University of Brasília
Institute of Exact Sciences
Department of Statistics

Master's Dissertation

# A Study of Non-Parametric Entropy Estimators for Analyzing Financial Data

**by**

**Helena Santos Brandão**

Brasília, 27 March of 2024

# A Study of Non-Parametric Entropy Estimators for Analyzing Financial Data

**by**

**Helena Santos Brandão**

Dissertation submitted to the Department of Statistics at the University of Brasília, as part of the requirements required to obtain Master Degree in Statistics.

Advisor: Prof. Dr. Raul Yukihiro Matsushita

Brasília, 27 March of 2024

Dissertation submitted to the Graduate Program in Statistics of the Departament of Statistics at the University of Brasília, as part of the requirements required to obtain the Master Degree in Statistics.

Approved by:

Prof. Dr. Raul Yukihiro Matsushita
Advisor, PPGEST/UnB

Prof. Dr. Eraldo Sergio Barbosa Da Silva
PPGECO/UFSC

Prof. Dr. Antônio Eduardo Gomes
PPGEST/UnB

Prof. Dr. Eduardo Yoshio Nakano
PPGEST/UnB

*Don’t go through life, grow through life*.

(Eric Butterworth)

To my pets Atena and Monet.

# Acknowledgments

I want to thank my family - Jorge, Iolanda, Gabi, Marcos, and Otávio - especially my parents - Jorge and Iolanda - for all the emotional and academic support. I would also like to thank my friends and my boyfriend João for their all support; to my puppies Athena and Monet, who stayed close to me, always giving much love; to my aunt Neca for being so affectionate and kind to me; to my colleague Cristiano who became a true friend during my first work experience.

I would also like to thank all the professors who contributed to my learning process up to the present moment, mainly my advisor during graduation, Maria Teresa, who taught me so much during this period, and my advisor, Raul, who agreed to help me with this project.

Finally, I would like to thank everyone who has helped me in some way throughout my academic and personal journey up to the present moment.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

# Resumo Expandido

## Um estudo sobre estimação não paramétrica de entropia diferencial para a análise de dados financeiros


**Resumo**

Esta dissertação é composta por uma coletânea de quatro artigos que abordam a estimação da entropia diferencial com aplicações em dados financeiros. A obtenção de um estimador robusto e dotado de propriedades satisfatórias para tal medida mostra-se de suma importância para sua aplicação. Nesse contexto, os objetivos do presente estudo incluem realizar uma revisão abrangente dos estimadores de entropia diferencial não paramétricos e propor aprimoramentos na escolha e otimização de seu uso, visando encontrar um estimador mais adequado para dados financeiros, os quais frequentemente apresentam distribuições com caudas pesadas e mudanças de regimes.


**Introdução**

A entropia diferencial é uma medida de incerteza aplicada a distribuições contínuas, sendo uma extensão da entropia de Shannon do caso discreto para o contínuo. Sua definição é dada por $H(X) = -\mathbb{E}\ln f(X)$, onde $f$ representa a função de densidade de uma variável aleatória contínua $X$ (Cover e Thomas, 2006).

A estimação da entropia depende da disponiblidade de um estimador da função de densidade de probabilidade da distribuição populacional. Em geral, tal distribuição é desconhecida e, por isso, métodos não paramétricos para sua estimação encontram-se disponíveis na literatura (Beirlant et al., 1997).

Como os conceitos de variabilidade, risco e incerteza não são necessariamente sinônimos, a aplicação dessa medida tem ganhado espaço em áreas como finanças (Zhou, Cai e Tong, 2013; Ormos e Zibriczky, 2014). Tipicamente, dados financeiros são complexos, podendo ser gerados por distribuições de caudas pesadas ou por uma mistura delas. Por isso, é preciso estudar as propriedades estatísticas desses estimadores tradicionais nesse ambiente turbulento e, se for o caso, propor estimadores mais apropriados para esse tipo de dados.

Nesse cenário, o presente estudo apresenta uma breve revisão da literatura sobre a estimação não paramétrica da entropia diferencial - concentrando a atenção nos métodos mais conhecidos e de fácil implementação computacional, como o da distância do vizinho mais próximo, o espaçamento amostral e o estimador por função kernel. Em seguida, considerando suas boas propriedades estatísticas e sua possibilidade de escolha de uma função kernel que ainda não foi testada na literatura, o estudo sugere uma melhoria no estimador kernel da entropia diferencial. Experimentos computacionais foram implementados sob diversos cenários ao longo do desenvolvimento deste trabalho para a observação empírica das propriedades estatísticas dos estimadores — em especial, viés, variância e erro quadrático médio. Esses cenários contemplam diversas características das distribuições tais como cauda leve ou pesada, simétrica ou assimétrica, suporte truncado ou ilimitado, dentre outras. A aplicação é ilustrada com séries históricas de registros diários de taxas de câmbio em relação ao dólar norte-americano e de preços de ações negociadas em bolsas de valores.

**Metodologia**

Basicamente, os estimadores não paramétricos da entropia diferencial se diferenciam de acordo com o método para a obtenção de estimativas da função de densidade $f(x)$. Entre esses métodos destacam-se o da função kernel, do espaçamento amostral (histograma empírico), e o da distância de vizinhos mais próximos. Dependendo de algumas condições de regularidade da distribuição populacional, esses métodos apresentam boas propriedades estatísticas (Beirlant et al., 1997). O problema é que dados reais — em especial os da área de finanças — podem ser gerados por processos complexos, resultando em distribuições sem as condições de regularidade que garantem as boas propriedades estatísticas desses estimadores.

Os estudos de Monte Carlo, que consideram cenários de amostragens de populações com caudas leves ou pesadas, simétricas ou assimétricas, com suportes truncados ou ilimitados, permitem comparar empiricamente a qualidade dos estimadores da entropia diferencial comumente encontrados na literatura em termos do vício, do desvio padrão do estimador e do erro quadrático médio.

Este trabalho também propõe uma melhoria metodológica do estimador da entropia diferencial mediante utilização da função kernel de Pareto. Essa função apresenta a vantagem de acomodar tanto a situação de amostragens de populações de caudas leves como pesadas. Além disso, como consequência natural da escolha desse kernel, a estimativa do expoente de Pareto para o cálculo da entropia propicia também uma indicação sobre o provável tipo de cauda populacional: pesada, leve ou muito leve (delta de Dirac).

Há outros elementos da teoria de informação que foram investigados neste trabalho, como a entropia bivariada e a informação mútua, estando relatados em outras partes desta dissertação.

**Resultados**

No estudo inicial, os estimadores baseados em Distâncias do Vizinho Mais Próximo (NN), Espaçamentos de Amostra (SS) e no Estimador de Kernel usando a distribuição Gaussiana (KE) foram comparados por meio de simulações de amostras provenientes de diferentes distribuições - variando não apenas a família da distribuição, mas também os parâmetros e o tamanho da amostra. Cada um desses cenários foi replicado 100 vezes - o viés, o desvio padrão e o Erro Quadrático Médio (EQM) foram então computados. Os resultados desse experimento estão resumidos na Tabela 1.

**Tabela 1:** Resumo dos resultados indicando qual foi o melhor estimador em cada cenário.

| **Distribuição** | **Viés** | **Desvio Padrão** | **EQM** |
|---|---|---|---|
| **Normal** | NN | KE | KE |
| **Caudas pesadas** | NN | KE | SS |
| **Assimétricas** | NN | KE | SS |
| **Beta** | — | KE | — |

Com respeito ao aperfeçoamento do estimador da entropia diferencial com base na função kernel de Pareto, foram realizadas simulações de Monte Carlo para comparar as propriedades de estimadores de entropia baseados em kernel, empregando diferentes funções de kernel[1], e a escolha otimizada da largura de banda concentrou-se na estimativa de entropia, em vez da estimativa da função de densidade. Os métodos de vizinho mais próximo e espaçamento de amostra foram utilizados como referência. Um estudo para o caso bivariado também foi realizado e seus resultados foram análogos aos observados no caso univariado. A Tabela 2 resume os resultados obtidos por essa análise.

[1] Os estimadores Pareto, Gaussian, Laplace e Cauchy são todos baseados no estimador kernel utilizando suas respectivas distribuições como funções de kernel.

**Tabela 2:** Resumo indicando qual o melhor e o pior estimador para cada medida.

| Medida | Melhor estimador | Pior estimador |
|---|---|---|
| **EQM** | Pareto | NND and Gaussian/Laplace |
| **Variância** | Pareto/Cauchy | NND |
| **Viés** | NND | Cauchy and Gaussian/Laplace |

**Aplicações**

Para ilustrar possíveis aplicações dessas medidas, foram analisados os retornos diários de várias bases de dados financeiros reais. Essa análise incluiu as taxas de câmbio históricas[2] de vinte e três moedas de diferentes países - cujas informações são baseadas em dados coletados pelo Federal Reserve Bank of New York - e valores de máximo diários de ações de algumas empresas disponíveis no Yahoo Finance[3], com informações provenientes do Yahoo Finance.

A seguinte expressão foi usada para calcular os retornos diários para cada conjunto de dados:

$$r_t = \ln(X_t) - \ln(X_{t-1}), \quad t = 1, ..., n,$$

onde o índice $t$ indica o momento em que o retorno é calculado; e $X_t, X_{t-1}$ representam os valores históricos nos momentos $t$ e $t-1$, respectivamente.

As Figuras 1 e 2 representam, respectivamente, o desvio padrão em relação à exponencial da estimativa de entropia baseada em Kernel (usando a função kernel gaussiana) dos dados históricos de taxa de câmbio[4] e a evolução histórica de duas taxas de câmbio - real brasileiro

[2] O intervalo de tempo considerado para as taxas de câmbio históricas varia em cada caso de acordo com sua respectiva disponibilidade - dados estão disponíveis em `https://www.federalreserve.gov/releases/h10/hist/`.

[3] Os valores históricos das ações estão disponíveis em `https://finance.yahoo.com/` e contemplam um intervalo de cinco anos (de 21 de novembro de 2018 a 21 de novembro de 2023).

[4] Dados foram coletados em junho de 2023.

e franco suíço[5] - em relação ao dólar norte-americano (usando a função kernel de Pareto). A Figura 3 representa a correlação total (com a função kernel de Pareto) entre diferentes ações de empresas. Em todos os casos, os retornos diários foram considerados como a medida de interesse.

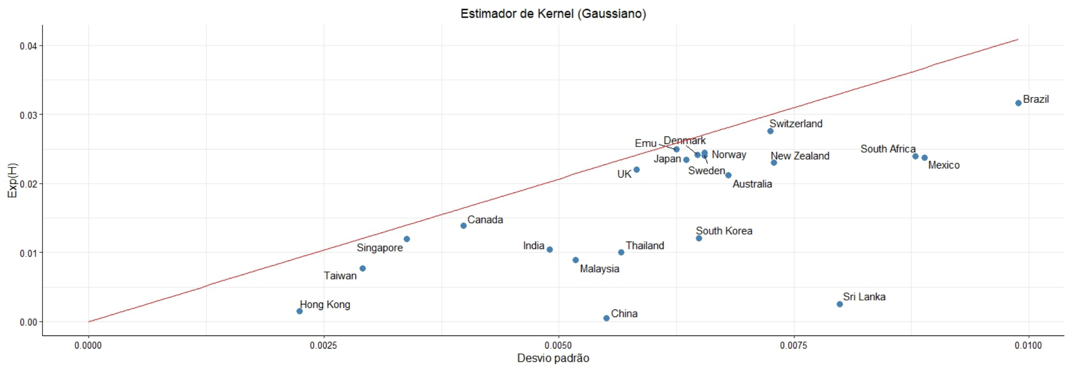


**Figura 1:** Exponencial da entropia contra o desvio padrão de algumas taxas de câmbio. A linha vermelha indica os valores correspondentes com distribuição normal.

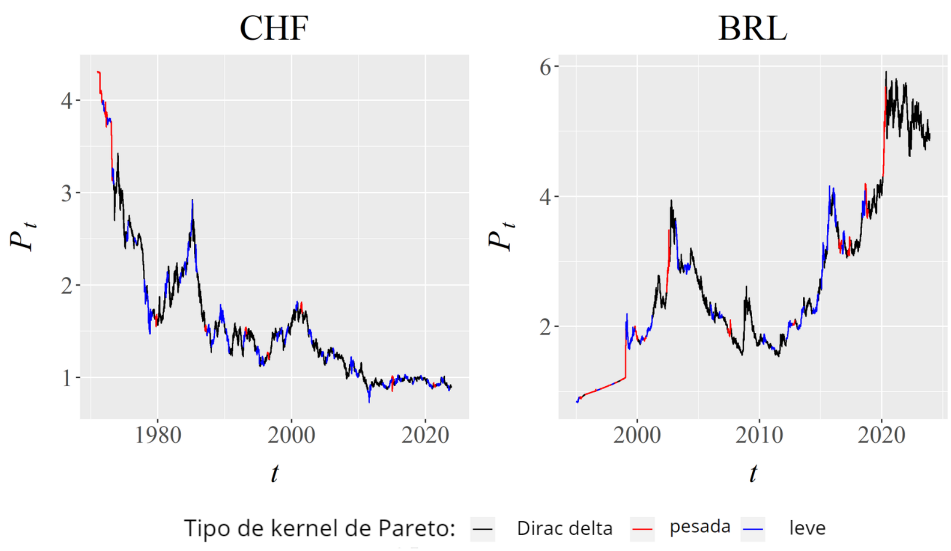


**Figura 2:** Evolução histórica de taxas de câmbio.

A partir da da Figura 1 nota-se que algumas localidades formam aglomerados, enquanto outras - tais como a moeda brasileira e a de Hong Kong - aparentam possuir comportamento

[5]Franco suíço (CHF): $n = 13.260$ observações de 5-Jan-1971 a 16-Nov-2023. Real brasileiro (BRL): $n = 7.260$ de 4-Jan-1995 a 13-Dec-2023.

divergente. Observa-se também que o desvio padrão e a entropia não fornecem necessariamente informações redundantes, destacando a importância do uso da entropia como medida de incerteza ou risco, especialmente quando a distribuição não possui segundo momento definido. A Figura 2 demonstra como a utilização do estimador kernel com núcleo baseado na distribuição de Pareto pode ser útil na detecção de diferentes regimes estocásticos na evolução histórica de uma taxa de câmbio como, por exemplo, a crise energética dos anos 1970 (CHF) e o Plano Real no início dos anos 1990 (BRL).

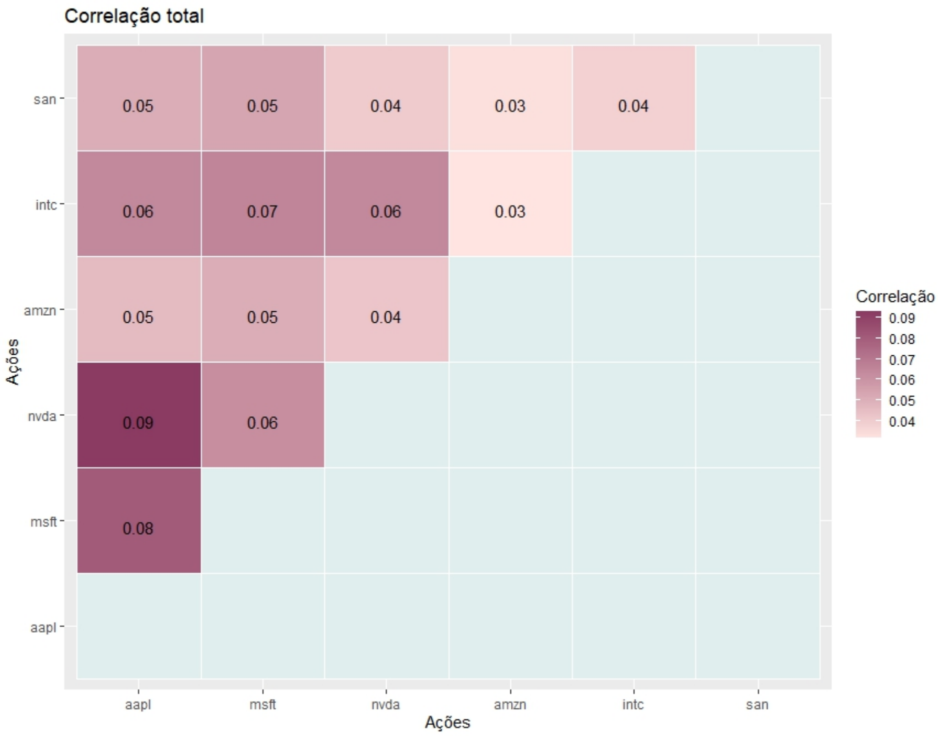


**Figura 3:** Correlação total de diferentes ações.

A análise da correlação total entre as empresas selecionadas, representada na Figura 3, fornece insights sobre suas inter-relações. A Apple Inc. demonstra uma maior dependência com a Microsoft Corporation e a NVIDIA Corporation, enquanto as interações envolvendo a Samsung Electronics e a Amazon.com exibem menores valores de correlação total.

**Conclusões**

A partir dos resultados obtidos, destaca-se a influência da distribuição dos dados - especialmente no que tange ao comportamento de sua cauda - para a estimação da entropia diferencial. A utilização de um estimador de kernel baseado no núcleo de Pareto mostrou-se como uma ferramenta robusta para enfrentar as complexidades inerentes a tais distribuições. Além disso, observou-se que diversas aplicações em contextos financeiros da medida de entropia podem ser implemetadas para fornecer informações relevantes acerca dos dados para complementar - ou até mesmo retificar - outras medidas de risco comumente utilizadas.

# Abstract

In financial risk, the conventional approach has typically linked risk to the variance of a variable, such as the return of a stock or portfolio. By recognizing the constraints of this conventional method and the need for various risk metrics, alternative measures have been developed to address downside risk or extreme outcomes specifically. One such complementary metric is the uncertainty measure, which enables us to capture and describe different aspects of risk, going beyond traditional notions of variability alone. Obtaining a robust estimator with desirable properties for entropy is crucial for its practical application. In particular, our study aims to conduct a comprehensive review of non-parametric differential entropy estimators and then propose adjustments regarding the choice and optimization of their use to find an estimator with convenient properties for application in financial data, which are often characterized by distributions with heavy tails. We also conducted real-data applications to illustrate the use of the proposed measures.

**Key-words**: differential entropy; nonparametric differential entropy estimator; kernel density estimator; heavy-tailed kernel; cluster analysis; stochastic regime.

# Contents

**1 Preface** **1**

1.1 Motivation . . . 1

1.2 Entropy . . . 2

1.3 Objectives . . . 3

**2 Non-parametric entropy estimation: an initial study** **4**

2.1 Introduction . . . 4

2.2 Method . . . 5

2.2.1 Computational Experiments . . . 8

2.2.2 Financial examples . . . 11

2.3 Data availability . . . 14

2.4 Conclusion . . . 15

**3 Pareto Kernel-based differential Entropy Estimation: overview** **16**

3.1 Introduction . . . 16

3.2 Method . . . 17

3.2.1 Differential entropy definitions . . . 17

3.2.2 Theoretical entropy for univariate and bivariate t-distribution . . . 18

3.2.3 Estimators for univariate random variables . . . 18

3.2.4 Estimators for bivariate random variables . . . 20

3.2.5 Estimators used in the study . . . . . . . . . . . . . . . . . . . . . . . 20
3.2.6 Mutual information . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
3.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
3.3.1 Univariate distributions . . . . . . . . . . . . . . . . . . . . . . . . . . 22
3.3.2 Bivariate distributions . . . . . . . . . . . . . . . . . . . . . . . . . . 25
3.4 Application . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
3.5 Data availability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
3.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

**4 Differential Entropy Estimation with a Paretian Kernel 30**

4.1 Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
4.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
4.3 Paretian kernel entropy estimator . . . . . . . . . . . . . . . . . . . . . . 32
4.4 A Monte Carlo Study . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
4.5 Illustration . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
4.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

**5 Final Considerations 45**

**References 46**

# List of Tables


1 Resumo dos resultados indicando qual foi o melhor estimador em cada cenário. x

2 Resumo indicando qual o melhor e o pior estimador para cada medida. . . . . . xi

2.1 Normal distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

2.2 Beta distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.3 Heavy-tailed distributions . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.4 Asymmetrical distributions . . . . . . . . . . . . . . . . . . . . . . . . 10

2.5 Summary indicating which was the best estimator in each scenario. . . . . . . . 10

2.6 Exchange rates currencies. . . . . . . . . . . . . . . . . . . . . . . . . 11

3.1 Comparison between different entropy estimators using multiple samples from Student t-distributions. . . . . . . . . . . . . . . . . . . . . . . . . . 22

3.2 Summary of each measure's best and worst estimator - univariate scenario. . . . 24

3.3 Comparison between different entropy estimators using multiple samples from bivariate Student t-distributions. . . . . . . . . . . . . . . . . . . . . 25

3.4 Abbreviation of each dataset. . . . . . . . . . . . . . . . . . . . . . . . 27

4.1 Examples of light-tailed and heavy-tailed kernels, where $u \in \mathbb{R}$. . . . . . . . . 34

4.2 Estimating differential entropy of an univariate Student's $t$ distribution with $\nu$ degrees of freedom: Mean squared error (MSE), variance ($\mathbb{V}[\hat{H}(\hat{f})]$), squared bias ($B^2$) from 500 Monte Carlo replications of different kernel entropy estimators, and NND and SS estimators as benchmarks. . . . . . . . . . . . . . . . . . . 42
4.3 Blocks with differential entropies estimated with heavy-tailed kernel . . . . . . 44

# List of Figures


1 Exponencial da entropia contra o desvio padrão de algumas taxas de câmbio. . . xii

2 Evolução histórica de taxas de câmbio. . . . . . . . . . . . . . . . . . . . . . . . xii

3 Correlação total de diferentes ações. . . . . . . . . . . . . . . . . . . . . . . . . xiii

2.1 Entropy and standard deviation of some countries' exchange rates. . . . . . . . 13

2.2 Entropy and standard deviation of some companies' stock values. . . . . . . . . 14

3.1 Comparison of estimators for each degree of freedom. . . . . . . . . . . . . . . 23

3.2 Comparison values for bivariate random variables. . . . . . . . . . . . . . . . . 26

3.3 Mutual information between different assets. . . . . . . . . . . . . . . . . . . . 28

3.4 Total correlation between different assets. . . . . . . . . . . . . . . . . . . . . . 29

4.1 Cross-validation loss function $CV(h)$ *versus* $h$ to estimate the differential entropy of a Cauchy distribution with scale parameter $\sigma = 3$ using light-tailed (Gaussian and Pareto) and heavy-tailed (Cauchy and Laplace) kernel functions from samples of size $n = 500$. The dashed line indicates the true population entropy, $H(f) = \ln(4\pi\sigma)$. Discontinuities occur when using a light-tailed kernel to estimate the entropy of this Cauchy distribution. . . . . . . . . . . . 36

4.2 **Left:** Mean squared errors (MSE) of kernel differential entropy estimates of a Student's $t$ distribution with $\nu$ degrees of freedom from 500 Monte Carlo replications. **Right:** Average values of $\alpha$ for different values of $\nu$ when using Pareto's kernel from the univariate Monte Carlo experiments (Table 4.2). . . . . . . . . . . . . . . . . . . . . . . . . . . 37

4.3 Historical evolution of foreign exchange rates against the US dollar. Swiss franc (CHF): $n = 13,260$ observations from 5-Jan-1971 to 16-Nov-2023. Brazilian real (BRL): $n = 7,260$ from 4-Jan-1995 to 13-Dec-2023. We find Pareto's kernel types suggesting different regimes generally coherent with known historical facts such as the 1970s energy crisis (CHF) and the Real Plan in the early 1990s (BRL). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

4.4 Log returns from data shown in Figure 4.3, with the Pareto's kernel types suggesting different stochastic regimes. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

4.5 $m$ estimated differential entropies *versus* sample variances calculated for each block of size $b = 60$, where $m = 221$ for CHF and $m = 121$ for BRL data. The red line refers to the Gaussian differential entropy given by $H(f) = 0.5 \ln(2\pi\sigma^2) + 0.5$ . . . . . . . . . . . . . . 41

4.6 $m$ estimated differential entropies *versus* estimated $\ln \alpha$ for each block of size $b = 60$, where $m = 221$ for CHF and $m = 121$ for BRL data. We find three clusters of $\alpha$ discriminated by dashed vertical lines: $\ln \alpha \leq \ln 2$ (heavy-tailed kernel), $\ln 2 < \ln \alpha \leq 5$ (light-tailed kernel) and we consider $\ln \alpha > 5$ as an approximation of the Dirac's delta kernel. . . . . . . . . . . 43

# Abbreviations and Acronyms

| | |
|---|---|
| NN | Estimator based on Nearest Neighbor Distances |
| SS | Estimator based on Nearest Neighbor Distances |
| KE | Estimator based on Kernel estimator using the Gaussian smoothing kernel distribution |

# List of Symbols and Notations

| | |
|---|---|
| $X$ | continuous random variable |
| $n$ | sample size |
| $f(x)$ | probability density function of $X$ |
| $H(X)$ | differential entropy of $X$ |
| $K(.)$ | Kernel function |
| $df$ | degree of freedom |

# Chapter 1

# Preface

## 1.1 Motivation

A financial portfolio is a compilation of investments where stakeholders grapple with decision-making challenges based on associated risks. Traditionally, financial risk has often been linked to the variability of certain variables, such as the return of a stock or portfolio.

Markowitz's mean-variance model, which assumes that asset returns follow a normal distribution, has been acknowledged as a pioneering model for portfolio selection. However, this assumption imposes limitations, as financial data can exhibit skewness, leptokurtosis, and fat tails. Indeed, alternative metrics like Value at Risk (VaR), Conditional Value at Risk (CVaR), and Drawdown have emerged to address these limitations, concentrating on downside risk or extreme outcomes.

Continuing to explore innovative methods is crucial for developing effective risk management strategies. An optimal risk management approach often involves leveraging diverse risk metrics. We emphasize adopting a dynamic and comprehensive approach not strictly tied to a specific distribution. Incorporating complementary measures, such as uncertainty, enhances the robustness of this strategy.

## 1.2 Entropy

Differential entropy measures the uncertainty or randomness associated with a continuous probability distribution. Higher values correspond to a greater unpredictability or surprise in the data. Differential entropy may assume negative values for certain distributions. Interpreting negative differential entropy can be less intuitive; however, it is commonly viewed as an indication of how much the distribution diverges from an idealized uniform distribution.

Cover and Thomas (2006) define the differential entropy $H(X)$ of a continuous random variable $X$ with density $f(x)$ as:

$$H(X) = -\,\mathbb{E}\ln f(X) = -\int_S f(x)\ln f(x)dx, \tag{1.1}$$

where $S$ is the support set of the random variable $X$.

In order to estimate its value, it is then necessary to estimate $f(x)$. Silverman (1986) emphasizes that the non-parametric approach requires fewer stringent assumptions about the distribution of observations, enabling the data to articulate its inherent patterns without undue constraints.

This study's primary objective is to estimate the differential entropy $H(X)$ through a non-parametric approach by estimating the density function of the variable of interest. Numerous studies have explored the application of entropy in portfolios and assets, with Zhou, Cai, and Tong (2013) paper serving as a valuable reference.

## 1.3 Objectives

Among different approaches, our study aims to offer a particular review of plug-in type estimators for $H(X)$. This class of estimators depends on a density estimator $\hat{f}(x)$, such as the kernel density estimator (KDE), the nearest neighbor estimator (NNE), and the sample-spacings estimator of order $m$ (SSEm). Specifically, the first objective is to investigate how the plug-in estimators behave for different population distributions: light-tailed or heavy-tailed, symmetrical or skewed, and truncated or not-truncated distributions. We also consider different non-parametric procedures, including Monte Carlo and Jackknife (leave-one-out point) methods. The second one refers to kernel bandwidth optimization since the best choices for density and entropy estimations may differ. Finally, the third objective is to suggest how our approach may be useful in detecting different stochastic regimes in the historical evolution of real financial data.

This dissertation is organized as a collection of articles. Thus, the financial data used to illustrate each analysis are described in their respective ones.

# Chapter 2

# Non-parametric entropy estimation: an initial study

## 2.1 Introduction

Numerous studies within the financial domain have integrated the concept of entropy into their analyses. For instance, Ormos and Zibriczky (2014) delves into entropy as a measure of financial risk. However, effectively applying this measure across various domains necessitates acquiring a robust entropy estimator with favorable properties. In this context, the present study aims to comprehensively review well-established non-parametric entropy estimators, particularly those based on nearest neighbor distance, sample spacing, and Kernel Estimator.

Initially, we conducted computational experiments to empirically observe various aspects of the estimators, mainly focusing on bias and mean squared errors. For this purpose, we simulated multiple samples from distributions such as Normal, Beta, and other heavy-tailed and asymmetrical distributions (see Section 2.2.1). Additionally, we implemented practical applications (Section 2.2.2) of these estimators using real financial data, which validated their performance and emphasized how estimation can provide valuable insights into the analyzed data.

## 2.2 Method

Let $X$ be a continuous random variable with density $f(x)$. Cover and Thomas (2006) define the differential entropy $h(X)$ as:

$$H(X) = -\mathbb{E}\ln f(X) = -\int_S f(x)\ln f(x)dx.$$

Beirlant et al. (1997) provides an overview of various methods for non-parametric differential entropy estimation of a continuous random variable. In this study, three non-parametric estimators will be considered, with the subsequent theory and properties based on their work:

- **Estimates of entropy based on nearest neighbor distances**

  Singh et al. (2003) introduced a non-parametric estimator of entropy which is based on the $k^{th}$ nearest neighbor distances between the $n$ sample points - which provide a competing estimator to that proposed by Kozachenko and Leonenko (1987).

  In the present study, we will consider $k = 1$ and its applicability in only one dimension. Let $\rho_{n,i} = \min_{j \neq i, j \leq n} ||X_i - X_j||$, then the nearest neighbor estimate is given by:

  $$H_n = \frac{1}{n}\sum_{i=1}^{n} \ln(n\rho_{n,i}) + \ln(2) + \gamma,$$

  where $\gamma$ is the Euler constant. Singh et al. (2003) also established the asymptotic unbiasedness and consistency of the proposed estimator.

- **Estimates of entropy based on sample-spacings**

  Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the corresponding order statistics of a i.i.d. real valued random variables. The $m$-spacing estimate for fixed $m$ is then given by:

  $$H_{m,n} = \frac{1}{n}\sum_{i=1}^{n-m} \ln\left(\frac{n}{m}(X_{(i+m)} - X_{(i)})\right) - \psi(m) + \ln(m),$$

where $\psi(m) = -(\ln\Gamma(m))'$ is the digamma function.

Beirlant et al. (1997) points out that the corresponding density estimate is not consistent, which implies that there is an additional term correcting the asymptotic bias in $H_n$. Beirlant and van Zuijlen (1985) proved the consistency of $H_n$ for uniform $f$. Under some conditions, Hall (1984) apud Beirlant et al. (1997) proved its weak consistency, and several authors studied and proved its asymptotic normality.

- **Estimates of entropy based on Kernel Estimator**

  Let $f_n$ be a kernel density estimator given by:

$$f_n = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

  where $n$ is the number of observations in the data base, $K$ is the Kernel function (which satisfies the condition $\int_{-\infty}^{\infty} K(x)d(x) = 1$) and $h$ is the smooth parameter.

  For a non-negative Kernel function that satisfies the previous condition, it will follow at once from the definition that $f_n$ will itself be a probability density and will inherit all the continuity and differentiable properties of $K$ (Silverman, 1986).

  Ahmad and Lin (1976) apud Beirlant et al. (1997) proposed an entropy estimator:

$$H_n = -\frac{1}{n}\sum_{i=1}^{n} \ln f_n(X_i).$$

  They also showed its mean square consistency. Joe (1989) studied the entropy estimation for multivariate probability density functions based on a kernel density estimate. Under some conditions, the authors verified asymptotic bias and variance terms and observed that non-unimodal kernels can reduce the estimator mean square error. Motivated by this work, Hall and Morton (1993b) introduced different entropy estimators from Joe's and described their properties. They showed that the same interaction of tail behavior, smoothness, and dimensionality also determines the convergence rate of Joe's estimator. They studied histogram and kernel estimators of entropy and, in each case, suggested empirical methods for choosing the smoothing parameter.

Those estimators based on Nearest Neighbor Distances (NN), Sample-Spacings (SS), and Kernel Estimator using the Gaussian smoothing kernel distribution (KE) were then compared by simulating some samples from different distributions - varying not only the family distribution but also the parameters and the sample size. We replicated each scenario 100 times and computed the mean, the standard deviation, and the Mean Square Error (MSE). The Maximum Likelihood estimator (ML) was used as a reference to analyze the quality of the proposed non-parametric estimators.

### 2.2.1 Computational Experiments

**Table 2.1:** Normal distribution

| Distribution | Sample size | Theorical value | Bias | | | | Standard deviation | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ML | NN | SS | KE | ML | NN | SS | KE | ML | NN | SS | KE |
| N(0,0.2) | 500 | 0.6142 | 0.0025 | 0.0002 | 0.0114 | 0.0138 | 0.0332 | 0.0753 | 0.0485 | 0.0334 | 0.0011 | 0.0056 | 0.0025 | 0.0013 |
| | 1000 | | 0.0019 | 0.0035 | 0.0094 | 0.009 | 0.0238 | 0.0575 | 0.036 | 0.0236 | 0.0006 | 0.0033 | 0.0014 | 0.0006 |
| | 2000 | | 0.0008 | 0.0035 | 0.0001 | 0.0034 | 0.0142 | 0.0379 | 0.0231 | 0.0142 | 0.0002 | 0.0014 | 0.0005 | 0.0002 |
| N(0,1) | 500 | 1.4189 | 0.0014 | 0.0048 | 0.0072 | 0.0098 | 0.0332 | 0.07 | 0.0442 | 0.0333 | 0.0011 | 0.0049 | 0.002 | 0.0012 |
| | 1000 | | 0.0008 | 0.0036 | 0.0069 | 0.0076 | 0.0205 | 0.0494 | 0.0326 | 0.0209 | 0.0004 | 0.0024 | 0.0011 | 0.0005 |
| | 2000 | | 0.0009 | 0.0007 | 0.0035 | 0.0049 | 0.0154 | 0.0347 | 0.0223 | 0.0153 | 0.0002 | 0.0012 | 0.0005 | 0.0003 |
| N(0,5) | 500 | 2.2237 | 0.0016 | 0.0006 | 0.0092 | 0.0093 | 0.0363 | 0.0698 | 0.0514 | 0.0368 | 0.0013 | 0.0048 | 0.0027 | 0.0014 |
| | 1000 | | 0.0006 | 0.0012 | 0.0063 | 0.0078 | 0.025 | 0.0582 | 0.037 | 0.0249 | 0.0006 | 0.0034 | 0.0014 | 0.0007 |
| | 2000 | | 0.0014 | 0.0019 | 0.0009 | 0.0026 | 0.0146 | 0.0351 | 0.0236 | 0.0146 | 0.0002 | 0.0012 | 0.0006 | 0.0002 |
| N(-2,0.5) | 500 | 1.0724 | 0.0037 | 0.0054 | 0.0124 | 0.0147 | 0.0333 | 0.0719 | 0.0462 | 0.0336 | 0.0011 | 0.0051 | 0.0023 | 0.0013 |
| | 1000 | | 0.0018 | 0.0016 | 0.0075 | 0.0086 | 0.0241 | 0.0519 | 0.0358 | 0.0245 | 0.0006 | 0.0027 | 0.0013 | 0.0007 |
| | 2000 | | 0.0006 | 0.0005 | 0.0035 | 0.0047 | 0.015 | 0.0389 | 0.0259 | 0.0148 | 0.0002 | 0.0015 | 0.0007 | 0.0002 |

Analyzing the values in Table 2.1 [1] - relative to the Normal distribution - it is evident that the NN estimator generally exhibits a more negligible bias than other estimators. On the other hand, the KE estimator demonstrates superior results in terms of standard deviation and MSE compared to other non-parametric estimators. In certain instances, the results surpass those the ML estimator obtained.

Table 2.2 [1] presents values for the Beta distribution. The KE estimator shows the most favorable measurements in terms of standard deviation and MSE. Regarding bias, none of the suggested estimators seems to exhibit significantly superior performance compared to the others.

From Table 2.3 [1] - which represents the results for some heavy-tailed distributions - it can be noticed that the Kernel estimator produces the smallest standard deviation values. The NN and SS estimators yield the best results concerning bias and MSE, respectively. Similar patterns are observed regarding some asymmetrical distributions (Table 2.4). Finally, Table 2.5 summarizes the findings discussed thus far.

[1] The colored cells indicate - for the related column (Bias, Standard Deviation, and MSE) - which of the proposed estimators presents the best results; when the text color is white, it indicates that the result of the estimator was even better than the observed using ML estimator.

**Table 2.2:** Beta distribution

| Distribution | Sample size | Theorical value | Bias ML | Bias NN | Bias SS | Bias KE | Standard deviation ML | Standard deviation NN | Standard deviation SS | Standard deviation KE | MSE ML | MSE NN | MSE SS | MSE KE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Beta(5,5) | 500 | -0.4806 | 0.0189 | 0.019 | 0.0226 | 0.0125 | 0.0268 | 0.0642 | 0.0401 | 0.0281 | 0.0011 | 0.0044 | 0.0021 | 0.0009 |
| | 1000 | | 0.0136 | 0.0093 | 0.0089 | 0.0051 | 0.0186 | 0.0461 | 0.0294 | 0.0193 | 0.0005 | 0.0022 | 0.0009 | 0.0004 |
| | 2000 | | 0.012 | 0.0008 | 0.0022 | 0.0019 | 0.0128 | 0.0345 | 0.0224 | 0.0131 | 0.0003 | 0.0012 | 0.0005 | 0.0002 |
| Beta(5,1) | 500 | -0.8094 | 0.0121 | 0.0122 | 0.0137 | 0.0699 | 0.0397 | 0.0828 | 0.0596 | 0.0399 | 0.0017 | 0.0069 | 0.0037 | 0.0065 |
| | 1000 | | 0.0094 | 0.0017 | 0.0036 | 0.0654 | 0.0265 | 0.0526 | 0.0373 | 0.0273 | 0.0008 | 0.0027 | 0.0014 | 0.005 |
| | 2000 | | 0.0086 | 0.0021 | 0.0038 | 0.0594 | 0.0163 | 0.0388 | 0.0243 | 0.0174 | 0.0003 | 0.0015 | 0.0006 | 0.0038 |
| Beta(2,2) | 500 | -0.1251 | 0.0227 | 0.0085 | 0.0002 | 0.0093 | 0.0157 | 0.0718 | 0.0437 | 0.0181 | 0.0008 | 0.0052 | 0.0019 | 0.0004 |
| | 1000 | | 0.0264 | 0.0008 | 0.0031 | 0.0037 | 0.0113 | 0.0487 | 0.0284 | 0.0129 | 0.0008 | 0.0023 | 0.0008 | 0.0002 |
| | 2000 | | 0.0249 | 0.0017 | 0.0011 | 0.0048 | 0.0081 | 0.0349 | 0.0208 | 0.009 | 0.0007 | 0.0012 | 0.0004 | 0.0001 |
| Beta(2,5) | 500 | -0.4845 | 0.0241 | 0.0011 | 0.0094 | 0.001 | 0.0284 | 0.0731 | 0.046 | 0.0299 | 0.0014 | 0.0053 | 0.0022 | 0.0009 |
| | 1000 | | 0.0176 | 0.0009 | 0.0032 | 0.0067 | 0.0188 | 0.0498 | 0.0322 | 0.0202 | 0.0007 | 0.0025 | 0.001 | 0.0004 |
| | 2000 | | 0.0195 | 0.0044 | 0.0049 | 0.0044 | 0.0134 | 0.0379 | 0.0238 | 0.014 | 0.0006 | 0.0014 | 0.0006 | 0.0002 |

**Table 2.3:** Heavy-tailed distributions

| Distribution | Sample size | Theorical value | Bias ML | Bias NN | Bias SS | Bias KE | Standard deviation ML | Standard deviation NN | Standard deviation SS | Standard deviation KE | MSE ML | MSE NN | MSE SS | MSE KE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| t(1) | 500 | 2.531 | 0.0054 | 0.0139 | 0.0272 | 0.1439 | 0.0937 | 0.102 | 0.0852 | 0.0694 | 0.0087 | 0.0105 | 0.0079 | 0.0255 |
| | 1000 | | 0.0056 | 0.0024 | 0.0113 | 0.1102 | 0.062 | 0.0774 | 0.0655 | 0.0544 | 0.0038 | 0.0059 | 0.0044 | 0.0151 |
| | 2000 | | 0.0052 | 0.0029 | 0.0044 | 0.0808 | 0.0452 | 0.0555 | 0.0458 | 0.0372 | 0.0021 | 0.0031 | 0.0021 | 0.0079 |
| t(4) | 500 | 1.6818 | 0.0016 | 0.0102 | 0.0048 | 0.0212 | 0.0491 | 0.079 | 0.0587 | 0.0444 | 0.0024 | 0.0063 | 0.0034 | 0.0024 |
| | 1000 | | 0.0024 | 0.0081 | 0.0107 | 0.0201 | 0.0331 | 0.0513 | 0.0397 | 0.0309 | 0.0011 | 0.0027 | 0.0017 | 0.0013 |
| | 2000 | | 0.0003 | 0.0017 | 0.0025 | 0.0105 | 0.0273 | 0.041 | 0.0303 | 0.0233 | 0.0007 | 0.0017 | 0.0009 | 0.0006 |
| t(8) | 500 | 1.5475 | 0.0019 | 0.0083 | 0.0079 | 0.0199 | 0.0404 | 0.0751 | 0.0517 | 0.0369 | 0.0016 | 0.0057 | 0.0027 | 0.0017 |
| | 1000 | | 0.002 | 0.0051 | 0.0087 | 0.0122 | 0.0279 | 0.0503 | 0.0343 | 0.0255 | 0.0008 | 0.0025 | 0.0012 | 0.0008 |
| | 2000 | | 0.0015 | 0.0025 | 0.005 | 0.0104 | 0.0217 | 0.039 | 0.0275 | 0.0212 | 0.0005 | 0.0015 | 0.0008 | 0.0006 |
| t(12) | 500 | 1.5039 | 0.0011 | 0.0068 | 0.0183 | 0.0236 | 0.0365 | 0.0817 | 0.0575 | 0.0367 | 0.0013 | 0.0067 | 0.0036 | 0.0019 |
| | 1000 | | 0.0056 | 0.0101 | 0.0007 | 0.0063 | 0.0276 | 0.0593 | 0.0389 | 0.0253 | 0.0008 | 0.0036 | 0.0015 | 0.0007 |
| | 2000 | | 0.0013 | 0.0065 | 0.0019 | 0.0036 | 0.0181 | 0.0376 | 0.0246 | 0.0178 | 0.0003 | 0.0014 | 0.0006 | 0.0003 |
| C(0,0.5) | 500 | 1.8379 | 0.0075 | 0.0108 | 0.0066 | 0.1293 | 0.0761 | 0.1105 | 0.0925 | 0.0755 | 0.0058 | 0.0122 | 0.0085 | 0.0224 |
| | 1000 | | 0.0028 | 0.0055 | 0.0077 | 0.1078 | 0.0523 | 0.0649 | 0.0544 | 0.0471 | 0.0027 | 0.0042 | 0.003 | 0.0138 |
| | 2000 | | 0.0003 | 0.0034 | 0.0027 | 0.0797 | 0.0357 | 0.0504 | 0.0432 | 0.0374 | 0.0013 | 0.0025 | 0.0019 | 0.0077 |
| C(0,1) | 500 | 2.531 | 0.0023 | 0.0072 | 0.0131 | 0.1394 | 0.0701 | 0.0969 | 0.0822 | 0.0726 | 0.0049 | 0.0093 | 0.0069 | 0.0246 |
| | 1000 | | 0.0017 | 0.0068 | 0.0041 | 0.1013 | 0.048 | 0.0769 | 0.0668 | 0.0563 | 0.0023 | 0.0059 | 0.0044 | 0.0134 |
| | 2000 | | 0.0048 | 0.0066 | 0.0116 | 0.0879 | 0.0354 | 0.0556 | 0.0477 | 0.0403 | 0.0013 | 0.0031 | 0.0024 | 0.0093 |
| C(0,2) | 500 | 3.2242 | 0.0078 | 0.0011 | 0.0213 | 0.1489 | 0.0672 | 0.1145 | 0.0957 | 0.075 | 0.0045 | 0.013 | 0.0095 | 0.0277 |
| | 1000 | | 0.0067 | 0.0106 | 0.0138 | 0.1084 | 0.0544 | 0.0693 | 0.0597 | 0.053 | 0.003 | 0.0049 | 0.0037 | 0.0145 |
| | 2000 | | 0.0011 | 0.0069 | 0.0121 | 0.0854 | 0.034 | 0.0523 | 0.0437 | 0.0364 | 0.0011 | 0.0028 | 0.002 | 0.0086 |
| C(-2,1) | 500 | 2.531 | 0.8034 | 0.0101 | 0.0256 | 0.1504 | 0.0299 | 0.1037 | 0.0909 | 0.0764 | 0.6464 | 0.0108 | 0.0088 | 0.0284 |
| | 1000 | | 0.8048 | 0.0163 | 0.0226 | 0.1175 | 0.0194 | 0.0735 | 0.0635 | 0.0571 | 0.648 | 0.0056 | 0.0045 | 0.017 |
| | 2000 | | 0.8047 | 0.0017 | 0.0078 | 0.0851 | 0.0166 | 0.0505 | 0.0427 | 0.0348 | 0.6478 | 0.0025 | 0.0019 | 0.0084 |

**Table 2.4:** Asymmetrical distributions

| Distribution | Sample | Theorical | Bias | | | | Standard deviation | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | size | value | ML | NN | SS | KE | ML | NN | SS | KE | ML | NN | SS | KE |
| Exp(0.5) | 500 | 1.6931 | 0.0009 | 0.0097 | 0.0112 | 0.0847 | 0.0475 | 0.0841 | 0.0608 | 0.0504 | 0.0022 | 0.0071 | 0.0038 | 0.0097 |
| | 1000 | | 0.0042 | 0.0024 | 0.0029 | 0.0797 | 0.0321 | 0.0566 | 0.0407 | 0.034 | 0.001 | 0.0032 | 0.0016 | 0.0075 |
| | 2000 | | 0.002 | 0.002 | 0.0034 | 0.0665 | 0.0245 | 0.0356 | 0.0275 | 0.025 | 0.0006 | 0.0013 | 0.0008 | 0.005 |
| Exp(1) | 500 | 1 | 0.003 | <0.001 | 0.0095 | 0.0809 | 0.0467 | 0.0798 | 0.0621 | 0.0478 | 0.0022 | 0.0063 | 0.0039 | 0.0088 |
| | 1000 | | 0.0011 | 0.001 | 0.0047 | 0.074 | 0.034 | 0.053 | 0.039 | 0.0355 | 0.0011 | 0.0028 | 0.0015 | 0.0067 |
| | 2000 | | 0.0004 | 0.0053 | 0.0016 | 0.0695 | 0.0234 | 0.0398 | 0.029 | 0.0245 | 0.0005 | 0.0016 | 0.0008 | 0.0054 |
| Exp(1.5) | 500 | 0.5945 | 0.0003 | 0.0038 | 0.0038 | 0.0805 | 0.0494 | 0.0804 | 0.0588 | 0.0528 | 0.0024 | 0.0064 | 0.0034 | 0.0092 |
| | 1000 | | 0.0002 | 0.0045 | 0.005 | 0.0769 | 0.0346 | 0.0516 | 0.0388 | 0.0374 | 0.0012 | 0.0027 | 0.0015 | 0.0073 |
| | 2000 | | 0.0008 | 0.0035 | 0.0004 | 0.0689 | 0.0239 | 0.0395 | 0.0296 | 0.0246 | 0.0006 | 0.0016 | 0.0009 | 0.0053 |
| Exp(5) | 500 | -0.6094 | 0.004 | 0.0004 | 0.0044 | 0.0849 | 0.0528 | 0.0863 | 0.0614 | 0.0571 | 0.0028 | 0.0074 | 0.0037 | 0.0104 |
| | 1000 | | 0.0053 | 0.0093 | 0.0113 | 0.0707 | 0.0345 | 0.0593 | 0.044 | 0.0371 | 0.0012 | 0.0036 | 0.002 | 0.0064 |
| | 2000 | | 0.0007 | 0.002 | 0.003 | 0.0691 | 0.0236 | 0.0426 | 0.0321 | 0.0246 | 0.0006 | 0.0018 | 0.001 | 0.0054 |
| Gama(0.5,1) | 500 | 0.0906 | 0.0052 | 0.0097 | 0.0078 | 0.3604 | 0.0716 | 0.0949 | 0.0784 | 0.0655 | 0.0051 | 0.009 | 0.0061 | 0.1341 |
| | 1000 | | 0.0149 | 0.0014 | 0.0024 | 0.3312 | 0.0603 | 0.0699 | 0.0617 | 0.0465 | 0.0038 | 0.0048 | 0.0038 | 0.1119 |
| | 2000 | | 0.0083 | 0.0034 | 0.0025 | 0.3195 | 0.0392 | 0.0454 | 0.0395 | 0.0347 | 0.0016 | 0.0021 | 0.0016 | 0.1033 |
| Gama(3,0.5) | 500 | 2.5407 | 0.0047 | 0.0006 | 0.0093 | 0.0061 | 0.0348 | 0.0758 | 0.0512 | 0.0355 | 0.0012 | 0.0057 | 0.0027 | 0.0013 |
| | 1000 | | 0.0033 | 0.0043 | 0.0001 | 0.0037 | 0.0236 | 0.0583 | 0.0366 | 0.0241 | 0.0006 | 0.0034 | 0.0013 | 0.0006 |
| | 2000 | | 0.0003 | 0.0026 | 0.0016 | 0.0002 | 0.0164 | 0.0369 | 0.0234 | 0.0169 | 0.0003 | 0.0014 | 0.0005 | 0.0003 |
| Gama(9,0.5) | 500 | 3.1726 | 0.0036 | 0.0058 | 0.0137 | 0.0125 | 0.0321 | 0.0665 | 0.044 | 0.0321 | 0.001 | 0.0044 | 0.0021 | 0.0012 |
| | 1000 | | 0.0015 | 0.0025 | 0.0063 | 0.0071 | 0.0222 | 0.0489 | 0.0341 | 0.0223 | 0.0005 | 0.0024 | 0.0012 | 0.0005 |
| | 2000 | | 0.0034 | 0.0021 | 0.0055 | 0.007 | 0.0157 | 0.0369 | 0.0239 | 0.0157 | 0.0003 | 0.0014 | 0.0006 | 0.0003 |
| Gama(7.5,1) | 500 | 2.3804 | 0.0003 | 0.0005 | 0.0081 | 0.0088 | 0.0331 | 0.077 | 0.0534 | 0.0336 | 0.0011 | 0.0059 | 0.0029 | 0.0012 |
| | 1000 | | 0.0017 | 0.0027 | 0.0047 | 0.0064 | 0.0224 | 0.0492 | 0.0328 | 0.0225 | 0.0005 | 0.0024 | 0.0011 | 0.0005 |
| | 2000 | | 0.0015 | 0.0015 | 0.0041 | 0.0049 | 0.0155 | 0.0381 | 0.0253 | 0.0153 | 0.0002 | 0.0014 | 0.0006 | 0.0003 |

**Table 2.5:** Summary indicating which was the best estimator in each scenario.

| Distribution | Bias | Sd | MSE |
|---|---|---|---|
| **Normal** | NN | KE | KE |
| **Heavy-tailed** | NN | KE | SS |
| **Asymmetrical** | NN | KE | SS |
| **Beta** | - | KE | - |

### 2.2.2 Financial examples

Entropy can be applied to financial data analysis to produce insights and contribute to decision-making. It can be a helpful mechanism when used with other financial analysis techniques.

We analyzed the daily returns of some databases: historical exchange rates of twenty-three countries' currencies - which information is based on data collected by the Federal Reserve Bank of New York (Table 2.6[2] shows those countries currencies); and some historical stock values of some companies available in Yahoo Finance [3]; to observe how the proposed non-parametric estimators perform in a real financial data examples. The highest value observed for each day of those periods was considered.

**Table 2.6:** Exchange rates currencies.

| Country | Monetary unit | Country | Monetary unit |
|---|---|---|---|
| Australia | Dollar | New Zealand | Dollar |
| Brazil | Real | Norway | Krone |
| Canada | Dollar | Singapore | Dollar |
| China, P.R. | Yuan | South Africa | Rand |
| Denmark | Krone | South Korea | Won |
| EMU member countries | Euro | Sri Lanka | Rupee |
| Greece | Drachma | Sweden | Krona |
| Hong Kong | Dollar | Switzerland | Franc |
| India | Rupee | Taiwan | Dollar |
| Japan | Yen | Thailand | Baht |
| Malaysia | Ringgit | United Kingdom | Pound |
| Mexico | Peso | | |

[2] Venezuela data were not considered since its values were behaving very differently from other currencies.

[3] Adobe, Airbnb, Amazon, Apple, AstraZeneca, Chevron, Coca-Cola, Colgate - Palmolive, Disney, Eletrobrás, Intel, Johnson & Johnson, Mc Donalds, Microsoft, Netflix, Nike, Nvidea, Petróleo Brasileiro (Petrobras), Pfizer, SalesForce, Vale, Visa, and Walmart.

The time interval considered for the historical exchange rates varies from each case according to its respective availability; the Yahoo data has historical data from five years [4]. Besides comparing each estimator, the main point is to see if it is possible to identify some patterns or clusters across different assets using the entropy estimation that could not be observed using only the standard deviation measurement.

In this study, the following expression was chosen to compute the returns of each data:

$$r_t = \ln(X_t) - \ln(X_{t-1}), t = 1, ..., n,$$

where the index $t$ indicates the time where the return is calculated; and the $X_t, X_{t-1}$ represent the exchange rates in time $t$ and $t-1$ respectively.

Figures 2.1 and 2.2 illustrate the values of standard deviation and the exponential value of the entropy estimations [5] considering the daily returns for each data. The red line represents the corresponding values for a Normal distribution population.

For the historical exchange rates data, some locations form clusters (like Australia, Japan, Norway, New Zealand, Sweden, Emu, etc.). In contrast, others seem to behave very differently (for example, Brazil and Hong Kong). On the other hand, the historical company stock values behave more as expected (following the red line).

Figure 2.1 unveils additional scenarios where entropy estimations are similar while the standard deviation varies significantly and cases where the opposite occurs. This observation highlights the potential for misunderstandings in financial data analysis when relying solely on standard deviation computation – which can arise due to the lack of a well-defined second moment in certain distributions. In contrast, entropy estimation measures uncertainty or volatility in the data for most distributions, offering a more comprehensive perspective.

[4] Data was collected on 28/06/2023.

[5] In order to compute the SS estimator it was considered only the unique values and its spacings were multiplied by the number of occurrences of each element; an adjustment was also made to compute the NN estimator in order not to consider duplicated values when evaluating the minimum difference of a value concerning the rest of the data.

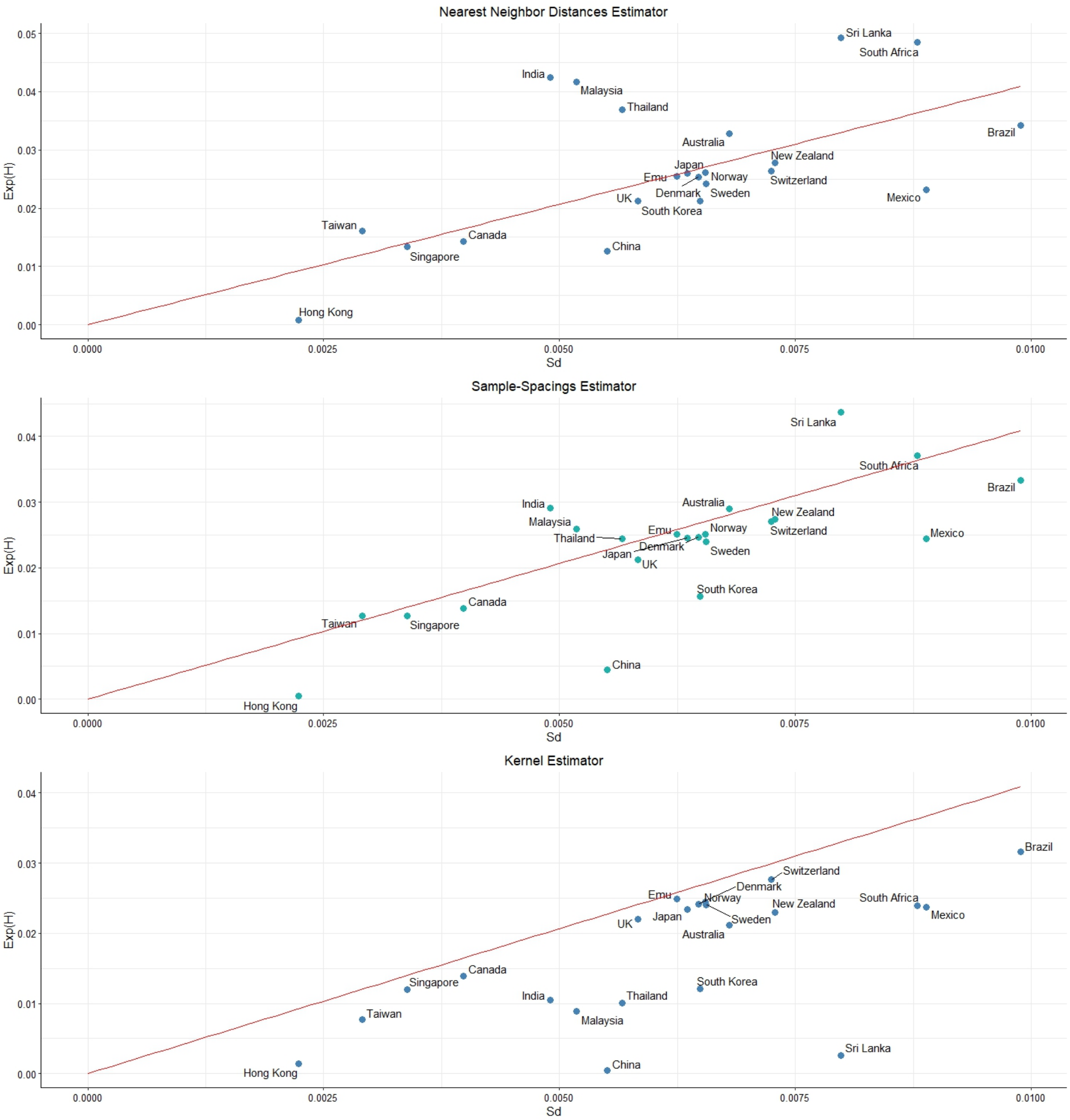


**Figure 2.1:** Entropy and standard deviation of some countries' exchange rates. The red line indicates the corresponding values with a normal distribution.

Figure 2.2 further illustrates that assets associated with various sectors (such as energy, health, technology, customer service, etc.) may exhibit similar behavior despite their diverse natures. This outcome is anticipated, considering the dataset comprises data from well-established and reputable companies in the market.

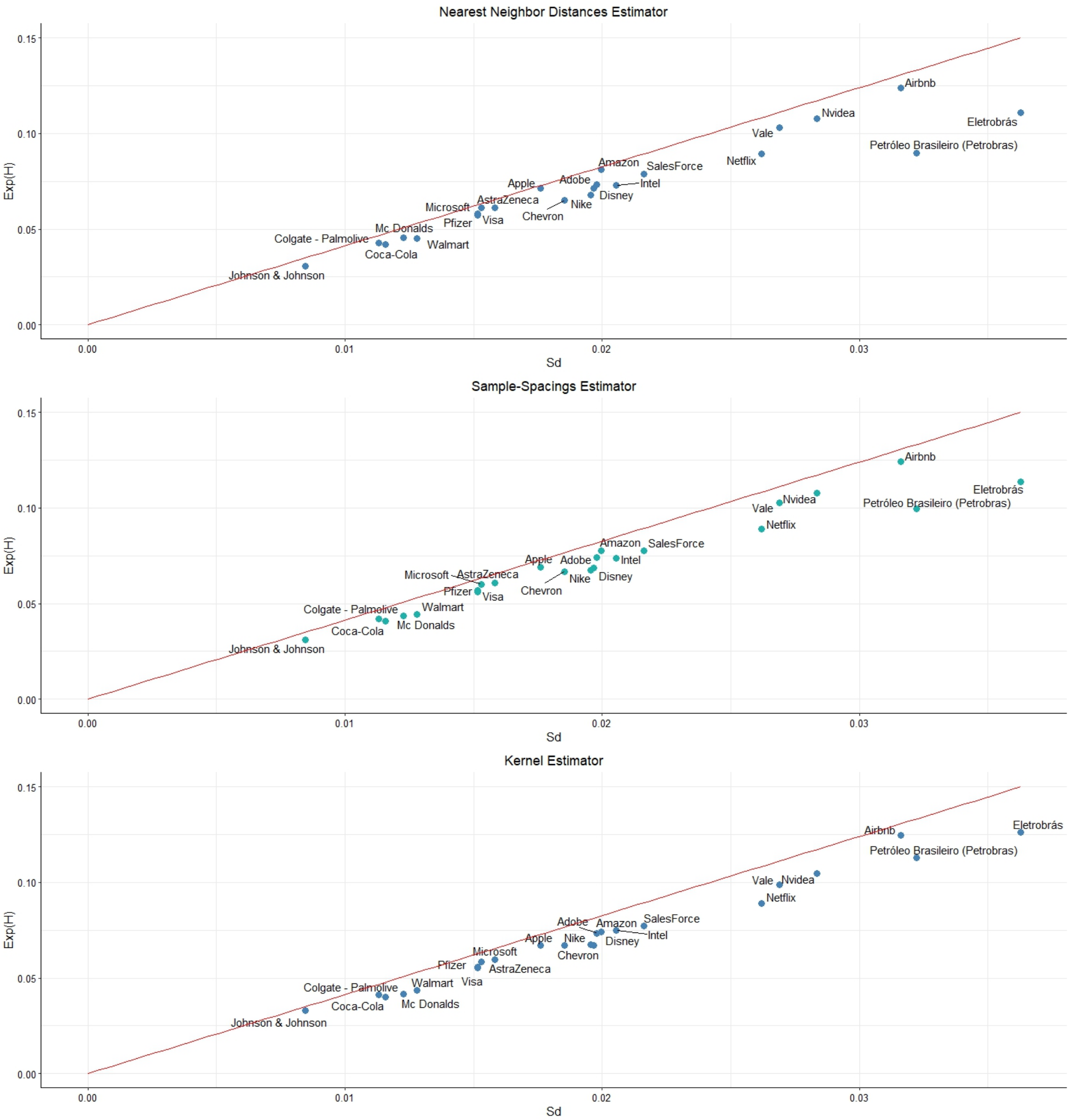


**Figure 2.2:** Entropy and standard deviation of some companies' stock values. The red line indicates the corresponding values with a normal distribution.

## 2.3 Data availability

Historical exchange rate data is available on `https://www.federalreserve.gov/releases/h10/hist/`.

Historical stock values are available on `https://finance.yahoo.com/`.

## 2.4 Conclusion

The Kernel estimator consistently yielded the smallest standard deviation values, and its Mean Square Error results appeared reasonable. Additionally, we observed that the Sample-Spacings (SS) and Nearest Neighbor (NN) estimators, designed for continuous random variables, encounter challenges when applied to real data with duplicated values (which can occur due to rounding issues).

Moreover, while the literature often emphasizes that the quality of a Kernel-based density estimate is mainly influenced by the smoothing parameter choice, with the choice of kernel playing a minor role (Scott, 1992), this study revealed a connection between the population distribution and the shape of the kernel function. Specifically, using a Gaussian Kernel yielded better results in Normal distribution experiments. In essence, the distribution's form can significantly impact the outcomes of non-parametric estimators. While these results are less than ideal, they underscore the need for adaptations in the Kernel estimator method to achieve improved outcomes that are not contingent on the population distribution.

Finally, the importance of entropy estimation's applicability becomes apparent through real data examples. Standard deviation sometimes fails to offer insightful information about data uncertainty or volatility. Therefore, entropy estimation proves valuable in revealing patterns across various assets.

# Chapter 3

# Pareto Kernel-based differential Entropy Estimation: overview

### 3.1 Introduction

Entropy has been proven to be an exciting tool for good portfolio diversification since its theoretical requirements are not so stringent, and it can be applied to a large number of distributions. Numerous studies have already explored this topic. For instance, Zhou, Cai, and Tong (2013) conducted a comprehensive review on entropy and its practical applications in the context of portfolios and asset pricing. Additionally, Ormos and Zibriczky (2014) delved into using entropy to measure financial risk.

There are already existing non-parametric entropy estimation methods for continuous random variables - some of which Beirlant et al. (1997) have made an overview. The Kernel-based entropy estimator is a promising option due to its favorable properties. Nevertheless, it faces a primary challenge related to the selection of bandwidth and kernel function. Typically, bandwidth is determined to optimize data density estimation, and the kernel function used is often a Gaussian distribution. This study seeks to introduce and empirically test a Kernel-based entropy estimator that does not depend on these predefined choices, potentially leading to an entropy estimator with superior qualities.

For this purpose, Monte Carlo simulations were performed to compare some properties of kernel-based entropy estimators employing different kernel functions. The optimization of bandwidth choice focused on entropy estimation rather than density function estimation. The nearest neighbor and sample-spacing methods were used as benchmarks. Finally, actual financial data was utilized to demonstrate the effectiveness of entropy as a metric for evaluating portfolios. This was achieved by calculating the mutual information between various assets. This measurement quantifies how much knowledge about the state of one asset reduces uncertainty regarding the others. The historical values were sourced from an online database, and the analysis considered a time horizon of five years.

## 3.2 Method

The study empirically compared the properties of several kernel-based entropy estimators utilizing different kernel functions - with the nearest neighbor and sample-spacing methods serving as benchmarks. For the kernel-based estimators, the bandwidth was selected to optimize the entropy estimate rather than the density function. Multiple Monte Carlo simulations were conducted for univariate and bivariate random variables, considering the Student's t-distribution with varying degrees of freedom. This included distributions with heavy tails, such as the Cauchy distribution ($df$ = 1) and distributions with infinite variance ($df$ = 2), as well as the normal distribution ($df$ = infinity). The theoretical entropy was also quantified. Subsequently, each estimator's bias, variance, and Mean Square Error (MSE) were computed.

### 3.2.1 Differential entropy definitions

Cover and Thomas (2006) provides the following definitions for entropy: let $X$ be a continuous random variable with density $f(x)$ and $X_1, ..., X_n$ a set of random variables with density $f(x_1, ..., x_n) = f(\mathbf{x})$, the differential entropy of $X$ and $X_1, ..., X_n$ is given respectively by:

$$H(X) = -\mathbb{E}\ln f(X) = -\int f(x)\ln f(x)dx,$$

$$H(X_1, ..., X_n) = -\int f(\mathbf{x})\ln f(\mathbf{x})d\mathbf{x}.$$

### 3.2.2 Theoretical entropy for univariate and bivariate t-distribution

**Student's t with $df$ degrees of freedom**

$$H(X) = \frac{df+1}{2}\left[\psi\left(\frac{df+1}{2}\right) - \psi\left(\frac{df}{2}\right)\right] + \ln\left[\sqrt{df}B(\frac{df}{2}, \frac{1}{2})\right],$$

where $\psi$ represents the digamma function and $B$ is the Beta function.

**Bivariate Student's t with independent marginals and equal degrees of freedom ($df$)**

$$H(X,Y) = 2H(X) = (df+1)\left[\psi\left(\frac{df+1}{2}\right) - \psi\left(\frac{df}{2}\right)\right] + 2\ln\left[\sqrt{df}B\left(\frac{df}{2}, \frac{1}{2}\right)\right],$$

where $\psi$ represents the digamma function and $B$ is the Beta function.

### 3.2.3 Estimators for univariate random variables

**Kernel-based entropy estimators**

Let $f_n$ be a kernel density estimator given by:

$$f_n = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where $n$ is the number of observations in the data base, $K$ is the Kernel function (which satisfies the condition $\int_{-\infty}^{\infty} K(x)d(x) = 1$) and $h$ is the smooth parameter.

Ahmad and Lin (1976) apud Beirlant et al. (1997) proposed an entropy estimator:

$$H_n = -\frac{1}{n}\sum_{i=1}^{n} \ln f_n(X_i).$$

**Sampling-spacings based entropy estimator**

Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the corresponding order statistics of a i.i.d. real valued random variables. The first-order estimate is given by:

$$H_n = \frac{1}{n}\sum_{i=1}^{n-1} \ln\left(n(X_{(i+1)} - X_{(i)})\right) - \psi(1),$$

where $\psi(m) = -(\ln\Gamma(m))'$ is the digamma function.

This can also be computed based on the empirical cumulative distribution function (this modified sample-spacings method allows ties):

$$H_n = \frac{1}{n}\sum_{i=1}^{n-1} \ln\left(\frac{n}{c_i}(X_{(i+1)} - X_{(i)})\right) - \psi\left(\frac{\sum_{j=1}^{n} c_j}{n}\right),$$

where $c_j$ is the corresponding counting of the observation $x_j$.

**Entropy Estimation Based on Nearest Neighbor Distances**

Singh et al. (2003) introduced a non-parametric entropy estimator that relies on the distances to the $k^{th}$ nearest neighbors among the $n$ sample points, offering an alternative to the method proposed by Kozachenko and Leonenko (1987). For this study, we consider $k = 1$ and restrict its application to one dimension.

The nearest neighbor then estimate is expressed as:

$$H_n = \frac{1}{n}\sum_{i=1}^{n} \ln(n\rho_{n,i}) + \ln(2) + \gamma,$$

where $\gamma$ denotes the Euler constant and $\rho_{n,i} = \min_{j\neq i, j\leq n} ||X_i - X_j||$.

### 3.2.4 Estimators for bivariate random variables

**Bivariate Kernel entropy estimation**

The extension for the bivariate case is given by:

$$\hat{f}(x,y) = \frac{1}{nh_xh_y}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_x}, \frac{y - Y_i}{h_y}\right),$$

where $k(u, v)$ is a bivariate density function, and $h_x$ and $h_y$ are the bandwidths for the marginal distributions of X and Y.

**Independent marginals**

When the marginals are independent, the joint differential entropy $H(X, Y)$ is given by:

$$H(X,Y) = H(X) + H(Y),$$

where $H(X)$ and $H(Y)$ are the differential entropies of the individual marginal distributions.

### 3.2.5 Estimators used in the study

Regarding those types of estimators, six different approaches for each case - univariate and bivariate case - were compared:

- **Gaussian**: differential entropy estimator based on the Gaussian kernel function;
- **Laplace**: differential entropy estimator based on the Laplacian kernel function;
- **Cauchy**: differential entropy estimator based on the Cauchy's kernel function;
- **Pareto**: differential entropy estimator based on the Pareto's kernel function;
- **NND**: differential entropy estimator based on the nearest neighbor distances;
- **SS**: differential entropy estimator based on the first-order sample-spacings.

### 3.2.6 Mutual information

The mutual information $I(X;Y)$ between two random variables $X$ and $Y$ with joint density $f(x,y)$ quantifies the reduction in uncertainty (entropy) about $X$ when $Y$ is known. In other words, it measures the amount of information that knowing the value of $Y$ provides about $X$. Cover and Thomas (2006) define its formula by:

$$\begin{aligned} I(X;Y) &= \int f(x,y) \ln \left( \frac{f(x,y)}{f(x)f(y)} \right) dxdy \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

#### Interpretation

The sign of mutual information reveals the nature of the association or dependence between $X$ and $Y$. Positive mutual information suggests a positive association, while negative mutual information suggests a negative association. A value of zero indicates independence between $X$ and $Y$.

#### Normalized mutual information

The normalization of the mutual information proves valuable when comparing relationship strengths across diverse variable pairs with varying individual entropies. Several normalization methods are available, and in this study, we considered the following:

$$\text{total correlation} := \tau(X,Y) = \frac{I(X,Y)}{\min(H(X), H(Y))};$$

This dimensionless metric ranges between 0 and 1. A value of 0 suggests negligible mutual information relative to the individual entropies. In contrast, a value of 1 indicates that the mutual information is maximally scaled concerning the minimum entropy of either $X$ or $Y$.

## 3.3 Results

### 3.3.1 Univariate distributions

The results of our experiment[1] for univariate distributions are presented in Table 3.1 and illustrated in Figure 3.1.

**Table 3.1:** Comparison between different entropy estimators using multiple samples from Student t-distributions.

| *df* | **Measure** | **Gaussian** | **Laplace** | **Cauchy** | **Pareto** | **NND** | **SS** |
|---|---|---|---|---|---|---|---|
| | MSE | 2.2172 | 0.7333 | 0.0104 | 0.01 | 0.0117 | 0.0094 |
| 1 | Var | 0.4316 | 0.2539 | 0.0082 | 0.0079 | 0.0117 | 0.0087 |
| | Bias | 1.3363 | 0.6924 | 0.0466 | 0.0452 | 0.0072 | 0.0269 |
| | MSE | 0.1453 | 0.0186 | 0.0038 | 0.0038 | 0.0072 | 0.0047 |
| 2 | Var | 0.0845 | 0.0111 | 0.0032 | 0.0033 | 0.0072 | 0.0044 |
| | Bias | 0.2466 | 0.087 | 0.0242 | 0.0224 | 0.003 | 0.0175 |
| | MSE | 0.0063 | 0.0027 | 0.0025 | 0.0021 | 0.0057 | 0.0032 |
| 4 | Var | 0.0048 | 0.0023 | 0.002 | 0.002 | 0.0057 | 0.0029 |
| | Bias | 0.0387 | 0.0203 | 0.0223 | 0.0115 | 0.0053 | 0.0159 |
| | MSE | 0.0014 | 0.0013 | 0.002 | 0.0013 | 0.0056 | 0.0026 |
| 15 | Var | 0.0012 | 0.0012 | 0.0012 | 0.0012 | 0.0056 | 0.0026 |
| | Bias | 0.0114 | 0.0109 | 0.0278 | 0.0103 | 0.0042 | 0.0078 |
| | MSE | 0.0012 | 0.0012 | 0.0019 | 0.0012 | 0.0053 | 0.0024 |
| 50 | Var | 0.0011 | 0.0011 | 0.0011 | 0.0011 | 0.0053 | 0.0024 |
| | Bias | 0.0095 | 0.0099 | 0.0286 | 0.0096 | 0.002 | 0.0083 |
| | MSE | 0.001 | 0.001 | 0.0018 | 0.001 | 0.0054 | 0.0024 |
| 120 | Var | 0.001 | 0.001 | 0.001 | 0.001 | 0.0054 | 0.0023 |
| | Bias | 0.0086 | 0.0092 | 0.0283 | 0.009 | 0.0004 | 0.0097 |

Examining the Mean Squared Error (MSE), the Pareto estimator consistently outperformed across all distribution scenarios. The NND estimator displayed elevated MSE values with larger

[1] In the univariate scenario, samples consisting of 500 observations were considered, and 500 Monte Carlo simulations were conducted.

degrees of freedom, while the Gaussian and Laplace estimators exhibited increased MSE for smaller degrees of freedom.

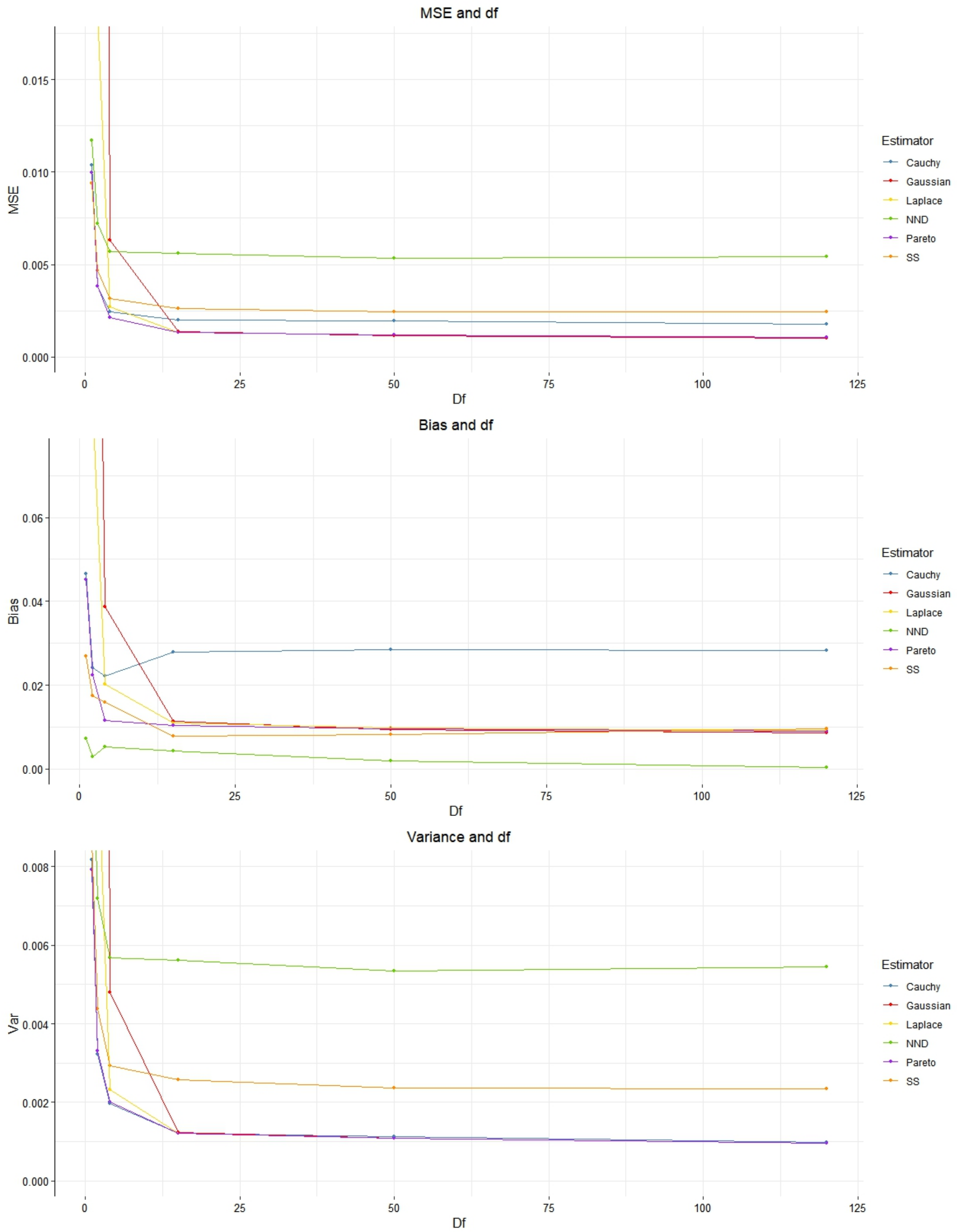


**Figure 3.1:** Comparison of estimators for each degree of freedom.

In terms of variance, the Pareto and Cauchy estimators consistently delivered favorable outcomes across all distributions, contrasting with the NND estimator's suboptimal performance in most scenarios.

Concerning bias, the NND estimator consistently demonstrated minimal values across all distributions. In contrast, the Cauchy estimator yielded less favorable results in most scenarios. The Pareto and SS estimators displayed intermediate performance, closely trailing the NND estimator.

Based on these considerations, it becomes evident that the Pareto estimator consistently yields satisfactory results regardless of the distribution. Table 3.2 summarizes these analyses.

**Table 3.2:** Summary of each measure's best and worst estimator - univariate scenario.

| Measure | Best Estimator | Worst Estimator |
|---|---|---|
| **MSE** | Pareto | NND (high $df$) and Gaussian/Laplace (small $df$) |
| **Var** | Pareto/Cauchy | NND |
| **Bias** | NND | Cauchy (high $df$) and Gaussian/Laplace (small $df$) |

### 3.3.2 Bivariate distributions

**Table 3.3:** Comparison between different entropy estimators using multiple samples from bivariate Student t-distributions.

| *df* | **Measure** | **Gaussian** | **Laplace** | **Cauchy** | **Pareto** | **NND** | **SS** |
|---|---|---|---|---|---|---|---|
| | MSE | 12.5928 | 2.6337 | 0.1002 | 0.098 | 0.1021 | 0.1084 |
| 1 | Var | 7.9102 | 1.2547 | 0.0712 | 0.0691 | 0.0935 | 0.0705 |
| | Bias | 2.1639 | 1.1743 | 0.1702 | 0.17 | 0.0927 | 0.1946 |
| | MSE | 0.3399 | 0.1353 | 0.0517 | 0.0461 | 0.0854 | 0.0569 |
| 2 | Var | 0.162 | 0.0656 | 0.0288 | 0.0308 | 0.0854 | 0.0472 |
| | Bias | 0.4217 | 0.264 | 0.1515 | 0.1234 | 0.0005 | 0.0983 |
| | MSE | 0.0694 | 0.0372 | 0.0465 | 0.0285 | 0.0486 | 0.0315 |
| 4 | Var | 0.0385 | 0.0235 | 0.0191 | 0.0202 | 0.0486 | 0.0254 |
| | Bias | 0.1756 | 0.117 | 0.1656 | 0.0915 | 0.0034 | 0.0776 |
| | MSE | 0.0174 | 0.0177 | 0.0455 | 0.0178 | 0.0468 | 0.0246 |
| 15 | Var | 0.0123 | 0.0119 | 0.0118 | 0.012 | 0.0458 | 0.0214 |
| | Bias | 0.0716 | 0.0766 | 0.1836 | 0.0759 | 0.0327 | 0.0563 |
| | MSE | 0.014 | 0.0154 | 0.0459 | 0.0158 | 0.0447 | 0.0239 |
| 50 | Var | 0.0093 | 0.0093 | 0.0096 | 0.0098 | 0.0447 | 0.0187 |
| | Bias | 0.0684 | 0.0781 | 0.1907 | 0.0776 | 0.0056 | 0.0723 |

The outcomes of our experiment [2] [3] for bivariate distributions are outlined in Table 3.3 and depicted in Figure 3.2. In this context, the results are analogous to those observed in the univariate scenario.

[2]In the bivariate scenario, samples comprised 100 observations, and 100 Monte Carlo simulations were performed. Compared to the univariate experiment, computational and time limitations necessitated reducing the number of simulations and sample sizes.

[3]For this experiment, for each $df$ it was considered independent and identically distributed marginals in order to make the samples simulations and to compute the entropy measures.

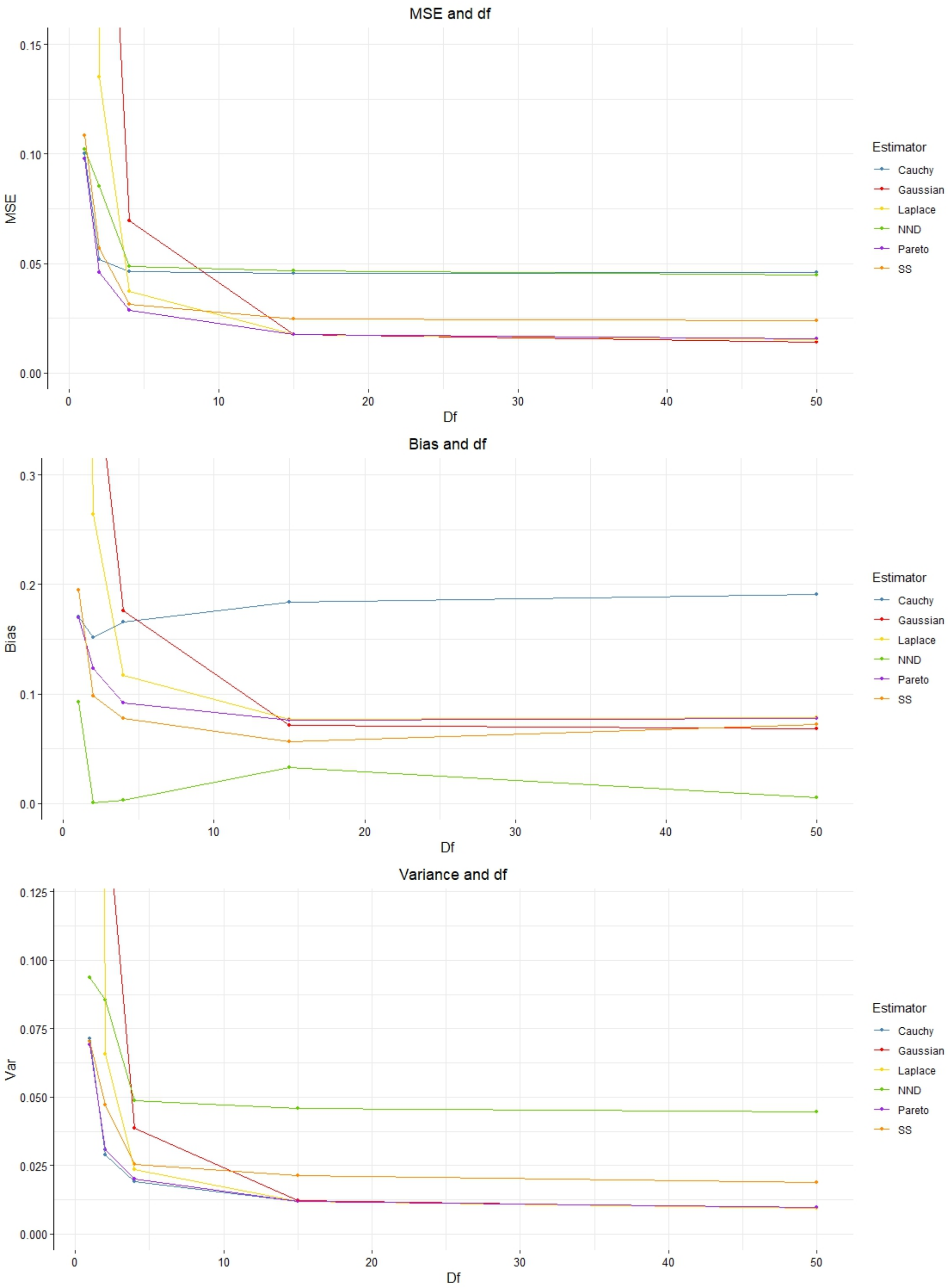


**Figure 3.2:** Comparison values for bivariate random variables.

## 3.4 Application

Mutual information in financial data analysis can offer valuable insights and enhance decision-making processes. When employed alongside other financial analysis techniques, it becomes a particularly beneficial tool for understanding complex relationships within the data.

The daily returns of selected databases were examined to analyze the practical application of this concept using real financial data. This analysis included historical stock values[4], with information sourced from Yahoo Finance (refer to Table 3.4 for the illustrated abbreviation used for each data). Each day's analysis incorporated the highest observed value within that specific timeframe.

**Table 3.4:** Abbreviation of each dataset.

| Data | Abbreviation |
|---|---|
| Apple Inc. | aapl |
| Amazon.com, Inc. | amzn |
| Intel Corporation | intc |
| NVIDIA Corporation | nvda |
| Microsoft Corporation | msft |
| Samsung Electronics | san |

The following expression was used to compute the daily returns for each dataset:

$$r_t = \ln(X_t) - \ln(X_{t-1}), t = 1, ..., n,$$

where the index $t$ indicates the time where the return is calculated; and the $X_t, X_{t-1}$ represent the historical values in time $t$ and $t-1$ respectively.

[4]Each historical data contains values from five years (Nov 21, 2018 - Nov 21, 2023).

Figures 3.3 and 3.4 represent the mutual information and the total correlation, respectively, computed across the daily returns from the datasets in Table 3.4.

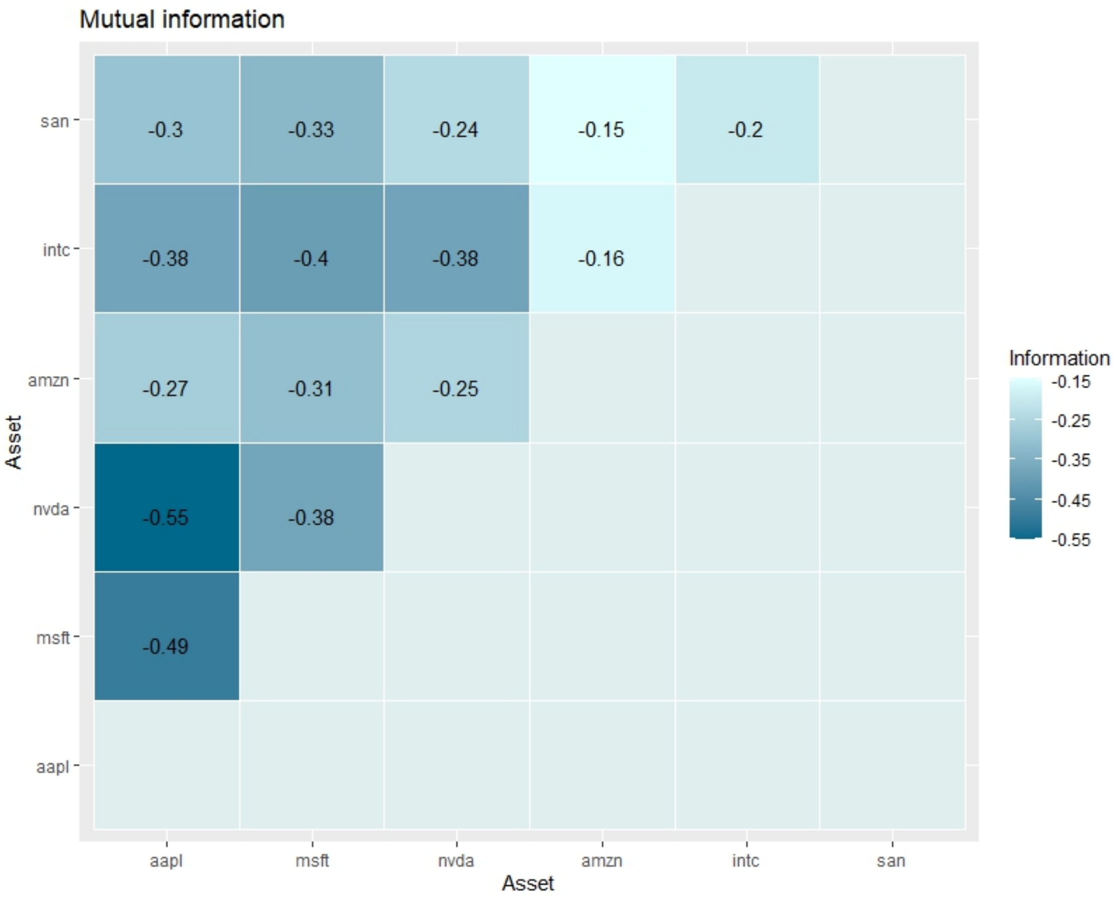


**Figure 3.3:** Mutual information between different assets.

Analyzing total correlation and mutual information among the selected companies provides insights into their interdependencies. Notably, Apple Inc. demonstrates relatively higher levels of dependence on information when combined with Microsoft Corporation and NVIDIA Corporation. While a total correlation of $0.09$ and $0.08$ might not be considered exceptionally high, it does indicate a noteworthy level of shared information. Conversely, interactions involving Samsung Electronics and Amazon.com show smaller values of total correlation, ranging from 0.03 to 0.05 - which may suggest a relatively lower degree of interdependence between those assets and the other companies in the dataset.

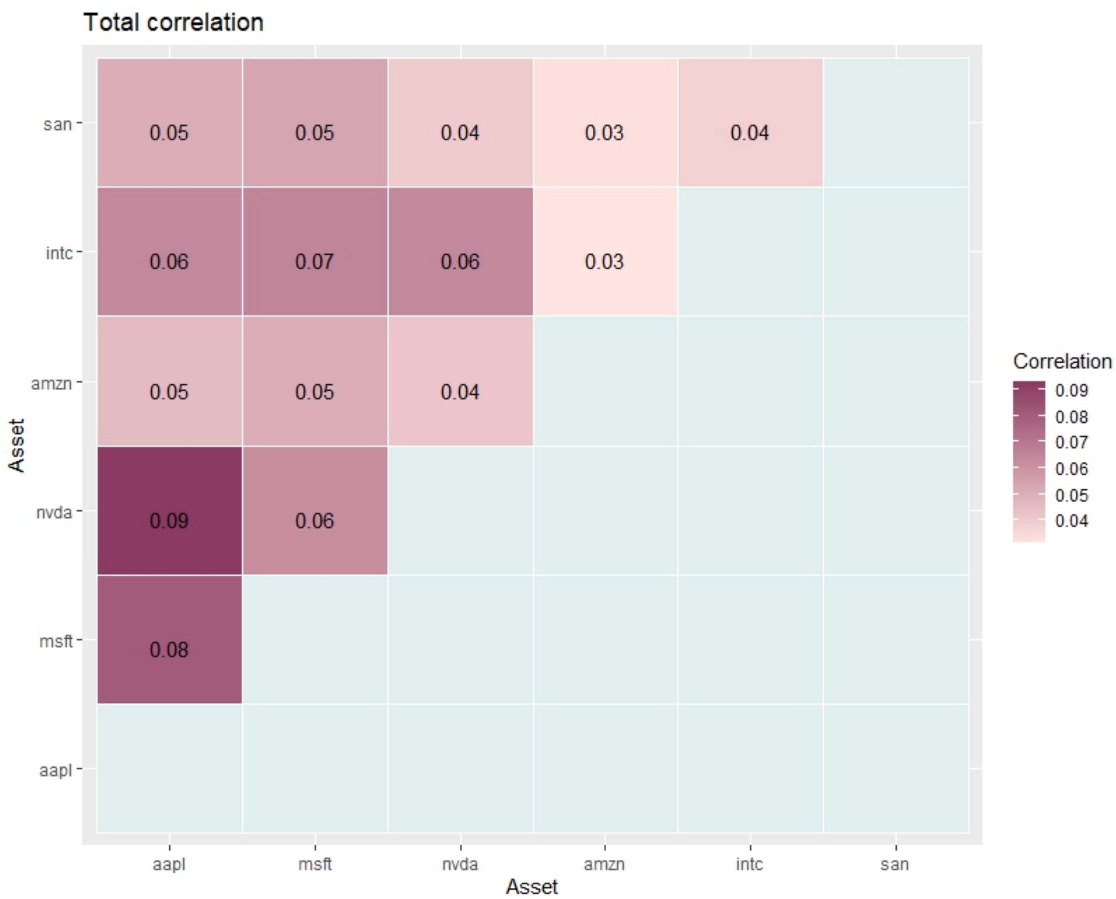


**Figure 3.4:** Total correlation between different assets.

## 3.5 Data availability

Historical exchange rate data and historical stock values are available on `https://finance.yahoo.com/`.

## 3.6 Conclusion

The findings indicate that, among the non-parametric estimators explored in the study, the Kernel estimator utilizing the Pareto Kernel function, optimized for the smoothing parameter with a specific emphasis on entropy estimation, exhibited superior overall properties.

# Chapter 4

# Differential Entropy Estimation with a Paretian Kernel

## 4.1 Abstract

Differential entropy extends the concept of entropy to continuous probability distributions, measuring the uncertainty associated with a continuous random variable. In financial data analysis, accurately estimating differential entropy is pivotal for understanding market dynamics and assessing risk. Traditional methods often fall short when dealing with the heavy-tailed distributions characteristic of financial returns. This paper introduces a novel approach to differential entropy estimation employing a Paretian kernel function adept at handling tail heaviness's intricacies. By incorporating an additional smoothing parameter, the Pareto exponent, our method offers flexibility in adjusting to light and heavy-tailed distributions. We compare our approach against established estimators through a comprehensive Monte Carlo simulation, demonstrating its superior performance in various scenarios. Applying our method to foreign exchange market data further illustrates its practical utility in identifying stochastic regimes and enhancing financial analysis. Our findings advocate for integrating the Paretian kernel estimator into the toolkit of financial analysts and researchers for a more nuanced understanding of market behavior.

## 4.2 Introduction

Differential entropy, a measure of uncertainty in continuous probability distributions in information theory, provides useful tools for analyzing financial data (Dionisio, Reis, and Coelho, 2008; Al-Yahyaee et al., 2019; Jaroonchokanan, Termsaithong, and Suwanna, 2022; Jizba and Korbel, 2014; Xu, Shang, and Zhang, 2019; Zhang et al., 2020; Smerlak, 2016; Xian, He, and Lai, 2016; Guo, Zhang, and Tian, n.d.; Jizba, Lavička, and Tabachová, 2021; Fernandes et al., 2023).

It is often estimated using statistical nonparametric methods when the underlying population distribution is unknown or complex (Beirlant et al., 1997; Chapeau-Blondeau and Rousseau, 2009; Govindan et al., 2007; Goel, Taneja, and Kumar, 2018).

Among the most widespread ones, the kernel density plugin-in estimator (KDE) of the differential entropy is an approach where the probability density function for calculating entropy is estimated by smoothing sample data points through a convenient kernel function. As an alternative method, the $k$-nearest neighbor (K-NN) technique allows the estimation of differential entropy by analyzing the distances between each data point and its nearest neighbors, inferring density from these proximities. Finally, partitioning-type methods involve dividing the data space and estimating probabilities in each segment to compute differential entropy. Generally, they are competing estimators under some smoothness, tail, and peak conditions Beirlant et al., 1997.

However, some issues arise when the underlying distributions are heavy-tailed. Regarding the KDE plugin-in estimator, choosing a suitable kernel function depends on the adverse influence of tail behavior of the underlying population distribution (Hall, 1987; Hall and Morton, 1993a). The K-NN method can be sensitive to the tail behavior of the distribution because the distances between a point and its $k$ nearest neighbors can be quite large, which may lead to higher bias. In partitioning methods, the spacing between samples within these partitions can affect the accuracy of the entropy estimation. However, consistent estimates can be produced from non-consistent spacing density estimates for large samples (Beirlant et al., 1997; Wachowiak et al., 2005).

As heavy-tailed distributions are ubiquitous in financial data (Mandelbrot, 2003; Matsushita et al., 2023), the improvement of the differential entropy estimation technique is of fundamental importance. It is known that a heavy tail kernel may counteract the adverse influence of tail behavior (Hall, 1987; Hall and Morton, 1993a). Therefore, this paper suggests a flexible KDE plugin-in estimator of the differential entropy using a Paretian kernel function, where Pareto's exponent accommodates tail heaviness. Thus, our approach allows us to deal with light and heavy tail situations.

We perform a Monte Carlo study to validate our results by considering the Student's $t$ distribution, which exhibits heavy or light tails depending on its parameter. We compare mean squared errors, biases, and variances of KDE, K-NN, and sample spacing estimators.

We illustrate our method by analyzing daily US dollar prices in Swiss franc (CHF) and Brazilian real (BRL) currencies collected by the Federal Reserve Bank of New York, identifying different stochastic regimes using differential entropy estimates with Pareto's kernel function.

The remainder of the paper is organized as follows. Our Paretian kernel entropy estimator is introduced in Section 4.3. A Monte Carlo study is carried out in Section 4.4, which aims to compare our method with KDE, K-NN, and sample spacing estimators as benchmarks. We illustrate our approach in Section 4.5 using daily data from foreign exchange markets. Finally, Section 4.6 concludes the study.

### 4.3 Paretian kernel entropy estimator

Letting $X$ be a continuous real-valued random variable governed by a density function $f$, its differential entropy is defined as

$$H(f) = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx = -\mathbb{E}[\ln f(X)], \tag{4.1}$$

where $f(x) \ln f(x) \equiv 0$ when $f(x) = 0$. However, in a practical situation where we estimate $f$ by $\hat{f}$, effectively we are dealing with the differential cross-entropy

$$H(\hat{f}) = -\int_{-\infty}^{\infty} f(x) \ln \hat{f}(x) dx = -\mathbb{E}[\ln \hat{f}(x)], \tag{4.2}$$

in which the expected uncertainty of $\hat{f}$ is obtained over the underlying distribution. The difference between $H(f)$ and $H(\hat{f})$ coincides with the Kullback–Leibler divergence from $\hat{f}$ to $f$,

$$D_{KL}(\hat{f}, f) = H(\hat{f}) - H(f) = \int_{\mathbb{R}} f(x) \ln \frac{f(x)}{\hat{f}(x)} dx. \tag{4.3}$$

Because $D_{KL}(\hat{f}, f)$ is a non-negative measure, the plug-in estimator $H(\hat{f})$ always tend to overestimate $H(f)$.

Now, given a random sample $X_1, \ldots, X_n$ drawn from a population $f$, the kernel estimator of $f$ is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{4.4}$$

where $h > 0$ is the smoothing bandwidth and $K$ is a Kernel function. Thus, we can consider the Kullback-Leibler loss as the empirical counterpart of (4.3), that is

$$\hat{D}_{KL}(\hat{f}_h, f) = \frac{1}{n} \sum_{j=1}^{n} \ln \frac{f(X_j)}{\hat{f}_h(X_j)}, \tag{4.5}$$

where the optimal bandwidth $\hat{h}$ can be found by minimizing (4.5) or, equivalently, minimize the empirical cross-entropy

$$\hat{H}(\hat{f}_h) = -\frac{1}{n} \sum_{j=1}^{n} \ln \hat{f}_h(X_j). \tag{4.6}$$

However, when $x = X_i$ in (4.4), as $K(0)$ does not depend on $h$, the bandwidth vanishes if we try to minimize (4.6). Because of this, we consider the cross-validation loss function

$$CV(h) = -\frac{1}{n}\sum_{j=1}^{n} \ln\left[\frac{1}{(n-1)h}\sum_{i\neq j} K\left(\frac{X_j - X_i}{h}\right)\right], \tag{4.7}$$

such that $\hat{h} = \arg\min_{h>0} CV(h)$ (Duin, 1976; Hall, 1987). By following this approach, the cross-validation estimate given by

$$\hat{H}(f_{\hat{h}}) = CV(\hat{h}) \tag{4.8}$$

is a consistent estimator of $H(f)$ depending on the tail properties of the kernel $K$ and the underlying density $f$ (Ivanov and Rozhkova, 1981; Hall and Morton, 1993a; Beirlant et al., 1997).

One of the contributions of this paper is to show empirically how the cross-validation loss function behaves depending on the tail properties of the kernel $K$ with a heavy-tailed population $f$. For this purpose, we consider the Gaussian and Laplace as light-tailed kernels, the Cauchy as a heavy-tailed kernel, and the Pareto as a switchable light/heavy-tailed kernel (Table 4.1).

**Table 4.1:** Examples of light-tailed and heavy-tailed kernels, where $u \in \mathbb{R}$.

| Distribution | Kernel | Tail |
|---|---|---|
| Gaussian | $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$ | light |
| Laplace | $K(u) = \frac{1}{2}\exp(-\lvert u\rvert)$ | light |
| Pareto | $K_\alpha(u) = \frac{\alpha}{2(1+\lvert u\rvert)^{\alpha-1}}$ | [r]heavy ($\alpha \leq 2$) |
| light ($2 < \alpha < \infty$) | | |
| Dirac delta ($\alpha = \infty$) | | |
| Cauchy | $K(u) = \frac{1}{\pi(1+u^2)}$ | heavy |

The Pareto's kernel is an interesting special case due to the exponent $\alpha$. Here, we take the Pareto's exponent as an extra smoothing parameter, defining the cross-validation loss as

$CV(h, \alpha)$ to get

$$(\hat{h}, \hat{\alpha}) = \underset{h>0, \alpha>0}{\arg\min}\, CV(h, \alpha).$$

This is a flexible and versatile kernel that can be either heavy-tailed ($\alpha \leq 2$) or light-tailed ($\alpha > 2$). As $\alpha \to \infty$, the Pareto's kernel approaches the Dirac delta function. Moreover, it accommodates two different degrees of heaviness: ($i$) $\int_{\mathbb{R}} uK(u)du = \infty$ for $\alpha \leq 1$; and ($ii$) $\int_{\mathbb{R}} uK(u)du < \infty$ and $\int_{\mathbb{R}} u^2K(u)du = \infty$, for $1 < \alpha \leq 2$. Thus, the tail heaviness of Pareto's kernel is left to be data-driven.

Figure 4.1 depicts how the $CV(h)$ function is influenced by the heaviness of the kernel function when estimating $H(f)$ of the Cauchy distribution. Discontinuities are observed when using Gaussian and Laplace's kernels, which are light-tailed. The opposite is observed with Cauchy and Pareto's ($\alpha = 1$) kernels. To confirm that heavy-tailed kernels help to deal with a heavy-tailed density $f$, we perform a Monte Carlo study in the next section.

## 4.4 A Monte Carlo Study

We performed 500 Monte Carlo replications to assess our suggested approach. In each one, we generated a sample of 500 deviations from a Student's $t$ distribution with degrees of freedom ($\nu$) ranging from 1 to 100, and we computed differential entropy estimates using the kernel functions from Table 4.1. We employ this distribution because it has heavier tails as $\nu$ decreases (Cauchy if $\nu = 1$) and the converse when $\nu$ increases (Gaussian if $\nu \to \infty$). When $\nu = 2$, we find a distribution with an undefined variance but a finite mean.

As a benchmark in this univariate study, we considered sample spacing (SS) and K-nearest Neighbor (KNN) estimates of the Student's t entropy $H(f)$, given by

$$H(f) = \frac{\nu+1}{2}\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] + \ln\left[\sqrt{\nu}B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right],$$

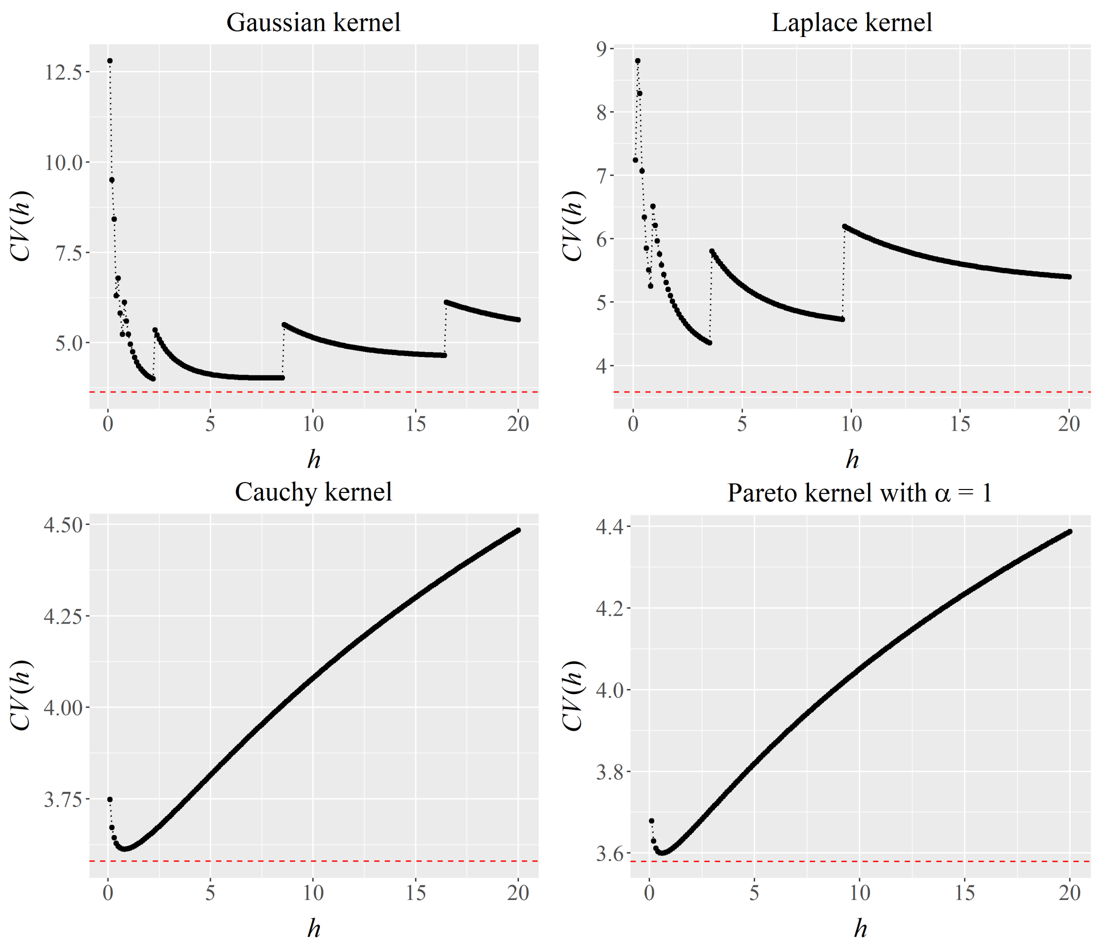


**Figure 4.1:** Cross-validation loss function $CV(h)$ *versus* $h$ to estimate the differential entropy of a Cauchy distribution with scale parameter $\sigma = 3$ using light-tailed (Gaussian and Pareto) and heavy-tailed (Cauchy and Laplace) kernel functions from samples of size $n = 500$. The dashed line indicates the true population entropy, $H(f) = \ln(4\pi\sigma)$. Discontinuities occur when using a light-tailed kernel to estimate the entropy of this Cauchy distribution.

where $\psi$ means the digamma function and $B$ is the beta function.

Table 4.2 shows the mean squared error, the variance, and the squared bias of the estimators from the Monte Carlo replications. Pareto's kernel generally offers the lowest mean squared errors among the competing estimators for all $\nu$, closely followed by the sample spacing and Cauchy's kernel entropy estimators. However, NND estimator performed poorly (Figure 4.2, left). Figure 4.2 (right) illustrates how the average $\bar{\alpha}$ evolves as a function of $\nu$. As expected, $\alpha$ increases as the distribution heaviness decreases ($\nu$ increases).

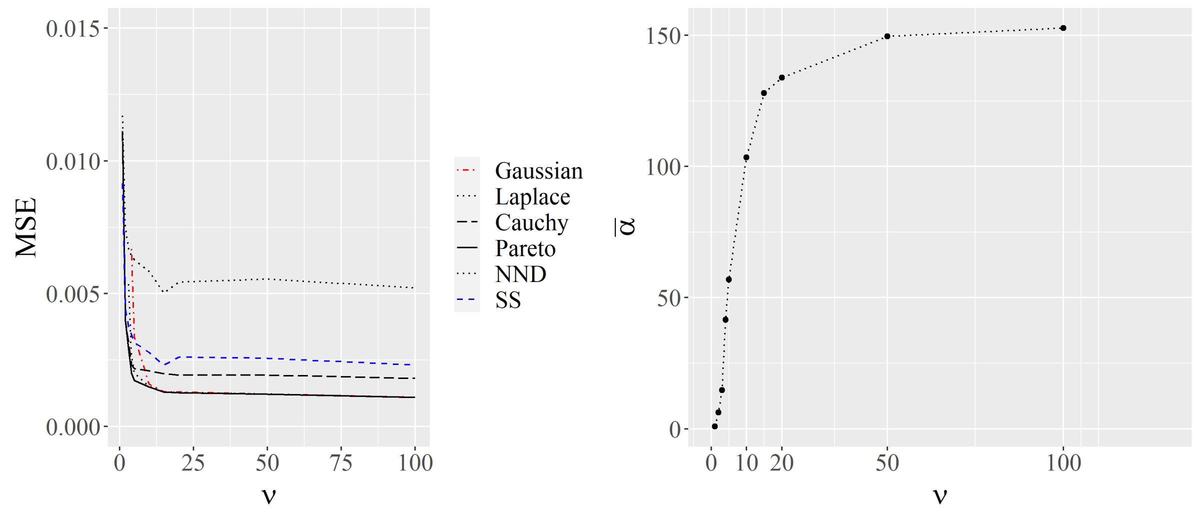


**Figure 4.2:** **Left:** Mean squared errors (MSE) of kernel differential entropy estimates of a Student's $t$ distribution with $\nu$ degrees of freedom from 500 Monte Carlo replications. **Right:** Average values of $\alpha$ for different values of $\nu$ when using Pareto's kernel from the univariate Monte Carlo experiments (Table 4.2).

## 4.5 Illustration

In our illustration, we aim to show how our approach may be useful in detecting different stochastic regimes in the historical evolution of an exchange rate. Consider the following as an example of a change in the exchange rate regime. In March 1973, several nations abandoned efforts to maintain the system of fixed exchange rates, beginning a period of transition from fixed to floating (market-determined) exchange rates between the US dollar and other major currencies.

We choose additional examples involving the Swiss Franc (CHF) and the Brazilian Real (BRL). Although the CHF is often referred to as a hard currency, on 15 January 2015, the Swiss National Bank suddenly announced that it would no longer hold the CHF at a fixed exchange rate against the euro. This resulted in a sharp drop in the rate from that day onwards (Matsushita et al., 2020).

In contrast,the BRL commonly experiences different regimes. The first began when Brazil implemented the Real Plan in 1993 as a monetary exchange rate-based stabilization policy, where the central element was an exchange rate anchor to reduce chronic inflation. The BRL currency was introduced in July 1994 under an exchange rate band strategy until 13 January 1999, when the first real currency crisis occurred. Just after that, the real-dollar rate was let to be market-determined. The BRL experienced a log-periodic bubble (28 May 2002 - 14 Jan 2003), possibly because of uncertainties in the Brazilian presidential election and the economic policy to be adopted by Lula, the elected president (Matsushita et al., 2006). Other factors occurred, such as the impeachment of President Dilma Roussef, the period of austerity during Michel Temer's short government, and the transition to Jair Bolsonaro's government, marking the country as a rollercoaster of economic and monetary policies.

Now, ignoring such historical facts, we apply our approach to detect them using only our data-driven approach with the following differential entropy estimation through the Pareto kernel. Given a time series of exchange rates $\{P_t\}$, we obtain the log-return as usual, $X_t = \ln P_t/P_{t-1}$, where $t$ is a time index varying from 2 to $n$, where $n$ denotes the sample size. In our illustration, we divided the time series of returns exactly into blocks of size $b$, so that $n = b \times m$. We deal with $m$ time series of size $b$. For each block, using Pareto's kernel, we calculate the differential entropy with the estimated $\alpha$, and we classify the block according to its magnitude. From Table 4.1. A heavy-tailed kernel occurs if $\ln \alpha \leq \ln 2$. Empirically, we find it reasonable to classify $\ln 2 < \ln \alpha \leq \ln 5$ as the light-tailed kernel case and $\ln \alpha > 5$ as an approximation of the Dirac's delta kernel.

The interpretation is straightforward. Price movements are random and unpredictable in perfectly efficient markets because they instantly incorporate all available information, leaving no room for predictable patterns or arbitrage opportunities. When $\alpha$ is high, the kernel function at a certain point $X_t$ depends almost nothing on the other sample points of the data block. We interpret it as economic efficiency because of a high degree of randomness in the data series in measuring entropy. At the opposite extreme, a small parameter ($\alpha \leq 2$ ) means that the kernel function at a certain point $X_t$ depends on almost all the other sample points of the block. We interpret it as a lack of economic efficiency, with a high degree of dependence on the entropy estimation. Thus, smaller $\alpha$'s are related to turbulent periods, while higher $\alpha$ means regular periods with smaller tails. Interestingly, higher/lower $\alpha$ does not imply more or less uncertainty or volatility because it is a kernel's parameter.

The data sets employed in this illustration were taken from the Federal Reserve website (`http://www.federalreserve.gov/releases/H10/hist/`). They refer to a currency value in US dollar terms collected by the Federal Reserve Bank of New York from a sample of market participants. We ignore gaps from weekends and holidays, concentrating our analysis on trading days.

We choose $b = 60$, which gives $m = 221$ and 121 for the CHF and BRL data, respectively. Figure 4.3 shows the BRL/USD and CHF/USD time series of exchange rates (CHF from 5-Jan-1971 to 16-Nov-2023 and BRL from 4-Jan-1995 to 13-Dec-2023). Interestingly, although the kernel types are obtained from blocks of log returns (Figure 4.4), we may find a general picture consistent with some known historical facts that affected price movements. Figure 4.5 depicts differential entropy estimates against sample variances from each data block. It shows that some entropies estimated with low $\alpha$ (heavy tail kernel) are further away from the Gaussian line. However, we also find blocks with small variances and entropies with small $\alpha$. It reinforces that tail heaviness, uncertainty, and volatility are not synonymous, which must be considered when measuring financial risks.

Table 4.3 lists the heavy-tailed blocks (13 CHF and 21 BRL blocks), which align with some facts described at the beginning of this section. Finally, Figure 4.6 shows differential entropy estimates against estimated values of $\ln \alpha$, revealing the existence of three kernel types. The heavy-tailed case is established theoretically because the Pareto distribution has no variance when $\ln \alpha \leq \ln 2$. Classifying $\ln \alpha > 5$ as an approximation of Dirac's delta kernel is reasonable because a cluster of points is formed after $\ln \alpha = 5$.

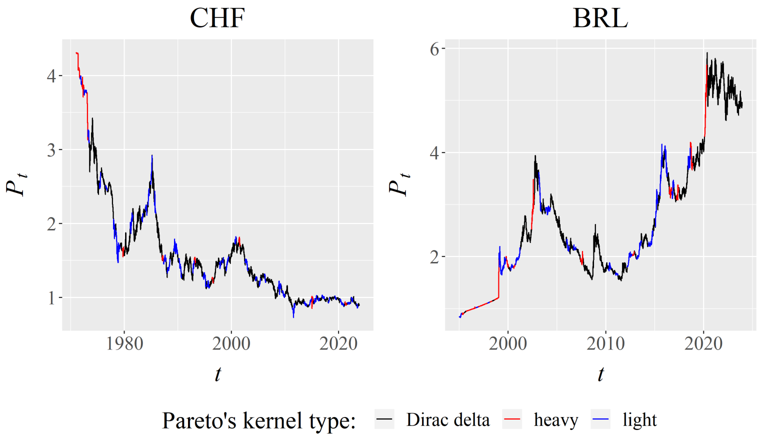


**Figure 4.3:** Historical evolution of foreign exchange rates against the US dollar. Swiss franc (CHF): $n = 13,260$ observations from 5-Jan-1971 to 16-Nov-2023. Brazilian real (BRL): $n = 7,260$ from 4-Jan-1995 to 13-Dec-2023. We find Pareto's kernel types suggesting different regimes generally coherent with known historical facts such as the 1970s energy crisis (CHF) and the Real Plan in the early 1990s (BRL).

## 4.6 Conclusion

This study underscores the crucial role of tail behavior in estimating differential entropy, especially within financial data analysis. Our findings advocate adopting a Paretian kernel-based estimator as a robust tool that accommodates the complexities inherent in financial distributions. This method enhances the accuracy of differential entropy estimation and provides a nuanced understanding of financial market dynamics through the lens of entropy. Future research could explore integrating this approach with other financial modeling techniques to further unravel the intricate behaviors of markets and enhance risk assessment methodologies. Emphasizing the adaptability and efficacy of the Paretian kernel, our research contributes to the advancement of statistical tools in financial analysis.

### Acknowledgement

Financial support from FAPDF (Grant number 00193-00001860/2023-17), CNPq (Grant numbers 301879/2022-2 and 311548/2022-9) and Capes (Finance Code 001) is acknowledged.

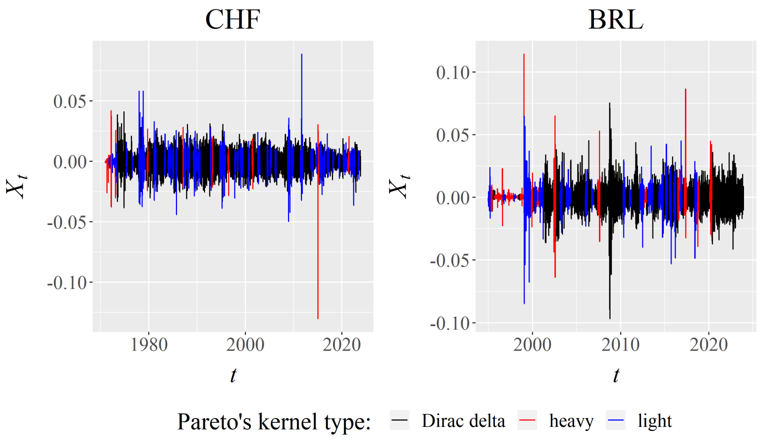


**Figure 4.4:** Log returns from data shown in Figure 4.3, with the Pareto's kernel types suggesting different stochastic regimes.

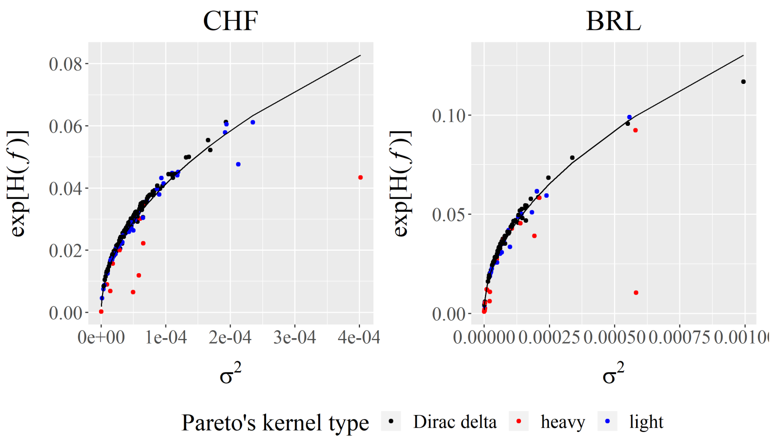


**Figure 4.5:** $m$ estimated differential entropies *versus* sample variances calculated for each block of size $b = 60$, where $m = 221$ for CHF and $m = 121$ for BRL data. The red line refers to the Gaussian differential entropy given by $H(f) = 0.5\ln(2\pi\sigma^2) + 0.5$

**Table 4.2:** Estimating differential entropy of an univariate Student's $t$ distribution with $\nu$ degrees of freedom: Mean squared error (MSE), variance ($\mathbb{V}[\hat{H}(\hat{f})]$), squared bias ($B^2$) from 500 Monte Carlo replications of different kernel entropy estimators, and NND and SS estimators as benchmarks.

| $\nu$ | measures | Gaussian | Laplace | Cauchy | Pareto | NND | SS |
|---|---|---|---|---|---|---|---|
| | MSE | 0.412719 | 0.641096 | 0.011115 | 0.010730 | 0.011699 | 0.009152 |
| 1 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.087178 | 0.218827 | 0.008426 | 0.008142 | 0.011698 | 0.008751 |
| | $B^2$ | 0.325541 | 0.422269 | 0.002689 | 0.002587 | 0.000001 | 0.000401 |
| | MSE | 0.109086 | 0.020924 | 0.003946 | 0.003965 | 0.007454 | 0.004392 |
| 2 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.052868 | 0.011148 | 0.003133 | 0.003244 | 0.007418 | 0.004269 |
| | $B^2$ | 0.056218 | 0.009776 | 0.000813 | 0.000721 | 0.000036 | 0.000123 |
| | MSE | 0.022403 | 0.005356 | 0.003139 | 0.002872 | 0.006675 | 0.003738 |
| 3 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.013795 | 0.003496 | 0.002342 | 0.002416 | 0.006642 | 0.003671 |
| | $B^2$ | 0.008608 | 0.001860 | 0.000797 | 0.000456 | 0.000033 | 0.000067 |
| | MSE | 0.006678 | 0.002625 | 0.002320 | 0.002006 | 0.006418 | 0.003446 |
| 4 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.005023 | 0.002214 | 0.001868 | 0.001890 | 0.006396 | 0.003155 |
| | $B^2$ | 0.001655 | 0.000411 | 0.000452 | 0.000116 | 0.000023 | 0.000292 |
| | MSE | 0.003356 | 0.002017 | 0.002174 | 0.001735 | 0.006271 | 0.003170 |
| 5 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.002538 | 0.001702 | 0.001561 | 0.001586 | 0.006261 | 0.003021 |
| | $B^2$ | 0.000818 | 0.000316 | 0.000613 | 0.000149 | 0.000010 | 0.000149 |
| | MSE | 0.001600 | 0.001521 | 0.002097 | 0.001493 | 0.005846 | 0.002784 |
| 10 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.001462 | 0.001411 | 0.001411 | 0.001404 | 0.005846 | 0.002678 |
| | $B^2$ | 0.000138 | 0.000109 | 0.000686 | 0.000089 | 0.000000 | 0.000106 |
| | MSE | 0.001310 | 0.001307 | 0.001986 | 0.001304 | 0.005027 | 0.002321 |
| 15 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.001212 | 0.001210 | 0.001247 | 0.001218 | 0.005024 | 0.002251 |
| | $B^2$ | 0.000097 | 0.000097 | 0.000739 | 0.000086 | 0.000002 | 0.000071 |
| | MSE | 0.001305 | 0.001279 | 0.001939 | 0.001271 | 0.005441 | 0.002608 |
| 20 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.001227 | 0.001198 | 0.001210 | 0.001197 | 0.005421 | 0.002451 |
| | $B^2$ | 0.000078 | 0.000081 | 0.000729 | 0.000074 | 0.000020 | 0.000158 |
| | MSE | 0.001217 | 0.001225 | 0.001942 | 0.001224 | 0.005545 | 0.002572 |
| 50 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.001164 | 0.001160 | 0.001195 | 0.001162 | 0.005531 | 0.002433 |
| | $B^2$ | 0.000052 | 0.000065 | 0.000747 | 0.000063 | 0.000014 | 0.000138 |
| | MSE | 0.001098 | 0.001105 | 0.001820 | 0.001102 | 0.005222 | 0.002311 |
| 100 | $\mathbb{V}[\hat{H}(\hat{f})]$ | 0.001049 | 0.001042 | 0.001061 | 0.001041 | 0.005220 | 0.002201 |
| | $B^2$ | 0.000050 | 0.000063 | 0.000759 | 0.000062 | 0.000003 | 0.000110 |

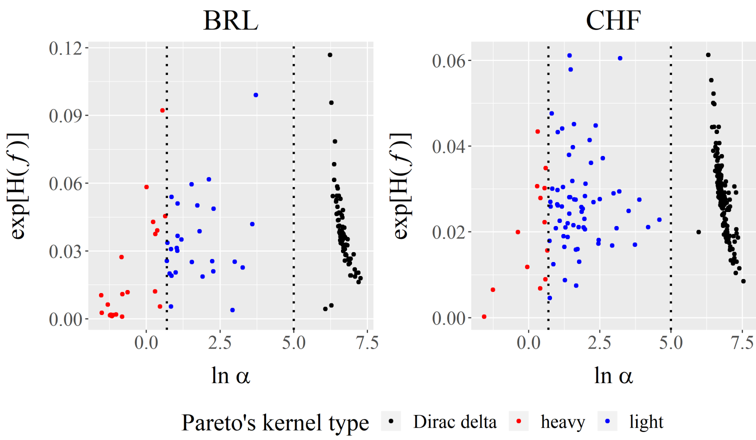


**Figure 4.6:** $m$ estimated differential entropies *versus* estimated $\ln \alpha$ for each block of size $b = 60$, where $m = 221$ for CHF and $m = 121$ for BRL data. We find three clusters of $\alpha$ discriminated by dashed vertical lines: $\ln \alpha \leq \ln 2$ (heavy-tailed kernel), $\ln 2 < \ln \alpha \leq 5$ (light-tailed kernel) and we consider $\ln \alpha > 5$ as an approximation of the Dirac's delta kernel.

**Table 4.3:** Blocks with differential entropies estimated with heavy-tailed kernel

| Swiss franc (CHF) | | | |
|---|---|---|---|
| block | period | $\alpha$ | $H(f)$ |
| 1 | 1971-01-05 — 1971-03-31 | 0.21 | -8.35 |
| 2 | 1971-04-01 — 1971-06-24 | 1.50 | -4.99 |
| 3 | 1971-06-25 — 1971-09-27 | 1.80 | -4.72 |
| 4 | 1971-12-28 — 1972-03-21 | 0.95 | -4.44 |
| 5 | 1972-03-22 — 1972-06-14 | 0.29 | -5.03 |
| 6 | 1972-12-08 — 1973-03-07 | 1.75 | -3.81 |
| 7 | 1979-08-23 — 1979-11-20 | 1.34 | -3.49 |
| 8 | 1987-01-29 — 1987-04-23 | 1.75 | -3.50 |
| 9 | 1993-01-15 — 1993-04-12 | 1.81 | -3.36 |
| 10 | 1996-05-16 — 1996-08-09 | 0.68 | -3.91 |
| 11 | 2001-05-18 — 2001-08-13 | 1.51 | -3.58 |
| 12 | 2014-12-30 — 2015-03-26 | 1.37 | -3.14 |
| 13 | 2021-03-31 — 2021-06-23 | 1.92 | -4.16 |
| Brazilian real (BRL) | | | |
| block | period | $\alpha$ | $H(f)$ |
| 1 | 1995-03-29 — 1995-06-21 | 0.44 | -4.52 |
| 2 | 1995-09-18 — 1995-12-13 | 0.32 | -6.43 |
| 3 | 1995-12-14 — 1996-03-18 | 0.22 | -5.93 |
| 4 | 1996-03-19 — 1996-06-11 | 0.30 | -6.34 |
| 5 | 1996-06-12 — 1996-09-05 | 0.27 | -5.08 |
| 6 | 1996-12-04 — 1997-03-03 | 0.29 | -6.49 |
| 7 | 1997-03-04 — 1997-05-27 | 0.30 | -6.53 |
| 8 | 1997-05-28 — 1997-08-20 | 0.31 | -6.69 |
| 9 | 1997-08-21 — 1997-11-17 | 0.36 | -6.23 |
| 10 | 1998-02-17 — 1998-05-11 | 0.44 | -6.97 |
| 11 | 1998-08-05 — 1998-10-29 | 1.58 | -5.22 |
| 12 | 1998-10-30 — 1999-01-28 | 0.22 | -4.56 |
| 13 | 1999-10-15 — 2000-01-10 | 0.43 | -3.60 |
| 14 | 2000-06-30 — 2000-09-25 | 0.53 | -4.44 |
| 15 | 2002-05-31 — 2002-08-23 | 1.72 | -2.38 |
| 16 | 2007-06-05 — 2007-08-28 | 1.90 | -3.09 |
| 17 | 2012-08-30 — 2012-11-27 | 1.34 | -4.41 |
| 18 | 2016-06-30 — 2016-09-23 | 1.35 | -3.28 |
| 19 | 2017-03-23 — 2017-06-15 | 1.44 | -3.24 |
| 20 | 2018-08-27 — 2018-11-21 | 1.26 | -3.15 |
| 21 | 2020-02-11 — 2020-05-05 | 1.01 | -2.84 |

# Chapter 5

# Final Considerations

The Gaussian Kernel estimator consistently demonstrated favorable outcomes, especially when sampling from light-tailed populations. However, challenges were observed with the Sample-Spacings (SS) and Nearest Neighbor (NN) estimators when dealing with real data containing ties. Additionally, while the literature often emphasizes the smoothing parameter's influence on Kernel-based density estimates, this study revealed a connection between the population distribution and the kernel function's shape. Despite these findings indicating the need for adaptations in the Kernel estimator method, its general applicability remains valuable. It is an alternative measure to the standard deviation, which may not always provide insightful information about data uncertainty.

We introduce a groundbreaking approach to estimating differential entropy in financial data analysis, which is critical for understanding market dynamics and assessing risk. Traditional methods often struggle with the heavy-tailed distributions characteristic of financial returns. We propose a novel estimator that employs a Paretian kernel function adept at handling the complexities of tail heaviness. Among the non-parametric estimators studied, our Kernel estimator utilizing the Pareto Kernel function stood out for its superior overall properties, particularly when optimized for the smoothing parameter focused on entropy estimation of light or heavy-tailed distributions. Thus, our approach improves upon existing techniques by incorporating an additional smoothing parameter, the Pareto exponent, which allows for enhanced flexibility in adjusting to light and heavy-tailed distributions.

Through Monte Carlo simulations, we have demonstrated the superior performance of our approach against established estimators across different scenarios. Furthermore, applying our method to foreign exchange market data underscores its practical utility in identifying stochastic regimes, offering a more nuanced understanding of market behavior. Our findings strongly advocate for the integration of the Paretian kernel estimator into the toolkit of financial analysts and researchers, proposing a significant shift towards more accurate and adaptable tools in financial analysis.

Subsequent studies may investigate integrating this approach with other financial modeling techniques to uncover complex market behaviors and improve risk assessment methodologies.

# References


Al-Yahyaee, Khamis Hamed et al. (2019). "Testing for the Granger-causality between returns in the U.S. and GIPSI stock markets". *Physica A: Statistical Mechanics and its Applications* 531, p. 120950. DOI: https://doi.org/10.1016/j.physa.2019.04.186.

Beirlant, Jan and van Zuijlen, M.C.A (1985). "The empirical distribution function and strong laws for functions of order statistics of uniform spacings". *Journal of Multivariate Analysis* 16.3, pp. 300–317.

Beirlant, Jan et al. (1997). "Nonparametric entropy estimation: An overview". *International Journal of Mathematical and Statistical Sciences* 6.1, pp. 17–39.

Chapeau-Blondeau, François and Rousseau, David (2009). "The minimum description length principle for probability density estimation by regular histograms". *Physica A: Statistical Mechanics and its Applications* 388.18, pp. 3969–3984. DOI: https://doi.org/10.1016/j.physa.2009.05.044.

Cover, Thomas M. and Thomas, Joy A. (2006). *Elements of information theory*. 2nd. John Wiley & Sons.

Dionisio, Andreia, Reis, A. Heitor, and Coelho, Luis (2008). "Utility function estimation: The entropy approach". *Physica A: Statistical Mechanics and its Applications* 387.15, pp. 3862–3867. DOI: https://doi.org/10.1016/j.physa.2008.02.072.

Duin (1976). "On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions". *IEEE Transactions on Computers* C-25.11, pp. 1175–1179. DOI: 10.1109/TC.1976.1674577.

Fernandes, Leonardo H. S. et al. (2023). “Assessment of Sector Bond, Equity Indices and Green Bond Index Using Information Theory Quantifiers and Clusters Techniques”. *Fractals* 31.1, pp. 2350017–537. DOI: 10.1142/S0218348X23500172.

Goel, Ritu, Taneja, H.C., and Kumar, Vikas (2018). “Measure of entropy for past lifetime and k-record statistics”. *Physica A: Statistical Mechanics and its Applications* 503, pp. 623–631. DOI: https://doi.org/10.1016/j.physa.2018.02.200.

Govindan, R.B. et al. (2007). “Revisiting sample entropy analysis”. *Physica A: Statistical Mechanics and its Applications* 376, pp. 158–164. DOI: https://doi.org/10.1016/j.physa.2006.10.077.

Guo, X., Zhang, H., and Tian, T. (n.d.). “Development of stock correlation networks using mutual information and financial big data”. *PLoS ONE* 13.4 (), e0195941. DOI: 10.1371/journal.pone.0195941.

Hall, P. and Morton, S.C. (1993a). “On the estimation of entropy”. *Annals of the Institute of Statistical Mathematics* 45, pp. 69–88.

Hall, Peter (1987). “On Kullback-Leibler Loss and Density Estimation”. *The Annals of Statistics* 15.4, pp. 1491–1519.

Hall, Peter and Morton, Sally C. (1993b). “On the estimation of the entropy.” *Ann. Inst. Statist. Math.* 45.1, pp. 69–88.

Ivanov, A. V. and Rozhkova, M. N. (1981). “On Properties of the Statistical Estimate of the Entropy of a Random Vector with a Probability Density”. *Problems of Information Transmission* 17.3, pp. 171–178.

Jaroonchokanan, Nawee, Termsaithong, Teerasit, and Suwanna, Sujin (2022). “Dynamics of hierarchical clustering in stocks market during financial crises”. *Physica A: Statistical Mechanics and its Applications* 607, p. 128183. DOI: https://doi.org/10.1016/j.physa.2022.128183.

Jizba, Petr and Korbel, Jan (2014). "Multifractal diffusion entropy analysis: Optimal bin width of probability histograms". *Physica A: Statistical Mechanics and its Applications* 413, pp. 438–458. DOI: `https://doi.org/10.1016/j.physa.2014.07.008`.

Jizba, Petr, Lavička, Hynek, and Tabachová, Zlata (2021). "Rényi Transfer Entropy Estimators for Financial Time Series". *Engineering Proceedings* 5.1. DOI: `10.3390/engproc2021005033`.

Joe, Harry (1989). "Estimation of entropy and other fuunctionals of a multivariate density." *Ann. Inst. Statist. Math.* 41.4, pp. 683–697.

Kozachenko, L. F. and Leonenko, N. N. (1987). "Sample estimate of entropy of a random vector." *Problems of Information Transmission* 23, pp. 95–101.

Mandelbrot, Benoit B. (2003). "Chapter 1 - Heavy Tails in Finance for Independent or Multifractal Price Increments". In: *Handbook of Heavy Tailed Distributions in Finance*. Ed. by Svetlozar T. Rachev. Vol. 1. Handbooks in Finance. Amsterdam: North-Holland, pp. 1–34. DOI: `https://doi.org/10.1016/B978-044450896-6.50003-0`.

Matsushita, Raul et al. (2006). "Log-periodic crashes revisited". *Physica A: Statistical Mechanics and its Applications* 364, pp. 331–335. DOI: `https://doi.org/10.1016/j.physa.2005.08.051`.

Matsushita, Raul et al. (2020). "Bypassing the truncation problem of truncated Lévy flights". *Physica A: Statistical Mechanics and its Applications* 559, p. 125035. DOI: `https://doi.org/10.1016/j.physa.2020.125035`.

Matsushita, Raul et al. (2023). "Retrodicting with the truncated Lévy flight". *Communications in Nonlinear Science and Numerical Simulation* 116, p. 106900. ISSN: 1007-5704. DOI: `https://doi.org/10.1016/j.cnsns.2022.106900`.

Ormos, Mihály and Zibriczky, Dávid (2014). "Entropy-Based Financial Asset Pricing". *PLOS ONE*.

Scott, David W. (1992). *Multivariate Density Estimation Theory, Practice, and Visualization*. John Wiley & Sons.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Springer.

Singh, Harshinder et al. (Feb. 2003). "Nearest Neighbor Estimates of Entropy". *American Journal of Mathematical and Management Sciences* 23.

Smerlak, Matteo (2016). "Thermodynamics of inequalities: From precariousness to economic stratification". *Physica A: Statistical Mechanics and its Applications* 441, pp. 40–50. DOI: https://doi.org/10.1016/j.physa.2015.09.001.

Wachowiak, Mark P. et al. (2005). "Estimation of generalized entropies with sample spacing". *Pattern Analysis and Applications* 8, pp. 95–101.

Xian, Lu, He, Kaijian, and Lai, Kin Keung (2016). "Gold price analysis based on ensemble empirical model decomposition and independent component analysis". *Physica A: Statistical Mechanics and its Applications* 454, pp. 11–23. DOI: https://doi.org/10.1016/j.physa.2016.02.055.

Xu, Meng, Shang, Pengjian, and Zhang, Sheng (2019). "Multiscale analysis of financial time series by Rényi distribution entropy". *Physica A: Statistical Mechanics and its Applications* 536, p. 120916. DOI: https://doi.org/10.1016/j.physa.2019.04.152.

Zhang, Yali et al. (2020). "Cumulative Tsallis entropy based on multi-scale permuted distribution of financial time series". *Physica A: Statistical Mechanics and its Applications* 548, p. 124388. DOI: https://doi.org/10.1016/j.physa.2020.124388.

Zhou, Rongxi, Cai, Ru, and Tong, Guanqun (2013). "Applications of Entropy in Finance: A Review". *School of Economics and Management*.