



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

**Seleção de ativos financeiros com base na estimativa
de probabilidades de *stress-strength*: o caso de
retornos seguindo distribuições GEV e TGEV**

por

Melquisadec de Souza Oliveira

Brasília, Julho de 2024

Seleção de ativos financeiros com base na estimativa de probabilidades de *stress-strength*: o caso de retornos seguindo distribuições GEV e TGEV

por

Melquisadec de Souza Oliveira

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

Coorientador: Prof. Dr. Felipe S. Quintino

Brasília, Julho de 2024

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 apresentado para:

Prof. Dr. Pushpa N. Rathie
Orientador, EST/UnB

Prof. Dr. Felipe S. Quintino
Co-orientador:, EST/UnB

Prof. Dr. Roberto Vila
Membro interno, EST/UnB

Prof. Dr. José Aílton Alencar Andrade
Membro externo, DEMA/UFC

Prof. Dra. Terezinha Késsia de Assis Ribeiro
Suplente, EST/UnB

Agradecimentos

Agradeço primeiramente a Deus por mais essa conquista, a conclusão desta dissertação marca um momento importante na minha jornada acadêmica. Este trabalho não seria possível sem o apoio, orientação e incentivo de várias pessoas às quais sou profundamente grato. A primeira pessoa a quem dedico essa dissertação e o meu carinho é a minha esposa Rebeca Klamerick Lima, que sempre me acompanhou, me deu conselhos e sempre me ajudou nesta jornada sem ela eu jamais teria conseguido.

Gostaria de expressar minha mais sincera gratidão aos meus orientadores, Prof. Dr. Pushpa Rathie e Prof. Dr. Felipe Quintino, por sua orientação, paciência e por ter me aceitado no grupo de estudo, logo no início do mestrado. sou extremamente grato pela oportunidade de aprender com especialistas tão dedicados e experientes. Externo também a minha gratidão aos pesquisadores deste grupo que participam com o professor Rathie e que de alguma forma colaboraram com seus conhecimentos e revisões neste trabalho: São eles o professor Dr. Tiago Alves da Fonseca, Dr. Luan Ozelim, professor Dr. Helton Saulo e professor Dr. Dióscoros Aguiar.

Agradeço também a banca avaliadora de qualificação composta pelos professores Dr. Roberto Vila e Dra. Terezinha Ribeiro do departamento de Estatística da universidade de Brasília, por suas contribuições neste trabalho, e desde já agradeço também ao professor Dr. José Ailton Alencar Andarade da Universidade Federal do Ceará que aceitou o convite juntamente com os professores já citados para compor minha banca de defesa final da dissertação.

Sou muito grato aos professores do Programa de Pós-Graduação em Estatística da universidade de Brasília que sempre demonstrou uma atenção e uma dedicação conosco durante esse processo de formação, gostaria também de agradecer aos meus colegas de curso e amigos, pelo

apoio constante, pelas discussões e pela amizade que tornaram esta jornada mais leve.

Aos meus familiares, especialmente aos meus pais Deli Neres e Antônia, por seu amor incondicional e compreensão ao longo desta trajetória. Vocês sempre foram minha fonte de força e inspiração, e sou eternamente grato por tudo que fizeram por mim.

Por fim, agradeço a todas as pessoas que, direta ou indiretamente, contribuíram para a realização deste trabalho eu os dedico a vocês. A todos, meu mais sincero muito obrigado.

Dedico esta dissertação ao meu querido pai Deli Neres de Oliveira, que sempre me motivou a lutar por aquilo em que acredito, inclusive para fazer o mestrado infelizmente ele não está mais aqui para ver esta conquista, mas sempre estará nas minhas memórias como um grande herói.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

A análise do comportamento estatístico dos retornos dos ativos mostrou-se uma abordagem interessante para realizar a seleção de ativos. Neste trabalho, exploramos uma abordagem de confiabilidade *stress-strength* para realizar a seleção de ativos com base em probabilidades do tipo $P(X < Y)$ quando tanto X quanto Y seguem a distribuição dos valores extremos generalizada (GEV) ou a distribuição transmutada GEV (TGEV). Derivamos novas relações analíticas e de forma fechada em termos da função de valor extremo \mathbb{H} , que foram obtidas sob mínimas restrições de parâmetros quando comparadas com resultados semelhantes na literatura. Para mostrar o desempenho de nossos resultados, incluímos estudos de simulações de Monte-Carlo e investigamos a aplicação da medida de confiabilidade $P(X < Y)$ na seleção de ativos financeiros com retornos caracterizados pelas distribuições X e Y . Assim, em vez de adotar a abordagem convencional de comparar os valores esperados de X e Y com base na teoria de portfólio moderna, exploramos a métrica $P(X < Y)$ como uma abordagem alternativa para avaliar retornos.

Palavras-chaves: Confiabilidade *stress-strength*; funções \mathbb{H} ; distribuição dos valores extremos generalizados, distribuição TGEV.

Abstract

Analyzing the statistical behavior of the assets' returns has shown to be an interesting approach to perform asset selection. In this work, we explore a stress-strength reliability approach to perform asset selection based on probabilities of the type $P(X < Y)$ when both X and Y follow a Generalized Extreme Value (GEV) or a Transmuted GEV (TGEV) distribution. We derive new analytical and closed form relations in terms of the extreme value \mathbb{H} -function, which have been obtained under fewer parameter restrictions while compared to similar results in the literature. To show the performance of our results, we include a Monte-Carlo simulation study and we investigate the application of the reliability measure $P(X < Y)$ in selecting financial assets with returns characterized by the distributions X and Y . Thus, instead of adopting the conventional approach of comparing the expected values of X and Y based on modern portfolio theory, we explore the metric $P(X < Y)$ as an alternative approach for assessing returns.

Keywords: stress-strength reliability; \mathbb{H} -functions; generalized extreme value distribution, TGEV distribution.

Sumário

1	Introdução	1
2	Preliminares	7
2.1	Teoria dos Valores Extremos	7
2.2	Funções Especiais	12
2.3	Probabilidade de Stress-Strength	14
2.4	Problemas Abordados	16
3	Seleção de ativos financeiros seguindo distribuição GEV	18
3.1	Resultados Analíticos	19
3.2	Estimação	24
3.3	Estudo de Simulações	26
3.4	Aplicações	29
3.4.1	Mercado de Ações Brasileiro	30
3.4.2	Fibras de carbono	36
4	Seleção de ativos financeiros via TGEV	39
4.1	Resultados Analíticos	40
4.2	Estimadores e Estudo de Simulações	48
4.3	Aplicações no Mercado de Ações	54
5	Conclusões	59
A	Artigo: Quintino et al. (2024)	60

B Artigo: Oliveira et al. (2024)	62
Referências Bibliográficas	64

Capítulo 1

Introdução

O trabalho de Markowitz (1959) permanece como um dos trabalhos pioneiros na teoria de portfólio. Rubinstein (2002) detalha as principais contribuições de Markowitz para a crescente teoria moderna de portfólio (Carteira de investimento). Além de fornecer uma descrição criteriosa do modelo de escolha de portfólio de M. (1952), o trabalho em Markowitz (1959) traçou direções de pesquisas futuras exploradas por outros cientistas, como a sugestão de substituir a variância pela semivariância como medida de risco, recomendações na maximização da utilidade logarítmica esperada do retorno e no delineamento de um modelo de mercado desenvolvido em profundidade por Sharpe (1963).

Contudo, o quadro de seleção de carteiras de Markowitz baseia-se nas premissas de que se pode medir a variância do retorno dos ativos, conforme sublinhado pela afirmação de que exemplos de variâncias rapidamente crescentes são de interesse sobretudo acadêmico (Markowitz, 1959). Essa restrição traz preocupações relacionadas à adequação da estrutura em um cenário em que distribuições de cauda pesada melhor modelam o retorno dos ativos, ao permitir volatilidades que variam rapidamente resultantes de eventos extremos.

Podemos encontrar evidências na literatura de que os dados financeiros são melhor modelados por processos α -estáveis (alternativa de cauda pesada ao movimento browniano (Bachelier, 1900)) ou por modelos de séries temporais de cauda pesada (Embrechts, Klüppelberg e Mikosch, 2013; Taleb, 2020). Embora neste trabalho nos baseemos na hipótese geral de que os retornos logarítmicos em dados financeiros seguem um processo α -estável com parâmetro $0 < \alpha < 2$ (M.,

1967) (o que implica que a variância do retorno é indefinida), sem perda de generalidade, usamos resultados da Teoria de Valores Extremos (TVE) em relação à distribuição de Valores Extremos Generalizados (GEV) como uma alternativa às distribuições α -estáveis (Jenkinson, 1955). Esta abordagem pode ser considerada válida uma vez que a distribuição GEV pode ter comportamento de cauda pesada, ou seja, não possui um decaimento exponencial nas caudas, dependendo do parâmetro de forma.

Do ponto de vista econômico, é bem conhecido que os retornos extremos das ações nos mercados de ações podem ter implicações importantes para a gestão do risco financeiro e vários estudos aplicaram com sucesso a distribuição GEV para modelar dados financeiros (Cirillo e Taleb, 2016). Por exemplo, Gettinby et al. (2004) caracterizou a distribuição de retornos extremos para um índice de ações do Reino Unido ao longo dos anos de 1975 a 2000. Os autores consideraram a adequação de diversas distribuições, sendo os máximos e mínimos semanais de retorno diário melhor modelados pela distribuição GEV e pelas distribuições Logística Generalizada. Para o caso do Reino Unido, a Logística Generalizada foi uma escolha melhor em geral. Por outro lado, o modelo GEV apresentou capacidade de modelagem semelhante, bem como apresenta algumas propriedades importantes. Hussain e Li (2015) estudaram a distribuição dos retornos diários extremos do Índice Composto da Bolsa de Valores de Xangai. Os autores modelaram os retornos do índice da Bolsa de Valores de Xangai com base nos dados de 1991 a 2013, o que indicou que a distribuição Logística Generalizada se ajusta melhor à série de mínimos e que a distribuição GEV se ajusta melhor à série de máximos dos retornos no mercado de ações chinês.

TVE é um ramo de Probabilidade e Estatística que trata da modelagem de eventos extremos relacionados a máximos e mínimos de amostras aleatórias independentes. As aplicações desta teoria são encontradas em finanças (Gencay e Selçuk, 2004), catástrofes naturais e falhas de equipamentos, entre outras. Os livros Embrechts, Klüppelberg e Mikosch (2013), Galambos (1987), Haan e Ferreira (2006) e Resnick (2008) fornecem ampla cobertura que permite um estudo detalhado sobre TVE.

Além disso, a TVE fornece uma base teórica e uma estrutura para lidar com desvios extremos da média das funções de distribuição (DFs), restringindo o comportamento dos DFs nas caudas.

A EVT se concentra no estudo das possíveis distribuições limitantes e suas propriedades para o máximo normalizado.

Especificamente, sejam X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias (VAs) independentes e identicamente distribuídas (*i.i.d.*) com função de distribuição comum F e defina $M_n = \max\{X_1, \dots, X_n\}$. A TVE estabelece propriedades de F e das possíveis funções de distribuição não degeneradas G que satisfazem

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad \forall x \in C(G), \quad (1.1)$$

para sequências de constantes $a_n \dots$ e $b_n \in \mathbb{R}$ ($n = 1, 2, \dots$), adequadamente escolhidas, onde $C(G)$ denota o conjunto de pontos de continuidade de G .

As possíveis funções de distribuição G satisfazendo (1.1) são conhecidas desde os trabalhos de Fisher e Tippett (1928) e têm sido extensivamente estudadas por diversos autores desde então. Essas funções de distribuição também são conhecidas como *leis de estabilidade do máximo* (ou distribuições max-estáveis) e só podem ser de três tipos bem conhecidos: Fréchet, Weibull Reversa ou Gumbel.

A TVE foi utilizada por Goncu et al. (2012) para modelar o comportamento de retorno extremo da Bolsa de Valores de Istambul, Turquia. Os autores consideraram as distribuições Gumbel, Fréchet e Weibull Reversa para modelar retornos extremos em diferentes horizontes de investimento. Os seus resultados indicam que quando o Valor em Risco (VaR) é calculado com as distribuições propostas, os resultados do *backtesting* indicam que a TVE proporciona uma gestão de risco superior em todos os subintervalos considerados em comparação com a estimativa do VaR sob o pressuposto de uma distribuição normal.

Para aplicações estatísticas, as distribuições max-estáveis podem ser resumidas em uma única função de distribuição chamada GEV. Especificamente, a distribuição GEV tem a função de distribuição acumulada (FDA) dada por

$$G(x) = \begin{cases} \exp\left(-\left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right), & 1 + \gamma x > 0 \text{ e } \gamma \neq 0, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & x \in \mathbb{R} \text{ e } \gamma = 0, \end{cases} \quad (1.2)$$

onde $\gamma > 0$ é o parâmetro de forma, $\mu \in \mathbb{R}$ é o parâmetro de locação e $\sigma > 0$ é o parâmetro de escala. Para mais detalhes da distribuição GEV e suas propriedades e aplicações, sugerimos a leitura de Haan e Ferreira (2006).

A distribuição GEV transmutada (TGEV) foi proposta por Aryal e Tsokos (2009) e teve propriedades teóricas estudadas e aplicações em modelagens feitas por Nascimento, Bourguignon e Leão (2016) e Otiniano, De Paiva e Neto (2019)

Essencialmente, a distribuição TGEV é uma modificação da distribuição GEV $G(x)$, e tem FDA dada por

$$F(x) = (1 + \lambda)G(x) - \lambda[G(x)]^2, \quad |\lambda| < 1. \quad (1.3)$$

Propriedades tais como momentos, quantis, comportamento de caudas e estatísticas de ordem foram estudadas por Otiniano, De Paiva e Neto (2019). Os autores também aplicaram essa distribuição na modelagem de retornos dos preços de ações. O parâmetro λ da distribuição TGEV permite maior flexibilidade para o ajuste da distribuição nos dados em situações onde os extremos podem ter uma tendência mais acentuada.

Nosso interesse neste trabalho é a confiabilidade (ou probabilidade) *stress-strength* (abreviada por SSR) que, em termos gerais, consiste no estudo da probabilidade de falha de um sistema ou componente a partir da comparação de um *stress* (tensão) aplicado com um *strength* (resistência) do sistema. Sejam o *stress* Y e o *strength* X VAs contínuas e independentes, com função de densidade de probabilidade (FDP) f_Y e FDA F_X , respectivamente. A probabilidade SSR (também chamada confiabilidade SSR) é definida como

$$R = P(X < Y) = \int_{-\infty}^{\infty} F_X(x)f_Y(x)dx. \quad (1.4)$$

Existem diversas aplicações desta teoria como em Engenharia e Manufatura, Aeroespacial e Defesa, Indústria Automotiva, Setor Energético, Saúde e Eletrônica, entre outros. Veja Kotz, Lumelskii e Pensky (2003) para mais detalhes.

A probabilidade SSR para as distribuições extremas de Fréchet, Weibull e Gumbel tem sido amplamente estudada na literatura. Nadarajah (2003a) considerou a classe de distribuições

de valores extremos e derivou as formas correspondentes para a confiabilidade R em termos de funções especiais. Os limites de confiança para R envolvendo modelos Weibull foram apresentados em Krishnamoorthy e Lin (2010). Kundu e Raqab (2009) propôs um estimador de máxima verossimilhança modificado de R e obteve a distribuição assintótica do estimador, que foram utilizados para construir o intervalo de confiança de R . Os resultados anteriores de R para distribuição Weibull foram generalizados por Nojosa e Rathie (2020), onde R foi expresso em termos de funções H . A estimativa bayesiana de R para distribuições de Fréchet e Weibull também foi explorada Abbas e Tang (2014) e Jia, Nadarajah e Guo (2017).

Ao aplicar essa metodologia SSR 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 duas diferentes ações. Em resumo, quando X e Y representam VAs de retorno e $R < 1/2$, o investidor deve escolher a ação da variável X . Se $R > 1/2$, ocorre o oposto. O caso $R = 1/2$ é inconclusivo. Diversos autores exploraram a estrutura de SSR na modelagem de dados do mercado financeiro (Lima et al., 2024; Rathie e Ozelim, 2017; Oliveira et al., 2024; Quintino et al., 2024).

O objetivo central do nosso trabalho foi propor a estimação da confiabilidade através de uma função \mathbb{H} , esta abordagem permite a estimação dos parâmetros sem a necessidade de impor restrições no espaço paramétrico da distribuição de X e Y abordada neste estudo. Os resultados apresentados aqui foram publicados nos artigos de Quintino et al. (2024) e Oliveira et al. (2024). Assim, a presente dissertação está organizada em quatro capítulos.

No Capítulo 2 apresentamos notações, conceitos e propriedades básicas que serão utilizadas no desenvolvimento dos capítulos seguintes. Em particular, apresentamos uma revisão das distribuições GEV e TGEV, assim como apresentamos as funções \mathbb{H} dos valores extremos, recentemente introduzida por Rathie et al. (2023). Conexões de tais funções com as bem conhecidas funções H também são abordadas. Para maiores informações sobre funções H veja (Mathai, Saxena e Haubold, 2010b).

No Capítulo 3, estudamos em detalhes os resultados apresentados no nosso artigo (Quintino et al., 2024), tema central deste trabalho. O capítulo é dividido em três seções de resultados. Na Seção 3.1, apresentamos a obtenção analítica da expressão de $R = P(X < Y)$ quando X e

Y possuem distribuição GEV independentes. Tais resultados são obtidos em termos da função dos valores extremos \mathbb{H} . Nas Seções 3.2 e 3.3, estimadores para R são apresentados e seus desempenhos são avaliados através de um estudo de simulações. Na Seção 3.4 apresentamos aplicação na modelagem de ativos financeiros.

No Capítulo 4, os resultados do Capítulo 3 são generalizados para o caso da distribuição TGEV. Resultados analíticos são apresentados na Seção 4.1. Um método de estimação dos parâmetros em duas etapas é proposto e a estimação da probabilidade R é avaliada através de um estudo de simulações na Seção 4.2, que confirmam os bons resultados do novo procedimento de estimação proposto. Comparação da modelagem via distribuições GEV e TGEV e aplicações da teoria de SSR na comparação de ativos financeiros é feita na Seção 4.3.

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

Os resultados obtidos neste trabalho para distribuição GEV foi publicado em formato de artigo na revista *AIMS Mathematics* veja Quintino et al. (2024) e os resultados obtidos para distribuição TGEV foi publicado em Oliveira et al. (2024) na Revista *Entropy*.

Capítulo 2

Preliminares

Neste capítulo, apresentamos conceitos que serão utilizados ao longo desta dissertação. Iniciamos a Seção 2.1 apresentando um referencial teórico da Teoria dos Valores Extremos (EVT). Também definimos a FDA e FDP para a distribuição GEV e para a distribuição TGEV.

A Seção 2.2 é dedicada a apresentação das funções especiais utilizadas nesse trabalho. São estas a função dos valores extremos \mathbb{H} , introduzida recentemente por Rathie et al. (2023), e a famosa classe de funções especiais \mathbb{H} (veja Mathai, Saxena e Haubold (2010a) para maiores detalhes).

A Seção 2.3 apresenta a fórmula da confiabilidade de *stress-strength* (SSR) que será utilizada nos capítulos seguintes. Na Seção 2.4 são apresentados os problemas abordados nos capítulos seguintes.

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

2.1 Teoria dos Valores Extremos

Os estudos sobre teoria dos valores extremos obteve contribuições significativas de diversos autores, o trabalho de Fisher e Tippett (1928) foi um dos pioneiros, realizando um trabalho fundamental em 1925 ao introduzir a distribuição de Fisher-Tippett-Gnedenko. Essa distribuição foi proposta como uma ferramenta para modelar dados que contenha valores extremos. Em 1958,

o trabalho de Gumbel (1958) desempenhou um papel crucial no avanço da teoria dos valores extremos ao propor a distribuição de Gumbel, destinada a modelar máximos e mínimos de valores extremos de um conjunto de dados. As contribuições de Gumbel foram particularmente relevantes nas aplicações práticas da teoria, sendo amplamente utilizadas em campos como hidrologia (veja Queiroz e Chaudhry (2010)) e meteorologia (veja Hartmann, Moala e Mendonça (2011)).

Outro marco significativo ocorreu em 1943, quando Gnedenko (1943) estabeleceu as bases teóricas da teoria dos extremos. Sua pesquisa concentrou-se na distribuição limite das somas parciais de variáveis independentes, identificando três tipos principais de distribuições de valores extremos: Gumbel, Fréchet e Weibull. Essa classificação proporcionou um arcabouço teórico essencial para compreender e modelar eventos extremos em diversas áreas. A EVT ganhou um grande impulso a partir da década de 1950 quando uma barragem na Holanda se rompeu provocando a morte de aproximadamente 1800 civis, a partir de então o governo holandês iniciou um comitê para estudos ligados a teoria dos valores extremos que visava abordagem de eventos raros. Porém a EVT não limita-se apenas a eventos climáticos podendo ser aplicadas, por exemplo, em áreas conforme visto em Lima et al. (2024), Rathie e Ozelim (2017) e Quintino et al. (2024).

A EVT fornece uma base teórica e uma estrutura para lidar com desvios extremos da média das funções de distribuição ao restringir o comportamento das caudas. A teoria se concentra no estudo das possíveis distribuições limites para o máximo normalizado e suas propriedades. Especificamente, sejam X_1, X_2, \dots, X_n uma sequência de VAs *i.i.d.* com função de distribuição comum F e defina $M_n = \max\{X_1, \dots, X_n\}$. A EVT se preocupa com as propriedades de F e das possíveis funções de distribuição não degeneradas G , ou seja, existe uma distribuição limite G que captura a variação e o comportamento desses máximos que satisfazem

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad \forall x \in C(G), \quad (2.1)$$

para sequências de constantes $a_n > 0$ and $b_n \in \mathbb{R}$ ($n = 1, 2, \dots$), adequadamente escolhidas, onde $C(G)$ denota o conjunto de pontos de continuidade de G .

As possíveis funções de distribuição G que satisfazem (2.1) são conhecidas desde Fisher e Tippett (1928) e têm sido extensivamente estudadas por vários autores desde então. Estas distribuições também são conhecidas como *leis max-estável* (ou distribuições de estabilidade do máximo) e só podem ser de três tipos bem conhecidos: Fréchet, Weibull ou Gumbel (veja Resnick (2008)).

Distribuição GEV

A distribuição GEV é uma ferramenta estatística essencial para modelar e entender eventos extremos em uma variedade de campos, desde hidrologia e climatologia até finanças e engenharia. Na hidrologia, é frequentemente utilizada para modelar a distribuição de eventos de precipitação extrema e o consequente escoamento em rios e bacias hidrográficas. Em climatologia, ajuda a entender e prever padrões de clima extremo, como tempestades e ondas de calor. Nas finanças, é aplicada para avaliar riscos de perdas extremas em investimentos e portfólios. Além disso, é aplicada para modelar eventos sísmicos e climáticos extremos (veja Pisarenko et al. (2014), Tolikas (2008) e Yoon, Kim e Yoo (2013)).

Originada da teoria dos valores extremos, introduzida por Jenkinson (1955), a distribuição GEV combina os três possíveis tipos de distribuição de valores extremos (Fréchet, Weibull e Gumbel) em uma única forma. A GEV é uma família de distribuições que descreve a probabilidade de ocorrência de eventos extremos em uma determinada distribuição de dados.

A distribuição é indexada por três parâmetros principais, (localização, escala e forma), a distribuição GEV permite capturar diferentes características de dados, como a tendência central, a variabilidade e a cauda da distribuição. Através da estimação desses parâmetros, é possível modelar uma ampla gama de cenários extremos, por exemplo, cheias, secas, picos de temperatura e eventos financeiros raros (veja Engeland, Hisdal e Frigessi (2004), Hasan, Radi e Kassim (2012), Morrison e Smith (2002) e Saputri et al. (2023)).

A distribuição GEV é obtida tomando FDA G definido como

$$G(x; \mu, \sigma, \gamma) = \exp \left\{ - \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}} \right\}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0, \quad (2.2)$$

em que $\mu \in \mathbb{R}$ é o parâmetro de localização, $\sigma \in \mathbb{R}_+$ é o parâmetro de escala e $\gamma \in \mathbb{R}$ é o parâmetro de forma. Denotaremos uma VA com distribuição GEV por $X \sim GEV(\mu, \sigma, \gamma)$. A FDP de X é dada por:

$$g(x; \mu, \sigma, \gamma) = G(x_i, \gamma_i, \mu, \sigma) \frac{1}{\sigma} \left[1 + \frac{\gamma}{\sigma} (x - \mu) \right]^{-\frac{1}{\gamma}-1}, \quad 1 + \frac{\gamma}{\sigma} (x - \mu) > 0. \quad (2.3)$$

A Figura 2.1 mostra o comportamento da FDP g para algumas escolhas de parâmetros. Observe que o parâmetro de locação desloca a curva para a direita ou esquerda, a escala controla a dispersão e o parâmetro de forma controla a taxa de decaimento da cauda e a forma da distribuição de acordo com o sinal deste parâmetro.

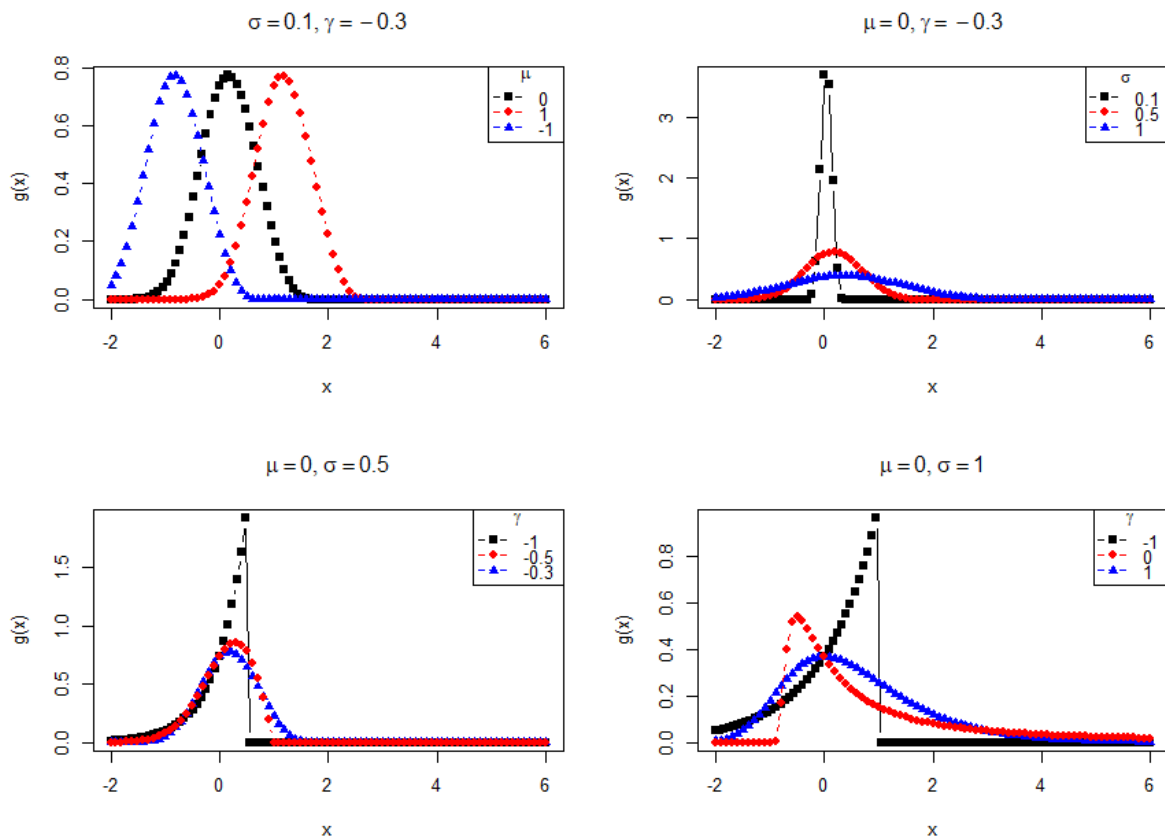


Figura 2.1: FDP da distribuição GEV

Distribuição GEV Trasmutada

A trasmutação é uma técnica que consiste em adicionar um novo parâmetro à distribuição que permitem maior flexibilidade na modelagem de dados, desde que a FDA possua inversa. A principal vantagem de adicionar o parâmetro λ na distribuição TGEV reside na sua capacidade de ajustar a cauda da distribuição de forma mais precisa, permitindo que capture melhor a frequência e a magnitude de eventos extremos raros. A distribuição TGEV foi proposta por Aryal e Tsokos (2009) e teve propriedades teóricas estudadas e aplicações em modelagens feitas por Nascimento, Bourguignon e Leão (2016) e Otiniano, De Paiva e Neto (2019). A TGEV é obtida da seguinte forma: dada a distribuição GEV $G(x)$, a função de distribuição trasmutada F é dada por:

$$F(x) = (1 + \lambda)G(x) - \lambda[G(x)]^2, \quad |\lambda| < 1. \quad (2.4)$$

A FDA e a FDP da distribuição TGEV são dadas respectivamente, por

$$F(x; \mu, \sigma, \gamma, \lambda) = \begin{cases} \exp(-w^{-1/\gamma}) [(1 + \lambda) - \lambda \exp(-w^{-1/\gamma})], & \gamma \neq 0; \\ \exp(-\exp(-(\frac{x-\mu}{\sigma})) [(1 + \lambda) - \lambda \exp(-(\frac{x-\mu}{\sigma})))] & , \gamma = 0, \end{cases} \quad (2.5)$$

e

$$f(x; \mu, \sigma, \gamma, \lambda) = \begin{cases} \left[\frac{(w)^{-1-1/\gamma} \exp(-w^{-1/\gamma})}{\sigma} \right] [(1 + \lambda) - 2\lambda \exp(-w^{-1/\gamma})], & \gamma \neq 0; \\ \frac{\exp(-(\frac{x-\mu}{\sigma})) \exp(-\exp(-(\frac{x-\mu}{\sigma})))}{\sigma} [(1 + \lambda) - 2\lambda \exp(-\exp(-(\frac{x-\mu}{\sigma})))] & , \gamma = 0, \end{cases} \quad (2.6)$$

onde $\mu \in \mathbb{R}$ é o parâmetro de localização, $\sigma \in \mathbb{R}_+$ é o parâmetro de escala e $\gamma \in \mathbb{R}$ é o parâmetro de forma e λ é um parâmetro no intervalo $[-1, 1]$. A distribuição está definida no suporte onde $w = 1 + (\gamma(x - \mu)/\sigma)$ e $\text{supp}(F) = \text{supp}(f) = \{x; w > 0\}$. Denotaremos por $X \sim TGEV(\mu, \sigma, \gamma, \lambda)$ uma VA com FDA (2.5). A Figura 2.2 mostra o comportamento da FDP da TGEV para algumas escolhas de parâmetros. A seleção apropriada de parâmetros pode ser explorada para representar os modelos extremos como casos particulares da distribuição TGEV, conforme mostrado na Tabela 2.1.

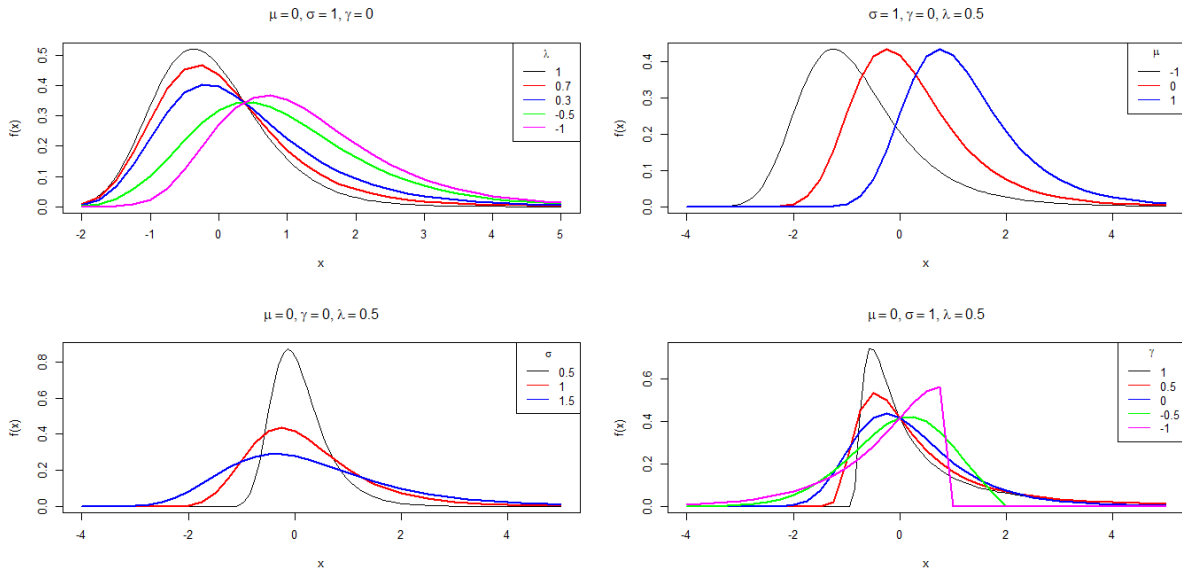


Figura 2.2: Densidade da TGEV para algumas escolhas de parâmetros.

Tabela 2.1: Casos particulares do modelo TGEV.

Distribution	FDA
GEV	$F(x; \mu, \sigma, \gamma, 0)$
Fréchet	$F((x - 1)/\gamma; 0, 1, \gamma, 0), \gamma > 0$
Reversed Weibull	$F(-(1 + x)/\gamma; 0, 1, \gamma, 0), \gamma < 0$
Gumbel	$F(x; 0, 1, 0, 0)$

2.2 Funções Especiais

A função \mathbb{H} foi criada para lidar com integrais complexas envolvendo exponenciais e polinômios, que são comuns em problemas de valores extremos. Em nosso estudo, utilizamos a função \mathbb{H} para reescrever a forma de estimar a confiabilidade sem impor restrições rígidas aos parâmetros das distribuições envolvidas. Recentemente, Rathie et al. (2023) introduziu a função de valor extremo \mathbb{H} como:

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^\infty y^{a_6} \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy, \quad (2.7)$$

onde $\Re(a_1), \Re(a_2), \Re(a_4) \in \mathbb{R}_+, a_3, a_5 \in \mathbb{C}$, nem $\Re(a_1)$ e $\Re(a_2)$ podem ser iguais a zero ao mesmo tempo, $\Re(a_6) > -1$ quando $a_1 \neq 0$ ou $a_1 = 0$ e $\text{ sinal}(a_3) = \text{ sinal}(a_5)$, $\Re(a_6) < -1$

quando $a_1 = 0$ e $\text{ sinal}(a_3) \neq \text{ sinal}(a_5)$. Nesta notação, \mathbb{R} , \mathbb{C} e \Re denotam os números reais, os números complexos e a parte real de um número complexo, respectivamente.

Note que (2.7) generaliza alguns casos importantes da função H (cf. (Mathai, Saxena e Haubold, 2010a)) definida por

$$H_{p,q}^{m,n} \left[z \left| \begin{array}{c} (a_1, A_1), \dots, (a_p, A_p) \\ (b_1, B_1), \dots, (b_q, B_q) \end{array} \right. \right] = \frac{1}{2\pi i} \int_L \frac{\prod_{k=1}^m \Gamma(b_k + B_k s) \prod_{j=1}^n \Gamma(1 - a_j - A_j s)}{\prod_{k=m+1}^q \Gamma(1 - b_k - B_k s) \prod_{j=n+1}^p \Gamma(a_j + A_j s)} z^{-s} ds, \quad (2.8)$$

onde $0 \leq m \leq q$, $0 \leq n \leq p$ (m e n não pode ser simultaneamente zero), $A_j > 0$ ($j = 1, \dots, p$), $B_k > 0$ ($k = 1, \dots, q$), a_j e b_k são números complexos tais que não há pólos de $\Gamma(b_k + B_k s)$ ($k = 1, \dots, m$) coincidem com pólos de $\Gamma(1 - a_j - A_j s)$ ($j = 1, \dots, n$). L é um contorno adequado $w - i\infty$ para $w + i\infty$, $w \in \mathbb{R}$, separando os pólos dos dois tipos mencionados acima. Para mais detalhes, consulte Mathai, Saxena e Haubold (2010b).

Um caso especial importante desta função é obtido tomando $a_4 = 0$, que representa um limite superior (ou inferior) para seu valor dependendo do sinal de a_5 . Este caso é, portanto, um valor extremo da função e pode ser escrito em termos da função H como proposto por Rathie et al. (2023):

$$\begin{aligned} \mathbb{H}(a_1, a_2, a_3, 0, a_5, a_6) &= \int_0^\infty y^{a_6} \exp\{-a_1 y - a_2^{a_5} y^{a_3 a_5}\} dy \\ &= \frac{1}{a_2^{(1+a_6)/a_3} a_3 a_5} H_{1,1}^{1,1} \left[a_1 a_2^{-1/a_3} \left| \begin{array}{c} (1 - \frac{(1+a_6)}{a_3 a_5}, \frac{1}{a_3 a_5}) \\ (0, 1) \end{array} \right. \right] \\ &= \frac{1}{a_1^{a_6+1}} H_{1,1}^{1,1} \left[\left(\frac{a_2}{a_1^{a_3}} \right)^{a_5} \left| \begin{array}{c} (-a_6, a_3 a_5) \\ (0, 1) \end{array} \right. \right], \end{aligned} \quad (2.9)$$

quando $\text{ sinal}(a_3) = \text{ sinal}(a_5)$ e:

$$\begin{aligned}
\mathbb{H}(a_1, a_2, a_3, 0, a_5, a_6) &= \frac{1}{a_2^{(1+a_6)/a_3} |a_3 a_5|} H_{0,2}^{2,0} \left[a_1 a_2^{-1/a_3} \mid \begin{array}{c} - \\ (0, 1), \left(\frac{1+a_6}{a_3 a_5}, \frac{1}{a_3 a_5} \right) \end{array} \right] \\
&= \frac{1}{a_1^{a_6+1}} H_{0,2}^{2,0} \left[\left(\frac{a_2}{a_1^{a_3}} \right)^{a_5} \mid \begin{array}{c} - \\ (0, 1), (1 + a_6, |a_3 a_5|) \end{array} \right], \quad (2.10)
\end{aligned}$$

de outra forma.

Temos mais dois casos especiais dessa função que serão utilizados para auxiliar na estimação dos parâmetros da distribuição TGEV abordados no Capítulo 4. Esses casos:

$$\mathbb{H}(a, b, c, 0, 1, 0) = \int_0^\infty \exp\{-ay - by^c\} dy = \frac{1}{b^{1/c} c} H_{1,1}^{1,1} \left[\frac{a}{b^{1/c}} \mid \begin{array}{c} \left(\frac{c-1}{c}, \frac{1}{c} \right) \\ (0, 1) \end{array} \right] \quad (2.11)$$

se $a > 0$, $b > 0$ e $c > 0$ e

$$\mathbb{H}(a, b, c, 0, 1, 0) = \int_0^\infty y^{d-1} \exp\{-ay - by^c\} dy = \frac{1}{a^d} H_{1,1}^{1,1} \left[ba^{-c} \mid \begin{array}{c} (1-d, c) \\ (0, 1) \end{array} \right], \quad (2.12)$$

se $a > 0$, $b > 0$, $c > 0$ e $d > 0$.

2.3 Probabilidade de Stress-Strength

Segundo Kotz e Nadarajah (2000) a confiabilidade em um sistema de componente é obtida através do cálculo da probabilidade de que a resistência (*Strength*) Y de um componente ou sistema seja maior que o estresse (*Stress*) X aplicado a ele. Sejam X e Y VAs com função de densidade conjunta $f_{X,Y}(x, y)$. Definimos a SSR como sendo

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

Quando X e Y são VAs independentes, a densidade conjunta $f_{X,Y}(x, y)$ é fatorada como

produto de FDPs marginais

$$f_{X,Y}(x, y) = f_X(x)f_Y(y),$$

onde $f_X(x)$ e $f_Y(y)$ denotam a FDP de X e Y , respectivamente. Nesse caso, obtemos

$$\begin{aligned} R &= \int_{-\infty}^{+\infty} \int_{-\infty}^y f_X(x)f_Y(y)dx dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^y f_X(x)dx \right) f_Y(y)dy \\ &= \int_{-\infty}^{+\infty} F_X(y)f_Y(y)dy, \end{aligned} \tag{2.13}$$

Para todo par (x, y) nos suportes das distribuições, note que F_X denota a FDA de X e f_Y denota a FDP de Y .

Exemplo 2.1. Considere a distribuição de Fréchet que possui FDA e FDP dadas, respectivamente, por

$$F(x; \mu, \sigma, \alpha) = \exp \left(- \left(\frac{x - \mu}{\sigma} \right)^{-\alpha} \right), \quad x > \mu, \quad \mu \in \mathbb{R}, \quad \sigma > 0, \quad \alpha > 0,$$

e

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

Sejam $X \sim \text{Fréchet}(\mu_x, \sigma_x, \alpha_x)$ e $Y \sim \text{Fréchet}(\mu_y, \sigma_y, \alpha_y)$ VAs independentes. A probabilidade SSR é dada por: $R = \mathbb{P}(X < Y) = \int_{-\infty}^{+\infty} F_X(u; \mu_x, \sigma_x, \alpha_x) f_Y(u; \mu_y, \sigma_y, \alpha_y) du$.

Suponha que $\mu_x = \mu_y = 0$. Então

$$R = \int_0^{\infty} \exp \left(- \left(\frac{u}{\sigma_x} \right)^{-\alpha_x} - \left(\frac{u}{\sigma_y} \right)^{-\alpha_y} \right) \frac{\alpha_y}{\sigma_y} \left(\frac{u}{\sigma_y} \right)^{-(\alpha_y+1)} du \tag{2.14}$$

Se $\alpha_x = \alpha_y = \alpha$ e substituindo $v = \left(\frac{u}{\sigma_y} \right)^{-\alpha}$, obtemos

$$\begin{aligned} R &= \int_0^{\infty} \exp \left(- \left(\frac{\sigma_y v^{-1/\alpha}}{\sigma_x} \right)^{-\alpha} - v \right) dv \\ &= \int_0^{\infty} \exp \left(- \left(\frac{\sigma_y}{\sigma_x} \right)^{-\alpha} v - v \right) dv \end{aligned}$$

Logo,

$$\begin{aligned} R &= \int_0^\infty \exp\left(-\left[\left(\frac{\sigma_y}{\sigma_x}\right)^{-\alpha} + 1\right]v\right) dv \\ &= \frac{\sigma_x^\alpha}{\sigma_x^\alpha + \sigma_y^\alpha}. \end{aligned} \quad (2.15)$$

Observe que assumir a igualdade de parâmetros $\alpha_x = \alpha_y = \alpha$ em (2.14) é essencial para obtermos a expressão de R em (2.15). Utilizando a função dos valores extremos \mathbb{H} dada em (2.7) poderíamos obter uma forma para R sem assumir restrição dos parâmetros. Substituindo $w = \left(\frac{u}{\sigma_y}\right)^{-\alpha_y}$ em (2.14), obtemos:

$$\begin{aligned} R &= \int_0^\infty \exp\left(-\left(\frac{\sigma_y}{\sigma_x}w^{-1/\alpha_y}\right)^{-\alpha_x} - \omega\right) dw \\ &= \int_0^\infty \exp\left(-\omega - \left(\frac{\sigma_x}{\sigma_y}\right)^{\alpha_x} w^{\frac{\alpha_x}{\alpha_y}}\right) dw \\ &= \mathbb{H}\left(1, \left(\frac{\sigma_x}{\sigma_y}\right)^{\alpha_x}, \frac{\alpha_x}{\alpha_y}, 0, 1, 0\right). \end{aligned} \quad (2.16)$$

Note ainda que a representação de R dada em (2.16) não é única, ou seja, outras possíveis representações podem ser obtidas mudando a escolha dos parâmetros, perceba que um procedimento de estimação para R é proposto baseado na estimação dos parâmetros de X e Y via máxima verossimilhança e estimando R via propriedade de invariância. A propriedade de invariância do estimador de máxima verossimilhança (EMV) afirma que se $\hat{\theta}$ é o estimador de máxima verossimilhança de um parâmetro θ , então, para qualquer função mensurável g de θ , o estimador de máxima verossimilhança de $g(\theta)$ é $g(\hat{\theta})$.

2.4 Problemas Abordados

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

1. Estendemos os resultados de Nadarajah (2003b), calculando analiticamente a probabilidade R (2.13) quando X e Y são duas seguindo distribuição GEV independentes.
2. Os resultados do tópico 1 são generalizados para a distribuição TGEV.

3. Métodos de estimação de R são avaliados através de um estudo de simulações.
4. Aplicações dessas novas metodologias são feitas no contexto de comparação e seleção de ativos financeiros.

Os tópicos 1,3 e 4 são abordados no Capítulo 3. No Capítulo 4 são apresentados e discutidos o tópico 2, assim como 3 e 4 voltam a ser abordados.

Capítulo 3

Seleção de ativos financeiros seguindo distribuição GEV

Neste capítulo, estudamos a probabilidade SSR $R = \mathbb{P}(X < Y)$ quando X e Y são VAs independentes com distribuição GEV. A FDA e FDP da distribuição GEV são dadas como em (2.2) e (2.3), ou seja,

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

e

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

Nos Teoremas 3.1.1 e 3.1.3 que serão apresentados a seguir, estabelecemos formas simples para R em termos da função \mathbb{H} dos valores extremos, definida em (2.7) como sendo

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^\infty y^{a_6} \exp \{ -a_1 y - (a_2 y^{a_3} + a_4)^{a_5} \} dy. \quad (3.2)$$

Em geral, resultados semelhantes na literatura requerem restrições dos parâmetros envolvidos ou transformações nos conjuntos de dados. Nossos resultados não precisam de tais restrições.

Na Seção 3.2, um procedimento de estimação para R é proposto baseado na estimação dos parâmetros que indexam as distribuição das VAs X e Y via máxima verossimilhança e estimando

R via propriedade de invariância.

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 *SSR* na modelagem de dados reais. No primeiro caso, utilizamos a metodologia *SSR* na construção de 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, fibras de carbono de diferentes tamanhos são comparadas.

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

3.1 Resultados Analíticos

Nessa seção, apresentamos formas simples para a probabilidade *SSR*. Nossos resultados estão separados para parâmetros de forma com mesmos sinais.

Teorema 3.1.1. Sejam Y e X VAs independentes, respectivamente, com distribuição $GEV(\mu_y, \sigma_y, \gamma_y)$ e $GEV(\mu_x, \sigma_x, \gamma_x)$, $\mu_j \in \mathbb{R}$, $\sigma_j \in \mathbb{R}_+$, $\gamma_j \in \mathbb{R}$ ($\gamma_j \neq 0$), $j \in \{x, y\}$. Então

- Para $\gamma_j > 0$, $j \in \{x, y\}$:

$$R = P(X < Y) = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right), \quad (3.3)$$

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

$$R = P(X < Y) = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right). \quad (3.4)$$

- Para $\gamma_j < 0$, $j \in \{x, y\}$:

$$R = P(X < Y) = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right), \quad (3.5)$$

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

$$R = P(X < Y) = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right). \quad (3.6)$$

Em particular, se $\mu_y - \frac{\sigma_y}{\gamma_y} = \mu_x - \frac{\sigma_x}{\gamma_x}$, obtemos

$$R = \frac{\gamma_x}{\gamma_y} \left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} H_{1,1}^{1,1} \left[\left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} \middle| \begin{matrix} (\frac{\gamma_y - \gamma_x}{\gamma_y}, \frac{\gamma_x}{\gamma_y}) \\ (0, 1) \end{matrix} \right]. \quad (3.7)$$

Demonstração. Defina $\mu_j \in \mathbb{R}$, $\sigma_j, \gamma_j \in \mathbb{R}_+$ ($j \in \{x, y\}$). Então

$$\begin{aligned} R &= P(X < Y) = \int_{-\infty}^{\infty} G(x; \mu_x, \sigma_x, \gamma_x) g(x; \mu_y, \sigma_y, \gamma_y) dx \\ &= \int_M^{+\infty} \exp \left\{ - \left[1 + \frac{\gamma_x}{\sigma_x} (x - \mu_x) \right]^{-\frac{1}{\gamma_x}} - \left[1 + \frac{\gamma_y}{\sigma_y} (x - \mu_y) \right]^{-\frac{1}{\gamma_y}} \right\} \\ &\quad \times \left[1 + \frac{\gamma_y}{\sigma_y} (x - \mu_y) \right]^{-\frac{1}{\gamma_y} - 1} \frac{1}{\sigma_y} dx, \end{aligned} \quad (3.8)$$

onde $M = \max \left\{ \mu_y - \frac{\sigma_y}{\gamma_y}, \mu_x - \frac{\sigma_x}{\gamma_x} \right\}$. substituindo $y = \left[1 + \frac{\gamma_y}{\sigma_y} (x - \mu_y) \right]^{-\frac{1}{\gamma_y}}$ e tomando $M = \mu_y - \frac{\sigma_y}{\gamma_y}$, podemos reescrever (3.8) como

$$R = \int_0^{+\infty} \exp \left\{ -y - \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) + \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} y^{-\gamma_y} \right]^{-1/\gamma_x} \right\} dy. \quad (3.9)$$

Logo, (3.3) segue de (3.2) e (3.9).

Para o caso em que $\gamma_j > 0$, $j \in \{x, y\}$ e $\mu_y - \frac{\sigma_y}{\gamma_y} \leq \mu_x - \frac{\sigma_x}{\gamma_x}$, é suficiente notar que $P(X < Y) = 1 - P(Y < X)$ e aplicar o resultado em (3.3) com subíndices trocados. Para os casos onde $\gamma_j < 0$, $j = 1, 2$, o mesmo raciocínio pode ser aplicado, apenas notando que neste caso x assume principalmente valores negativos. O caso onde $\gamma_j = 0$, $j = 1, 2$, pode ser obtido como um procedimento limitante e será explicitamente explorado posteriormente neste trabalho. Além disso, aplicando (2.12) com $\mu_y - \frac{\sigma_y}{\gamma_y} = \mu_x - \frac{\sigma_x}{\gamma_x}$, obtemos (3.7). \square

Observação 3.1.1. Em um cenário prático, as estimativas $(\hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x)$ devem ser obtidas. Então, se $sign(\hat{\gamma}_y) = sign(\hat{\gamma}_x)$, as condições $\mu_y - \frac{\sigma_y}{\gamma_y} \geq \mu_x - \frac{\sigma_x}{\gamma_x}$ ou $\mu_y - \frac{\sigma_y}{\gamma_y} \leq \mu_x - \frac{\sigma_x}{\gamma_x}$ devem ser verificados e a expressão R correspondente deve ser usado.

A seguir, consideramos alguns casos especiais de duas variáveis aleatórias independentes que seguem distribuição GEV. Temos a seguinte consequência do Teorema 3.1.1:

Teorema 3.1.2. Sejam $Y \sim GEV(\mu_y, \sigma_y, \gamma_y)$ e $X \sim GEV(\mu_x, \sigma_x, \gamma_x)$ VAs independentes, com $\gamma_y = \gamma_x = \gamma \in \mathbb{R}, \gamma \neq 0, \mu_y, \mu_x \in \mathbb{R}, \sigma_y, \sigma_x \in \mathbb{R}_+$ e $\mu_y - \mu_x = \frac{\sigma_y}{\gamma} - \frac{\sigma_x}{\gamma}$. Então temos

$$R = P(X < Y) = \left(\frac{\sigma_y^{1/\gamma}}{\sigma_y^{1/\gamma} + \sigma_x^{1/\gamma}} \right). \quad (3.10)$$

Por último, consideramos os casos de duas GEV independentes com $\gamma_y = \gamma_x = 0$.

Teorema 3.1.3. Sejam Y e X VAs independentes, respectivamente, com distribuição $GEV(\mu_y, \sigma_y, 0)$ and $GEV(\mu_x, \sigma_x, 0)$, $\mu_j \in \mathbb{R}, \sigma_j \in \mathbb{R}_+, j \in \{x, y\}$. Então

$$\begin{aligned} R &= \exp \left\{ \frac{\mu_y}{\sigma_y} \right\} \mathbb{H} \left(\exp \left\{ \frac{\mu_y}{\sigma_y} \right\}, \exp \left\{ \frac{\mu_x}{\sigma_x} \right\}, \frac{\sigma_y}{\sigma_x}, 0, 1, 0 \right) \\ &= \exp \left\{ \frac{\mu_y - \mu_x}{\sigma_y} \right\} \frac{\sigma_x}{\sigma_y} H_{1,1}^{1,1} \left[\exp \left\{ \frac{\mu_y - \mu_x}{\sigma_y} \right\} \mid \begin{matrix} (\frac{\sigma_y - \sigma_x}{\sigma_y}, \frac{\sigma_x}{\sigma_y}) \\ (0, 1) \end{matrix} \right]. \end{aligned} \quad (3.11)$$

Demonstração. Defina $\mu_j \in \mathbb{R}$ e $\sigma_j \in \mathbb{R}_+ (j \in \{x, y\})$. Então

$$\begin{aligned} R &= P(X < Y) = \int_{-\infty}^{\infty} G(x; \mu_x, \sigma_x, 0) g(x; \mu_y, \sigma_y, 0) dx \\ &= \int_{-\infty}^{\infty} \exp \left\{ -\exp \left\{ -\frac{x - \mu_x}{\sigma_x} \right\} - \exp \left\{ -\frac{x - \mu_y}{\sigma_y} \right\} \right\} \exp \left\{ -\frac{x - \mu_y}{\sigma_y} \right\} \frac{dx}{\sigma_y}. \end{aligned} \quad (3.12)$$

Substituindo $y = \exp\{-x/\sigma_y\}$, podemos reescrever (3.12) como

$$R = \exp \left\{ \frac{\mu_y}{\sigma_y} \right\} \int_0^{+\infty} \exp \left\{ -\exp \left\{ \frac{\mu_y}{\sigma_y} \right\} y - \exp \left\{ \frac{\mu_x}{\sigma_x} \right\} y^{\sigma_y/\sigma_x} \right\} dy. \quad (3.13)$$

Logo, (3.11) segue de (3.2) e (3.13).

□

Temos a seguinte consequência imediata do Teorema 3.1.3 que estabelece uma forma explícita para R quando $\sigma_1 = \sigma_2 = \sigma$.

Teorema 3.1.4. Seja $Y \sim GEV(\mu_y, \sigma_y, 0)$ e $X \sim GEV(\mu_x, \sigma_x, 0)$ VAs independente, com

$\sigma_y = \sigma_x = \sigma \in \mathbb{R}_+$. Então, temos

$$R = P(X < Y) = \frac{\exp\{\mu_y/\sigma\}}{\exp\{\mu_y/\sigma\} + \exp\{\mu_x/\sigma\}}. \quad (3.14)$$

Os resultados apresentados nos Teoremas 3.1.1 e 3.1.3 são mais gerais do que os apresentados na literatura. A função \mathbb{H} nos permite escrever a probabilidade R com o mínimo possível de restrições entre os parâmetros. A Tabela 3.1 lista alguns estudos relacionados e suas restrições de parâmetros.

Tabela 3.1: Lista de estudos relacionados à probabilidade de *stress-strength* e distribuição extremas.

$sign(\gamma)$	Distribuição	Referência	Restrição de parâmetro
0	Gumbel	Nadarajah (2003a)	$\sigma_y = \sigma_x$ ou $\sigma_y = 2\sigma_x$ ou $\sigma_x/\sigma_y > 1$
1	Fréchet	Nadarajah (2003a) Jia, Nadarajah e Guo (2017) Abbas e Tang (2014) Rathie et al. (2022)	$\mu_y = \mu_x$ e ($\gamma_y = \gamma_x$ ou $\gamma_x = 2\gamma_y$ ou $\gamma_x/\gamma_y = p/q$)* $\mu_y = \mu_x = 0$ $\mu_y = \mu_x = 0$ e $\gamma_y = \gamma_x$ $\mu_y = \mu_x = 0$
-1	Weibull (min)	Nadarajah (2003a) Krishnamoorthy e Lin (2010) Kundu e Raqab (2009) Nojosa e Rathie (2020)	$\mu_y = \mu_x$ e ($\gamma_y = \gamma_x$ ou $\gamma_x = 2\gamma_y$ ou $\gamma_x/\gamma_y = p/q$)* $\mu_y = \mu_x = 0$ e $\gamma_y = \gamma_x = \gamma$ $\mu_y = \mu_x = \mu$ e $\gamma_y = \gamma_x = \gamma$ $\mu_y = \mu_x = 0$

* p e q são inteiros coprimos.

Observação 3.1.2. O caso particular de GEV enquanto $sign(\gamma) = -1$ (Galambos, 1987) é chamado de Weibull invertida. A distribuição Weibull estudada pelos autores citados na Tabela 3.1 é obtida como o limite de um mínimo normalizado VAs i.i.d. Ou seja, a distribuição Weibull é obtida por

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{\min\{X_1, \dots, X_n\} - b_n}{a_n} \leq x\right) &= \lim_{n \rightarrow \infty} (1 - (1 - F(a_n x + b_n))^n) \\ &= 1 - \exp\left\{-\left(\frac{x - \mu}{\sigma}\right)^\gamma\right\}, \quad x \geq \mu, \end{aligned} \quad (3.15)$$

onde X_1, \dots, X_n são i.i.d. VAs de F e a_n e b_n são sequências adequadas de constantes (veja o Teorema 2.1.5 em Galambos (1987)).

Confiabilidade do Sistema Multicomponente

Seja X_1, \dots, X_n i.i.d. VAs com distribuição $GEV(\mu_x, \sigma_x, \gamma_x)$ e seja Y uma VA independente com distribuição $GEV(\mu_y, \sigma_y, \gamma_y)$. Seja $M_n = \max\{X_1, \dots, X_n\}$. Então, $P(M_n \leq u) = G^n(u; \mu_x, \sigma_x, \gamma_x)$ e obtendo

$$P(X_1 < Y, \dots, X_n < Y) = P(M_n \leq Y) = \int_{-\infty}^{\infty} G^n(u; \mu_x, \sigma_x, \gamma_x) g(u; \mu_y, \sigma_y, \gamma_y) du =: I_n. \quad (3.16)$$

Expressões fechadas para (3.16) são apresentados no resultado a seguir. Sua prova segue o mesmo passo a passo dos Teoremas 3.1.1 e 3.1.3 e será omitida.

Teorema 3.1.5. Seja X_1, \dots, X_n VAs i.i.d. com distribuição $GEV(\mu_x, \sigma_x, \gamma_x)$ e seja Y uma VA independente de X com distribuição $GEV(\mu_y, \sigma_y, \gamma_y)$. Então

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

$$P(X_1 < Y, \dots, X_n < Y) = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} n^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right] n^{-\gamma_x}, -\frac{1}{\gamma_x} \right), \quad (3.17)$$

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

$$P(X_1 < Y, \dots, X_n < Y) = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x} n^{-\gamma_y}, -\gamma_x, \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right] n^{-\gamma_y}, -\frac{1}{\gamma_y} \right). \quad (3.18)$$

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

$$P(X_1 < Y, \dots, X_n < Y) = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} n^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right] n^{-\gamma_x}, -\frac{1}{\gamma_x} \right), \quad (3.19)$$

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

$$P(X_1 < Y, \dots, X_n < Y) = 1 - \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\sigma_y \gamma_x} n^{-\gamma_y}, -\gamma_x, \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right] n^{-\gamma_y}, -\frac{1}{\gamma_y} \right). \quad (3.20)$$

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

$$P(X_1 < Y, \dots, X_n < Y) = \exp \left\{ \frac{\mu_y}{\sigma_y} \right\} \mathbb{H} \left(\exp \left\{ \frac{\mu_y}{\sigma_y} \right\}, \exp \left\{ \frac{\mu_x}{\sigma_x} \right\}, n, \frac{\sigma_y}{\sigma_x}, 0, 1 \right). \quad (3.21)$$

Observação 3.1.3. Em um contexto mais amplo de confiabilidade multicomponente k-out-of-n, considere VAs independentes Y, X_1, \dots, X_k com $Y \sim GEV(\mu_y, \sigma_y, \gamma_y)$ e $X_j \sim GEV(\mu_x, \sigma_x, \gamma_x)$, para $j = 1, \dots, k$ (X_j 's são i.i.d.). A confiabilidade para este tipo de modelo é dada por

$$\begin{aligned} R_{s,k} &= P(\text{pelo menos } s \text{ das } (X_1, \dots, X_k) \text{ excedem } Y) \\ &= \sum_{j=s}^k \binom{k}{j} \int_{-\infty}^{\infty} (1 - G(u; \mu_x, \sigma_x, \gamma_x))^j (G(u; \mu_x, \sigma_x, \gamma_x))^{k-j} g(u; \mu_y, \sigma_y, \gamma_y) du. \end{aligned}$$

Usando a expansão binomial obtemos

$$R_{s,k} = \sum_{j=s}^k \sum_{r=0}^j \binom{k}{j} \binom{j}{r} (-1)^{j-r} \int_{-\infty}^{\infty} (G(u; \mu_x, \sigma_x, \gamma_x))^{k-r} g(u; \mu_y, \sigma_y, \gamma_y) du. \quad (3.22)$$

Observe que os termos integrais em (3.22) são casos particulares de (3.16) desde que $n = k - r$.

Portanto,

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

3.2 Estimação

Esta seção trata da estimativa de parâmetros para $R = P(X < Y)$ dados duas VAs GEV independentes. A literatura apresenta alguns estimadores de máxima verossimilhança para R considerando formas explícitas de R obtidas por fortes restrições de parâmetros em distribuições de valores extremos (como em Abbas e Tang (2014), Jia, Nadarajah e Guo (2017) e Kundu e Raqab (2009)). Essas abordagens exigem que a estimativa dos parâmetros seja feita conjuntamente nas duas amostras.

No nosso caso, flexibilizamos quaisquer requisitos sobre ter os mesmos parâmetros entre diferentes amostras, já que lidamos com expressões de R em termos de funções \mathbb{H} , conforme os Teoremas 3.1.1 e 3.1.3.

Considere uma amostra $\mathbf{X} = (X_1, \dots, X_n)$ da PDF $g(\cdot; \mu, \sigma, \gamma)$ definida em (3.1). A função

de verossimilhana para $GEV(\mu, \sigma, \gamma)$  obtida por:

$$L(\mu, \sigma, \gamma; \mathbf{X}) = \prod_{i=1}^n g(X_i; \mu, \sigma, \gamma) \mathbb{1}_{[1+\gamma(X_i-\mu)/\sigma>0]}, \quad (3.23)$$

onde $\mathbb{1}_A$ denota a funo indicadora no conjunto A . Observe que $\prod_{i=1}^n \mathbb{1}_{[1+\gamma(X_i-\mu)/\sigma>0]} > 0$, se, e somente se, $X_i \in \text{supp } g(\cdot; \mu, \sigma, \gamma)$ para todos $i = 1, \dots, n$. Aqui, $\text{supp } g$ denota o suporte da funo g . Ento, se $\gamma \neq 0$, no somos capazes de obter o MLE explicitamente, ento um procedimento numrico adicional  necessrio na maximizao da verossimilhana (para uma discusso mais detalhada veja Embrechts, Klppelberg e Mikosch (2013)).

Observaco 3.2.1. Define-se $\mathbf{X} = (X_1, \dots, X_n)$ uma amostra aleatria seguindo distribuico $GEV(\mu_x, \sigma_x, \gamma_x)$ e $\mathbf{Y} = (Y_1, \dots, Y_m)$ uma amostra aleatria de $GEV(\mu_y, \sigma_y, \gamma_y)$, com $\gamma_j > 0$, $j = 1, 2$ e $\mu_y - \sigma_y/\gamma_y \geq \mu_x - \sigma_x/\gamma_x$ (ou $\gamma_j < 0$, $j = 1, 2$ e $\mu_y - \sigma_y/\gamma_y \leq \mu_x - \sigma_x/\gamma_x$). Sejam $\hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_i$ ($i = 1, 2$) as estimativas de $\mu_i, \sigma_i, \gamma_i$. Podemos estimar R pela propriedade de invarincia do MLE, como segue

$$\hat{R} = \mathbb{H} \left(1, \frac{\hat{\gamma}_x \hat{\sigma}_y}{\hat{\sigma}_x \hat{\gamma}_y}, -\hat{\gamma}_y, 1 + \frac{\hat{\gamma}_x}{\hat{\sigma}_x} \left(\hat{\mu}_y - \hat{\mu}_x - \frac{\hat{\sigma}_y}{\hat{\gamma}_y} \right), -\frac{1}{\hat{\gamma}_x} \right). \quad (3.24)$$

Alternativamente, sempre que $\gamma_j > 0$, $j \in \{x, y\}$ e $\mu_y - \sigma_y/\gamma_y \leq \mu_x - \sigma_x/\gamma_x$ (ou $\gamma_j < 0$, $j \in \{x, y\}$ e $\mu_y - \sigma_y/\gamma_y \geq \mu_x - \sigma_x/\gamma_x$), a mesma propriedade de invarincia pode ser aplicada, levando a:

$$\hat{R} = 1 - \mathbb{H} \left(1, \frac{\hat{\gamma}_y \hat{\sigma}_x}{\hat{\sigma}_y \hat{\gamma}_x}, -\hat{\gamma}_x, 1 + \frac{\hat{\gamma}_y}{\hat{\sigma}_y} \left(\hat{\mu}_x - \hat{\mu}_y - \frac{\hat{\sigma}_x}{\hat{\gamma}_x} \right), -\frac{1}{\hat{\gamma}_y} \right). \quad (3.25)$$

Isto se deve aos Teoremas 3.1.1 e 3.1.3 que descrevem R em termos da funo \mathbb{H} (que  uma integral, portanto uma funo contnua e mensurvel nas variveis X e Y com as devidas estimativas paramtricas).

Sempre que um nico conjunto de realizaes das variveis aleatrias envolvidas estiver disponvel, a abordagem MLE acima  de extrema importncia.  o caso, por exemplo, da seleo de ativos, quando uma nica srie temporal de retornos observados est disponvel para cada ativo.

Por outro lado, para ilustrar a adequação das expressões analíticas aqui derivadas, uma abordagem de simulação direta pode ser realizada. Nesse caso, diversas amostras de tamanho n podem ser extraídas de cada variável aleatória, que são então utilizadas para estimar o valor empírico de R e podem ser repetidas diversas vezes. Ambas as abordagens serão exploradas na próxima seção. As estimativas pontuais também podem ser comparadas com um estimador empírico que não depende da estimativa de parâmetros ou da escolha de um modelo probabilístico. Considerando o estimador:

$$\hat{R}_{NIP} = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j \leq y_j\}},$$

onde $\mathbb{1}_A$ denota a função do indicador no conjunto A e n é o tamanho da amostra.

3.3 Estudo de Simulações

Para avaliar a correção da expressão de forma fechada para R dada no Teorema 3.1.1, geramos N amostras de Monte-Carlo, cada uma das quais de tamanho n , das variáveis aleatórias $GEV(\mu_x, \sigma_x, \gamma_x)$ e $GEV(\mu_y, \sigma_y, \gamma_y)$. Nestes casos, os valores de $\mu_x, \sigma_x, \gamma_x, \mu_y, \sigma_y, \gamma_y$ são pré-especificados na Tabela 3.2.

Os resultados obtidos com a distribuição GEV com parâmetro de forma negativo são apresentados nas Tabelas 3.2 e 3.3 onde analisamos as estimativas \hat{R} , viés (Bias) e Raiz do erro quadrático médio (RMSE). Na Tabela 3.4 é apresentado o cenário com parâmetro de forma positiva.

Para a simulação, para cada linha das Tabelas 3.2 - 3.4 foi realizado o seguinte procedimento:

1. para cada amostra de Monte-Carlo, a estimativa \hat{R} é calculada empiricamente, simplesmente como $n^{-1} \sum_i \mathbb{1}(x_i < y_i)$, onde $\mathbb{1}(\cdot)$ é uma função indicadora que é 1 para argumentos verdadeiros e 0, caso contrário;
2. \hat{R}_{MC} é avaliado tomando a média amostral das amostras de Monte-Carlo \hat{R} ;
3. o Bias é calculado como a diferença entre o valor obtido pelo Teorema 3.1.1 e \hat{R}_{MC} . O mesmo acontece com o RMSE, que também considera o valor verdadeiro como aquele obtido analiticamente.

Como esperado, as Tabelas 3.2 e 3.3 ilustram que os resultados analíticos obtidos coincidem com os empíricos. Esta é uma evidência clara dos bons resultados das novas expressões aqui derivadas. Além disso, aumentar o tamanho da amostra n leva a estimativas mais precisas de R , indicando viés reduzido e maior consistência. Na Tabela 3.4 observamos o mesmo bom comportamento do estimador caracterizado por baixos viés e RMSE.

Tabela 3.2: Média, Bias, RMSE de \hat{R}_{MC} para parâmetro de forma negativo ($N = 100$ e $n = 100$).

μ_x	σ_x	γ_x	μ_y	σ_y	γ_y	R	\hat{R}_{MC}	Bias	RMSE
2.0	1.5	-1.0	0.0	0.5	-0.3	0.1147	0.1151	-0.0004	0.0330
0.0	1.5	-1.0	0.0	0.5	-0.3	0.4350	0.4280	0.0070	0.0502
0.4	1.5	-1.0	0.5	0.5	-0.3	0.4650	0.4591	0.0059	0.0525
2.0	1.0	-1.0	0.0	0.7	-0.3	0.0798	0.0796	0.0002	0.0241
0.0	1.0	-1.0	0.0	0.7	-0.3	0.5298	0.5277	0.0021	0.0466
0.4	1.0	-1.0	0.5	0.7	-0.3	0.5686	0.5669	0.0017	0.0492
2.0	1.5	-1.0	0.0	0.9	-0.3	0.1414	0.1387	0.0027	0.0389
0.0	1.5	-1.0	0.0	0.9	-0.3	0.5092	0.5129	-0.0037	0.0462
0.4	1.5	-1.0	0.5	0.9	-0.3	0.5371	0.5362	0.0009	0.0461
2.0	1.5	-1.0	0.0	0.5	-1.0	0.1015	0.1025	-0.0010	0.0313
0.0	1.5	-1.0	0.0	0.5	-1.0	0.3851	0.3873	-0.0022	0.0463
0.4	1.5	-1.0	0.5	0.5	-1.0	0.4116	0.4159	-0.0043	0.0511
2.0	1.0	-1.0	0.0	0.7	-1.0	0.0590	0.0573	0.0017	0.0270
0.0	1.0	-1.0	0.0	0.7	-1.0	0.4358	0.4275	0.0083	0.0530
0.4	1.0	-1.0	0.5	0.7	-1.0	0.4816	0.4807	0.0009	0.0474
2.0	1.5	-1.0	0.0	0.9	-1.0	0.1104	0.1120	-0.0016	0.0277
0.0	1.5	-1.0	0.0	0.9	-1.0	0.4190	0.4279	-0.0089	0.0474
0.4	1.5	-1.0	0.5	0.9	-1.0	0.4478	0.4500	-0.0022	0.0506
2.0	1.5	-1.5	0.0	0.5	-1.5	0.1237	0.1234	0.0003	0.0302
0.0	1.5	-1.5	0.0	0.5	-1.5	0.3715	0.3701	0.0014	0.0488
0.4	1.5	-1.5	0.5	0.5	-1.5	0.3989	0.4034	-0.0045	0.0472
2.0	1.0	-1.5	0.0	0.7	-1.5	0.0822	0.0797	0.0025	0.0276
0.0	1.0	-1.5	0.0	0.7	-1.5	0.4160	0.4134	0.0026	0.0403
0.4	1.0	-1.5	0.5	0.7	-1.5	0.4739	0.4794	-0.0055	0.0548
2.0	1.5	-1.5	0.0	0.9	-1.5	0.1271	0.1303	-0.0032	0.0325
0.0	1.5	-1.5	0.0	0.9	-1.5	0.3999	0.4068	-0.0069	0.0457
0.4	1.5	-1.5	0.5	0.9	-1.5	0.4329	0.4344	-0.0015	0.0452

Tabela 3.3: Média, Bias, RMSE de \hat{R}_{MC} para parâmetro de forma negativo (N=1000 e n=1000).

μ_x	σ_x	γ_x	μ_y	σ_y	γ_y	R	\hat{R}_{MC}	Bias	RMSE
2.0	1.5	-1.0	0.0	0.5	-0.3	0.1147	0.1147	-0.0000	0.0106
0.0	1.5	-1.0	0.0	0.5	-0.3	0.4350	0.4355	-0.0005	0.0151
0.4	1.5	-1.0	0.5	0.5	-0.3	0.4650	0.4645	0.0005	0.0156
2.0	1.0	-1.0	0.0	0.7	-0.3	0.0798	0.0798	0.0001	0.0084
0.0	1.0	-1.0	0.0	0.7	-0.3	0.5298	0.5299	-0.0001	0.0159
0.4	1.0	-1.0	0.5	0.7	-0.3	0.5686	0.5693	-0.0007	0.0151
2.0	1.5	-1.0	0.0	0.9	-0.3	0.1414	0.1414	-0.0000	0.0112
0.0	1.5	-1.0	0.0	0.9	-0.3	0.5092	0.5096	-0.0004	0.0162
0.4	1.5	-1.0	0.5	0.9	-0.3	0.5371	0.5374	-0.0003	0.0158
2.0	1.5	-1.0	0.0	0.5	-1.0	0.1015	0.1018	-0.0003	0.0098
0.0	1.5	-1.0	0.0	0.5	-1.0	0.3851	0.3852	-0.0001	0.0152
0.4	1.5	-1.0	0.5	0.5	-1.0	0.4116	0.4119	-0.0003	0.0152
2.0	1.0	-1.0	0.0	0.7	-1.0	0.0590	0.0587	0.0003	0.0074
0.0	1.0	-1.0	0.0	0.7	-1.0	0.4358	0.4366	-0.0008	0.0154
0.4	1.0	-1.0	0.5	0.7	-1.0	0.4816	0.4814	0.0002	0.0150
2.0	1.5	-1.0	0.0	0.9	-1.0	0.1104	0.1106	-0.0002	0.0101
0.0	1.5	-1.0	0.0	0.9	-1.0	0.4190	0.4196	-0.0006	0.0157
0.4	1.5	-1.0	0.5	0.9	-1.0	0.4478	0.4471	0.0008	0.0150
2.0	1.5	-1.5	0.0	0.5	-1.5	0.1237	0.1243	-0.0006	0.0103
0.0	1.5	-1.5	0.0	0.5	-1.5	0.3715	0.3722	-0.0007	0.0158
0.4	1.5	-1.5	0.5	0.5	-1.5	0.3989	0.3974	0.0014	0.0154
2.0	1.0	-1.5	0.0	0.7	-1.5	0.0822	0.0822	0.0000	0.0086
0.0	1.0	-1.5	0.0	0.7	-1.5	0.4160	0.4166	-0.0005	0.0160
0.4	1.0	-1.5	0.5	0.7	-1.5	0.4739	0.4736	0.0004	0.0162
2.0	1.5	-1.5	0.0	0.9	-1.5	0.1271	0.1265	0.0007	0.0105
0.0	1.5	-1.5	0.0	0.9	-1.5	0.3999	0.3998	0.0001	0.0155
0.4	1.5	-1.5	0.5	0.9	-1.5	0.4329	0.4323	0.0006	0.0151

Tabela 3.4: Média, Bias, RMSE de \hat{R}_{MC} para parâmetro de forma positivo ($N = 1000$ and $n = 100$).

μ_x	σ_x	γ_x	μ_y	σ_y	γ_y	R	\hat{R}_{MC}	Bias	RMSE
2.0	1.5	1.0	0.0	0.5	0.3	0.0617	0.0624	-0.0007	0.0244
0.0	1.5	1.0	0.0	0.5	0.3	0.4288	0.4306	-0.0018	0.0497
0.4	1.5	1.0	0.5	0.5	0.3	0.4491	0.4478	0.0012	0.0514
2.0	1.0	1.0	0.0	0.7	0.3	0.0906	0.0891	0.0015	0.0284
0.0	1.0	1.0	0.0	0.7	0.3	0.4443	0.4463	-0.0020	0.0507
0.4	1.0	1.0	0.5	0.7	0.3	0.4717	0.4727	-0.0010	0.0499
2.0	1.5	1.0	0.0	0.9	0.3	0.1299	0.1300	-0.0001	0.0332
0.0	1.5	1.0	0.0	0.9	0.3	0.4419	0.4423	-0.0004	0.0509
0.4	1.5	1.0	0.5	0.9	0.3	0.4611	0.4602	0.0008	0.0497
2.0	1.5	1.0	0.0	0.5	1.0	0.1469	0.1474	-0.0004	0.0359
0.0	1.5	1.0	0.0	0.5	1.0	0.4764	0.4774	-0.0010	0.0486
0.4	1.5	1.0	0.5	0.5	1.0	0.4947	0.4940	0.0007	0.0504
2.0	1.0	1.0	0.0	0.7	1.0	0.1846	0.1847	-0.0001	0.0382
0.0	1.0	1.0	0.0	0.7	1.0	0.4980	0.4985	-0.0006	0.0520
0.4	1.0	1.0	0.5	0.7	1.0	0.5240	0.5235	0.0004	0.0480
2.0	1.5	1.0	0.0	0.9	1.0	0.2144	0.2164	-0.0019	0.0431
0.0	1.5	1.0	0.0	0.9	1.0	0.4951	0.4925	0.0026	0.0497
0.4	1.5	1.0	0.5	0.9	1.0	0.5128	0.5124	0.0004	0.0480
2.0	1.5	1.5	0.0	0.5	1.5	0.1868	0.1883	-0.0015	0.0393
0.0	1.5	1.5	0.0	0.5	1.5	0.4906	0.4894	0.0013	0.0523
0.4	1.5	1.5	0.5	0.5	1.5	0.5076	0.5072	0.0004	0.0495
2.0	1.0	1.5	0.0	0.7	1.5	0.2237	0.2243	-0.0006	0.0414
0.0	1.0	1.5	0.0	0.7	1.5	0.5056	0.5052	0.0004	0.0501
0.4	1.0	1.5	0.5	0.7	1.5	0.5318	0.5310	0.0008	0.0496
2.0	1.5	1.5	0.0	0.9	1.5	0.2461	0.2451	0.0011	0.0444
0.0	1.5	1.5	0.0	0.9	1.5	0.5046	0.5042	0.0004	0.0491
0.4	1.5	1.5	0.5	0.9	1.5	0.5220	0.5226	-0.0006	0.0490

3.4 Aplicações

Nesta seção abordamos a aplicação do estudo de confiabilidade nos preços do fechamento de quatro ações de empresas brasileiras, foram elas, Banco do Brasil SA (BBAS3), Itau unibanco SA (ITUB4), Grupo Casas Bahia SA (VIA3) e Companhia Vale do Rio Doce SA (VALE3). Os dados foram obtidos da Bolsa de Valores Brasileira (BOVESPA).

Compararemos os log-retornos dos preços das ações dos *tickers*¹ BBAS3, ITUB4, VALE3 e VIA3. Essas ações são de empresas que representam uma ampla variedade de setores

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

econômicos: bancário (BBAS3, ITUB4), mineração (VALE3) e varejo (VIAA3). Daqui em diante, omitiremos o sufixo .SA presente nos tickers em análise. Os conjuntos de dados foram adquiridos utilizando o pacote *quantmod* dentro do software R (R Core Team, 2023), cobrindo o período analisado de 1 de janeiro de 2022 a 30 de abril de 2023, e possuem informações na moeda brasileira (R\$, BRL). É importante ressaltar que esses conjuntos de dados foram analisados neste Capítulo 3, e posteriormente no Capítulo 4 mostramos que a distribuição TGEV se ajusta melhor aos log-retornos do que a GEV, de acordo com os critérios de informação.

3.4.1 Mercado de Ações Brasileiro

A Figura 3.1 abaixo denota os preços dos fechamentos das ações no período avaliado.

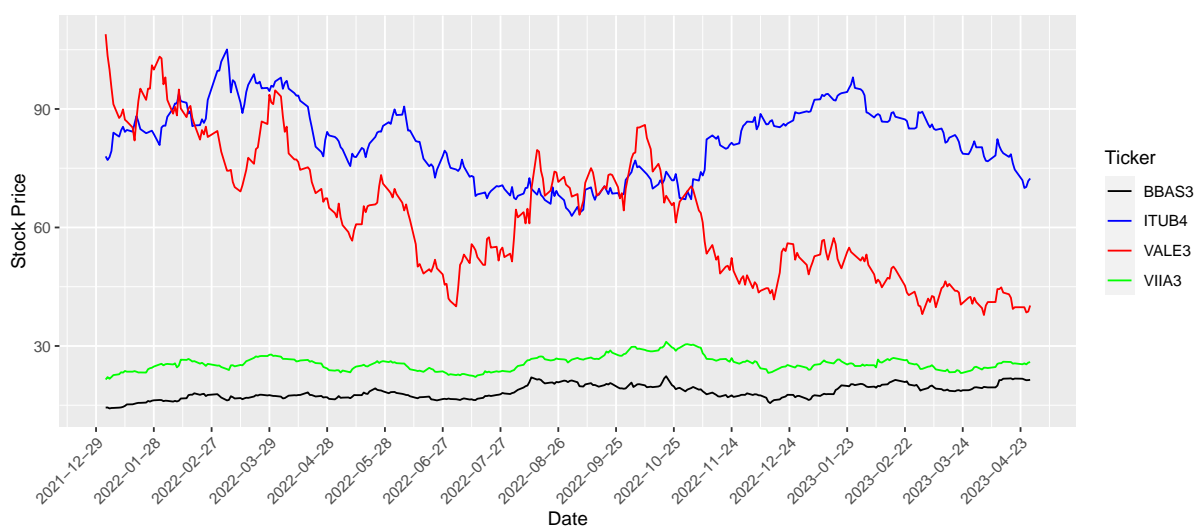


Figura 3.1: Preço do fechamento diários das ações em moeda brasileira (R\$, BRL).

Em geral, séries financeiras possuem autocorrelação. Neste trabalho utilizamos os incrementos logarítmizados das séries, para calcularmos os log retornos utilizando os seguintes procedimentos: seja X_t uma série temporal, os log retornos são obtidos através da seguinte expressão:

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

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

A aplicação do logaritmo aos incrementos é um procedimento para suavizar a serie. Além disso, medimos as correlações dos log-retornos utilizando os métodos de Pearson, Kendall e

Spearman, comparando pares de ações que não são correlacionadas. Os pares selecionados para comparação — VALE3-BBAS3, VALE3-ITUB4 e VALE3-VIIA3 — apresentaram medidas de correlação iguais ou inferiores a 0,25, conforme detalhado nas tabelas. 3.5, 3.6 e 3.7. A Figura 3.2 exibe a função de autocorrelação (ACF) dos log-retornos, indicando ausência de correlações temporais entre os retornos.

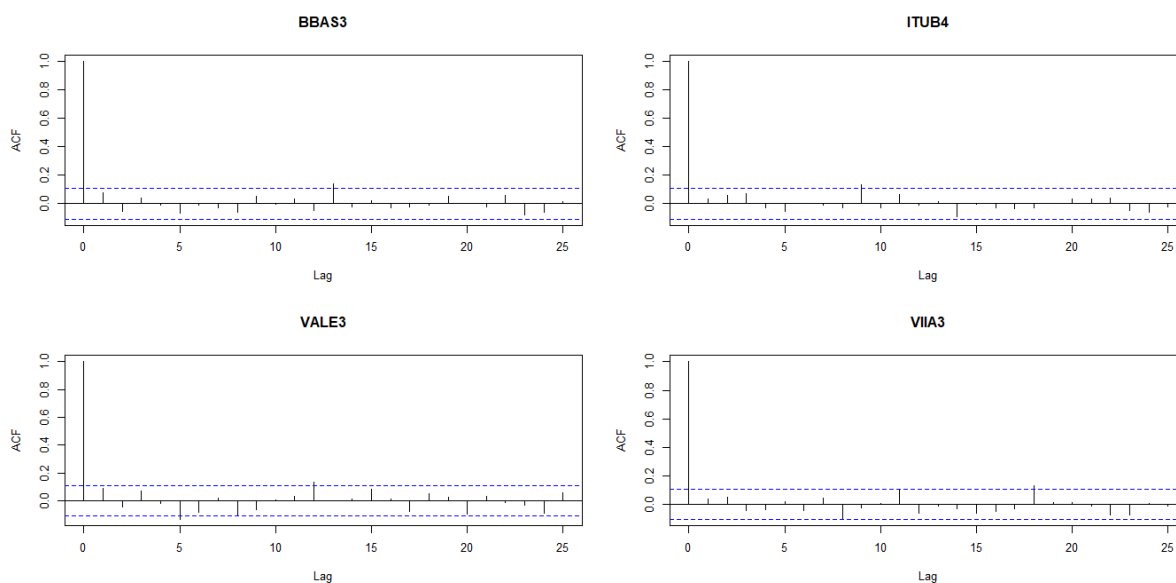


Figura 3.2: ACFs dos log-retornos para os ativos BBAS3, ITUB4, VIIA3 e VALE3. As linhas horizontais azuis nos gráficos são os limites $\pm 1.96 \sqrt{(n)}$.

Tabela 3.5: Matrix de correlação de Pearson.

	X_1	X_2	X_3	X_4
X_1	1.00	0.61	0.10	0.29
X_2		1.00	0.15	0.31
X_3			1.00	0.03
X_4				1.00

Tabela 3.6: Matrix de correlação de Spearman.

	X_1	X_2	X_3	X_4
X_1	1.00	0.63	0.15	0.31
X_2		1.00	0.22	0.34
X_3			1.00	0.06
X_4				1.00

Tabela 3.7: Matrix de correlação de Kendall.

	X_1	X_2	X_3	X_4
X_1	1.00	0.45	0.10	0.21
X_2		1.00	0.15	0.24
X_3			1.00	0.04
X_4				1.00

Visando estimar probabilidades do tipo $R = P(X < Y)$ via os Teoremas 3.1.1 e 3.1.3, necessita-se que X e Y sejam VAs independentes. Nesse sentido, analisamos as estruturas de dependência de X_1, \dots, X_4 utilizando as matrizes de correlação de Pearson, Kendal e Spearman, e os resultados são apresentados nas Tabelas 3.5, 3.6 e 3.7. Os pares escolhidos para serem comparados são $X_3 - X_2$, $X_3 - X_4$ e $X_3 - X_1$.

As estatísticas descritivas para os quatro conjuntos de dados são apresentadas na Tabela 3.8, mostrando uma certa simetria dos log-retornos em torno de zero e uma maior variabilidade para VIIA3. O tamanho da amostra foi de $n = 330$ e cada amostra informa o log-retorno diário do preço de fechamento das ações. Em todos os conjuntos de dados, a existência de valores extremos é uma característica consistente, o que está de acordo com a natureza dos dados financeiros.

Tabela 3.8: Estatísticas descritivas para o retorno logaritmo dos preços das ações.

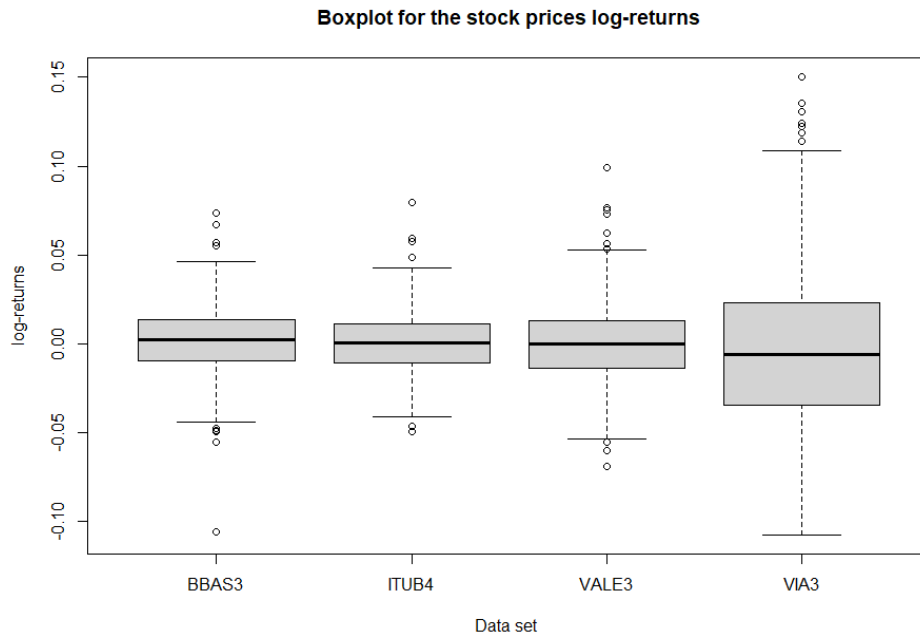
VA	Dados	Min	1st Qu.	Median	Mean	3rd Qu.	Max	n
X_1	BBAS3	-0.1057	-0.0097	0.0019	0.0012	0.0136	0.0736	330
X_2	ITUB4	-0.0492	-0.0105	0.0004	0.0006	0.0109	0.0794	330
X_3	VALE3	-0.0689	-0.0140	0.0001	-0.0002	0.0128	0.0989	330
X_4	VIIA3	-0.1075	-0.0344	-0.006	-0.0030	0.0231	0.1504	330

Denote X_1, X_2, X_3 e X_4 os log-retornos dos preços das ações de BBAS3, ITUB4, VALE3 e VIIA3, respectivamente. Os conjuntos de dados são obtidos diretamente através do software R (R Core Team, 2023) pelo seguinte comando:

```
quantmod::getSymbols("BBAS3.SA", src = "yahoo", auto.assign = FALSE,
from = '2022-01-01', to = '2023-04-30', return.class = 'xts').
```

Um boxplot é apresentado na Figura 3.3 mostrando a simetria em torno de zero dos log-retornos e que X_4 tem mais dispersão que os demais.

Figura 3.3: Boxplot dos retornos logarítmicos dos preços das ações BBAS3, ITUB4, VALE3 e VIIA3.



Para verificar se a adequação da distribuição aos dados utilizou-se o teste de Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1939). As estimativas de máxima verossimilhança (ML) e os p-valores de Kolmogorov-Smirnov (KS) são fornecidos na Tabela 3.9. A Figura 3.4 mostra histogramas e o ajuste do modelo GEV para X_1 , X_2 , X_3 e X_4 . Para cada conjunto de dados, o CDF empírico (ECDF) é comparado ao modelo teórico na Figura 3.5.

Embora o p-valor do teste de Kolmogorov-Smirnov (KS) seja pequeno para o BBAS3 (X_1), a análise gráfica não apresenta desvios consideráveis de que a distribuição adequa-se aos dados.

Tabela 3.9: Estimativas ML, Verossimilhança(ll_{max}) e p-valor-ks para o modelo GEV.

VA	Data set	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$-ll_{max}$	KS p-value
X_1	BBAS3	-0.0063	0.0219	-0.2535	-803.9168	0.0147
X_2	ITUB4	-0.0064	0.0165	-0.1545	-870.9392	0.4299
X_3	VALE3	-0.0095	0.0222	-0.1631	-774.0895	0.2331
X_4	VIIA3	-0.0217	0.0396	-0.1170	-567.6979	0.6996

Medidas de confiabilidade do tipo $R = P(X < Y)$ aparecem no processo de decisão de um investidor. Em resumo, quando X e Y representam VAs de lucro e $R < 1/2$, é aconselhável que o investidor escolha a ação da variável X . Se $R > 1/2$, ocorre o oposto. O caso $R = 1/2$ é inconclusivo. Assim, saber avaliar R com precisão é importante para apoiar o processo de decisão.

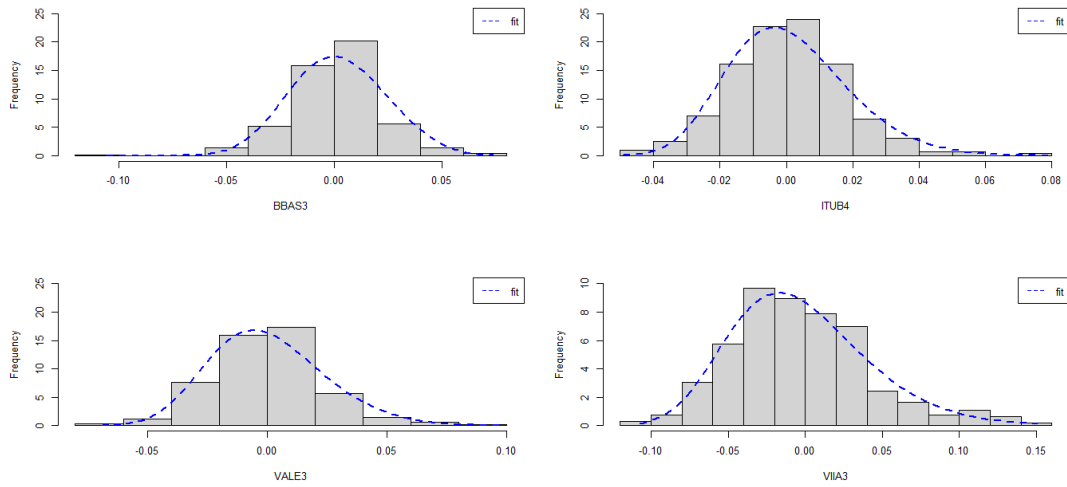


Figura 3.4: Histogramas e densidades para os log-retornos dos preços das ações.

Nesse sentido, a Tabela 3.10 apresenta as estimativas de $P(X_3 < X_1)$, $P(X_3 < X_2)$ e $P(X_3 < X_4)$ e o Intervalo de Confiança Bootstrap (CI) de 95%. Para todas as tabelas subsequentes, \hat{R} é a estimativa obtida ajustando primeiro o conjunto de dados a uma distribuição GEV e usando os resultados da Observação 3.24 (propriedade de invariância do MLE dos parâmetros e a expressão analítica obtida para R). Além disso, \hat{R}_{NP} é uma estimativa não paramétrica de R , que considera uma abordagem semelhante à do estudo de Simulação (usa uma função indicadora para comparar amostras das duas distribuições). Finalmente, \hat{R}_{Boot} é o valor médio da estimativa bootstrap de R usando a abordagem não paramétrica e o CI relatado é para tal média.

Tabela 3.10: Estimativa das probabilidades SSR e CI Bootstrap.

$R = P(X < Y)$	\hat{R}	\hat{R}_{NP}	\hat{R}_{Boot}	95% CI
$P(X_3 < X_1)$	0.5283	0.5242	0.5277	(0.4883; 0.5676)
$P(X_3 < X_2)$	0.5174	0.5242	0.5181	(0.4766; 0.5569)
$P(X_3 < X_4)$	0.4506	0.4364	0.4500	(0.4055; 0.4944)

As estimativas pontuais de R indicam que, no período analisado, a VALE3 teria preferência sobre BBAS3 e ITUB4, enquanto VVIA3 teria preferência sobre VALE3. Por outro lado, as estimativas do Bootstrap IC indicam que apenas o caso $P(X_3 < X_4)$ foi conclusivo, pois 0.5 não pertence ao intervalo de confiança estimado.

É importante comparar diferentes distribuições como modelos candidatos para a modelagem de log-retornos. Considerando resultados anteriores da literatura Gettinby et al. (2004), comparamos o desempenho das distribuições GEV e distribuição Logística Generalizada (GL) como modelos para

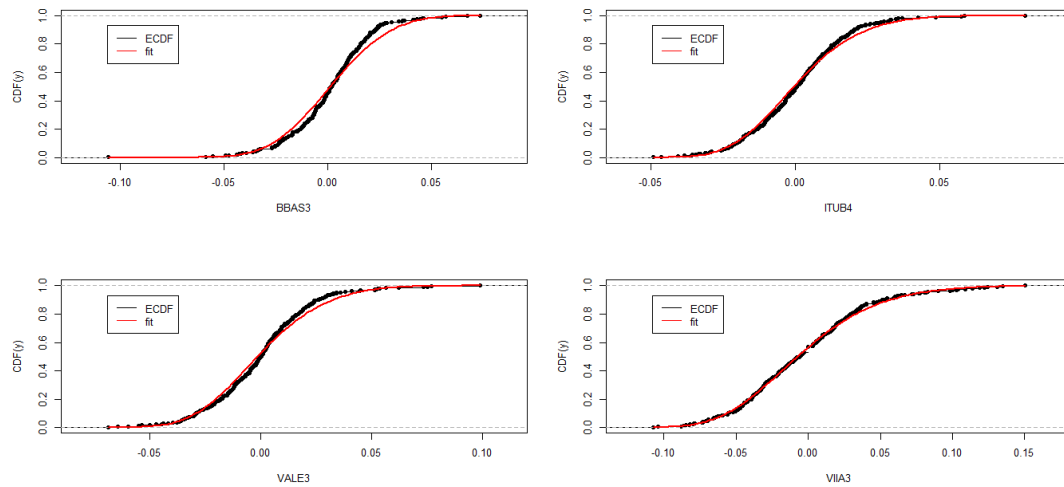


Figura 3.5: ECDF para os modelos GEV.

retornos diários.

Tabela 3.11: Comparação entre as verossimilhanças para os modelos GEV e GL.

VA	Dados	GEV	GL
X_1	BBAS3	-803.92	- 833.43
X_2	ITUB4	-870.94	- 882.01
X_3	VALE3	-774.09	- 787.81
X_4	VVIA3	-567.70	- 569.60

É possível ver que tanto a distribuição GEV quanto a distribuição Logística Generalizada forneceram capacidades de modelagem bastante semelhantes (aproximadamente os mesmos valores de log-verossimilhança).

Considerando Log-Retornos Máximos Semanais

Em vez de considerar os log retornos logarítmicos dos preços de fechamento, como foi feito nas análises anteriores, pode-se estudar como se comportam os máximos (ou mínimos) semanais. Isto tem uma interpretação econômica direta: um proxy para maiores lucros (ou maiores perdas, ou seja, défices e valores em risco) e tem sido explorado na literatura (veja Gettinby et al. (2004), Hussain e Li (2015) e Goncu et al. (2012))

Seguindo estudos anteriores Hussain e Li (2015), o caso em que são considerados os retornos máximos para as variáveis aleatórias GEV. A Tabela 3.12 mostra as estimativas ML para os parâmetros considerando o modelo GEV para tal caso.

Tabela 3.12: Estimativas ML, Verossimilhança máxima (ll_{max}) e p-valor do teste KS para os log-retornos dos máximos semanais e suas modelagens via GEV.

VA	Dados	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$-ll_{max}$	p-valor KS
X_1	BBAS3	-0.0128	0.0393	-0.1157	-116.3016	0.8076
X_2	ITUB4	-0.091	0.0397	-0.4319	-126.5530	0.8292
X_3	VALE3	-0.0197	0.0426	-0.2003	-114.0183	0.9365
X_4	VIIA3	-0.0464	0.0750	-0.1339	-72.9854	0.5378

A Tabela 3.12 indica que a distribuição GEV é adequada para todos os casos considerados, o que é esperado uma vez que a teoria dos valores extremos prevê tal convergência. Agora que os parâmetros foram obtidos, a Tabela 3.13 apresenta os cálculos de confiabilidade.

Tabela 3.13: Estimativa das probabilidades *stress-strength* e IC Bootstrap para os máximos semanais

$R = P(X < Y)$	\hat{R}	\hat{R}_{NP}	\hat{R}_{Boot}	95% CI
$P(X_3 < X_1)$	0.5453	0.5522	0.5413	(0.4483; 0.6334)
$P(X_3 < X_2)$	0.5409	0.5522	0.5483	(0.4421; 0.6358)
$P(X_3 < X_4)$	0.4364	0.4328	0.4350	(0.3379; 0.5350)

Como $R = 0,5$ está dentro dos intervalos de confiança, a métrica torna-se inconclusiva. É possível perceber, por outro lado, que se \hat{R} for considerado como uma métrica por si só, refletiria justamente a alta volatilidade do VIIA3, que sofreu severa instabilidade e perdas durante a janela de tempo analisada.

Além da aplicação direta na seleção de ativos financeiros, algumas aplicações de engenharia também podem se beneficiar das novas relações aqui definidas. Uma aplicação é ilustrada na próxima subseção.

3.4.2 Fibras de carbono

A confiabilidade da resistência ao estresse também pode ser aplicada à modelagem e comparação de fibras de carbono de comprimentos de 20 mm e 10 mm. Os dados representam os dados de resistência medidos em GPa (Gigapascal), para fibras simples de carbono testadas sob tensão. Os dados são frequentemente utilizados na literatura e também são apresentados abaixo (por exemplo, Nojosa e Rathie (2020)).

A Tabela 3.16 e a Figura 3.6 mostram que as estatísticas descritivas de X e Y nas quais é possível observar que Y (fibras de carbono de comprimento 10mm) tendem a têm valores de resistência maiores que X (fibras de carbono de comprimento 20mm).

Tabela 3.14: Fibras de carbono de comprimento 20mm:

1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958
1.966	1.977	2.006	2.021	2.027	2.055	2.063	2.098	2.140	2.179
2.224	2.240	2.253	2.270	2.272	2.274	2.359	2.382	2.382	2.426
2.434	2.435	2.478	2.490	2.511	2.514	2.535	2.554	2.566	2.570
2.301	2.301	2.586	2.629	2.633	2.642	2.648	2.684	2.697	2.726
2.770	2.773	2.800	2.809	2.818	2.821	2.848	2.880	2.954	3.012
3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.585	

Tabela 3.15: Fibras de carbono de comprimento 10mm:

1.901	2.132	2.203	2.228	2.257	2.350	2.361	2.396	2.397	2.445
2.454	2.474	2.518	2.522	2.525	2.532	2.575	2.614	2.616	2.618
2.624	2.659	2.675	2.738	2.740	2.856	2.917	2.928	2.937	2.937
2.977	2.996	3.030	3.125	3.139	3.145	3.220	3.223	3.235	3.243
3.264	3.272	3.294	3.332	3.346	3.377	3.408	3.435	3.493	3.501
3.537	3.554	3.562	3.628	3.852	3.871	3.886	3.971	4.024	4.027
4.225	4.395	5.020							

Tabela 3.16: Estatística descritivas para fibras de carbono de tamanho 20mm (X) e 10mm (Y).

VA	Dados	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n
X	Fibras de carbono de 20 mm	1.312	2.098	2.478	2.451	2.773	3.585	69
Y	Fibra de carbono de 10 mm	1.901	2.554	2.996	3.059	3.421	5.020	63

As estimativas de ML e os valores p de KS para o modelo GEV são apresentados na Tabela 3.17 e o bom ajuste pode ser observado na Figura 3.7.

As estimativas de confiabilidade de *stress-strength* e IC Bootstrap 95% são $\hat{R} = 0,774$ e $(0,699, 0,851)$, respectivamente. É fácil concluir que como $P(X < Y) > 1/2$ e 0.5 não está dentro do IC, que $X < Y$ estatisticamente e que Y é maior que X , ou seja, as fibras com comprimento de 10mm possuem valores de resistência estatisticamente maiores que os de comprimento de 20mm.

Tabela 3.17: Estimativas ML e Kolmogorov-Smirnov p-valor (KS) para o modelo GEV.

Dados	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	p-valor KS
X	2.2781	0.4956	-0.2851	0.9978
Y	2.7904	0.5245	-0.0747	0.8216

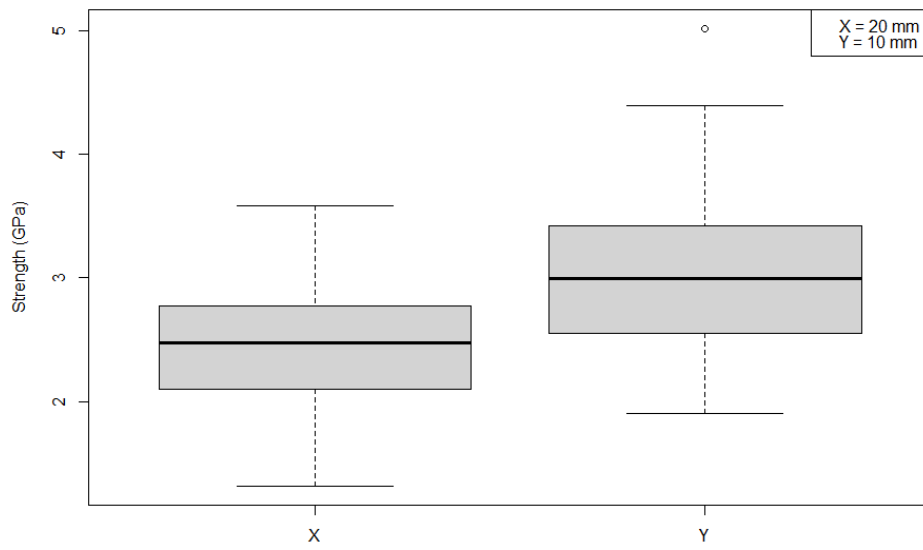


Figura 3.6: Boxplot para fibras de carbono de tamanho 20mm (X) e 10mm (Y).

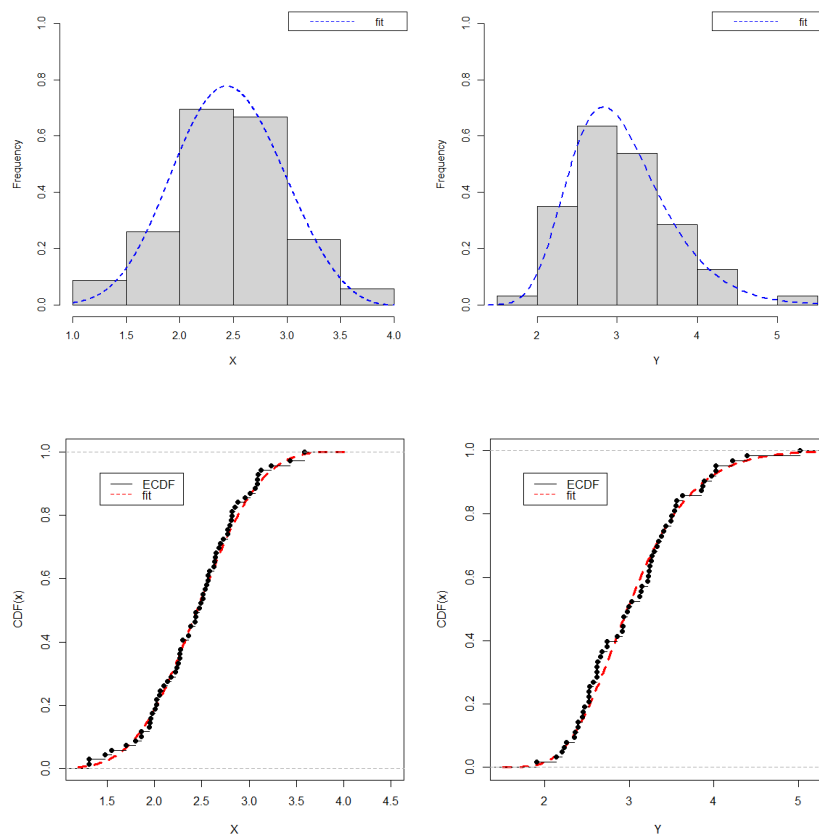


Figura 3.7: Histograma, ECDF e modelo GEV ajustado para fibras de carbono.

Capítulo 4

Seleção de ativos financeiros via TGEV

Neste capítulo estudamos a probabilidade SSR $R = \mathbb{P}(X < Y)$ quando X e Y são VAs independentes com distribuição GEV transmutada (TGEV). A distribuição TGEV é obtida da seguinte forma: dada a distribuição GEV $G(x)$, a função de distribuição transmutada F é dada por:

$$F(x) = (1 + \lambda)G(x) - \lambda[G(x)]^2, \quad |\lambda| < 1. \quad (4.1)$$

Assim, as funções FDA e FDP são dadas pelas Equações (2.5) e (2.6), ou seja,

$$F(x; \mu, \sigma, \gamma, \lambda) = \begin{cases} \exp(-w^{-1/\gamma}) [(1 + \lambda) - \lambda \exp(-w^{-1/\gamma})], & \gamma \neq 0; \\ \exp(-\exp(-(\frac{x-\mu}{\sigma}))) [(1 + \lambda) - \lambda \exp(-(\frac{x-\mu}{\sigma}))], & \gamma = 0, \end{cases}$$

e

$$f(x; \mu, \sigma, \gamma, \lambda) = \begin{cases} \left[\frac{(w)^{-1-1/\gamma} \exp(-w^{-1/\gamma})}{\sigma} \right] [(1 + \lambda) - 2\lambda \exp(-w^{-1/\gamma})], & \gamma \neq 0; \\ \frac{\exp(-(\frac{x-\mu}{\sigma})) \exp(-\exp(-(\frac{x-\mu}{\sigma})))}{\sigma} [(1 + \lambda) - 2\lambda \exp(-\exp(-(\frac{x-\mu}{\sigma})))], & \gamma = 0. \end{cases} \quad (4.2)$$

onde $w = 1 + (\gamma(x - \mu)/\sigma)$ e $\text{supp}(F) = \text{supp}(f) = \{x; w > 0\}$, $\mu \in \mathbb{R}$ é o parâmetro de localização, $\sigma \in \mathbb{R}_+$ é o parâmetro de escala e $\gamma \in \mathbb{R}$ é o parâmetro de forma e por fim λ é um parametro no intervalo $[-1, 1]$.

No Teorema 3.1.1 estabelecemos formas de R em termos da função \mathbb{H} dos valores extremos definida

em (2.7):

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^\infty y^{a_6} \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy. \quad (4.3)$$

Em geral, os resultados obtidos são generalizações dos resultados do Capítulo 3. Na Seção 4.2, um procedimento de estimação para R é proposto baseado na estimação dos parâmetros de X e Y via máxima verossimilhança e estimando R via propriedade de invariância.

Apresentamos ainda a avaliação do estimador proposto por meio de um estudo de simulações. Na Seção 4.1, propomos e validamos a aplicação da metodologia SSR na modelagem de dados reais.

Utilizando as mesmas bases descritas na Subseção 3.4.1, mostramos que os resultados da modelagem dos retornos dos ativos financeiros podem ser melhorados quando utilizamos a distribuição TGEV ao invés da GEV.

Por fim, destacamos que os resultados apresentados nesse capítulo foram submetidos em Oliveira et al. (2024).

4.1 Resultados Analíticos

Nesta seção, a confiabilidade de duas variáveis aleatórias independentes seguindo distribuição TGEV é derivada em termos de funções \mathbb{H} . Além disso, com restrições de parâmetros adequadas, também são obtidas expressões mais simples em termos da função \mathbb{H} . Primeiramente, consideramos o caso de duas VAs TGEV independentes com $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) \neq 0$

Teorema 4.1.1. Sejam X e Y VAs independentes, respectivamente, com distribuição, $TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e $TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$, $\mu_j \in \mathbb{R}$, $\sigma_j \in \mathbb{R}_+$, $\gamma_j \in \mathbb{R}$ ($\gamma_j \neq 0$), $\lambda_j \in [-1, 1]$, $j \in \{x, y\}$ denotamos $R = P(X < Y)$. Então

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

$$\begin{aligned} R &= (1 + \lambda_x)(1 + \lambda_y) \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right) \\ &\quad - 2\lambda_y(1 + \lambda_x) \mathbb{H} \left(2, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right) \\ &\quad - \lambda_x(1 + \lambda_y) \mathbb{H} \left(1, 2^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 2^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right], -\frac{1}{\gamma_x}, 0 \right) \\ &\quad + 2\lambda_x \lambda_y \mathbb{H} \left(2, 2^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 2^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right], -\frac{1}{\gamma_x}, 0 \right), \quad (4.4) \end{aligned}$$

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

$$\begin{aligned}
 R &= 1 - (1 + \lambda_y)(1 + \lambda_x) \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right) \\
 &+ 2\lambda_x(1 + \lambda_y) \mathbb{H} \left(2, \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right) \\
 &+ \lambda_y(1 + \lambda_x) \mathbb{H} \left(1, 2^{-\gamma_y} \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 2^{-\gamma_y} \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right], -\frac{1}{\gamma_y}, 0 \right) \\
 &- 2\lambda_y \lambda_x \mathbb{H} \left(2, 2^{-\gamma_y} \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 2^{-\gamma_y} \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right], -\frac{1}{\gamma_y}, 0 \right). \quad (4.5)
 \end{aligned}$$

• Quando $\gamma_j < 0, j \in \{x, y\}$:

$$\begin{aligned}
 R &= (1 + \lambda_x)(1 + \lambda_y) \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right) \\
 &- 2\lambda_y(1 + \lambda_x) \mathbb{H} \left(2, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right), -\frac{1}{\gamma_x}, 0 \right) \\
 &- \lambda_x(1 + \lambda_y) \mathbb{H} \left(1, 2^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 2^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right], -\frac{1}{\gamma_x}, 0 \right) \\
 &+ 2\lambda_x \lambda_y \mathbb{H} \left(2, 2^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, 2^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right], -\frac{1}{\gamma_x}, 0 \right), \quad (4.6)
 \end{aligned}$$

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

$$\begin{aligned}
 R &= 1 - (1 + \lambda_y)(1 + \lambda_x) \mathbb{H} \left(1, \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right) \\
 &+ 2\lambda_x(1 + \lambda_y) \mathbb{H} \left(2, \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right), -\frac{1}{\gamma_y}, 0 \right) \\
 &+ \lambda_y(1 + \lambda_x) \mathbb{H} \left(1, 2^{-\gamma_y} \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 2^{-\gamma_y} \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right], -\frac{1}{\gamma_y}, 0 \right) \\
 &- 2\lambda_y \lambda_x \mathbb{H} \left(2, 2^{-\gamma_y} \frac{\gamma_y \sigma_x}{\gamma_x \sigma_y}, -\gamma_x, 2^{-\gamma_y} \left[1 + \frac{\gamma_y}{\sigma_y} \left(\mu_x - \mu_y - \frac{\sigma_x}{\gamma_x} \right) \right], -\frac{1}{\gamma_y}, 0 \right). \quad (4.7)
 \end{aligned}$$

Demonstração. Defina $S = \text{supp } F_X \cap \text{supp } f_Y$. Então

$$S = \begin{cases} (M, \infty), & \text{se } \text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = 1, \\ (-\infty, m), & \text{se } \text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = -1, \\ \mathbb{R}, & \text{se } \text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = 0, \end{cases} \quad (4.8)$$

onde $M = \max\{\mu_x - \sigma_x/\gamma_x, \mu_y - \sigma_y/\gamma_y\}$ e $m = \min\{\mu_x - \sigma_x/\gamma_x, \mu_y - \sigma_y/\gamma_y\}$.

Observe que

$$\begin{aligned}
R &= \int_{-\infty}^{\infty} F_X(u; \mu_x, \sigma_x, \gamma_x, \lambda_x) f_Y(u; \mu_y, \sigma_y, \gamma_y, \lambda_y) du \\
&= \int_S \left(\exp(-w_x^{-1/\gamma_x}) \left[(1 + \lambda_x) - \lambda_x \exp(-w_x^{-1/\gamma_x}) \right] \right. \\
&\quad \times \left. \left[\frac{(w_y)^{-1-1/\gamma_y} \exp(-w_y^{-1/\gamma_y})}{\sigma_y} \right] \left[(1 + \lambda_y) - 2\lambda_y \exp(-w_y^{-1/\gamma_y}) \right] \right) du, \quad (4.9)
\end{aligned}$$

onde $w_i = 1 + \frac{\gamma_i}{\sigma_i}(u - \mu_i)$, $i \in \{x, y\}$. Temos quatro casos a considerar:

1. $\gamma_x > 0$ e $\gamma_y > 0$

(a) $\mu_y - \sigma_y/\gamma_y \geq \mu_x - \sigma_x/\gamma_x$;

(b) $\mu_y - \sigma_y/\gamma_y < \mu_x - \sigma_x/\gamma_x$;

2. $\gamma_x < 0$ e $\gamma_y < 0$

(a) $\mu_y - \sigma_y/\gamma_y < \mu_x - \sigma_x/\gamma_x$;

(b) $\mu_y - \sigma_y/\gamma_y \geq \mu_x - \sigma_x/\gamma_x$.

Vamos considerar o caso 1(a). Substituindo $v = w_y^{-1/\gamma_y}$, segue da Equação (4.9) que

$$\begin{aligned}
R &= (1 + \lambda_x)(1 + \lambda_y) \int_0^{\infty} \exp \left(-v - \left[\frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} v^{-\gamma_y} + 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right]^{-1/\gamma_x} \right) dv \\
&\quad - 2(1 + \lambda_x)\lambda_y \int_0^{\infty} \exp \left(-2v - \left[\frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} v^{-\gamma_y} + 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right]^{-1/\gamma_x} \right) dv \\
&\quad - \lambda_x(1 + \lambda_y) \int_0^{\infty} \exp \left(-v - 2 \left[\frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} v^{-\gamma_y} + 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right]^{-1/\gamma_x} \right) dv \\
&\quad + 2\lambda_x \lambda_y \int_0^{\infty} \exp \left(-2v - 2 \left[\frac{\gamma_x \sigma_y}{\sigma_x \gamma_y} v^{-\gamma_y} + 1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right]^{-1/\gamma_x} \right) dv \quad (4.10)
\end{aligned}$$

Portando, (4.4) segue de (4.3) e (4.10). Para o caso 1(b), basta notar que

$$P(X < Y) = 1 - P(Y < X)$$

e aplicar o resultado em (4.4) com subíndices intercambiados. Para os casos 2(a) e 2(b), o mesmo raciocínio pode ser aplicado. \square

Observação 4.1.1. Observe que se tomarmos $\lambda_x = \lambda_y = 0$, a distribuição GEV é obtida para X e Y , então o Teorema 4.1.1 generaliza o Teorema 3.1.1 apresentado no capítulo anterior.

Observação 4.1.2. Num cenário prático, a estimativa $(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\lambda}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\lambda}_y)$ deve ser obtida. Então se $\text{ sinal}(\hat{\gamma}_x) = \text{ sinal}(\hat{\gamma}_y) \neq 0$, então a condição $\mu_y - \frac{\sigma_y}{\gamma_y} \geq \mu_x - \frac{\sigma_x}{\gamma_x}$ or $\mu_y - \frac{\sigma_y}{\gamma_y} < \mu_x - \frac{\sigma_x}{\gamma_x}$ deve ser verificada e a expressão R correspondente deve ser usada.

Observação 4.1.3. Segue-se de (2.11) que, se $\mu_x - \sigma_x/\gamma_x = \mu_y - \sigma_y/\gamma_y$ e $\text{ sinal}(\gamma_x) = \text{ sinal}(\gamma_y) \neq 0$, então (4.4) pode ser escrito em termos da função H como:

$$\begin{aligned}
 R = & (1 + \lambda_x)(1 + \lambda_y) \frac{\gamma_x}{\gamma_y} \left(\frac{\sigma_y \gamma_x}{\sigma_x \gamma_y} \right)^{1/\gamma_y} H_{1,1}^{1,1} \left[\left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} \mid \begin{matrix} (1 - \frac{\gamma_x}{\gamma_y}, \frac{\gamma_x}{\gamma_y}) \\ (0, 1) \end{matrix} \right] \\
 & - 2\lambda_y(1 + \lambda_x) \frac{\gamma_x}{\gamma_y} \left(\frac{\sigma_y \gamma_x}{\sigma_x \gamma_y} \right)^{1/\gamma_y} H_{1,1}^{1,1} \left[2 \left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} \mid \begin{matrix} (1 - \frac{\gamma_x}{\gamma_y}, \frac{\gamma_x}{\gamma_y}) \\ (0, 1) \end{matrix} \right] \\
 & - \lambda_x(1 + \lambda_y) \frac{\gamma_x}{\gamma_y} 2^{-\gamma_x/\gamma_y} \left(\frac{\sigma_y \gamma_x}{\sigma_x \gamma_y} \right)^{1/\gamma_y} H_{1,1}^{1,1} \left[2^{-\gamma_x/\gamma_y} \left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} \mid \begin{matrix} (1 - \frac{\gamma_x}{\gamma_y}, \frac{\gamma_x}{\gamma_y}) \\ (0, 1) \end{matrix} \right] \\
 & + \lambda_x \lambda_y \frac{\gamma_x}{\gamma_y} 2^{1-\gamma_x/\gamma_y} \left(\frac{\sigma_y \gamma_x}{\sigma_x \gamma_y} \right)^{1/\gamma_y} H_{1,1}^{1,1} \left[2^{1-\gamma_x/\gamma_y} \left(\frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} \right)^{1/\gamma_y} \mid \begin{matrix} (1 - \frac{\gamma_x}{\gamma_y}, \frac{\gamma_x}{\gamma_y}) \\ (0, 1) \end{matrix} \right]. \quad (4.11)
 \end{aligned}$$

Em particular, usando um caso especial da função H como visto em Mathai, Saxena e Haubold (2010a), se $\mu_x - \frac{\sigma_x}{\gamma_x} = \mu_y - \frac{\sigma_y}{\gamma_y}$, $\frac{\sigma_y}{\sigma_x} < 2^{-\gamma_x}$ e $\gamma_x = \gamma_y \neq 0$, temos:

$$\begin{aligned}
 R = & (1 + \lambda_x)(1 + \lambda_y) \left[1 + \left(\frac{\sigma_y}{\sigma_x} \right)^{-1/\gamma_x} \right]^{-1} - 2(1 + \lambda_x)\lambda_y \left[2 + \left(\frac{\sigma_y}{\sigma_x} \right)^{-1/\gamma_x} \right]^{-1} \\
 & - \lambda_x(1 + \lambda_y) \left[1 + 2 \left(\frac{\sigma_y}{\sigma_x} \right)^{-1/\gamma_x} \right]^{-1} + \lambda_x \lambda_y \left[1 + \left(\frac{\sigma_y}{\sigma_x} \right)^{-1/\gamma_x} \right]^{-1}. \quad (4.12)
 \end{aligned}$$

Por fim, consideramos os casos de duas distribuições independentes TGEV com $\gamma_x = \gamma_y = 0$.

Teorema 4.1.2. Seja $X \sim TGEV(\mu_x, \sigma_x, 0, \lambda_x)$ e $Y \sim TGEV(\mu_y, \sigma_y, 0, \lambda_y)$ VAs independente com $\mu_j \in \mathbb{R}$, $\sigma_j > 0$, $\lambda_j \in [-1, 1]$, $j \in \{x, y\}$. Então

$$\begin{aligned}
R &= (1 + \lambda_x)(1 + \lambda_y) \mathbb{H} \left(1, \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, 0 \right) \\
&\quad - 2(1 + \lambda_x)\lambda_y \mathbb{H} \left(2, \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, 0 \right) \\
&\quad - \lambda_x(1 + \lambda_y) \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) \mathbb{H} \left(1, \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, \frac{\sigma_y}{\sigma_x} \right) \\
&\quad + 2\lambda_x\lambda_y \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) \mathbb{H} \left(2, \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, \frac{\sigma_y}{\sigma_x} \right). \tag{4.13}
\end{aligned}$$

Em particular, se tomarmos $\sigma_x = \sigma_y$, obtemos a forma explícita para

$$R = \frac{(1 + \lambda_x)(1 + \lambda_y)}{1 + \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right)} - \frac{2(1 + \lambda_x)\lambda_y}{2 + \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right)} - \frac{\lambda_x(1 + \lambda_y) \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right)}{\left(1 + \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) \right)^2} + \frac{2\lambda_x\lambda_y \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right)}{\left(2 + \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) \right)^2}.$$

Demonstração. Observe que

$$\begin{aligned}
R &= \int_{-\infty}^{+\infty} F_X(u; \mu_x, \sigma_x, \gamma_x, \lambda_x) f_Y(u; \mu_y, \sigma_y, \gamma_y, \lambda_y) du \\
&= \int_{-\infty}^{+\infty} \exp \left(- \exp \left(- \frac{u - \mu_x}{\sigma_x} \right) \right) \left[(1 + \lambda_x) - \lambda_x \exp \left(- \frac{u - \mu_x}{\sigma_x} \right) \right] \\
&\quad \times \frac{\exp \left(- \frac{u - \mu_y}{\sigma_y} \right) \exp \left(- \exp \left(- \frac{u - \mu_y}{\sigma_y} \right) \right)}{\sigma_y} \left[(1 + \lambda_y) - 2\lambda_y \exp \left(- \exp \left(- \frac{u - \mu_y}{\sigma_y} \right) \right) \right] du. \tag{4.14}
\end{aligned}$$

Substituindo $v = \exp \left(- \frac{u - \mu_y}{\sigma_y} \right)$, podemos reescrever (4.14) como

$$\begin{aligned}
R &= (1 + \lambda_x)(1 + \lambda_y) \int_0^\infty \exp \left(-v - \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) v^{\sigma_y/\sigma_x} \right) dv \\
&\quad - 2(1 + \lambda_x)\lambda_y \int_0^\infty \exp \left(-2v - \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) v^{\sigma_y/\sigma_x} \right) dv \\
&\quad - \lambda_x(1 + \lambda_y) \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) \int_0^\infty v^{\sigma_y/\sigma_x} \exp \left(-v - \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) v^{\sigma_y/\sigma_x} \right) dv \\
&\quad + 2\lambda_x\lambda_y \exp \left(- \frac{\mu_y - \mu_x}{\sigma_x} \right) \int_0^\infty v^{\sigma_y/\sigma_x} \exp \left(-2v - \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right) v^{\sigma_y/\sigma_x} \right) dv. \tag{4.15}
\end{aligned}$$

Por isso, (4.13) segue de (4.3) e (4.15). □

Observação 4.1.4. Segue-se de (2.11) e (2.12) e pela equação (4.13) podemos reescrever em termos da

função H como

$$\begin{aligned}
 R = & (1 + \lambda_x)(1 + \lambda_y) \frac{\sigma_x}{\sigma_y} \exp\left(\frac{\mu_y - \mu_x}{\sigma_y}\right) H_{1,1}^{1,1} \left[\exp\left(\frac{\mu_y - \mu_x}{\sigma_y}\right) \middle| \begin{matrix} (\frac{\sigma_y - \sigma_x}{\sigma_y}, \frac{\sigma_x}{\sigma_y}) \\ (0, 1) \end{matrix} \right] \\
 & - 2(1 + \lambda_x)\lambda_y \frac{\sigma_x}{\sigma_y} \exp\left(\frac{\mu_y - \mu_x}{\sigma_y}\right) H_{1,1}^{1,1} \left[2 \exp\left(\frac{\mu_y - \mu_x}{\sigma_y}\right) \middle| \begin{matrix} (\frac{\sigma_y - \sigma_x}{\sigma_y}, \frac{\sigma_x}{\sigma_y}) \\ (0, 1) \end{matrix} \right] \\
 & - \lambda_x(1 + \lambda_y) \exp\left(\frac{\mu_x - \mu_y}{\sigma_x}\right) H_{1,1}^{1,1} \left[\exp\left(\frac{\mu_x - \mu_y}{\sigma_x}\right) \middle| \begin{matrix} (-\frac{\sigma_y}{\sigma_x}, \frac{\sigma_y}{\sigma_x}) \\ (0, 1) \end{matrix} \right] \\
 & + \lambda_x\lambda_y \exp\left(\frac{\mu_x - \mu_y}{\sigma_x}\right) 2^{-\frac{\sigma_y}{\sigma_x}} H_{1,1}^{1,1} \left[2^{-\frac{\sigma_y}{\sigma_x}} \exp\left(\frac{\mu_x - \mu_y}{\sigma_x}\right) \middle| \begin{matrix} (-\frac{\sigma_y}{\sigma_x}, \frac{\sigma_y}{\sigma_x}) \\ (0, 1) \end{matrix} \right]. \quad (4.16)
 \end{aligned}$$

Estudo para o caso de Multicomponentes

Seja X_1, \dots, X_n VAs i.i.d com distribuição $TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e Y seja uma VA com distribuição independente $TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$. Seja $M_n = \max\{X_1, \dots, X_n\}$. Então,

$$P(M_n \leq u) = F^n(u; \mu_x, \sigma_x, \gamma_x, \lambda_x)$$

e temos

$$P(X_1 < Y, \dots, X_n < Y) = P(M_n \leq Y) = \int_{-\infty}^{\infty} F^n(u; \mu_x, \sigma_x, \gamma_x, \lambda_x) f(u; \mu_y, \sigma_y, \gamma_y, \lambda_y) du =: I_n. \quad (4.17)$$

No contexto mais amplo, considere VAs independente Y, X_1, \dots, X_k com $Y \sim TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$ e $X_j \sim TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$, para $j = 1, \dots, k$. A confiabilidade no sistema multicomponente no modelo stress-strength é obtido por

$$\begin{aligned}
 R_{s,k} &= P(\text{pelo menos } s \text{ das } (X_1, \dots, X_k) \text{ excedem } Y) \\
 &= \sum_{j=s}^k \binom{k}{j} \int_{-\infty}^{\infty} (1 - F(u; \mu_x, \sigma_x, \gamma_x, \lambda_x))^j (F(u; \mu_x, \sigma_x, \gamma_x, \lambda_x))^{k-j} f(u; \mu_y, \sigma_y, \gamma_y, \lambda_y) du.
 \end{aligned}$$

Usando expansão binomial, obtemos

$$R_{s,k} = \sum_{j=s}^k \sum_{r=0}^j \binom{k}{j} \binom{j}{r} (-1)^{j-r} \int_{-\infty}^{\infty} (F(u; \mu_x, \sigma_x, \gamma_x, \lambda_x))^{k-r} f(u; \mu_y, \sigma_y, \gamma_y, \lambda_y) du. \quad (4.18)$$

Note que a integral em termos (4.18) é a mesmo que (4.17) quando $n = k - r$. Logo,

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

Expressões explícitas para (4.17) são apresentados no resultado a seguir:

Teorema 4.1.3. Seja X_1, \dots, X_n VAs i.i.d. com distribuição $TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e Y uma VA independente com distribuição $TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$. Então

- Quando $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = 1$:

$$I_n = \sum_{l=0}^n \binom{n}{l} (-1)^{n-l} (1 + \lambda_x)^l \lambda_x^{n-l} \left[(1 + \lambda_y) \tilde{I}_{n,l} - 2\lambda_y \hat{I}_{n,l} \right], \quad (4.19)$$

onde

$$\tilde{I}_{n,l} = \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} (2n - l)^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right] (2n - l)^{-\gamma_x}, -\frac{1}{\gamma_x}, 0 \right)$$

e

$$\hat{I}_{n,l} = \mathbb{H} \left(2, (2n - l)^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, (2n - l)^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y} \right) \right], -\frac{1}{\gamma_x}, 0 \right),$$

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

- Quando $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = -1$, a expressão (4.19) se mantém desde que $\mu_x - \frac{\sigma_x}{\gamma_x} \geq \mu_y - \frac{\sigma_y}{\gamma_y}$.
- Quando $\gamma_x = \gamma_y = 0$:

$$I_n = \sum_{l=0}^n \binom{n}{l} (-1)^{n-l} (1 + \lambda_x)^l \lambda_x^{n-l} \exp \left(\frac{(n-l)(\mu_x - \mu_y)}{\sigma_x} \right) \left[(1 + \lambda_y) \tilde{J}_{n,l} - 2\lambda_y \hat{J}_{n,l} \right], \quad (4.20)$$

onde

$$\tilde{J}_{n,l} = \mathbb{H} \left(1, n \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, \frac{(n-l)\sigma_y}{\sigma_x} \right)$$

e

$$\hat{J}_{n,l} = \mathbb{H} \left(2, n \exp \left(\frac{\mu_x - \mu_y}{\sigma_x} \right), \frac{\sigma_y}{\sigma_x}, 0, 1, \frac{(n-l)\sigma_y}{\sigma_x} \right).$$

Demonstração. Para simplificar as notações, denotamos

$$F_X(u) = F(u; \mu_x, \sigma_x, \gamma_x, \lambda_x),$$

$$F_Y(u) = F(u; \mu_y, \sigma_y, \gamma_y, \lambda_y),$$

$$G_X(u) = G(u; \mu_x, \sigma_x, \gamma_x, \lambda_x),$$

e

$$G_Y(u) = G(u; \mu_y, \sigma_y, \gamma_y, \lambda_y).$$

Segue-se de (4.1) e (4.17) por isso

$$\begin{aligned} I_n &= \int_{-\infty}^{+\infty} F_X^n(u) f_Y(u) du \\ &= \int_{-\infty}^{+\infty} [(1 + \lambda_x) G_X(u) - \lambda_x G_X^2(u)]^n f_Y(u) du. \end{aligned}$$

Por expansão binomial

$$I_n = \sum_{l=0}^n \binom{n}{l} (-1)^{n-l} (1 + \lambda_x)^l \lambda_x^{n-l} \int_{-\infty}^{+\infty} G_X(u)^{2n-l} f_Y(u) du. \quad (4.21)$$

Observe que $f_y(u) = F'_y(u)$, que implica

$$\begin{aligned} \int_{-\infty}^{+\infty} G_X(u)^{2n-l} f_Y(u) du &= \int_{-\infty}^{+\infty} G_X(u)^{2n-l} [(1 + \lambda_y) g_y(u) - 2\lambda_y G_Y(u) g_Y(u)] du \\ &= (1 + \lambda_y) \int_{-\infty}^{+\infty} G_X(u)^{2n-l} g_Y(u) du \\ &\quad - 2\lambda_y \int_{-\infty}^{+\infty} G_X(u)^{2n-l} G_Y(u) g_Y(u) du. \end{aligned} \quad (4.22)$$

se $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) = 1$, segue de (3.17) em que

$$\begin{aligned} &\int_{-\infty}^{+\infty} G_X(u)^{2n-l} g_Y(u) du \\ &= \mathbb{H} \left(1, \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x} (2n-l)^{-\gamma_x}, -\gamma_y, \left[1 + \frac{\gamma_x}{\sigma_x} (\mu_y - \mu_x - \frac{\sigma_y}{\gamma_y}) \right] (2n-l)^{-\gamma_x}, -\frac{1}{\gamma_x}, 0 \right), \end{aligned} \quad (4.23)$$

desde que $\mu_x - \frac{\sigma_x}{\gamma_x} \leq \mu_y - \frac{\sigma_y}{\gamma_y}$. Se $\text{sign}(\gamma_x) = \text{sign}(\gamma_y) = -1$, (3.19) implica (4.23) desde que $\mu_x - \frac{\sigma_x}{\gamma_x} \geq \mu_y - \frac{\sigma_y}{\gamma_y}$.

Observe que o intervalo de integração pode ser simplificado usando os resultados para a interseção dos suportes de G_X e g_Y , tais que:

$$S = \text{supp } G_X \cap \text{supp } g_Y = \begin{cases} (M, +\infty), & \gamma_x > 0 \text{ e } \gamma_y > 0, \\ (-\infty, m), & \gamma_x < 0 \text{ e } \gamma_y < 0, \\ \mathbb{R}, & \gamma_x = \gamma_y = 0, \end{cases}$$

onde $M = \max\{\mu_x - \frac{\sigma_x}{\gamma_x}, \mu_y - \frac{\sigma_y}{\gamma_y}\}$ e $m = \min\{\mu_x - \frac{\sigma_x}{\gamma_x}, \mu_y - \frac{\sigma_y}{\gamma_y}\}$. Então, se $\gamma_x > 0$ e $\gamma_y > 0$ (caso $\gamma_x < 0$ e $\gamma_y < 0$ é análogo), temos que $\tilde{I} := \int_{-\infty}^{+\infty} G_x(u)^{2n-l} G_y(u) g_y(u) du$ é obtido por

$$\begin{aligned} \tilde{I} &= \int_M^{+\infty} \exp \left\{ -(2n-l) \left[1 + \frac{\gamma_x}{\sigma_x} (u - \mu_x) \right]^{-1/\gamma_x} - 2 \left[1 + \frac{\gamma_y}{\sigma_y} (u - \mu_y) \right]^{-1/\gamma_y} \right\} \\ &\quad \times \left[1 + \frac{\gamma_y}{\sigma_y} (u - \mu_y) \right]^{-1/\gamma_y - 1} \frac{du}{\sigma_y}. \end{aligned}$$

Substituindo $v = \left[1 + \frac{\gamma_y}{\sigma_y} (u - \mu_y) \right]^{-1/\gamma_y}$, obtemos

$$\begin{aligned} &-2\lambda_y \int_{-\infty}^{+\infty} G_x(u)^{2n-l} G_y(u) g_y(u) du \tag{4.24} \\ &= -2\lambda_y \mathbb{H} \left(2, (2n-l)^{-\gamma_x} \frac{\gamma_x \sigma_y}{\gamma_y \sigma_x}, -\gamma_y, (2n-l)^{-\gamma_x} \left[1 + \frac{\gamma_x}{\sigma_x} \left(\mu_y - \mu_x - \frac{\sigma_y}{\gamma_x} \right) \right], -\frac{1}{\gamma_x}, 0 \right). \end{aligned}$$

Logo, (4.19) segue de (4.21), (4.22), (4.23) e (4.24). Por outro lado, quando $\gamma_x = \gamma_y = 0$, a prova segue o mesmo raciocínio do caso da prova do Teorema 4.1.2, apenas considerando a expansão binomial no processo. Esta prova será omitida por questão de analogia. \square

4.2 Estimadores e Estudo de Simulações

Esta seção apresentamos a estimação de $R = P(X < Y)$ dado duas variáveis aleatórias independentes TGEV. A literatura apresenta alguns estimadores de máxima verossimilhança (MLEs) para R considerando formas explícitas de R obtidas após restrições severas nos parâmetros das distribuições de valores extremos. Essas abordagens exigem que a estimação dos parâmetros seja feita em conjunto nas duas amostras e

também requerem uma série de transformações para serem aplicadas corretamente aos componentes TGEV.

Para a distribuição TGEV, temos dois casos a considerar: $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) \neq 0$ e $\gamma_x = \gamma_y = 0$. O primeiro caso exige que $\mu_y - \sigma_y/\gamma_y \geq \mu_x - \sigma_x/\gamma_x$ ou $\mu_y - \sigma_y/\gamma_y < \mu_x - \sigma_x/\gamma_x$ (Teorema 4.1.1). Por outro lado, se $\gamma_x = \gamma_y = 0$, liberamos quaisquer restrições nos parâmetros para as expressões, pois uma fórmula única pode ser usada para obter R em termos de funções \mathbb{H} (Teorema 4.1.2).

Seja $X \sim TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e $Y \sim TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$ VAs independentes com $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) \neq 0$. Observe que $R = R(\boldsymbol{\theta})$, onde $\boldsymbol{\theta} = (\mu_x, \sigma_x, \gamma_x, \lambda_x, \mu_y, \sigma_y, \gamma_y, \lambda_y)$. Assim, seja $\mathbf{x} = (X_1, \dots, X_n)$ seja uma amostra aleatória seguindo distribuição $TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e $\mathbf{y} = (Y_1, \dots, Y_m)$ seja uma amostra aleatória de $TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$, independente de \mathbf{X} e com $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) \neq 0$. Seja $\hat{\boldsymbol{\theta}} = (\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\lambda}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\lambda}_y)$ as estimativas de $\boldsymbol{\theta}$. Como o Teorema 4.1.1 descreve R em termos de integrais (portanto, funções contínuas e mensuráveis), somos capazes de estimar R simplesmente como $\hat{R} = R(\hat{\boldsymbol{\theta}})$ devido à propriedade de invariância do MLE.

Considere a FDP $f(\cdot; \mu, \sigma, \gamma, \lambda)$ definida na equação (4.2). A função de verossimilhança é dada por:

$$L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \prod_{j=1}^n f(X_j; \mu_x, \sigma_x, \gamma_x, \lambda_x) \prod_{i=j}^m f(Y_j; \mu_y, \sigma_y, \gamma_y, \lambda_y). \quad (4.25)$$

Quando $\gamma_x = \gamma_y = 0$, o suporte de f não depende de parâmetros desconhecidos e o Teorema 4.1.2 não requer restrições nos parâmetros. A função de log-verossimilhança é dada por:

$$\begin{aligned} l(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = & -n\sigma_x - m\sigma_y - \sum_{j=1}^n \frac{x_j - \mu_x}{\sigma_x} - \sum_{k=1}^m \frac{y_k - \mu_y}{\sigma_y} \\ & - \sum_{j=1}^n \exp\left(-\frac{x_j - \mu_x}{\sigma_x}\right) + \sum_{j=1}^n \log\left[\left(1 + \lambda_x\right) - 2\lambda_x \exp\left(-\exp\left(-\frac{x_j - \mu_x}{\sigma_x}\right)\right)\right] \\ & - \sum_{k=1}^m \exp\left(-\frac{y_k - \mu_y}{\sigma_y}\right) + \sum_{k=1}^m \log\left[\left(1 + \lambda_y\right) - 2\lambda_y \exp\left(-\exp\left(-\frac{y_k - \mu_y}{\sigma_y}\right)\right)\right]. \end{aligned} \quad (4.26)$$

Assim, o MLE pode ser obtida pela função de log-verossimilhança (4.26), igualando seu gradiente a zero e encontrando seus pontos críticos.

Quando $\text{sinal}(\gamma_x) = \text{sinal}(\gamma_y) \neq 0$, o suporte de f depende do parâmetro desconhecido $(\mu_j, \sigma_j, \gamma_j)$, $j \in x, y$. Portanto, não somos capazes de obter a MLE explicitamente, sendo necessário um procedimento numérico adicional para realizar a maximização da função de verossimilhança. Isso é semelhante ao que

acontece com a distribuição GEV (para uma discussão mais detalhada consulte Embrechts, Klüppelberg e Mikosch (2013)). A função de verossimilhança torna-se:

$$\begin{aligned}
 L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) &= \sigma_x^{-n} \sigma_y^{-m} \exp \left\{ - \sum_{j=1}^n w_{x,j}^{-1/\gamma_x} \right\} \exp \left\{ - \sum_{k=1}^m w_{y,k}^{-1/\gamma_y} \right\} \\
 &\times \prod_{j=1}^n \frac{[(1 + \lambda_x) - 2\lambda_x \exp(-w_{x,j}^{-1/\gamma_x})]}{w_{x,j}^{1+1/\gamma_x}} \mathbb{1}_{(0,\infty)}(w_{x,j}) \\
 &\times \prod_{k=1}^m \frac{[(1 + \lambda_y) - 2\lambda_y \exp(-w_{y,k}^{-1/\gamma_y})]}{w_{y,k}^{1+1/\gamma_y}} \mathbb{1}_{(0,\infty)}(w_{y,k}), \tag{4.27}
 \end{aligned}$$

onde $w_{x,j} = 1 + \gamma_x(x_j - \mu_x)/\sigma_x$ e $w_{y,k} = 1 + \gamma_y(y_k - \mu_y)/\sigma_y$. Observe que $\prod_{j=1}^n \mathbb{1}_{(0,\infty)}(w_{x,j}) > 0$ se e somente se $w_{x,j} \in (0, \infty)$ para todo $j = 1, \dots, n$. Uma restrição semelhante deve ser observada para $w_{y,k}$. Procedimentos numéricos devem ser aplicados para superar a falta de uma expressão explícita de MLE.

Introduzimos um método alternativo para estimar $\boldsymbol{\theta}$ através de um processo de duas etapas, conforme descrito abaixo:

Algoritmo 4.2.1. Dadas as amostras \mathbf{x} e \mathbf{y} ,

Passo 1 Estimamos $(\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x)$ e $(\hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y)$ utilizando MLE para a distribuição GEV¹;

Passo 2 Os parâmetros (λ_x, λ_y) são estimados determinando

$$(\hat{\lambda}_x, \hat{\lambda}_y) = \arg \max_{[-1,1]^2} L(\lambda_x, \lambda_y; \mathbf{x}, \mathbf{y}),$$

onde $L(\lambda_x, \lambda_y; \mathbf{x}, \mathbf{y})$ é dada em (4.25) utilizando os parâmetros estimados no Passo 1 como sendo valores conhecidos.

Para escolher entre os Teoremas 4.1.1 e 4.1.2 para obter $R(\hat{\boldsymbol{\theta}})$, precisamos verificar se:

(a) $\hat{\gamma}_x \approx \hat{\gamma}_y \approx 0$;

ou,

(b) $\hat{\mu}_y - \hat{\sigma}_y/\hat{\gamma}_y \geq \hat{\mu}_x - \hat{\sigma}_x/\hat{\gamma}_x$ ou $\hat{\mu}_y - \hat{\sigma}_y/\hat{\gamma}_y < \hat{\mu}_x - \hat{\sigma}_x/\hat{\gamma}_x$.

¹Esta estimativa pode ser realizada utilizando o pacote *extRemes* dentro do software R R Core Team, 2023.

Apesar dessa verificação adicional, espera-se que o tempo computacional necessário para o Algoritmo 4.2.1 seja menor do que o necessário para maximizar diretamente as equações (4.25) e (4.26).

O Algoritmo 4.2.2 descreve a abordagem utilizada para obter intervalos de confiança (ICs) para as estimativas de R .

Algoritmo 4.2.2. Seja (\mathbf{x}, \mathbf{y}) uma amostra de tamanho n e M um número inteiro positivo que denota o número de repetições de Bootstrap.

Passo 1 Gerar amostras de Bootstrap $(\mathbf{x}, \mathbf{y})_i$.

Passo 2 Calcular as estimativas $\hat{\theta}_i = (\hat{\mu}_x, \hat{\sigma}_x, \hat{\gamma}_x, \hat{\lambda}_x, \hat{\mu}_y, \hat{\sigma}_y, \hat{\gamma}_y, \hat{\lambda}_y)_i$ com base em $(\mathbf{x}, \mathbf{y})_i$. Neste caso, os parâmetros de cada amostra de Bootstrap são estimados individualmente usando o Algoritmo 4.2.1.

Passo 3 Obter $\hat{R}_i = R(\hat{\theta}_i)$ usando o Teorema 4.1.1 ou 4.1.2.

Passo 4 Repetir os Passos 1 a 3 M vezes.

Passo 5 O intervalo de confiança aproximado de $100(1-\alpha)\%$ para \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 cumulativa de \hat{R} .

Para o problema de seleção de ativos usando o SSR apresentado na Seção 4.3, apenas uma série temporal de retornos observados está disponível para cada ativo. Portanto, a abordagem de MLE acima é de extrema importância. Para ilustrar a adequação das expressões analíticas aqui derivadas, apresentamos um estudo de simulação. Nesse caso, várias amostras de tamanho n podem ser retiradas de cada variável aleatória, que são então usadas para estimar o valor de R e podem ser repetidas várias vezes.

Apresentamos um estudo envolvendo simulações de Monte Carlo que analisam o desempenho do estimador $\hat{R} = R(\hat{\theta})$. Para avaliar o desempenho do estimador $\hat{R} = R(\hat{\theta})$, fixamos vários valores dos parâmetros $\mu_x, \sigma_x, \gamma_x, \lambda_x, \mu_y, \sigma_y, \gamma_y, \lambda_y$, e então geramos $N = 100$ amostras de Monte Carlo, cada uma delas com tamanho $n = 100$, das variáveis aleatórias $X \sim TGEV(\mu_x, \sigma_x, \gamma_x, \lambda_x)$ e $Y \sim TGEV(\mu_y, \sigma_y, \gamma_y, \lambda_y)$. Analisamos as estimativas \hat{R} , (Bias) e erro quadrático médio (RMSE).

Amostras aleatórias da distribuição TGEV podem ser geradas pelo método da inversão de quantis. Recorde que a função u-quantil da distribuição TGEV foi apresentada por Otiniano, De Paiva e Neto

(2019) e, para conveniência do leitor, é apresentada a seguir Conforme descrito por:

$$F^{-1}(U) = \begin{cases} \frac{\sigma}{\gamma} \left\{ -1 + \left[-\log \left(\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda U}}{2\lambda} \right) \right]^{-\gamma} \right\}, & \gamma \neq 0, \\ \mu + \sigma \left\{ -\log \left[-\log \left(\frac{1+\lambda-\sqrt{(1+\lambda)^2-4\lambda U}}{2\lambda} \right) \right] \right\} & \gamma = 0, \end{cases}$$

onde U é uma VA uniforme em $[0, 1]$.

Para as simulações, em cada linha das Tabelas 4.1 e 4.2 foi realizado o seguinte procedimento:

1. para cada amostra de Monte-Carlo, a estimativa $\hat{R} = R(\hat{\theta})$ é calculada;
2. \hat{R}_{MC} é avaliado tomando a média amostral das amostras de Monte-Carlo \hat{R} ;
3. o Bias é calculado como a diferença entre o valor teórico R e \hat{R}_{MC} . O mesmo acontece com o RMSE, que também considera o valor verdadeiro como aquele obtido analiticamente.

A distribuição TGEV com parâmetros de forma negativa é tratada nas Tabelas 4.1 e 4.3, enquanto as Tabelas 4.2 e 4.4 tratam de parâmetros de forma positiva. Em ambos os casos, o estimador apresenta bom comportamento com Bias e RMSE baixos. Além disso, aumentar o número de replicatas N leva às mesmas conclusões (ver Tabela 4.4).

Tabela 4.1: Bias e RMSE do \hat{R}_{MC} ($N = 100$ e $n = 100$) com parâmetro de forma negativo.

μ_x	σ_x	γ_x	λ_x	μ_y	σ_y	γ_y	λ_y	R	\hat{R}_{MC}	Bias	RMSE
0	0.7	-0.1	0.3	0	0.5	-0.2	0.1	0.50945	0.50151	-0.00793	0.00030
0	0.7	-0.1	0.3	0.5	0.5	-0.2	0.1	0.71367	0.70374	-0.00993	0.00034
0	0.7	-0.1	0.3	0	0.7	-0.2	0.1	0.52511	0.51976	-0.00534	0.00031
0	0.7	-0.1	0.3	0.5	0.7	-0.2	0.1	0.70569	0.70066	-0.00503	0.00026
0	0.7	-0.1	0.3	0	0.5	-0.4	0.1	0.49137	0.48478	-0.00659	0.00029
0	0.7	-0.1	0.3	0.5	0.5	-0.4	0.1	0.69977	0.69183	-0.00794	0.00030
0	0.7	-0.1	0.3	0	0.7	-0.4	0.1	0.50809	0.50191	-0.00618	0.00030
0	0.7	-0.1	0.3	0.5	0.7	-0.4	0.1	0.69163	0.68617	-0.00546	0.00033
0	0.7	-0.1	0.3	0	0.5	-0.2	0.2	0.49602	0.48709	-0.00893	0.00036
0	0.7	-0.1	0.3	0.5	0.5	-0.2	0.2	0.70372	0.69562	-0.00810	0.00034
0	0.7	-0.1	0.3	0	0.7	-0.2	0.2	0.50870	0.50454	-0.00416	0.00029
0	0.7	-0.1	0.3	0.5	0.7	-0.2	0.2	0.69299	0.68870	-0.00429	0.00031
0	0.7	-0.1	0.3	0	0.5	-0.4	0.2	0.47863	0.47081	-0.00782	0.00030
0	0.7	-0.1	0.3	0.5	0.5	-0.4	0.2	0.68979	0.68537	-0.00442	0.00027
0	0.7	-0.1	0.3	0	0.7	-0.4	0.2	0.49230	0.49000	-0.00230	0.00033
0	0.7	-0.1	0.3	0.5	0.7	-0.4	0.2	0.67876	0.67348	-0.00529	0.00030

Tabela 4.2: Bias e RMSE do \hat{R}_{MC} ($N = 100$ e $n = 100$) com parâmetro de forma positivo.

μ_x	σ_x	γ_x	λ_x	μ_y	σ_y	γ_y	λ_y	R	\hat{R}_{MC}	Bias	RMSE
0	0.7	0.1	0.3	0	0.5	0.2	0.1	0.53183	0.52584	-0.00600	0.00030
0	0.7	0.1	0.3	0.5	0.5	0.2	0.1	0.71971	0.71210	-0.00761	0.00029
0	0.7	0.1	0.3	0	0.7	0.2	0.1	0.54076	0.53666	-0.00410	0.00030
0	0.7	0.1	0.3	0.5	0.7	0.2	0.1	0.71272	0.70832	-0.00440	0.00030
0	0.7	0.1	0.3	0	0.5	0.4	0.1	0.54736	0.54049	-0.00686	0.00029
0	0.7	0.1	0.3	0.5	0.5	0.4	0.1	0.73060	0.72135	-0.00926	0.00031
0	0.7	0.1	0.3	0	0.7	0.4	0.1	0.55478	0.55053	-0.00425	0.00023
0	0.7	0.1	0.3	0.5	0.7	0.4	0.1	0.72470	0.71852	-0.00618	0.00027
0	0.7	0.1	0.3	0	0.5	0.2	0.2	0.51767	0.51685	-0.00082	0.00028
0	0.7	0.1	0.3	0.5	0.5	0.2	0.2	0.71043	0.70462	-0.00581	0.00030
0	0.7	0.1	0.3	0	0.7	0.2	0.2	0.52393	0.52110	-0.00284	0.00028
0	0.7	0.1	0.3	0.5	0.7	0.2	0.2	0.70105	0.69495	-0.00610	0.00029
0	0.7	0.1	0.3	0	0.5	0.4	0.2	0.53280	0.52937	-0.00343	0.00031
0	0.7	0.1	0.3	0.5	0.5	0.4	0.2	0.72130	0.71108	-0.01021	0.00031
0	0.7	0.1	0.3	0	0.7	0.4	0.2	0.53775	0.53576	-0.00199	0.00027
0	0.7	0.1	0.3	0.5	0.7	0.4	0.2	0.71325	0.71132	-0.00193	0.00024

Tabela 4.3: Bias e RMSE do \hat{R}_{MC} ($N = 1000$ e $n = 100$) com parâmetro de forma negativo.

μ_x	σ_x	γ_x	λ_x	μ_y	σ_y	γ_y	λ_y	R	\hat{R}_{MC}	Bias	RMSE
0	0.7	-0.1	0.3	0	0.5	-0.2	0.1	0.50945	0.50135	-0.00809	0.00031
0	0.7	-0.1	0.3	0.5	0.5	-0.2	0.1	0.71367	0.70749	-0.00618	0.00027
0	0.7	-0.1	0.3	0	0.7	-0.2	0.1	0.52511	0.51830	-0.00680	0.00032
0	0.7	-0.1	0.3	0.5	0.7	-0.2	0.1	0.70569	0.69953	-0.00616	0.00028
0	0.7	-0.1	0.3	0	0.5	-0.4	0.1	0.49137	0.48216	-0.00921	0.00033
0	0.7	-0.1	0.3	0.5	0.5	-0.4	0.1	0.69977	0.69122	-0.00855	0.00032
0	0.7	-0.1	0.3	0	0.7	-0.4	0.1	0.50809	0.50197	-0.00612	0.00028
0	0.7	-0.1	0.3	0.5	0.7	-0.4	0.1	0.69163	0.68582	-0.00581	0.00031
0	0.7	-0.1	0.3	0	0.5	-0.2	0.2	0.49602	0.48986	-0.00616	0.00030
0	0.7	-0.1	0.3	0.5	0.5	-0.2	0.2	0.70372	0.69768	-0.00604	0.00028
0	0.7	-0.1	0.3	0	0.7	-0.2	0.2	0.50870	0.50619	-0.00250	0.00027
0	0.7	-0.1	0.3	0.5	0.7	-0.2	0.2	0.69299	0.68945	-0.00354	0.00027
0	0.7	-0.1	0.3	0	0.5	-0.4	0.2	0.47863	0.47359	-0.00504	0.00030
0	0.7	-0.1	0.3	0.5	0.5	-0.4	0.2	0.68979	0.68290	-0.00689	0.00031
0	0.7	-0.1	0.3	0	0.7	-0.4	0.2	0.49230	0.48817	-0.00413	0.00031
0	0.7	-0.1	0.3	0.5	0.7	-0.4	0.2	0.67876	0.67517	-0.00359	0.00030

Tabela 4.4: Bias e RMSE do \hat{R}_{MC} ($N = 1000$ e $n = 100$) com parâmetro de forma positivo.

μ_x	σ_x	γ_x	λ_x	μ_y	σ_y	γ_y	λ_y	R	\hat{R}_{MC}	Bias	RMSE
0	0.7	0.1	0.3	0	0.5	0.2	0.1	0.53183	0.52561	-0.00623	0.00031
0	0.7	0.1	0.3	0.5	0.5	0.2	0.1	0.71971	0.71151	-0.00820	0.00031
0	0.7	0.1	0.3	0	0.7	0.2	0.1	0.54076	0.53576	-0.00500	0.00031
0	0.7	0.1	0.3	0.5	0.7	0.2	0.1	0.71272	0.70567	-0.00705	0.00029
0	0.7	0.1	0.3	0	0.5	0.4	0.1	0.54736	0.54108	-0.00628	0.00031
0	0.7	0.1	0.3	0.5	0.5	0.4	0.1	0.73060	0.72414	-0.00646	0.00027
0	0.7	0.1	0.3	0	0.7	0.4	0.1	0.55478	0.54712	-0.00765	0.00030
0	0.7	0.1	0.3	0.5	0.7	0.4	0.1	0.72470	0.71812	-0.00658	0.00032
0	0.7	0.1	0.3	0	0.5	0.2	0.2	0.51767	0.51462	-0.00305	0.00029
0	0.7	0.1	0.3	0.5	0.5	0.2	0.2	0.71043	0.70328	-0.00715	0.00030
0	0.7	0.1	0.3	0	0.7	0.2	0.2	0.52393	0.52199	-0.00194	0.00032
0	0.7	0.1	0.3	0.5	0.7	0.2	0.2	0.70105	0.69579	-0.00526	0.00031
0	0.7	0.1	0.3	0	0.5	0.4	0.2	0.53280	0.52864	-0.00416	0.00031
0	0.7	0.1	0.3	0.5	0.5	0.4	0.2	0.72130	0.71402	-0.00728	0.00029
0	0.7	0.1	0.3	0	0.7	0.4	0.2	0.53775	0.53627	-0.00148	0.00028
0	0.7	0.1	0.3	0.5	0.7	0.4	0.2	0.71325	0.70860	-0.00464	0.00030

4.3 Aplicações no Mercado de Ações

Para avaliar os resultados teóricos propostos na Seção 4.1, apresentamos aplicações do SSR na seleção de ativos. Medidas do tipo $P(X < Y)$ podem servir como guia para selecionar ativos financeiros ao gerenciar uma carteira. A abordagem envolve modelar os log-retornos dos preços das ações por meio da distribuição TGEV.

No Capítulo 3, mostramos que a distribuição GEV se ajusta adequadamente aos dados. Nosso interesse agora é determinar se o uso da distribuição TGEV melhora o ajuste com a adição do parâmetro extra λ . Para alcançar isso, empregamos o método de estimação em duas etapas, descrito no Algoritmo 4.2.1. Considerando o diferente número de parâmetros dos modelos GEV e TGEV, para conduzir uma análise comparativa entre esses modelos, utilizamos os critérios de informação *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) e *Efficient Determination Criterion* (EDC). As estimativas dos parâmetros para os log-retornos dos preços das ações estão apresentadas na Tabela 4.5, enquanto a Tabela 4.6 mostra que todos os critérios indicam uma melhoria no ajuste ao usar a distribuição TGEV, em comparação com a GEV.

Tabela 4.5: Estimativas dos parâmetros para os preços dos log-retornos das ações: BBAS3, ITUB4, VVIIA3 e VALE3.

Ações	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
BBAS3	0.0103	-0.0063	0.0219	-0.2535
ITUB4	-0.0373	-0.0064	0.0165	-0.1545
VALE3	-0.0088	-0.0095	0.0222	-0.1631
VIIA3	-0.0058	-0.0217	0.0396	-0.1170

Tabela 4.6: Critérios de informação e p-valores do teste KS para os modelos GEV e TGEV.

Ações	Distribuição	AIC	BIC	EDC	KS p-valor
BBAS3	TGEV	-1615.84	-1654.23	-1621.90	0.0155
	GEV	-1613.83	-1642.63	-1618.38	0.0133
ITUB4	TGEV	-1749.33	-1787.72	-1755.39	0.3303
	GEV	-1747.28	-1776.07	-1751.82	0.4422
VALE3	TGEV	-1556.20	-1594.59	-1562.26	0.2388
	GEV	-1554.18	-1582.97	-1558.73	0.2313
VIIA3	TGEV	-1143.34	-1181.73	-1149.40	0.6851
	GEV	-1141.40	-1170.19	-1145.94	0.7091

A adequação das distribuições TGEV ajustadas pode ser avaliada por meio de métodos de avaliação gráfica. Isso inclui traçar a FDP teórica sobre o histograma (Figura 4.1), comparar a FDA teórica contra a FDA empírica (ECDF) (Figura 4.2), e examinar os gráficos quantil-quantil (QQ) normais dos resíduos (Figura 4.3). Embora o teste Kolmogorov-Smirnov (KS) rejeite o ajuste TGEV para os dados BBAS3, uma análise visual do histograma e do ECDF não descarta a adequação da distribuição TGEV. Além disso, o teste KS tende a ser excessivamente sensível, especialmente para tamanhos amostrais médios a grandes, levando à sua responsividade mesmo a desvios mínimos, o que pode explicar essa discrepância.

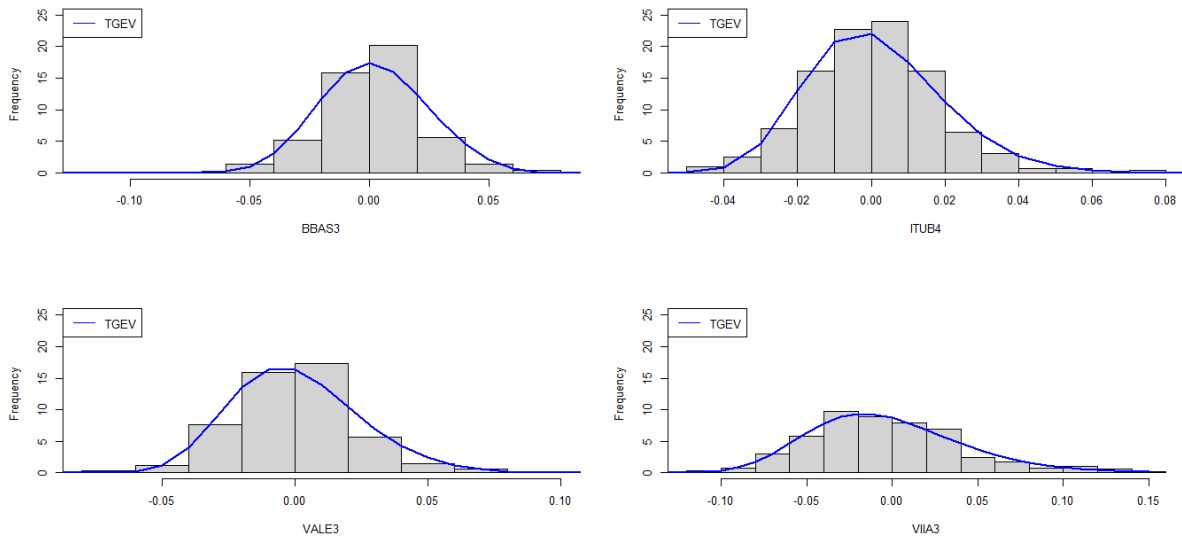


Figura 4.1: Histogramas e densidades TGEV ajustadas para os log-retornos dos preços das ações.

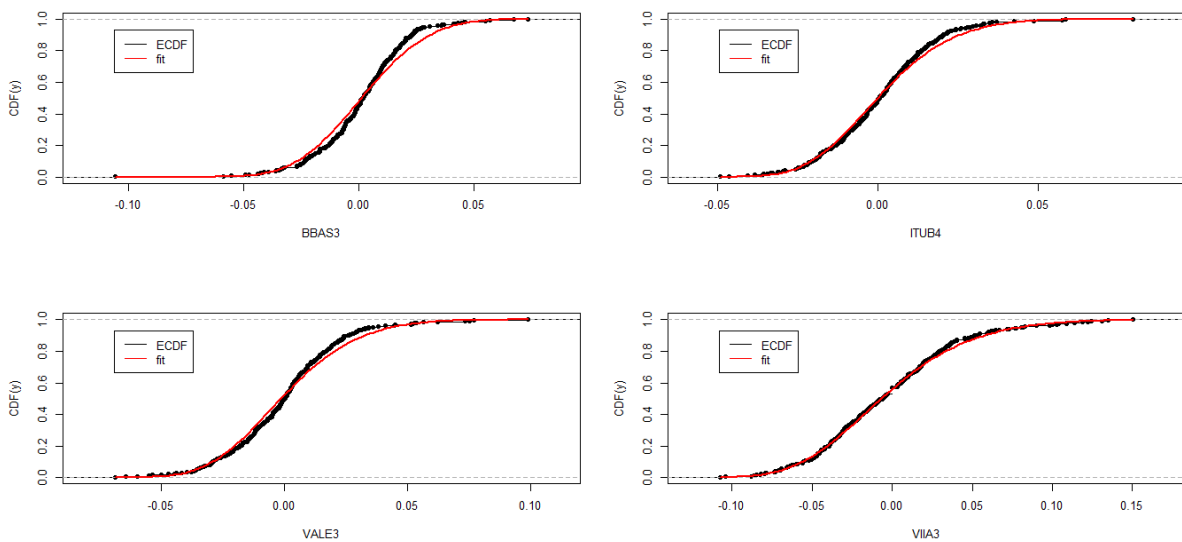


Figura 4.2: ECDF ajustadas para os modelos TGEV log-retornos dos preços das ações.

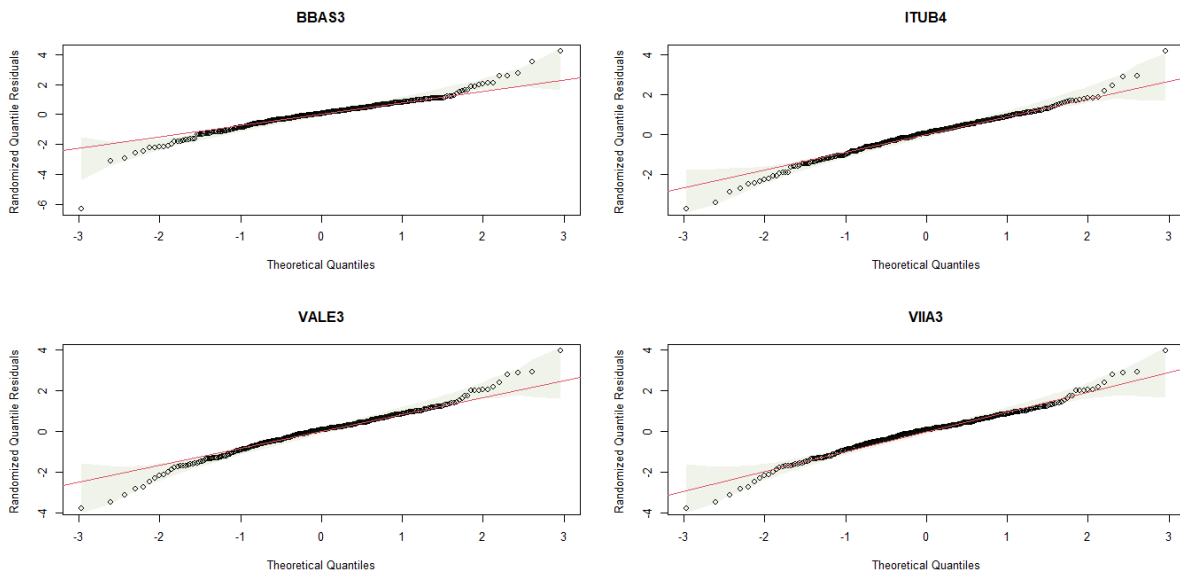


Figura 4.3: Gráfico QQ normal exibindo resíduos dos modelos TGEV ajustados.

As medidas de confiabilidade, denotadas como $R = P(X < Y)$, desempenham um papel fundamental no processo de tomada de decisão de um investidor. Para simplificar, quando X e Y simbolizam lucro a partir de retornos logarítmicos e $R < 1/2$, o investidor tende a favorecer a seleção do ativo financeiro correspondente a X . Por outro lado, se $R > 1/2$, o investidor inclina-se para a escolha oposta. No entanto, quando $R = 1/2$, a decisão se torna inconclusiva. Nesse sentido, a Tabela 4.7 apresenta as estimativas de $P(X < Y)$ e os ICs de Bootstrap de 95%, obtidos por \hat{R} e pelo Algoritmo 4.2.2.

Tabela 4.7: Estimativas da probabilidade *Stress-strength* e IC de bootstrap para log-retornos seguindo distribuição TGEV.

X	Y	\hat{R}	95% CI
VALE3	BBAS3	0.53	(0.40; 0.59)
VALE3	ITUB4	0.52	(0.40; 0.59)
VALE3	VIIA3	0.45	(0.38; 0.52)

Utilizando a distribuição GEV, estimativas de confiabilidade \hat{R}^{GEV} de 0,54, 0,54 e 0,43 para os pares VALE3-BBAS3, VALE3-ITUB4 e VALE3-VIIA3 foram obtidas no Capítulo 3, respectivamente. Estes valores assemelham-se aos descritos na Tabela 4.7 para distribuição TGEV. Em relação aos intervalos de confiança, houve redução no tamanho do intervalo para o último par de ações, enquanto os tamanhos permaneceram consistentes para os demais.

As estimativas pontuais também podem ser comparadas com um estimador empírico que não depende

da estimativa de parâmetros ou da escolha de um modelo probabilístico. Considerando o estimador:

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

onde $\mathbb{1}_A$ denota a função do indicador no conjunto A e n é o tamanho da amostra. As estimativas obtidas são, respectivamente, 0,55, 0,55 e 0,43 para os pares apresentados na Tabela 4.7, mostrando a proximidade das estimativas paramétricas e não paramétricas.

Capítulo 5

Conclusões

Apesar de suas limitações significativas, a teoria moderna de portfólio de Markowitz ainda é utilizada por muitos profissionais devido à sua simplicidade de fácil utilização. Dessa forma, estudar abordagens alternativas que também sejam de fácil compreensão é de extrema importância.

Nesta dissertação, estudamos a confiabilidade *stress-strength* $R = P(X < Y)$ quando X e Y seguem distribuições GEV de três parâmetros. Em resumo, quando X e Y representam VAs 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. Assim, expressões exatas para R foram obtidas em termos da função de valor extremo III com restrições mínimas de parâmetros.

Simulações de Monte-Carlo atestam o desempenho das expressões analíticas de R obtidas em termos de funções especiais aqui derivadas. 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, também foi descrita uma aplicação de engenharia, onde fibras de carbono testadas sob tensão foram modeladas no sentido de confiabilidade *stress-strength*.

Foram estendidos os estudos supracitados para distribuições TGEV. Uma abordagem de estimativa em duas etapas foi proposta e Simulações de Monte-Carlo atestam o desempenho das expressões analíticas também obtidas para a distribuição TGEV, por ser um modelo com maior parâmetros obteve-se um custo computacional maior do que obtido para distribuição GEV, porém ambos os modelos apresentaram boas estimativas de R , atestando assim bons ajuste de estimativas pontuais e intervalar para ambos os casos. Uma abordagem Bayesiana para estimação de R e modelos de séries temporais para comparação de ativos financeiros são sugestões de trabalhos futuros.

Apêndice A

Artigo: Quintino et al. (2024)

Autores: Felipe S. Quintino, Melquisadec Oliveira, Pushpa N. Rathie, Luan C. S. M. Ozelim and Tiago A. da Fonseca

Título: “Asset selection based on estimating stress-strength probabilities: The case of returns following three-parameter generalized extreme value distributions”

Revista: AIMS Mathematics

Ano: 2024



Research article

Asset selection based on estimating stress-strength probabilities: The case of returns following three-parameter generalized extreme value distributions

Felipe S. Quintino¹, Melquisadec Oliveira¹, Pushpa N. Rathie¹, Luan C. S. M. Ozelim^{2,*} and Tiago A. da Fonseca³

¹ Department of Statistics, University of Brasília, Brasília, DF 70.910-900, Brazil

² Department of Civil and Environmental Engineering, University of Brasília, Brasília, DF 70.910-900, Brazil

³ Gama Engineering College, University of Brasília, Brasília, DF 72.444-240, Brazil

* **Correspondence:** Email: luanoz@gmail.com.

Abstract: Analyzing the statistical behavior of the assets' returns has shown to be an interesting approach to perform asset selection. In this work, we explore a stress-strength reliability approach to perform asset selection based on probabilities of the type $P(X < Y)$ when both X and Y follow a generalized extreme value (GEV) distribution with three parameters. At first, we derive new analytical and closed form relations in terms of the extreme value \mathbb{H} -function, which have been obtained under fewer parameter restrictions compared to similar results in the literature. To show the performance of our results, we include a Monte-Carlo simulation study and we investigate the application of the reliability measure $P(X < Y)$ in selecting financial assets with returns characterized by the distributions X and Y . Therefore, rather than the conventional approach of comparing the expected values of X and Y based on modern portfolio theory, we delve into the metric $P(X < Y)$ as an alternative parameter for assessing better returns.

Keywords: stress-strength reliability; \mathbb{H} -functions; generalized extreme value distribution

Mathematics Subject Classification: 91-10, 62N05, 33C60, 33F05

1. Introduction

The work of Markowitz [1] stands as one of the pioneering works on portfolio theory [2] and details Markowitz's major contributions to the rising modern portfolio theory. Besides providing criterious description of Markowitz's model of portfolio choice [3], the work in [1] traced future research directions explored by other scientists like the suggestion to replace variance with semi-variance as

Apêndice B

Artigo: Oliveira et al. (2024)

Autores: Melquisadec Oliveira, Felipe S. Quintino, Dióscoros Aguiar, Pushpa N. Rathie, Helton Saulo, Tiago A. da Fonseca e Luan C. S. M. Ozelim








Título: “On the stress-strength reliability of transmuted GEV random variables with applications to financial assets selection”

Revista: Entropy

Ano: 2024

Article

On the Stress–Strength Reliability of Transmuted GEV Random Variables with Applications to Financial Assets Selection

Melquisadec Oliveira ¹, Felipe S. Quintino ¹, Dióscoros Aguiar ², Pushpa N. Rathie ¹, Helton Saulo ^{1,*},
Tiago A. da Fonseca ³ and Luan Carlos de Sena Monteiro Ozelim ⁴

¹ Department of Statistics, University of Brasília, Brasília 70910-900, DF, Brazil; melquisadec.oliveira@gmail.com (M.O.); felipes.quintino2@gmail.com (F.S.Q.); pushpanrathie@yahoo.com (P.N.R.)

² Instituto de Ciências Exatas e Tecnológicas, Universidade Federal de Jataí, Jataí 75801-615, GO, Brazil; dioscoros.junior@ufj.edu.br

³ Gama Engineering College, University of Brasília, Brasília 72444-240, DF, Brazil; fonsecafga@unb.br

⁴ Department of Civil and Environmental Engineering, University of Brasília, Brasília 70910-900, DF, Brazil; luanoz@gmail.com

* Correspondence: heltonsaulo@unb.br

Abstract: In reliability contexts, probabilities of the type $R = P(X < Y)$, where X and Y are random variables, have shown to be useful tools to compare the performance of these stochastic entities. By considering that both X and Y follow a transmuted generalized extreme-value (TGEV) distribution, new analytical relationships were derived for R in terms of special functions. The results hereby obtained are more flexible when compared to similar results found in the literature. To highlight the applicability and correctness of our results, we conducted a Monte-Carlo simulation study and investigated the use of the reliability measure $P(X < Y)$ to select among financial assets whose returns were characterized by the random variables X and Y . Our results highlight that R is an interesting alternative to modern portfolio theory, which usually relies on the contrast of involved random variables by a simple comparison of their means and standard deviations.

Keywords: stress–strength reliability; extreme-value \mathbb{H} -function; TGEV distribution; assets selection

MSC: 60E05; 62Exx; 62Fxx



Citation: Oliveira, M.; Quintino, F.S.; Aguiar, D.; Rathie, P.N.; Saulo, H.; Fonseca, T.A.d.; Ozelim, L.C.d.S.M. On the Stress–Strength Reliability of Transmuted GEV Random Variables with Applications to Financial Assets Selection. *Entropy* **2024**, *26*, 441. <https://doi.org/10.3390/e26060441>

Academic Editors: Yousri Slaoui and Solym Manou-Abi

Received: 4 April 2024

Revised: 13 May 2024

Accepted: 14 May 2024

Published: 23 May 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

Consider a component with a strength Y and subjected to a stress X . The component fails if the stress X exceeds the component strength Y ; otherwise, it works properly. For independent components, the stress–strength reliability (SSR) R , also referred to as stress–strength probability, is given by:

$$R = P(X < Y) = \int_{-\infty}^{+\infty} F_X(x) f_Y(x) dx, \quad (1)$$

where F_X and f_Y denote, respectively, the cumulative distribution function (CDF) of X and the probability density function (PDF) of Y .

Although R was initially applied in the context of engineering, the interest in such metric spreads to several areas, such as household financial fragility [1], stock market modeling [2], asset selection [3], among others. We refer the reader to [4] for further details on stress–strength models.

The choice of an appropriate distribution to model both X and Y directly influences the calculation and estimation of R . In Finance, we have strong evidence that asset returns are better modeled by either α -stable processes (heavy-tailed alternative to Brownian motion [5]) or by heavy-tailed time series models [6,7]. The Extreme-Value Theory (EVT) made

Referências Bibliográficas

- Abbas, K. e Tang, Y. (2014). “Objective Bayesian analysis of the Frechet stress–strength model”. *Statistics & Probability Letters* 84, pp. 169–175.
- Aryal, G. R. e Tsokos, C. P. (2009). “On the transmuted extreme value distribution with application”. *Nonlinear Analysis: Theory, Methods & Applications* 71.12, e1401–e1407.
- Bachelier, L. (1900). “Theorie de la Speculation, Doctor Thesis, Annales Scientifiques Ecole Normale Superieure III-17”. *The random character of stock market prices*.
- 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.
- Embrechts, P., Klüppelberg, C. e Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*. Vol. 33. Springer Science & Business Media.
- Engeland, K., Hisdal, H. e Frigessi, A. (2004). “Practical extreme value modelling of hydrological floods and droughts: a case study”. *Extremes* 7, pp. 5–30.
- Fisher, R. e Tippett, L. C. (1928). “Limiting forms of the frequency distribution of the largest or smallest member of a sample”. *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 24. 2. Cambridge University Press, pp. 180–190.
- Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Wiley series in probability and mathematical statistics. R.E. Krieger Publishing Company. ISBN: 9780898749571.
- Gencay, R. e Selçuk, F. (2004). “Extreme value theory and Value-at-Risk: Relative performance in emerging markets”. *International Journal of forecasting* 20.2, pp. 287–303.
- 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.

- Gnedenko, B. (1943). “Sur la distribution limite du terme maximum d’une serie aleatoire”. *Annals of mathematics*, pp. 423–453.
- 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.
- Gumbel, EJ (1958). “Statistics of Extremes”. *Columbia University Press*.
- Haan, L. e Ferreira, A. (2006). *Extreme value theory: an introduction*. Vol. 3. Springer.
- Hartmann, M., Moala, F. A e Mendonça, MA. (2011). “Estudo das precipitações máximas anuais em Presidente Prudente”. *Revista Brasileira de Meteorologia* 26, pp. 561–568.
- Hasan, H., Radi, NF. A. e Kassim, S. (2012). “Modeling of extreme temperature using generalized extreme value (GEV) distribution: A case study of Penang”. *Proceedings of the world congress on engineering*. Vol. 1, pp. 181–186.
- 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.
- Kolmogorov, A. N. (1933). “Sulla determinazione empirica di una legge di distribuzione”. *Giornale dell’Istituto Italiano degli Attuari* 4, pp. 83–91.
- 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.

- Lima, R.K., Quintino, F. S., Fonseca, T. A da, Ozelim, L. CSM., Rathie, P. N. e Saulo, H. (2024). “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.
- M., Benoit (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).
- M., Harry (1952). “Portfolio Selection”. *The Journal of Finance* 7.1, pp. 77–91. DOI: 10.1111/j.1540-6261.1952.tb01525.x. URL: <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- Markowitz, H.M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons. ISBN: 978-0-300-01372-6.
- Mathai, A., Saxena, R. e Haubold, H. (2010a). *The H-function: Theory and Applications*. Springer Science & Business Media.
- Mathai, A. M., Saxena, R. K. e Haubold, H. J. (2010b). *The H-function: theory and applications*. Springer Science & Business Media.
- Morrison, J. E e Smith, J. A. (2002). “Stochastic modeling of flood peaks using the generalized extreme value distribution”. *Water resources research* 38.12, pp. 41–1.
- Nadarajah, S. (2003a). “Reliability for extreme value distributions”. *Mathematical and computer modelling* 37.9-10, pp. 915–922.
- (2003b). “Reliability for extreme value distributions”. *Mathematical and Computer Modelling* 37.9-10, pp. 915–922.
- Nascimento, F., Bourguignon, M. e Leão, J. (2016). “Extended generalized extreme value distribution with applications in environmental data”. *Hacettepe Journal of Mathematics and Statistics* 45.6, pp. 1847–1864.
- Nojosa, R. e Rathie, P. N. (2020). “Stress–strength reliability models involving generalized gamma and Weibull distributions”. *International Journal of Quality & Reliability Management* 37.4, pp. 538–551.

- Oliveira, M., Quintino, F. S., Aguiar, D., Rathie, P. N., Saulo, H., Fonseca, T. A. da e Ozelim, L. CSM. (2024). “On stress-strength reliability for the transmuted GEV distribution with applications in asset selection”. *Entropy*.
- Otiniano, C., De Paiva, B. e Neto, D. (2019). “The transmuted GEV distribution: properties and application”. *Communications for Statistical Applications and Methods* 26.3, pp. 239–259.
- Pisarenko, VF., Sornette, A, Sornette, D. e Rodkin, MV. (2014). “Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory”. *Pure and Applied Geophysics* 171, pp. 1599–1624.
- Queiroz, M. M. F. de e Chaudhry, F. H. (2010). “Análise de eventos hidrológicos extremos, usando-se a distribuição GEV e momentos LH”. *Revista Brasileira de Engenharia Agrícola e Ambiental* 14.2, pp. 133–138.
- Quintino, F. S., Oliveira, M., Rathie, P. N., Ozelim, L. CSM. 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 Math* 9, pp. 2345–2368.
- 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/>.
- Rathie, P. N., Freitas, S., Nojosa, R., Mendes, A. e Silva, T. (2022). “Stress-Strength Reliability Models Involving H-Function Distributions”. *Journal of Ramanujan Society of Mathematics and Mathematical Sciences* 9, pp. 217–234.
- Rathie, P. N. e Ozelim, L. CSM. (2017). “Exact and approximate expressions for the reliability of stable Lévy random variables with applications to stock market modelling”. *Journal of Computational and Applied Mathematics* 321, pp. 314–322.
- Rathie, P. N., Ozelim, L. CSM., Quintino, S. F. e Fonseca, T. A. (2023). “On the Extreme Value H-Function”. *Stats* 6.3, pp. 802–811. ISSN: 2571-905X. DOI: 10.3390/stats6030051. URL: <https://www.mdpi.com/2571-905X/6/3/51>.
- Resnick, S. (2008). *Extreme values, regular variation, and point processes*. Vol. 4. Springer Science & Business Media.
- Rubinstein, M. (2002). “Markowitz’s “portfolio selection”: A fifty-year retrospective”. *The Journal of finance* 57.3, pp. 1041–1045.

- Saputri, P. D., Prastyo, D. D., Oktaviana, P. P., Azmi, U. e Siswono, G. O. (2023). “Rare Event Classification Based on Binary Generalized Extreme Value-Additive Models”. *2023 6th International Conference on Information and Communications Technology (ICOIACT)*. IEEE, pp. 269–274.
- Sharpe, W. F. (1963). “A simplified model for portfolio analysis”. *Management science* 9.2, pp. 277–293.
- Smirnov, N. V. (1939). “On the estimation of the discrepancy between empirical curves of distribution for two independent samples”. *Bulletin of Moscow University, Mathematics Series 2*, pp. 3–16.
- Taleb, NN. (2020). “Statistical Consequences of Fat Tails (Technical Incerto Collection)”. *Scribe Media: Austin, TX, USA*.
- Tolikas, K. (2008). “Value-at-risk and extreme value distributions for financial returns”. *The Journal of Risk* 10.3, pp. 31–77.
- Yoon, P., Kim, T. e Yoo, C. (2013). “Rainfall frequency analysis using a mixed GEV distribution: a case study for annual maximum rainfalls in South Korea”. *Stochastic Environmental Research and Risk Assessment* 27, pp. 1143–1153.