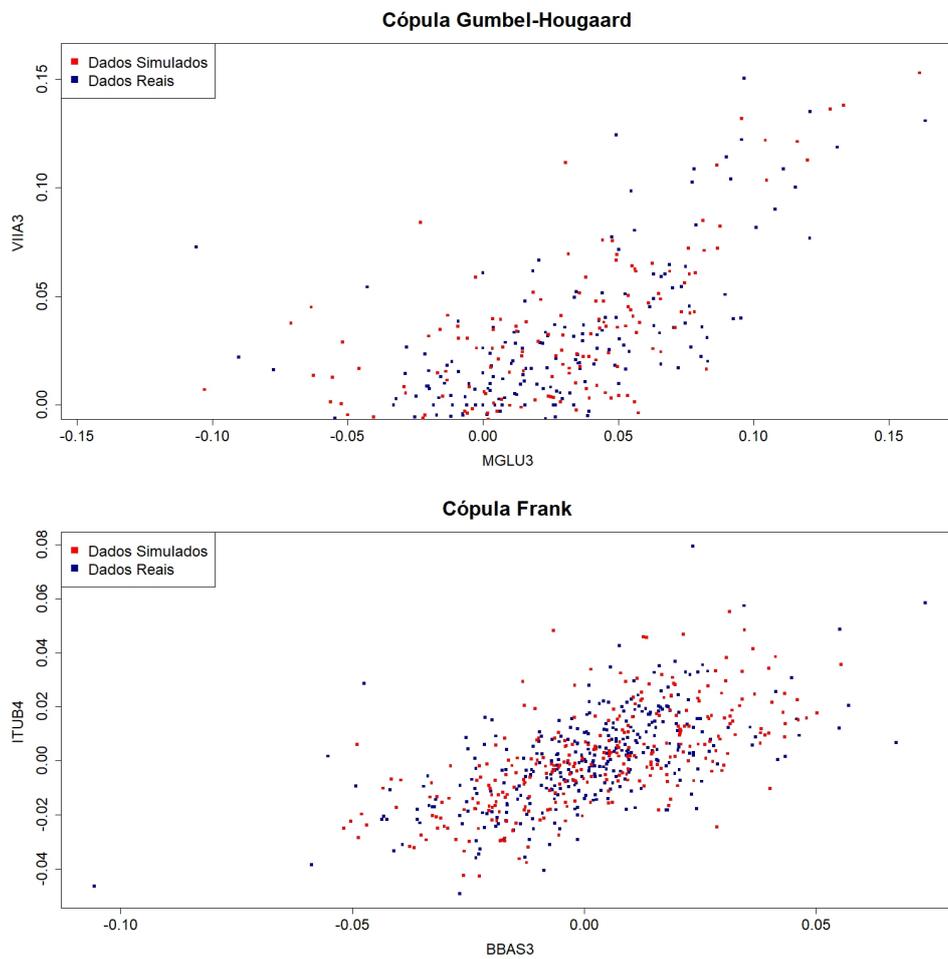


Figura 3.7: Dados reais e simulados segundo os modelos ajustados.

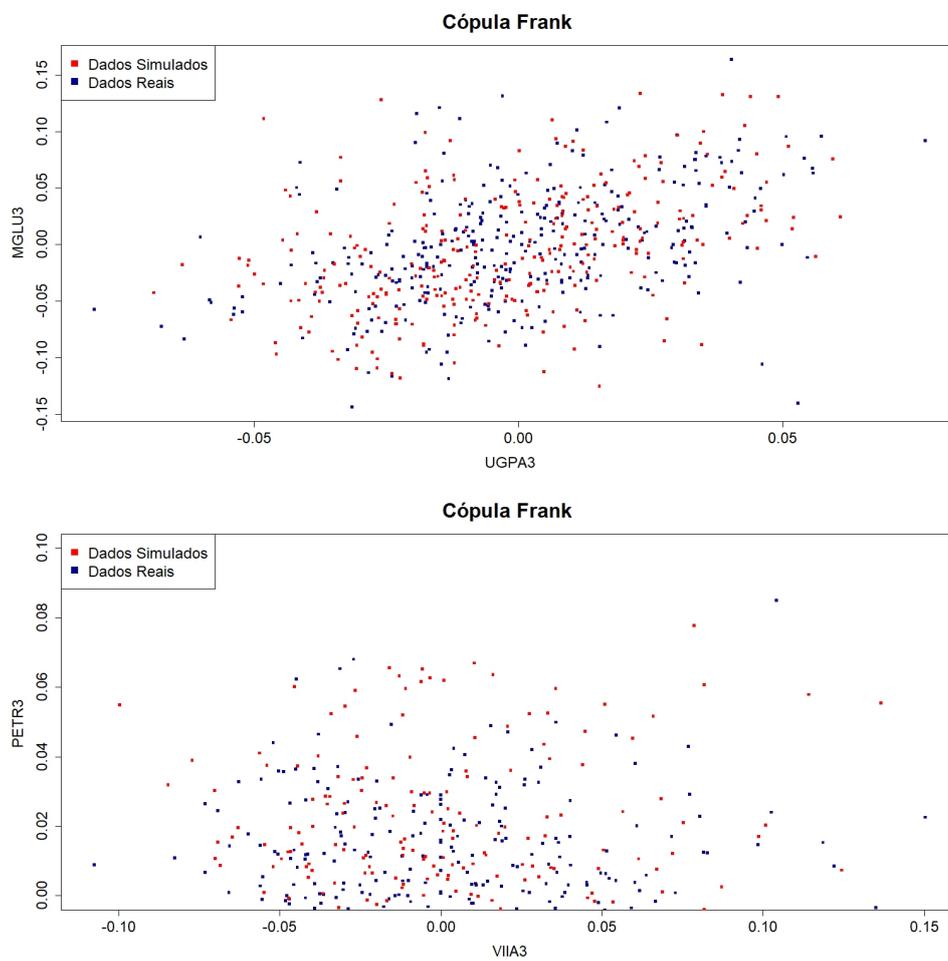


Figura 3.8: Dados reais e simulados segundo os modelos ajustados.

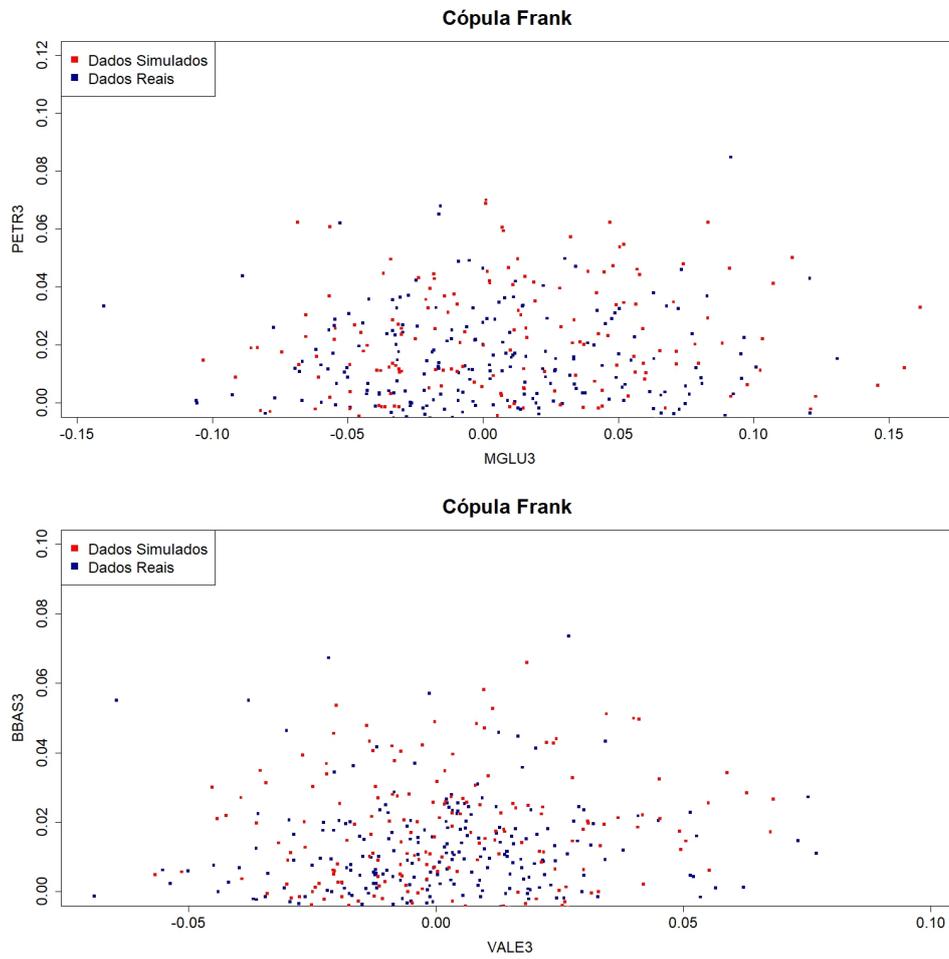


Figura 3.9: Dados reais e simulados segundo os modelos ajustados.

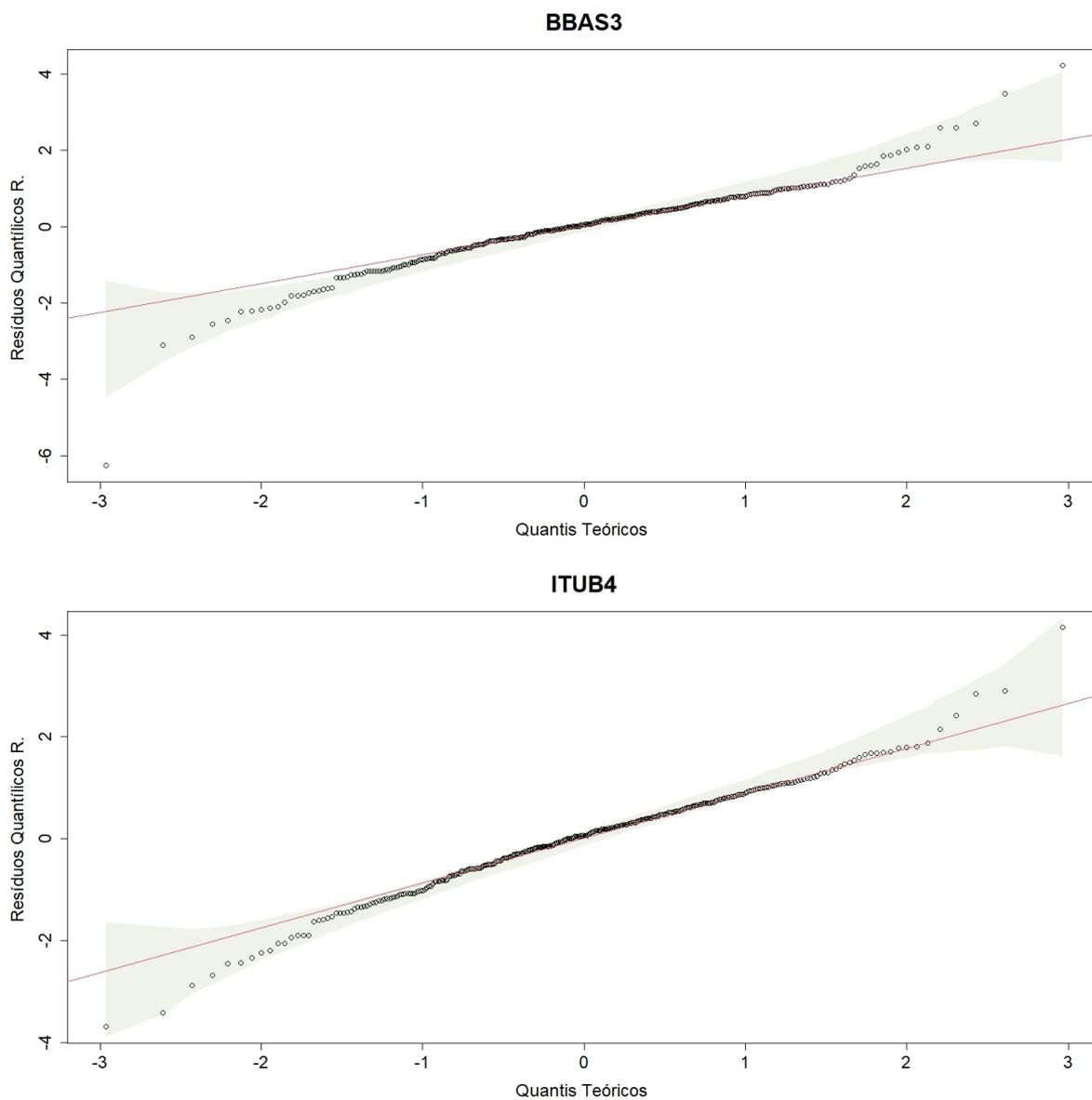


Figura 3.10: Gráficos QQ normais dos ajustes univariados dos log-retornos ajustados das instituições bancárias BBAS3 e ITUB4.

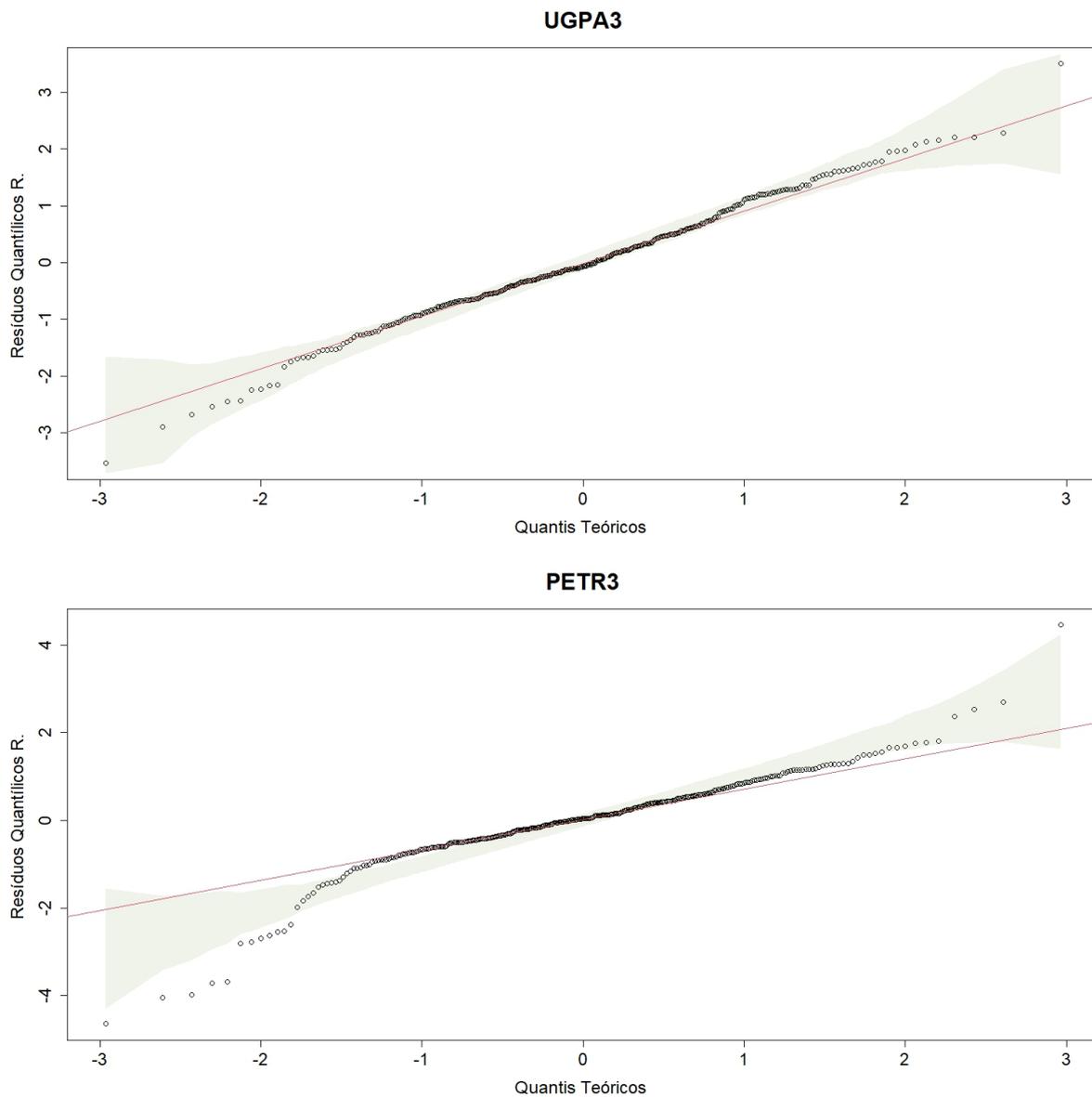


Figura 3.11: Gráficos QQ normais dos ajustes univariados dos log-retornos ajustados das empresas de petróleo e gás UGPA3 e PETR3.

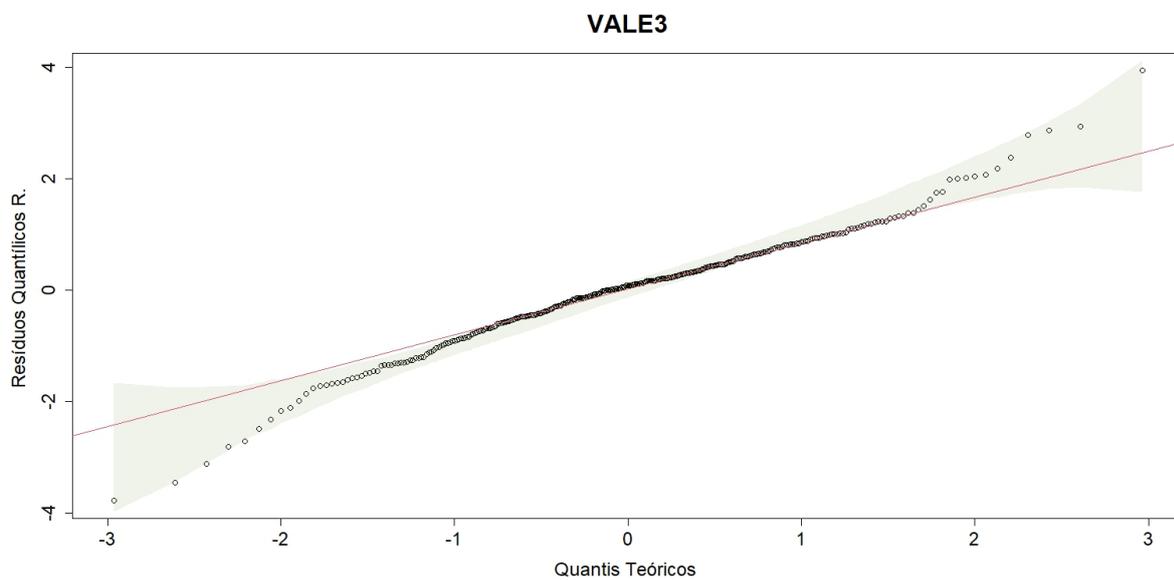


Figura 3.12: Gráficos QQ normais dos ajustes univariados dos log-retornos ajustados da empresa de mineração VALE3.

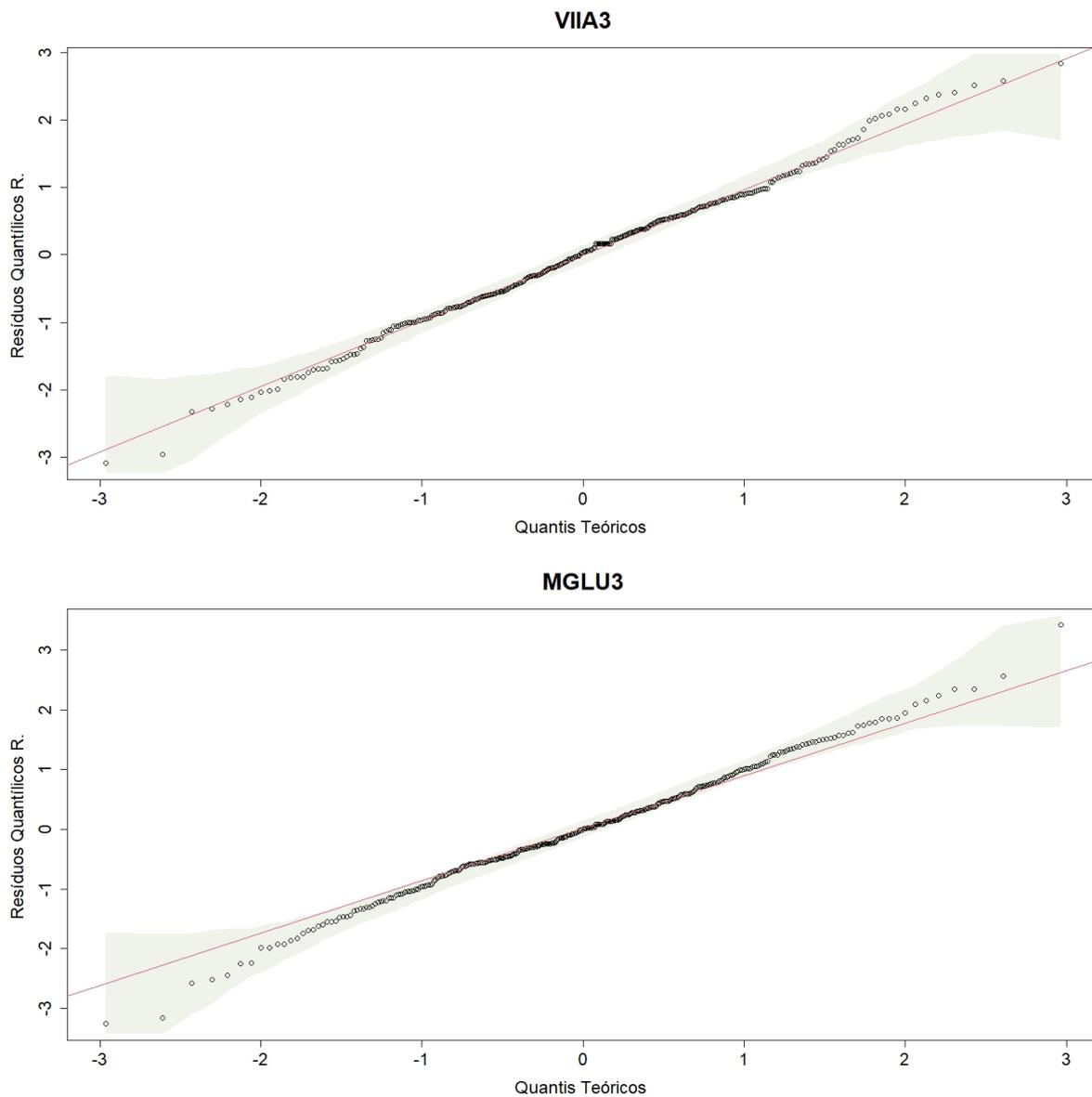


Figura 3.13: Gráficos QQ normais dos ajustes univariados dos log-retornos ajustados das empresas de varejo VIIA3 e MGLU3.

Tabela 3.8: Seleção de modelo para marginais com AIC, BIC e EDC.

Dados	Distribuição	AIC	BIC	EDC
BBAS3	Normal	1638.59	1657.78	1641.62
	GEV	1613.83	1642.63	1618.38
ITUB4	Normal	1750.27	1769.47	1753.30
	GEV	1747.28	1776.07	1751.82
MGLU3	Normal	1043.51	1062.70	1046.54
	GEV	1046.06	1074.86	1050.61
VIA3	Normal	1119.85	1139.05	1122.88
	GEV	1141.40	1170.19	1145.94
VALE3	Normal	1556.49	1575.69	1559.52
	GEV	1554.18	1582.97	1558.73
PETR3	Normal	1427.54	1446.74	1430.57
	GEV	1411.09	1439.88	1415.63
UGPA3	Normal	1498.53	1517.73	1501.56
	GEV	1499.31	1528.11	1503.86

3.4.2 Dados de renda e consumo

Para avaliar outra aplicabilidade de SSR $R = P(X < Y)$, foi utilizado um conjunto de dados públicos de renda e consumo. Os dados² são provenientes da Pesquisa do Banco da Itália sobre Renda e Patrimônio das Famílias do ano de 2008.

A renda é composta por:

- renda do trabalho (salários líquidos, remunerações e benefícios adicionais);
- pensões e transferências líquidas (pensões, pagamentos atrasados e outras transferências);
- renda líquida de autônomo (renda de autônomo e renda empresarial); e
- renda de propriedade (seja de imóveis ou de ativos financeiros).

O consumo é composto pelos anos de despesas duráveis e não duráveis. As despesas duráveis

²Disponível em https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/documentazione/ricerca/ricerca.html?min_anno_publicazione=2008&max_anno_publicazione=2008, acessado em: 09/11/2023

representam o equilíbrio entre bens comprados e vendidos, enquanto as despesas não duráveis são compostas por gastos mensais (aluguéis, alimentos, itens não alimentícios, etc.) e anuais.

O mesmo conjunto de dados também foi utilizado em Domma e Giordano (2012). No entanto, esses autores transformaram os dados de consumo e, até onde sabemos, não conseguimos obter os mesmos dados transformados seguindo os procedimentos descritos em Domma e Giordano (2012). Portanto, optamos por usar os dados brutos da pesquisa, disponíveis como os conjuntos de dados RFAM08 (renda como variável Y) e RISFAM08 (consumo como variável X), de acordo com o arquivo de descrição dos dados da Pesquisa sobre Renda e Consumo das Famílias de 2008³. A amostra filtrada resultou em 7977 domicílios, após a remoção de entradas cuja renda ou consumo são negativos ou não disponíveis.

A Tabela 3.9 apresenta as estatísticas descritivas para a renda (Y) e o consumo (X). As Figuras 3.14 e 3.15 mostram dados com caudas pesadas e o ajuste da distribuição GEV aos dados por meio da densidade e ECDF dos dados. Os parâmetros estimados são apresentados na Tabela 3.10. O ajuste da cópula de Frank e marginais GEV também é avaliado na Figura 3.16.

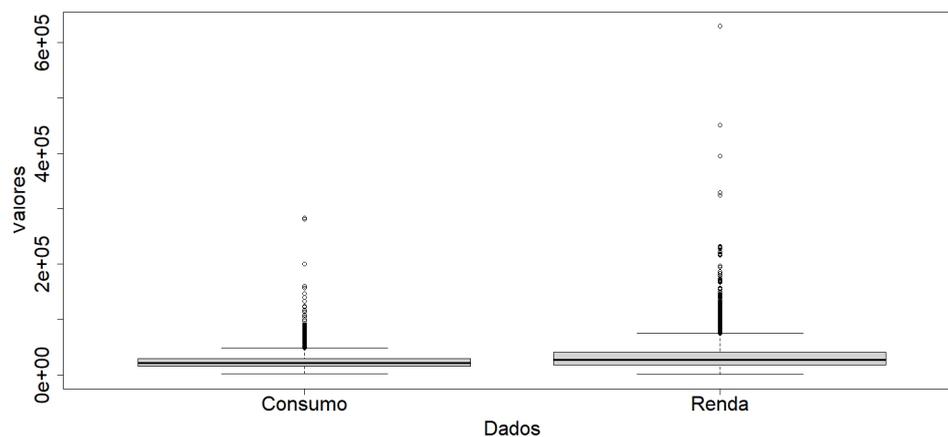


Figura 3.14: Boxplot para consumo e renda.

³Disponível em https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/documentazione/documenti/2008/eng_Legen08.pdf, acessado em: 09/11/2023

Tabela 3.9: Estatísticas descritivas para o consumo X e a renda Y .

Dados	Min.	Q_1	Mediana	Média	Q_3	Max.	Desvio Padrão	Assimetria
X	1680	15600	20760	23836	28800	283100	13779.52	3.7685
Y	65.7	17890.5	26784.2	32424.4	40587.0	629339.7	24333.25	5.2347

Tabela 3.10: Parâmetros estimados da distribuição GEV.

Dados	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
X	17753.46	8226.34	0.13
Y	21993.26	12807.62	0.20

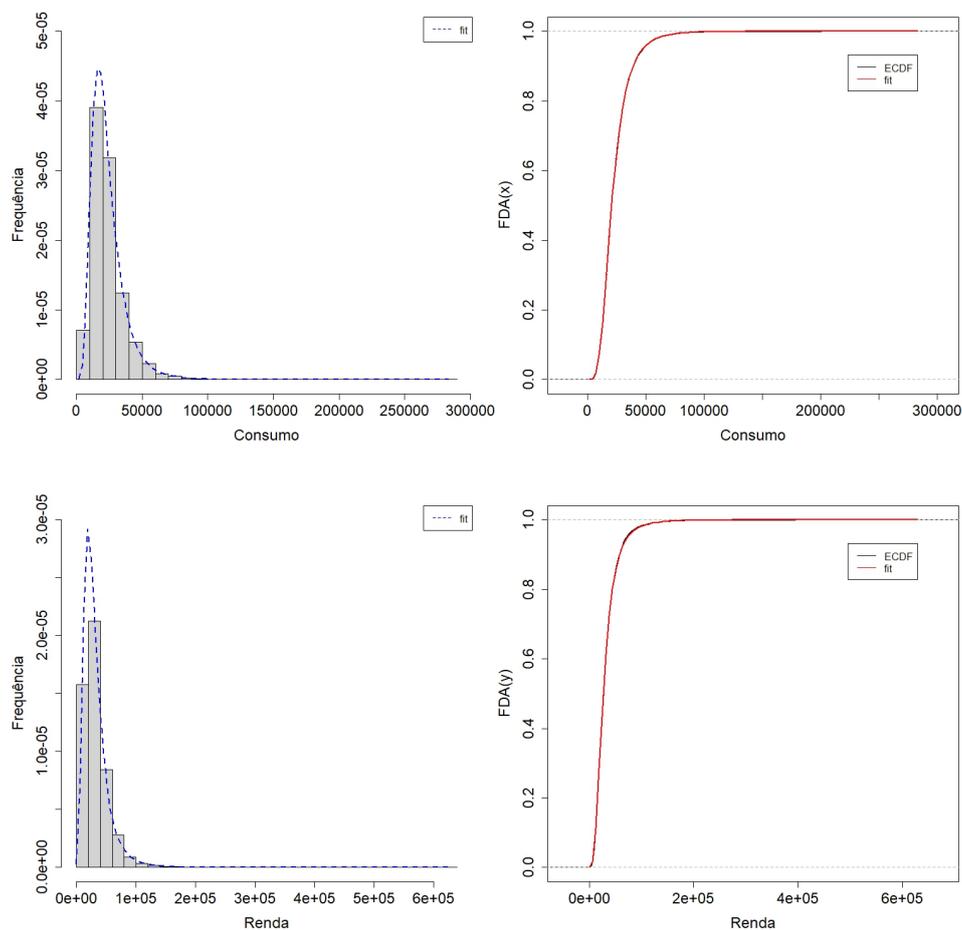
**Figura 3.15:** Densidades marginais ajustadas (esquerda) e ECDF (direita) para consumo (topo) e renda (parte inferior).

Tabela 3.11: Estimativas de dependência de cópula e seleção de modelo com AIC, BIC e EDC.

Cópula	$\hat{\theta}$	AIC	BIC	EDC
Gumbel-Hougaard	0.40	-8410.24	-8298.50	-8393.51
Frank	9.14	-8983.35	-8871.60	-8966.62
Clayton	2.21	-7313.24	-7201.49	-7296.50

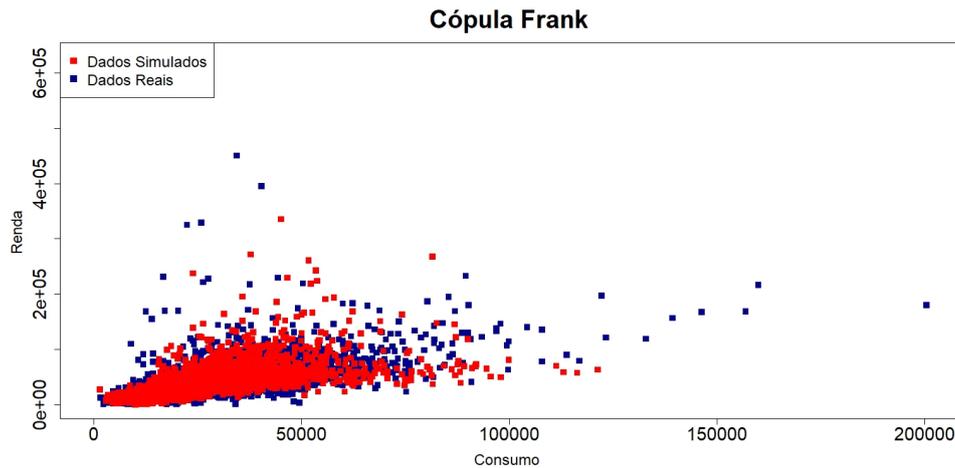


Figura 3.16: Dados reais e simulados.

Apesar da distribuição GEV modelar efetivamente as marginais, a independência não pode ser considerada neste contexto, tornando inviável a aplicação direta dos resultados de Quintino et al. (2024). Conseqüentemente, a abordagem preferida neste cenário envolve a modelagem da dependência por meio de cópulas. Com base na Tabela 3.11, a cópula de Frank emerge como a melhor escolha para modelar conjuntamente as variáveis X e Y , pois demonstra o critério de informação mais favorável, em comparação com as cópulas de Gumbel-Hougaard e Clayton. Ao estimar R através de (3.7), o valor resultante é $\hat{R} = 0,7841$, acompanhado por um IC de 95% de $(0,7562; 0,7957)$. Isso sugere uma probabilidade de aproximadamente 79% para uma família típica encerrar o ano com saldo positivo de renda.

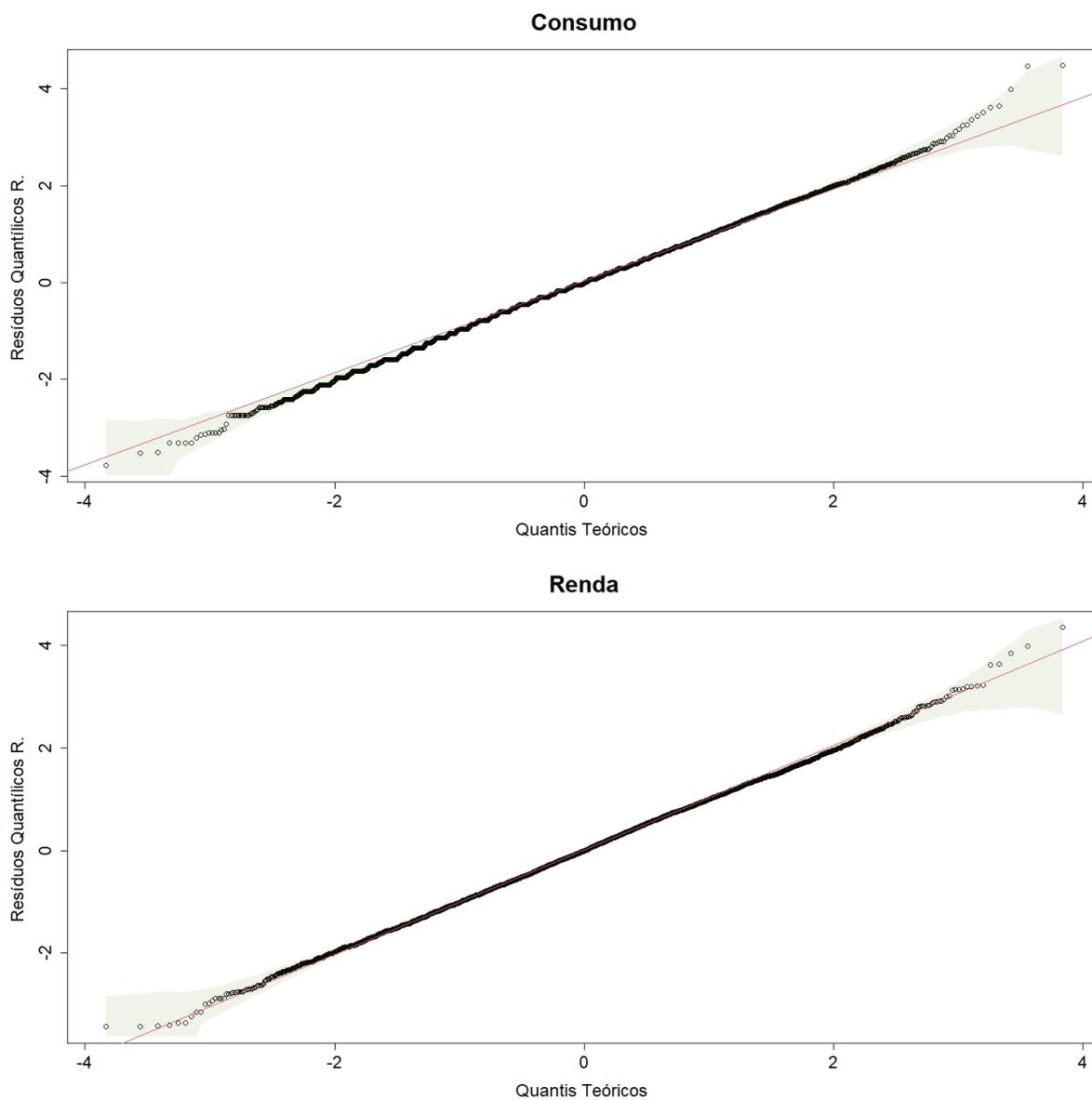


Figura 3.17: Gráficos QQ normais para os resíduos dos ajustes marginais.

Dado que os dados de X e Y são positivos, podemos fazer uma comparação entre a utilização da distribuição GEV e outras distribuições familiares, como Weibull e Gamma, conforme ilustrado na Tabela 3.12.

Tabela 3.12: Seleção de modelo para marginais com AIC, BIC e EDC.

Data set	Model	AIC	BIC	EDC
X	Gamma	170462.98	170494.91	170467.76
	Weibull	171812.76	171844.69	171817.54
	GEV	169926.30	169974.19	169933.47
Y	Gamma	177934.22	177966.14	177939.00
	Weibull	178862.71	178894.64	178867.49
	GEV	169926.30	169974.19	169933.47

3.4.3 Fluxos mínimos mensais de água.

Ramos et al. (2019) analisaram cinco conjuntos de dados reais relacionados aos fluxos mínimos mensais de água (m^3/s) no rio Piracicaba, localizado no estado de São Paulo, Brasil. Eles obtiveram os conjuntos de dados junto ao Departamento de Recursos Hídricos e à agência gestora de recursos hídricos do Estado de São Paulo, no período de 1960 a 2014.

Considerando que a matriz elétrica do Brasil é principalmente composta por fontes hidráulicas (cerca de 62% do total⁴), é importante compreender as mudanças mensais nos fluxos de água para considerar fontes alternativas de geração de eletricidade. Ao longo dos meses, pode ser necessário usar fontes de combustíveis fósseis para complementar a quantidade de energia gerada necessária para todo o país. Dessa forma, medidas de confiabilidade do tipo $P(X < Y)$ podem ser usadas para estimar a quantidade de combustíveis fósseis que precisam ser estocados para um determinado mês, de forma comparativa.

Assim, a Tabela 3.13 apresenta estatísticas descritivas para os fluxos mínimos mensais de água em setembro (Y) e julho (X). A Figura 3.18 mostra as PDFs e CDFs ajustadas da distribuição GEV, bem como as empíricas. Os gráficos QQplots na Figura 3.19 indicam uma boa qualidade de ajuste. Além disso, os parâmetros estimados são apresentados na Tabela 3.14.

⁴Disponível em <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>, acessado em: 24/12/2023

Tabela 3.13: Estatísticas descritivas para X em julho e Y em setembro.

Dados	Min.	Q_1	Mediana	Média	Q_3	Max.	Desvio Padrão	Assimetria
X	7.26	11.40	13.49	25.41	25.14	174.94	30.97	3.35
Y	6.18	11.12	16.44	28.28	32.91	153.78	29.32	2.45

Tabela 3.14: Parâmetros estimados da distribuição GEV.

Dados	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
X	12.7542	5.8039	0.7114
Y	13.3178	8.4558	0.7448

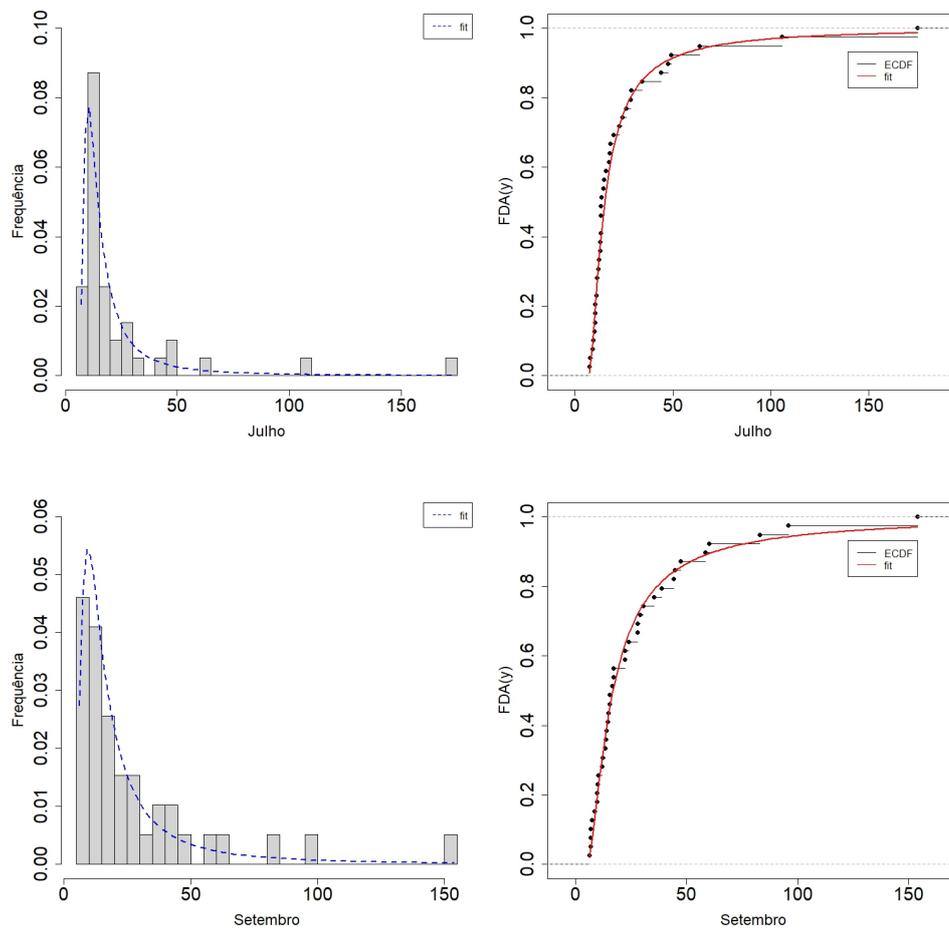


Figura 3.18: PDFs do modelo GEV ajustadas (esquerda) e CDF empírica (direita) para o fluxo mínimo mensal em julho (topo) e setembro (parte inferior).

Refirido-se à Tabela 3.15, onde a cópula de Gumbel-Hougaard surge como a escolha ideal para modelar conjuntamente as variáveis X e Y , pois demonstra o critério de informação mais favorável, em comparação com as cópulas de Frank e Clayton. Ao estimar R através de (3.7), o valor resultante é $\hat{R} = 0,5384$, acompanhado por um intervalo de confiança de 95% de $(0,3918; 0,6899)$. Isso sugere que a necessidade de contingência em ambos os meses é aproximadamente a mesma (os fluxos mínimos são estatisticamente equivalentes).

Tabela 3.15: Estimativas de dependência de cópula e seleção de modelo com AIC, BIC e EDC.

Copula	$\hat{\theta}$	AIC	BIC	EDC
Gumbel-Hougaard	0.87	11.87	49.16	16.05
Frank	1.13	12.55	49.84	16.73
Clayton	0.10	13.73	51.02	17.91

Como feito em subseções anteriores, ao observar que os conjuntos de dados X e Y são positivos, é possível realizar uma análise comparativa dos ajustes marginais da distribuição GEV com aqueles dos modelos Gamma e Weibull, como ilustrado na Tabela 3.16.

Tabela 3.16: Seleção de modelo para marginais com AIC, BIC e EDC.

Data set	Model	AIC	BIC	EDC
X	Gamma	377.69	388.35	378.89
	Weibull	376.39	387.05	377.59
	GEV	342.17	358.15	343.96
Y	Gamma	338.44	349.09	339.63
	Weibull	340.89	351.54	342.08
	GEV	327.42	343.40	329.21

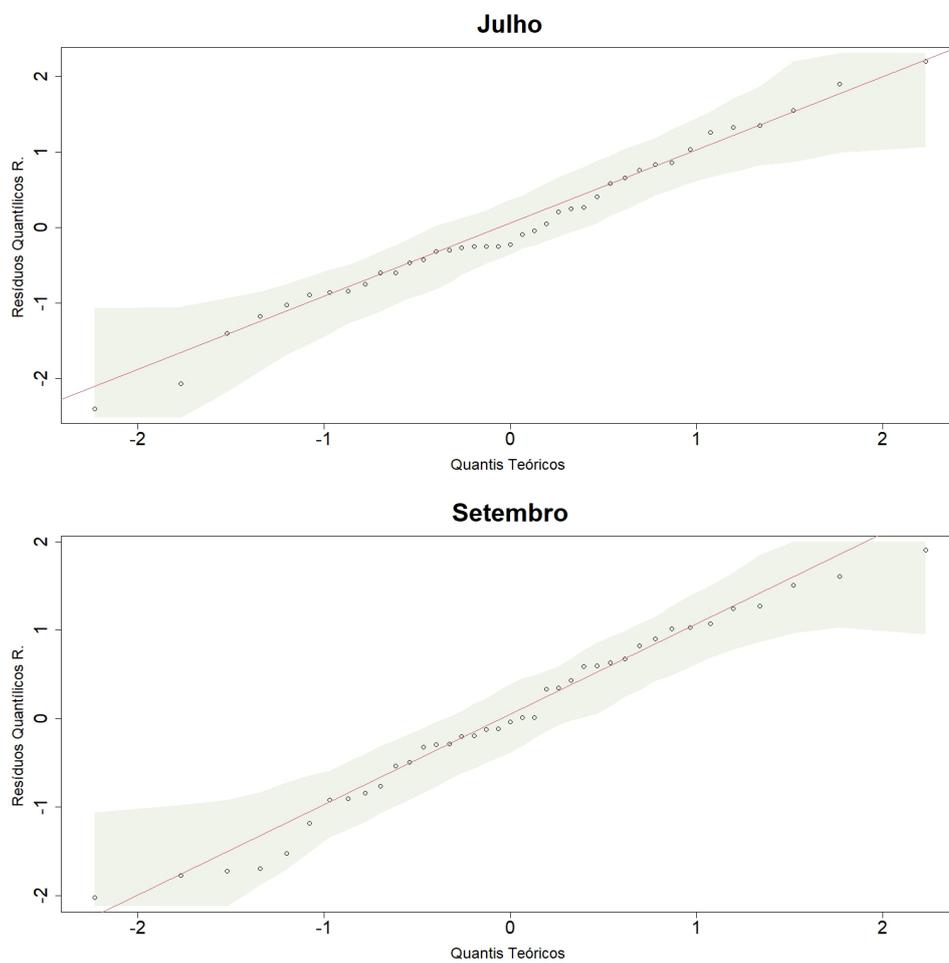


Figura 3.19: Gráficos QQ normais para os resíduos do fluxo mínimo mensal

Capítulo 4

MSSR com Marginais de Distribuições de Valores Extremos: Teoria e Aplicação a Dados Hidrológicos

4.1 Introdução

Neste capítulo, estudamos a confiabilidade de *stress-strength* multicomponente (MSSR), que é $R_{s,k} = P(\text{pelo menos } s \text{ das } (X_1, \dots, X_k) \text{ excedem } Y)$ quando X_1, \dots, X_n e Y são v.a.'s independentes. Sejam X_1, \dots, X_n v.a.'s i.i.d. com FDA $G_X(\cdot)$, e Y uma v.a. independente com FDP $f_Y(\cdot)$. Defina $M_n = \max\{X_1, \dots, X_n\}$. Então, $P(M_n \leq u) = G_X^n(u)$ e temos

$$I_n = P(M_n \leq Y) = \int_{-\infty}^{\infty} G_X^n(u) f_Y(u) du. \quad (4.1)$$

Na Seção 4.2 apresentamos uma forma fechada da probabilidade MSSR. Expressões fechadas para (4.1) são fornecidas para cada família das distribuições extremas descritas na Seção 2.3. Na Seção 4.3, propomos e validamos a aplicação da metodologia R na modelagem de dados reais. No primeiro caso, avaliamos dados que representam a capacidade mensal de água do reservatório

Shasta na Califórnia, EUA, durante os meses de agosto, setembro e dezembro de 1980 a 2015. Na segunda aplicação, avaliamos dados de fluxos mínimos mensais de água do Piracicaba.

4.2 Confiabilidade do sistema multicomponente

Num contexto mais amplo, a confiabilidade MSSR é dada por

$$\begin{aligned} R_{s,k} &= P(\text{pelo menos } s \text{ de } (X_1, \dots, X_k) \text{ exceder } Y) \\ &= \sum_{j=s}^k \binom{k}{j} \int_{-\infty}^{\infty} (1 - G_X(u))^j (G_X(u))^{k-j} f_Y(u) du. \end{aligned}$$

Usando uma expansão binomial, obtemos

$$R_{s,k} = \sum_{j=s}^k \sum_{r=0}^j \binom{k}{j} \binom{j}{r} (-1)^{j-r} \int_{-\infty}^{\infty} (G_X(u))^{k-r} f_Y(u) du. \quad (4.2)$$

Definição (4.2), (4.1) pode ser escrita como:

$$R_{s,k} = \sum_{j=s}^k \sum_{r=0}^j \binom{k}{j} \binom{j}{r} (-1)^{j-r} I_{k-r}. \quad (4.3)$$

Estamos interessados em obter $R_{s,k}$ para as leis estáveis l-max e as seis leis p-max estáveis. A partir de (4.3), basta ter uma expressão fechada para I_n definida em (4.1).

Recentemente, Rathie et al. (2023) introduziu a função de valor extremo \mathbb{H} como:

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^{\infty} y^{a_6} \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy, \quad (4.4)$$

onde $\Re(a_1), \Re(a_2), \Re(a_4) \in \mathbb{R}_+, a_3, a_5 \in \mathbb{C}$, em ambos $\Re(a_1)$ e $\Re(a_2)$ pode ser igual a zero ao mesmo tempo, $\Re(a_6) > -1$ quando $a_1 \neq 0$ ou $a_1 = 0$ e $\text{sign}(a_3) = \text{sign}(a_5)$, $\Re(a_6) < -1$ quando $a_1 = 0$ e $\text{sign}(a_3) \neq \text{sign}(a_5)$. Aqui, \mathbb{R}, \mathbb{C} e \Re denotam os números reais, os números

complexos e a parte real de um número complexo, respectivamente.

A função de valor extremo \mathbb{H} é uma ferramenta importante para obter expressões fechadas de I_n para cada distribuição de valor extremo.

4.2.1 Uma expressão alternativa para MSSR

O MSSR (4.3) pode ser reescrito como

$$R_{s,k} = 1 - \sum_{\omega=k-s+1}^k \binom{k}{\omega} \int_{-\infty}^{+\infty} (F_X(u))^{\omega} (1 - F_X(u))^{k-\omega} f_Y(u) du. \quad (4.5)$$

De fato,

$$\begin{aligned} R_{s,k} &= P(\text{pelo menos } s \text{ de } (X_1, \dots, X_k) \text{ excede } Y) \\ &= 1 - P(\text{pelo menos } s \text{ de } (X_1, \dots, X_k) \text{ não exceda } Y) \\ &= \int_{-\infty}^{+\infty} \left[1 - \sum_{\omega=k-s+1}^k \binom{k}{\omega} (F_X(u))^{\omega} (1 - F_X(u))^{k-\omega} \right] f_Y(u) du, \end{aligned}$$

desde $\int_{-\infty}^{+\infty} f_Y(u) du = 1$.

Observe que

$$1 = (F_X(u) + (1 - F_X(u)))^k = \sum_{n=0}^k \binom{k}{n} (F_X(u))^n (1 - F_X(u))^{k-n}.$$

Por isso,

$$1 - \sum_{\omega=k-s+1}^k \binom{k}{\omega} (F_X(u))^{\omega} (1 - F_X(u))^{k-\omega} = \sum_{\omega=0}^k \binom{k}{\omega} (F_X(u))^{\omega} (1 - F_X(u))^{k-\omega}.$$

Defina $k - w = j$. Observando que $\binom{k}{k-j} = \binom{k}{w}$, obtemos

$$1 - \sum_{w=k-s+1}^k \binom{k}{w} (F_X(u))^w (1 - F_X(u))^{k-w} = \sum_{j=s}^{k-s} \binom{k}{s} (F_X(u))^{k-j} (1 - F_X(u))^j,$$

o que implica a igualdade entre (4.3) e (4.5).

Nas próximas subsecções, lembraremos a FDA e a FDP das leis estáveis l- e p-max apresentadas no Capítulo 2 e obtemos a expressão correspondente de I_n em cada caso.

4.2.2 Leis estáveis l-máximas

Fréchet

A seguir são apresentados a CDF e a PDF da distribuição Fréchet, respectivamente,

$$F(x; \mu, \sigma, \alpha) = \begin{cases} 0, & x < \mu, \\ \exp \left\{ - \left(\frac{x-\mu}{\sigma} \right)^{-\alpha} \right\}, & x \geq \mu, \end{cases}$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma} \left(\frac{x - \mu}{\sigma} \right)^{-\alpha-1} F(x; \mu, \sigma, \alpha) \mathbb{1}_{(\mu, \infty)}, \quad (4.6)$$

onde $\mathbb{1}_A$ denota a função indicadora no conjunto A .

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição Fréchet($\mu_x, \sigma_x, \gamma_x$) e Y uma v.a. independente com distribuição Fréchet($\mu_y, \sigma_y, \gamma_y$). Se $\mu_y \geq \mu_x$, então

$$I_n = \mathbb{H} \left(1, n^{-1/\alpha_x} \frac{\sigma_y}{\sigma_x}, -\frac{1}{\alpha_y}, n^{-1/\alpha_x} \frac{\mu_y - \mu_x}{\sigma_x}, -\alpha_x, 0 \right), \quad (4.7)$$

De fato,

$$\begin{aligned} I_n &= P(X_1 \leq Y, \dots, X_n \leq Y) \\ &= P(M_n \leq Y) \\ &= \int_{-\infty}^{+\infty} F_X^n(u; \mu_x, \sigma_x, \gamma_x) f_Y(u; \mu_y, \sigma_y, \gamma_y) du. \end{aligned}$$

Defina $M = \max\{\mu_x, \mu_y\}$. Então,

$$I_n = \int_M^{+\infty} \exp\left\{-n \left(\frac{u - \mu_x}{\sigma_x}\right)^{-\alpha_x}\right\} \frac{\alpha_y}{\sigma_y} \left(\frac{u - \mu_y}{\sigma_y}\right)^{-\alpha_y-1} \exp\left\{-\left(\frac{u - \mu_y}{\sigma_y}\right)^{-\alpha_y}\right\} du.$$

Substituindo $v = \left(\frac{u - \mu_y}{\sigma_y}\right)^{-\alpha_y}$, se $M = \mu_y$, obtemos

$$I_n = \int_0^{\infty} \exp\left\{-v - n \left(\frac{\sigma_y v^{-1/\alpha_y} + \mu_y - \mu_x}{\sigma_x}\right)^{-\alpha_x}\right\} dv. \quad (4.8)$$

Portanto, (4.7) segue de (4.4) e (4.8).

Weibull reversa

A seguir são apresentados a FDA e a FDP da distribuição Weibull reversa, respectivamente,

$$F(x; \mu, \sigma, \alpha) = \begin{cases} \exp\left\{-\left(\frac{\mu-x}{\sigma}\right)^\alpha\right\}, & x < \mu, \\ 1, & x \geq \mu, \end{cases}$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha}{\sigma} \left(\frac{\mu-x}{\sigma}\right)^{\alpha-1} F(x; \mu, \sigma, \alpha) \mathbb{1}_{(-\infty, \mu)}(x). \quad (4.9)$$

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição RWeibull($\mu_x, \sigma_x, \gamma_x$) e Y uma v.a. independente com distribuição RWeibull($\mu_y, \sigma_y, \gamma_y$). Se $\mu_x \geq \mu_y$, então

$$I_n = \mathbb{H}\left(1, n^{1/\alpha_x} \frac{\sigma_y}{\sigma_x}, \frac{1}{\alpha_y}, n^{1/\alpha_x} \frac{\mu_x - \mu_y}{\sigma_x}, \alpha_x, 0\right). \quad (4.10)$$

Gumbel

A seguir são apresentados a FDA e a FDP da distribuição Gumbel, respectivamente,

$$F(x; \mu, \sigma, \alpha) = \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \right\},$$

e

$$f(x; \mu, \sigma, \alpha) = \frac{1}{\sigma} \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \right\}.$$

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição Gumbel(μ_x, σ_x) e Y uma v.a. independente com distribuição Gumbel(μ_y, σ_y). Então,

$$I_n = \exp \left\{ \frac{\mu_y}{\sigma_y} \right\} \mathbb{H} \left(\exp \left\{ \frac{\mu_y}{\sigma_y} \right\}, n \exp \left\{ \frac{\mu_x}{\sigma_x} \right\}, \frac{\sigma_y}{\sigma_x}, 0, 1, 0 \right). \quad (4.11)$$

Distribuição GEV

A seguir são apresentados a FDA e a FDP da distribuição GEV, respectivamente,

$$F(x; \mu, \sigma, \gamma) = \exp \left\{ - \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}} \right\}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0, \quad (4.12)$$

e

$$f(x; \mu, \sigma, \gamma) = F(x, \gamma, \mu, \sigma) \frac{1}{\sigma} \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}-1}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0. \quad (4.13)$$

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição $GEV(\mu_x, \sigma_x, \gamma_x)$ e Y uma v.a. independente com distribuição $GEV(\mu_y, \sigma_y, \gamma_y)$. Então:

- Quando $\gamma_j > 0, j \in x, y$:

$$I_n = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} n^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right] n^{-\gamma_x}, -\frac{1}{\gamma_x} \right), \quad (4.14)$$

desde que $\mu_y - \frac{\sigma_y}{\gamma_y} \geq \mu_x - \frac{\sigma_x}{\gamma_x}$. Quando $\mu_y - \frac{\sigma_y}{\gamma_y} \leq \mu_x - \frac{\sigma_x}{\gamma_x}$:

$$I_n = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x} n^{-\gamma_y}, -\gamma_x, \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right] n^{-\gamma_y}, -\frac{1}{\gamma_y} \right). \quad (4.15)$$

- Quando $\gamma_j < 0, j \in x, y$:

$$I_n = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} n^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right] n^{-\gamma_x}, -\frac{1}{\gamma_x} \right), \quad (4.16)$$

desde que $\mu_y - \frac{\sigma_y}{\gamma_y} \leq \mu_x - \frac{\sigma_x}{\gamma_x}$. Quando $\mu_y - \frac{\sigma_y}{\gamma_y} \geq \mu_x - \frac{\sigma_x}{\gamma_x}$:

$$I_n = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x} n^{-\gamma_y}, -\gamma_x, \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right] n^{-\gamma_y}, -\frac{1}{\gamma_y} \right). \quad (4.17)$$

- Quando $\gamma_y = \gamma_x = 0$:

$$I_n = \exp \left\{ \frac{\mu_y}{\sigma_y} \right\} \mathbb{H} \left(\exp \left\{ \frac{\mu_y}{\sigma_y} \right\}, \exp \left\{ \frac{\mu_x}{\sigma_x} \right\}, n, \frac{\sigma_y}{\sigma_x}, 0, 1 \right). \quad (4.18)$$

4.2.3 Leis p-max estáveis

Caso 1: log-Fréchet

Seja X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_1(\alpha_x, \beta_x, \gamma_x)$ e Y são uma v.a. independente com distribuição $H_1(\alpha_y, \beta_y, \gamma_y)$, $\alpha_i, \beta_i, \gamma_i \in (0, \infty), i \in \{x, y\}$. Então, no caso H_1 , FDA, FDP e I_n são dados, respectivamente, por

$$H_1(x; \alpha, \beta, \gamma) = \begin{cases} 0, & \text{se } \gamma x^\beta < 1, \\ \exp\{-(\log(\gamma x^\beta))^{-\alpha}\}, & \text{se } \gamma x^\beta \geq 1, \end{cases}$$

$$h_1(x; \alpha, \beta, \gamma) = \exp\{-[\log \gamma x^\beta]^{-\alpha}\} \frac{\alpha \beta}{x} [\log(\gamma x^\beta)]^{-\alpha-1} \mathbb{1}_{(\gamma^{-1/\beta}, \infty)}(x),$$

e

$$I_n = \mathbb{H} \left(1, n^{-1/\alpha_x} \frac{\beta_x}{\beta_y}, -\frac{1}{\alpha_y}, n^{-1/\alpha_x} \left(\log \gamma_x - \frac{\beta_x}{\beta_y} \log \gamma_y \right), -\alpha_x, 0 \right),$$

desde que $\gamma_y^{-1/\beta_y} \geq \gamma_x^{-1/\beta_x}$.

Caso 2: log-Weibull

Seja X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_2(\alpha_x, \beta_x, \gamma_x)$ e Y são uma v.a. independente com distribuição $H_2(\alpha_y, \beta_y, \gamma_y)$, $\alpha_i, \beta_i, \gamma_i \in (0, \infty)$, $i \in \{x, y\}$. Então, no caso H_2 , FDA, FDP e I_n são dados, respectivamente, por

$$H_2(x; \alpha) = \begin{cases} 0, & \text{se } \gamma x^\beta < 0, \\ \exp\{-(-\log(\gamma x^\beta))^\alpha\}, & \text{se } 0 \leq \gamma x^\beta < 1, \\ 1, & \text{se } \gamma x^\beta \geq 1, \end{cases}$$

$$h_2(x; \alpha, \beta, \gamma) = \exp\{-[-\log(\gamma x^\beta)]^\alpha\} [-\log(\gamma x^\beta)]^{\alpha-1} \frac{\beta\alpha}{x} \mathbb{1}_{(0, \gamma^{-1/\beta})}(x),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$,

$$I_n = H\left(1, n^{1/\alpha_x} \frac{\beta_x}{\beta_y}, \frac{1}{\alpha_y}, n^{1/\alpha_x} \left(-\log \gamma_x + \frac{\beta_x}{\beta_y} \log \gamma_y\right), \alpha_x, 0\right),$$

desde que $\gamma_y^{-1/\beta_y} \leq \gamma_x^{-1/\beta_x}$.

Caso 3: log-Fréchet inversa

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_3(\alpha_x, \beta_x, \gamma_x)$ e Y uma v.a. independente com distribuição $H_3(\alpha_y, \beta_y, \gamma_y)$, onde $\alpha_i, \beta_i, \gamma_i \in (0, \infty)$, $i \in \{x, y\}$. Então, no caso H_3 , a FDA, FDP e I_n são dadas, respectivamente, por:

$$H_3(x; \alpha) = \begin{cases} 0, & \text{se } -\gamma(-x)^\beta < -1, \\ \exp\{-(-\log(\gamma(-x)^\beta))^{-\alpha}\}, & \text{se } -1 \leq -\gamma(-x)^\beta < 0, \\ 1, & \text{se } -\gamma(-x)^\beta \geq 0, \end{cases}$$

$$h_3(x; \alpha, \beta, \gamma) = \exp\{-[-\log(\gamma(-x)^\beta)]^{-\alpha}\} \frac{\alpha\beta}{-x} [-\log(\gamma(-x)^\beta)]^{-\alpha-1} \mathbb{1}_{(-1, 0)}(x),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$, e

$$I_n = H \left(1, n^{-\frac{1}{\alpha_x}} \frac{\beta_x}{\beta_y}, -\frac{1}{\alpha_y}, n^{-\frac{1}{\alpha_x}} \left(-\log \gamma_x + \frac{\beta_x}{\beta_y} \log \gamma_y \right), -\alpha_x, 0 \right),$$

desde que $\gamma_y^{-1/\beta_y} \leq \gamma_x^{-1/\beta_x}$.

Caso 4: log-Weibull inversa

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_4(\alpha_x, \beta_x, \gamma_x)$ e Y uma v.a. independente com distribuição $H_4(\alpha_y, \beta_y, \gamma_y)$, onde $\alpha_i, \beta_i, \gamma_i \in (0, \infty)$, $i \in \{x, y\}$. Então, no caso H_4 , a FDA, FDP e I_n são dadas, respectivamente, por:

$$H_4(x; \alpha) = \begin{cases} \exp\{-(\log(\gamma(-x)^\beta))^\alpha\}, & \text{se } -\gamma(-x)^\beta < -1, \\ 1, & \text{se } -\gamma(-x)^\beta \geq -1, \end{cases}$$

$$h_4(x; \alpha, \beta, \gamma) = \exp\{-[\log(\gamma(-x)^\beta)]^\alpha\} [\log(\gamma(-x)^\beta)]^{\alpha-1} \frac{\alpha\beta}{-x} \mathbb{1}_{(-\infty, -1)}(x),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$, e

$$I_n = \mathbb{H} \left(1, n^{\frac{1}{\alpha_x}} \frac{\beta_x}{\beta_y}, \frac{1}{\alpha_y}, n^{\frac{1}{\alpha_x}} \left(\log \gamma_x - \frac{\beta_x}{\beta_y} \log \gamma_y \right), \alpha_x, 0 \right),$$

desde que $\gamma_y^{1/\beta_y} \leq \gamma_x^{1/\beta_x}$.

Caso 5: Fréchet padrão

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_5(\beta_x, \gamma_x)$ e Y uma v.a. independente com distribuição $H_5(\beta_y, \gamma_y)$, onde $\beta_i, \gamma_i \in (0, \infty)$, $i \in \{x, y\}$. Então, no caso H_5 , a FDA, FDP e I_n são dadas, respectivamente, por:

$$H_5(x) = \begin{cases} 0, & \text{se } \gamma x^\beta < 0, \\ \exp\{-(\gamma x^\beta)^{-1}\} & \text{se } \gamma x^\beta \geq 0, \end{cases}$$

$$h_5(x; \beta, \gamma) = \exp\{-(\gamma x^\beta)^{-1}\} \frac{\beta}{\gamma} x^{-\beta-1} \mathbb{1}_{[0, \infty)}(x),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$, e

$$I_n = \frac{1}{\gamma_y} \mathbb{H} \left(\gamma_y^{-1}, n\gamma_x^{-1}, \frac{\beta_x}{\beta_y}, 0, 1, 0 \right).$$

Caso 6: Weibull padrão

Sejam X_1, \dots, X_n v.a.'s i.i.d. com distribuição $H_6(\beta_x, \gamma_x)$ e Y uma v.a. independente com distribuição $H_6(\beta_y, \gamma_y)$, onde $\alpha_i, \beta_i, \gamma_i \in (0, \infty)$, $i \in \{x, y\}$. Então, no caso H_6 , a FDA, FDP e I_n são dadas, respectivamente, por:

$$H_6(x) = \begin{cases} \exp\{-\gamma(-x)^\beta\}, & \text{se } -\gamma(-x)^\beta < 0, \\ 1 & \text{se } -\gamma(-x)^\beta \geq 0, \end{cases}$$

$$h_6(x; \beta, \gamma) = \exp\{-\gamma(-x)^\beta\} \gamma \beta (-x)^{\beta-1} \mathbb{1}_{(-\infty, 0)}(x),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}_+$, e

$$I_n = \gamma_y \mathbb{H} \left(\gamma_y, n\gamma_x, \frac{\beta_x}{\beta_y}, 0, 1, 0 \right).$$

4.3 Aplicações a dados hidrológicos

Nesta seção, discutimos duas aplicações utilizando dados reais. Avaliamos a validade do modelo para todos os casos e identificamos quais distribuições abordadas neste capítulo se ajustam adequadamente aos conjuntos de dados considerados.

4.3.1 Aplicação aos dados hidrológicos

Para avaliar a estrutura proposta, modelamos conjuntos de dados hidrológicos como distribuições de valores extremos e os comparamos no sentido MSSR. Denote X_1, X_2, X_3, X_4 e X_5 os dados de resistência observados e Y os dados de tensão observados, respectivamente. Os dados são frequentemente utilizados na literatura e também são apresentados abaixo (por exemplo, Lio et al. (2024), Lio et al. (2022) e Wang et al. (2021)).

$$\mathbf{X} = \begin{pmatrix} X_1 & X_2 & X_3 & X_4 & X_5 \\ 0.4238 & 0.5579 & 0.7262 & 0.8112 & 0.8296 \\ 0.2912 & 0.3634 & 0.3719 & 0.4637 & 0.4785 \\ 0.5381 & 0.5612 & 0.7226 & 0.7449 & 0.7540 \\ 0.5249 & 0.6060 & 0.6686 & 0.7159 & 0.7552 \\ 0.3451 & 0.4253 & 0.4688 & 0.7188 & 0.7420 \\ 0.2948 & 0.3929 & 0.4616 & 0.6139 & 0.7951 \end{pmatrix} \text{ e } Y = \begin{pmatrix} 0.7009 \\ 0.6532 \\ 0.4589 \\ 0.7183 \\ 0.5310 \\ 0.7665 \end{pmatrix}.$$

Os dados representam a capacidade mensal de água do reservatório Shasta na Califórnia, EUA, durante os meses de agosto, setembro e dezembro de 1980 a 2015. Neste caso, consideramos $k = 5$ componentes de força que são obtidos da seguinte forma: $Y_1 = 0.7009$ é a capacidade de dezembro de 1980, e $X_1 = (0.4238, 0.2912, 0.5381, 0.7183, 0.5310, 0.7665)$ são as capacidades de agosto de 1981 a 1985; $Y_2 = 0.6532$ é a capacidade de dezembro de 1986, $X_2 = (0.5579, 0.3634, 0.5612, 0.6060, 0.4253, 0.3929)$ são as capacidades de agosto de 1987 a 1991, e assim por diante. Para mais detalhes sobre o conjunto de dados, consulte Wang et al. (2021) e as referências lá contidas.

Com o objetivo de estimar probabilidades do tipo $R_{s,k}$ conforme dado em (4.3) através das expressões I_n fornecidas na seção anterior, precisamos analisar a melhor distribuição de valores extremos para ajustar os dados. É importante comparar diferentes distribuições como modelos

candidatos para a modelagem do conjunto de dados. Considerando resultados anteriores na literatura (Wang et al., 2021), comparamos o desempenho das leis estáveis l-max e p-max com as distribuições Rayleigh como modelos para a capacidade mensal de água. Estimativas de máxima verossimilhança (ML) e p-valor do teste Kolmogorov-Smirnov (KS) estão apresentados na Tabela 4.1. Para cada conjunto de dados, a Função de Distribuição Empírica (ECDF) é comparada aos modelos teóricos nas Figuras 4.1 e 4.2.

Tabela 4.1: Parâmetros estimados da distribuição GEV e valor p do teste Kolmogorov-Smirnov.

VA	Modelos	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	KS p-valor
X	Fréchet	0.29	0.12	0.66	4×10^{-3}
X	Gumbel	0.49	0.15	-	0.34
X	Weibull	0.83	0.20	0.60	2×10^{-3}
X	GEV	0.54	0.18	-0.59	0.47
X	Rayleigh	0.24	7.44	-	0.28
VA	Modelos	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	KS p-valor
X	H_1	18.85	0.18	3.15	0.29
X	H_2	1.34	2.27	1.40	0.31
X	H_5	3.17	11.38	-	0.49
VA	Modelos	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	KS p-valor
Y	Fréchet	0.46	0.17	1.01	0.18
Y	Gumbel	0.58	0.11	-	0.71
Y	Weibull	0.77	0.12	1.00	0.95
Y	GEV	0.64	0.14	-1.07	0.98
Y	Rayleigh	0.41	15.82	-	0.68
VA	Modelos	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	KS p-valor
Y	H_1	0.08	18.29	1539843.25	0.03
Y	H_2	0.63	7.69	7.72	0.51
Y	H_5	5.41	20.43	-	0.64

É possível observar que Gumbel, GEV, H_2 e Rayleigh forneceram capacidades de modelagem bastante semelhantes. Os critérios de informação na Tabela 4.2 mostram que H_2 é o melhor modelo para os conjuntos de dados. Portanto, considerando $s = 3$, $k = 5$ e os modelos Gumbel, GEV e H_2 em (4.3), a Tabela 4.3 mostra as estimativas para o MSSR da capacidade mensal de água do Reservatório Shasta na Califórnia durante o período analisado.

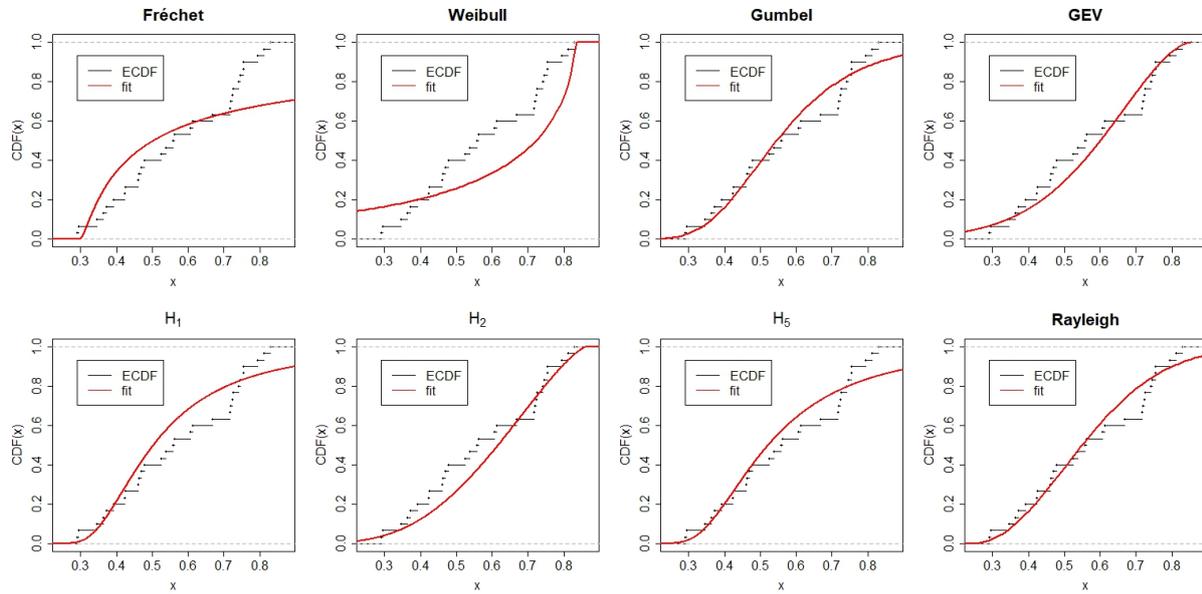


Figura 4.1: Modelos ajustados e ECDF para *strength* (X).

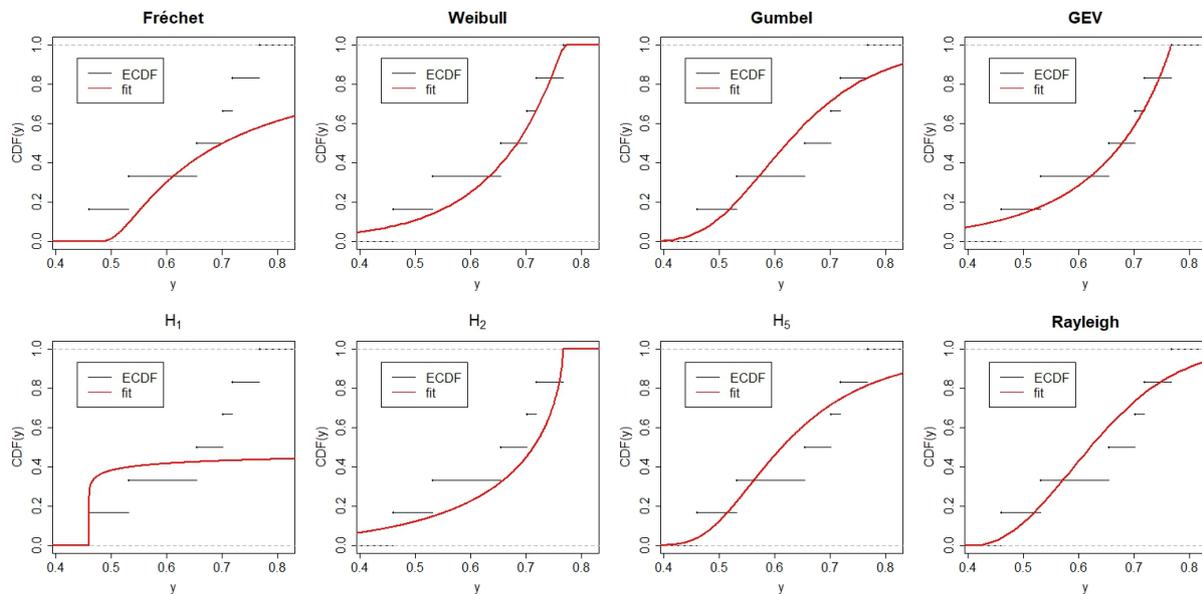


Figura 4.2: Modelos ajustados e ECDF para *stress* (Y).

Tabela 4.2: Seleção de modelo com AIC, BIC e EDC.

VA	Modelos	AIC	BIC	EDC
X	Fréchet	19.83	27.03	20.50
X	Gumbel	-18.68	-13.88	-18.23
X	Weibull	-1.76	5.44	-1.09
X	GEV	-23.01	-15.81	-22.34
X	H_1	-16.28	-23.48	-16.95
X	H_2	-30.76	-37.96	-31.43
X	H_5	-17.21	-22.01	-17.66
X	Rayleigh	-22.05	-17.25	-21.60
Y	Fréchet	-1.92	0.46	-3.17
Y	Gumbel	-6.38	-4.80	-7.22
Y	Weibull	-9.57	-7.19	-10.82
Y	GEV	-13.62	-11.24	-14.87
Y	H_1	-11.61	-13.99	-10.36
Y	H_2	-17.97	-20.34	-16.72
Y	H_5	-9.55	-11.13	-8.71
Y	Rayleigh	-7.36	-5.78	-8.20

Tabela 4.3: Estimativas de R_{sk} para alguns modelos de valores extremos com $s = 3$ e $k = 5$.

Modelos	$\hat{R}_{3,5}$
Gumbel	0.3036
GEV	0.3151
H_2	0.3049

Neste contexto, Y refere-se o nível de água em um determinado mês de um reservatório na Califórnia. As variáveis $X_1 \cdots X_5$ são os níveis do reservatório em meses diferentes. Neste trabalho, baseados em Lio et al. (2022), Lio et al. (2024) e Wang et al. (2021), consideramos Y como referência e calculamos a probabilidade de em pelo menos $s = 3$ dos $k = 5$ meses excederem o nível do reservatório registrado em Y . Se a probabilidade R_{sk} for menor que 0,5 não houve nenhum X que excedeu Y , caso contrário há pelo menos um X que excede o nível do reservatório registrado em Y . Em termos práticos essa metodologia pode ser utilizada para calcular a probabilidade de um determinado mês Y e então calcular a probabilidade de três meses ou mais exceder o nível do reservatório registrado em Y , os resultados obtidos na Tabela 4.3

demonstra que nenhum mês X excedeu o nível registrado em Y .

Capítulo 5

Conclusões

Apesar de suas limitações significativas, a teoria moderna de portfólio de Markowitz continua sendo amplamente utilizada em diversos estudos, devido à sua simplicidade e facilidade de uso. Por isso, é crucial estudar abordagens alternativas que sejam igualmente fáceis de compreender.

No Capítulo 3, estudamos a confiabilidade *stress-strength* $R = P(X < Y)$ quando X e Y seguem distribuições GEV dependentes. Para tratar a dependência entre as variáveis X e Y , foi utilizada uma estrutura de cópulas, empregando as cópulas de Gumbel-Hougaard, Frank e Clayton. Em resumo, quando X e Y representam v.a.'s de retorno e $R < 1/2$, é aconselhável que o investidor escolha a variável X . Se $R > 1/2$, ocorre o oposto. O caso $R = 1/2$ é inconclusivo.

Simulações de Monte-Carlo atestam o desempenho das expressões analíticas de R . Ao aplicar nossa metodologia a dados financeiros do mundo real, poderíamos orientar um procedimento de seleção de ações calculando $P(X < Y)$ quando X e Y representam retornos de ações. Além disso, foram descritas aplicações da confiabilidade *stress-strength* em dados de Renda e Consumo das Famílias, extraídos do banco de dados da Itália de 2008. Investigamos a frequência com que as famílias apresentaram um consumo anual superior à sua renda. Também foi realizada uma aplicação utilizando dados de vazão mínima mensal do rio Piracicaba, no Brasil.

No Capítulo 4, apresentamos novas expressões analíticas para a probabilidade de *stress-strength* com multicomponentes, quando as marginais seguem distribuições das famílias de leis

estáveis l-máximas e p-máximas estáveis. Uma aplicação a conjuntos de dados reais foi realizada para demonstrar o desempenho dessas expressões na modelagem dos dados.

Por fim, destacamos que, no estudo de simulações dos estimadores da probabilidade *stress-strength*, observou-se que por diversas vezes, o estimador não paramétrico performou melhor que o estimador de máxima verossimilhança. Para futuros trabalhos, recomendamos a busca por melhorias na estimação via máxima verossimilhança.

Apêndice A

Artigo: Lima et al. (2024)

Autores: Rebeca Klamerick Lima, Felipe Sousa Quintino, Tiago A. da Fonseca, Luan Carlos de Sena Monteiro Ozelim, Pushpa Narayan Rathie and Helton Saulo

Título: “Assessing the impact of copula selection on reliability measures of type $P(X<Y)$ with generalized extreme value marginals”

Revista: Modelling, MDPI

Ano: 2024

Article

Assessing the Impact of Copula Selection on Reliability Measures of Type $P(X < Y)$ with Generalized Extreme Value Marginals

Rebeca Klamerick Lima ¹, Felipe Sousa Quintino ¹, Tiago A. da Fonseca ²,
Luan Carlos de Sena Monteiro Ozelim ^{3,*}, Pushpa Narayan Rathie ¹ and Helton Saulo ¹

¹ Department of Statistics, University of Brasilia, Brasilia 70910-900, Brazil; rebeca.klamerick@aluno.unb.br (R.K.L.); felipe.quintino@unb.br (F.S.Q.); pushpanrathie@yahoo.com (P.N.R.); heltonsauro@unb.br (H.S.)

² Gama Engineering College, University of Brasilia, Brasilia 72444-240, Brazil; fonsecafga@unb.br

³ Department of Civil and Environmental Engineering, University of Brasilia, Brasilia 70910-900, Brazil

* Correspondence: luanoz@gmail.com or ozelim@unb.br

Abstract: In reliability studies, we are interested in the behaviour of a system when it interacts with its surrounding environment. To assess the system's behaviour in a reliability sense, we can take the system's intrinsic quality as strength and the outcome of interactions as stress. Failure is observed whenever stress exceeds strength. Taking Y as a random variable representing the stress the system experiences and random variable X as its strength, the probability of not failing can be taken as a proxy for the reliability of the component and given as $P(Y < X) = 1 - P(X < Y)$. This way, in the present paper, it is considered that X and Y follow generalized extreme value distributions, which represent a family of continuous probability distributions that have been extensively applied in engineering and economic contexts. Our contribution deals with a more general scenario where stress and strength are not independent and copulas are used to model the dependence between the involved random variables. In such modelling framework, we explored the proper selection of copula models characterizing the dependence structure. The Gumbel–Hougaard, Frank, and Clayton copulas were used for modelling bivariate data sets. In each case, information criteria were considered to compare the modelling capabilities of each copula. Two economic applications, as well as an engineering one, on real data sets are discussed. Overall, an easy-to-use methodological framework is described, allowing practitioners to apply it to their own research projects.

Keywords: stress–strength reliability; GEV distribution; bivariate copulas



Citation: Lima, K.R.; Quintino, F.S.; da Fonseca, T.A.; de Sena Monteiro Ozelim, L.C.; Rathie, P.N.; Saulo, H. Copula selection on $P(X < Y)$ with generalized extreme value marginals. *Modelling* **2024**, *5*, 180–200. <https://doi.org/10.3390/modelling5010010>

Academic Editor: Jürgen Pilz

Received: 27 December 2023

Revised: 18 January 2024

Accepted: 23 January 2024

Published: 28 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The probability of failure of a system or component can be calculated by statistically comparing the applied stress to its strength. Let X (strength) and Y (stress) be continuous random variables (RVs) with joint probability density function (PDF) $f_{X,Y}(\cdot, \cdot)$. The stress–strength probability (or reliability) is defined as

$$R = P(X < Y) = \int_{-\infty}^{\infty} \int_{-\infty}^y f_{X,Y}(x, y) dx dy. \quad (1)$$

There are several applications of this framework, such as in asset selection [1], household financial fragility [2] and engineering [3], among others. See Kotz et al. [4] for more details.

Equation (1) can only be evaluated when the analytical representation of the joint PDF is known (or any other equivalent statistical formula which can be transformed into $f_{X,Y}$); thus, properly assessing this joint formulation is of utmost importance in stress–strength applications of type $P(X < Y)$. Finding an accurate representation of $f_{X,Y}$ involves two steps:

Referências Bibliográficas

- Abbas, K. e Tang, Y. (2014). “Objective Bayesian analysis of the Frechet stress–strength model”. *Statistics & Probability Letters* 84, pp. 169–175.
- Akaike, H. (1974). “A new look at the statistical model identification”. *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.
- Akram, S. S. e Yousif, A. H. (2022). “Bayesian Estimation of Reliability for Multicomponent Stress-Strength Model Based on Topp-Leone Distribution”. *Wasit Journal of Pure sciences* 1.3, pp. 90–104.
- Bachelier, L. (1900). “Theorie de la Speculation, Doctor Thesis, Annales Scientifiques Ecole Normale Speriure III-17”. *The random character of stock market prices* 17, pp. 21–86.
- Bhattacharyya, G. K. e Johnson, R. A. (1974). “Estimation of reliability in a multicomponent stress-strength model”. *Journal of the American Statistical Association* 69.348, pp. 966–970.
- Cherubini, U., Luciano, E. e Vecchiato, W. (2004). *Copulas and Their Applications in Finance*. Hoboken, NJ, USA: John Wiley & Sons.
- Cherubini, U., Mulinacci, S. e Gobbi, F. (2009). *Copula Methods in Finance*. Hoboken, NJ, USA: John Wiley & Sons.
- Cirillo, P. e Taleb, N. N. (2016). “Expected shortfall estimation for apparently infinite-mean models of operational risk”. *Quantitative Finance* 16.10, pp. 1485–1494.
- Dagum, C. (1977). “A new model of personal income distribution: specification and estimation”. *Économie appliquée* 30.3, pp. 413–437. DOI: 10.3406/ecoap.1977.4213.

- Domma, F. e Giordano, S. (2012). “A stress–strength model with dependent variables to measure household financial fragility”. *Statistical Methods & Applications* 21, pp. 375–389.
- Embrechts, P., Klüppelberg, C. e Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*. Vol. 33. Springer Science & Business Media.
- Frees, E. W. e Valdez, E. A. (1998). “Understanding relationships using copulas”. *North American Actuarial Journal* 2.1, pp. 1–25.
- Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Wiley series in probability and mathematical statistics. R.E. Krieger Publishing Company. ISBN: 9780898749571.
- Genest, C. e Rivest, LP (1993). “Statistical Inference Procedures for Bivariate Archimedean Copulas”. *Journal of the American Statistical Association* 88.423, 1034–1043. ISSN: 1537-274X. DOI: 10.1080/01621459.1993.10476372. URL: <http://dx.doi.org/10.1080/01621459.1993.10476372>.
- Gettinby, G. D., Sinclair, C. D., Power, D. M. e Brown, R. A. (2004). “An analysis of the distribution of extreme share returns in the UK from 1975 to 2000”. *Journal of Business Finance & Accounting* 31.5-6, pp. 607–646.
- Ghosh, S. (2010). “Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula”. *Hydrological Processes* 24.24, pp. 3558–3567.
- Goncu, A., Akgul, A. K., Imamoğlu, O., Tiryakioğlu, M. e Tiryakioğlu, M. (2012). “An analysis of the extreme returns distribution: the case of the Istanbul Stock Exchange”. *Applied Financial Economics* 22.9, pp. 723–732.
- Gudendorf, G. e Segers, J. (2010). “Extreme-value copulas”. *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*. Springer, pp. 127–145.
- Haan, L. e Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Vol. 3. Springer.
- Hussain, S. I. e Li, S. (2015). “Modeling the distribution of extreme returns in the Chinese stock market”. *Journal of International Financial Markets, Institutions and Money* 34, pp. 263–276.

- Jenkinson, A. F. (1955). “The frequency distribution of the annual maximum (or minimum) values of meteorological elements”. *Quarterly Journal of the Royal Meteorological Society* 81.348, pp. 158–171.
- Jia, X., Nadarajah, S. e Guo, B. (2017). “Bayes estimation of $P(Y < X)$ for the Weibull distribution with arbitrary parameters”. *Applied Mathematical Modelling* 47, pp. 249–259.
- Kasper, T. M., Fuchs, S. e Trutschnig, W. (2021). “On weak conditional convergence of bivariate Archimedean and extreme value copulas, and consequences to nonparametric estimation”. *Bernoulli* 27.4, pp. 2217–2240.
- Kolev, N., Anjos, U. e Mendes, B. V. M. (2006). “Copulas: A Review and Recent Developments”. *Stochastic Models* 22.4, pp. 617–660.
- Kong, X. M., Huang, G. H., Fan, Y. R. e Li, Y. P. (2015). “Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi river, China”. *Stochastic Environmental Research and Risk Assessment* 29, pp. 833–846.
- Kotz, S., Lumelskii, Y. e Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. World Scientific.
- Kotz, S. e Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific.
- Krishnamoorthy, K. e Lin, Y. (2010). “Confidence limits for stress–strength reliability involving Weibull models”. *Journal of Statistical Planning and Inference* 140.7, pp. 1754–1764.
- Kundu, D. e Raqab, M. (2009). “Estimation of $R = P(Y < X)$ for three-parameter Weibull distribution”. *Statistics & Probability Letters* 79.17, pp. 1839–1846.
- Levy, H. (1969). “The EDC: A New Approach to Portfolio Selection”. *The Journal of Financial and Quantitative Analysis* 4.4, pp. 105–119.
- Lima, R. K., Quintino, F. S., Fonseca, T. A. da, Ozelim, L.C.S.M., Rathie, P. N. e Saulo, H. (2024a). “Assessing the Impact of Copula Selection on Reliability Measures of Type $P(X < Y)$ with Generalized Extreme Value Marginals”. *Modelling* 5.1, pp. 180–200. DOI: 10.5281/modelling5010010.

- Lima, R. K., Quintino, F. S., Fonseca, T. A. da, Ozelim, L.C.S.M., Rathie, P. N. e Saulo, H. (2024b). *Codes for Assessing the impact of copula selection on reliability measures of the type $P(X < Y)$ with generalized extreme value marginals*. Versão 1.0.0. URL: https://github.com/eip-unb/Copulas_GEV.
- Lio, Y., Chen, DG, Tsai, TR e Wang, L. (2024). “The Reliability Inference for Multicomponent Stress–Strength Model under the Burr X Distribution”. *AppliedMath* 4.1, pp. 394–426.
- Lio, Y., Tsai, TR, Wang, L. e Cecilio Tejada, I P (2022). “Inferences of the Multicomponent Stress–Strength Reliability for Burr XII Distributions”. *Mathematics* 10.14, p. 2478.
- Mai, JF e Scherer, M. (2017). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*. 2nd. Vol. 6. Series in Quantitative Finance 6. World Scientific Publishing Company.
- Mandelbrot, B. (1967). “The Variation of Some Other Speculative Prices”. *The Journal of Business* 40.4, pp. 393–413. ISSN: 00219398, 15375374. URL: <http://www.jstor.org/stable/2351623> (acesso em 13/09/2023).
- Mohan, N. e Ravi, S. (1993). “Max domains of attraction of univariate and multivariate p-max stable laws”. *Theory of Probability & Its Applications* 37.4, pp. 632–643.
- Nadarajah, S. (2003). “Reliability for extreme value distributions”. *Mathematical and Computer Modelling* 37.9-10, pp. 915–922.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag, New York.
- Pantcheva, E. (1985). “Limit theorems for extreme order statistics under nonlinear normalization”. Em: *Stability Problems for Stochastic Models*. Springer, pp. 284–309.
- Quintino, F. S., Oliveira, M., Rathie, P. N., Ozelim, L. C. S. M. e Fonseca, T. A. da (2024). “Asset selection based on estimating stress-strength probabilities: The case of returns following three-parameter generalized extreme value distributions”. *AIMS Mathematics* 9.1, pp. 2345–2368. ISSN: 2473-6988. DOI: 10.3934/math.2024116.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

- Ramos, P. L., Louzada, F., Ramos, E. e Dey, S. (2019). “The Fréchet distribution: Estimation and application - An overview”. *Journal of Statistics and Management Systems* 23.3, 549–578. ISSN: 2169-0014. DOI: 10.1080/09720510.2019.1645400.
- Rathie, P. N., Ozelim, L. C. S. M., Quintino, F. e Fonseca, T. A da (2023). “On the Extreme Value H -Function”. *Stats* 6.3, pp. 802–811.
- Rathie, P. N., Ozelim, L.C.S.M. e Andrade, B. B. de (2021). “Portfolio Management of Copula-Dependent Assets Based on $P(Y < X)$ Reliability Models: Revisiting Frank Copula and Dagum Distributions”. *Stats* 4.4, pp. 1027–1050.
- Resnick, S. (2008). *Extreme Values, Regular Variation, and Point Processes*. Vol. 4. Springer Science & Business Media.
- Schwarz, G. (1978). “Estimating the Dimension of a Model”. *The Annals of Statistics* 6.2, pp. 461–464.
- Sklar, A. (1959). “Fonctions de répartition à n dimensions et leurs marges”. *Publications de l’Institut Statistique de l’Université de Paris* 8, pp. 229–231.
- Smith, R. L. (1985). “Maximum likelihood estimation in a class of nonregular cases”. *Biometrika* 72.1, pp. 67–90.
- Surles, J. G. e Padgett, W. J. (1998). “Inference for $P(Y < X)$ in the Burr type X model”. *Journal of Applied Statistical Sciences* 7.4, pp. 225–238.
- Taleb, N. N. (2020). *Statistical Consequences of Fat Tails (Technical Incerto Collection)*.
- Wang, L., Lin, H., Ahmadi, K. e Lio, Y. (2021). “Estimation of stress-strength reliability for multicomponent system with Rayleigh data”. *Energies* 14.23, p. 7917.