

2023-05-01

Volatility Modeling Of Time Series Using Fractal And Self-Similarity Models

William Kubin
University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd



Part of the [Applied Mathematics Commons](#), and the [Mathematics Commons](#)

Recommended Citation

Kubin, William, "Volatility Modeling Of Time Series Using Fractal And Self-Similarity Models" (2023). *Open Access Theses & Dissertations*. 3808.

https://scholarworks.utep.edu/open_etd/3808

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.

VOLATILITY MODELING OF TIME SERIES USING
FRACTAL AND SELF-SIMILARITY
MODELS

WILLIAM KUBIN

Doctoral Program in Computational Science

APPROVED:

Maria C. Mariani, Ph.D., Chair

Osei K. Tweneboah, Ph.D., Co-Chair

Thompson Sarkodie-Gyan, Ph.D.

Granville Sewell, Ph.D.

Elsa Villa, Ph.D.

Stephen Crites, Ph.D.
Dean of the Graduate School

©Copyright

by

William Kubin

2023

to my

FATHER, Mr. Hawk and MOTHER, Mama Maggie

with love

VOLATILITY MODELING OF TIME SERIES USING
FRACTAL AND SELF-SIMILARITY
MODELS

by

WILLIAM KUBIN

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

May 2023

Acknowledgements

My sincere thanks go to my advisor, Dr. Maria C. Mariani from the Mathematics Department at The University of Texas at El Paso, for her valuable insights, feedback, patience, and continuous support that helped me to shape my research project and refine my writing skills.

I would also like to express my appreciation to the members of my dissertation committee, Dr. Thompson Sarkodie-Gyan, Dr. Granville Sewell, Dr. Elsa Villa, and Dr. Osei K. Tweneboah for their participation in this project and for taking the time out of their busy schedules.

I am grateful to the faculty and staff of the Computational Science Program and Mathematics Department at The University of Texas at El Paso for their expertise, and dedication. They have provided me with the resources necessary to complete my degree and pursue a career as a computational scientist.

I would like to extend my heartfelt gratitude to my parents for their unwavering love, support and guidance throughout my upbringing and trust that allowed me to pursue my academic goals. Their encouragement has been invaluable in helping me reach this point.

I am grateful for the invaluable teaching and guidance provided by Mr. Isaac Nketsiah of Moree Star of Hope, Mr. Victor K. Egyir of St. Augustine's College, and Mr. Francis Eshun of Academy of Christ The King Senior Secondary School. Thanks to their mentorship, my passion for mathematics and computational modeling has flourished.

Thank you all for your contributions and kindness. I will always cherish the memories and lessons learned from this remarkable journey.

NOTE: This thesis was submitted to my Supervising Committee on the April 31, 2023.

Abstract

The study uses various methods to compare the scaling parameters and long-term memory behavior of financial and geophysical time series. The Cantor Detrended Fluctuation Analysis (CDFA) method is proposed to provide more accurate estimates of Hurst exponents. The CDFA method is applied to real-time series and the results are verified. The study also analyzes the memory behavior of daily Covid-19 cases before and after the announcement of effective vaccines. Low and high-frequency data's influence on the Hurst Index estimation is investigated, and a new PCDFFA method is proposed. The stability of the Dow Jones Industrial Average is analyzed using a multi-scale normalized diffusion entropy and conditional diffusion entropy. The study aims to investigate memory behavior in time series using deep learning techniques in future work.

Table of Contents

	Page
Acknowledgements	v
Abstract	vi
Table of Contents	vii
List of Tables	x
List of Figures	xi
Chapter	
1 Introduction	1
2 Methodology	7
2.1 Variance Scaling Method (VSM)	7
2.1.1 Detrended Fluctuation Analysis (DFA)	7
2.1.2 Q-Order Detrended Fluctuation Analysis (Q-DFA)	8
2.2 Probability Scaling Methods (PSM)	10
2.2.1 Diffusion Entropy Analysis (DEA)	10
2.2.2 Conditional Diffusion Entropy (CDE)	13
2.3 Q-Order Diffusion Entropy Analysis (Q-DEA)	14
2.4 Multi Scale Conditional Diffusion Entropy (MS-CDE)	15
2.5 Truncated Lévy Flight (TLF)	16
2.6 Relationship Between VSM and TLF	19
2.6.1 DFA and TLF	19
2.6.2 Q-DFA and TLF	20
2.7 Relationship Between PSM and TLF	20
2.7.1 DEA and TLF	20
2.7.2 Q-DEA and TLF	21
2.8 Cantor Detrended Fluctuation Analysis (CDFA)	22

2.8.1	The Cantor set	25
2.8.2	Definition	25
2.8.3	Algorithm of the CDFA	26
2.9	Parallel Framework for CDFA (PCDFA)	27
2.9.1	Hardware & Software Environment	27
3	Data	29
3.1	Financial Time Series	29
3.2	Geophysical Time Series	31
3.3	Biomedical Time Series	32
3.4	Covid-19 Time Series	34
4	Applications	38
4.1	Relationships Between DFA, DEA, and TLF	38
4.2	CDFA Applied to Biomedical Signals	39
4.3	Q-CDFA Applied to Covid-19 Time Series	42
4.3.1	Before November 2020	42
4.3.2	After November 2020	43
4.4	PCDFA Applied to Financial Time Series	46
4.4.1	Results and Discussions	46
4.4.2	Speedup Ratio of PCDFA	47
4.4.3	Parallel Efficiency of PCDFA	49
4.4.4	Parallel Overhead of PCDFA	50
4.5	Q-DEA for Dow Jones Industrial Average (DJI)	51
4.6	CDE and MS-CDE of DJI from April 2013 to April 2021	52
5	Conclusion	55
	References	58
	Appendix A	67
	Appendix B	79
	Appendix C	83

Curriculum Vitae 88

List of Tables

3.1	Financial Market Data for Analysis.	30
4.1	Scaling exponents of various financial markets	38
4.2	Scaling exponents of Volcanic Data from the two seismic stations	39
4.3	DFA's Hurst Exponents for White noise time series	40
4.4	DFA's Hurst Exponents for Monofractal time series	40
4.5	DFA's Hurst Exponents for Multifractal time series	41
4.6	Comparison of Scaling Exponents of DFA(H) & CDFA(H^c) & TLF (α) on Noise-like Time series	41
4.7	Q-DFA scaling exponents from Covid cases before Vaccines	43
4.8	Q-DFA scaling exponents from Covid cases after Vaccines	44
4.9	PCDFA Time Analysis of Dow Jones Industrial Average (in seconds)	46
4.10	PCDFA Time Analysis of Bitcoin Data (in seconds)	47
4.11	PCDFA Speedup Ratios of Low- & High-Frequency DJI & BTC Data	48
4.12	PCDFA Parallel Efficiency Analysis of DJI & BTC	49
4.13	PCDFA Parallel Overhead Analysis of DJI & BTC	51

List of Figures

2.1	Fractal behavior of a Ternary Cantor Set.	25
2.2	A flow chart of PC DFA for 4 available processors	28
3.1	Plots of trading days closings and actual returns of HSI	30
3.2	Plots of trading days closings and actual returns of SP500	31
3.3	Daily Close Prices Plot of DJI	32
3.4	Daily Close Prices Plot of BTC	33
3.5	Volcanic Eruption 1 and 2 Plots	34
3.6	Volcanic Eruption 7 and 8 Plots	35
3.7	Biomedical time series plots	36
3.8	Evolution of daily US cumulative and log-difference of Covid-19 cases	37
4.1	Plot of q-order Hurst exponents Before and After the Vaccines	45
4.2	Plot of q-order singularity spectrum Before and After the Vaccines	46
4.3	Q-order Diffusion Entropy as a function of diffusion time constructed based on different q values of the Rényi family of entropies using DJI	52
4.4	Daily Close Prices Plot of DJI. The gray region corresponds to the Covid-19 US Stock Market Crash of February/March 2020	53
5.1	TLF model fits for HSI	67
5.2	TLF model fits for SP500 USA	68
5.3	DFA Plot for HSI	69
5.4	DFA for SP500 USA	70
5.5	DEA for HSI	71
5.6	DEA for SP500 USA	72
5.7	TLF model fit for Volcanic Eruption 2	73

5.8	TLF model fit for Volcanic Eruption 8	74
5.9	DFA for Volcanic Eruption 2	75
5.10	DFA for Volcanic Eruption 8	76
5.11	DEA for Volcanic Eruption 2	77
5.12	DEA for Volcanic Eruption 8	78
5.13	Speedup Ratio plot of PC DFA for low-to-high frequency DJI data	79
5.14	Speedup Ratio plot of PC DFA for low-to-high frequency BTC data	80
5.15	Efficiency plot of PC DFA for low-to-high frequency DJI data	81
5.16	Efficiency plot of PC DFA for low-to-high frequency BTC data	82
5.17	Monthly Conditional Entropy of DJI at Time Scale $t=2$	83
5.18	Monthly Conditional Entropy of DJI at Time Scale $t=4$	83
5.19	Monthly Conditional Entropy of DJI at Time Scale $t=8$	84
5.20	Monthly Conditional Entropy of DJI at Time Scale $t=16$	84
5.21	Monthly Conditional Entropy of DJI at Time Scale $t=32$	85
5.22	Monthly Conditional Entropy of DJI at Time Scale $t=64$	85
5.23	Monthly Conditional Entropy of DJI at Time Scale $t=128$	86
5.24	Monthly Conditional Entropy of DJI at Time Scale $t=256$	86
5.25	Monthly Conditional Entropy of DJI at Time Scale $t=512$	87
5.26	Monthly Conditional Entropy of DJI at Time Scale $t=1024$	87

Chapter 1

Introduction

A fractal is a geometric object with similar statistical properties at all scales. Successive magnifications of a fractal make it look the same as the original fractal shape. This pattern, which appears on smaller and smaller scales, is often known as self-similarity in fractal mathematics [1, 2]. A self-similarity process is a type of stochastic process that exhibits self-similarity phenomena. Self-similar processes occur in many areas of applied mathematics, such as fractals, chaos theory, long-term memory processes, and spectral analysis. They can exhibit long-term dependencies, also referred to as long memory persistence. This phenomenon describes events in which the dependencies of a time series decrease more slowly than exponential decay (typically power decay) [3].

Definition 1.0.1 (Stochastic Process). *A stochastic process $X(t)$ is self-similar if there exists a constant $H > 0$ such that for any scaling factor $a > 0$, the processes $\{X_{at}\}_{t \geq 0}$ and $\{a^H X_t\}_{t \geq 0}$ have the same laws in the sense of a finite-dimensional distribution. The constant H is called the self-similarity exponent of process X .*

The variances of time series ranging from financial markets to geophysical phenomena such as seismic energy released from volcanic eruptions are stochastic. It changes over time and increases during periods of high volatility. This function is modeled by evaluating trends over shorter periods to determine fluctuations. There are scaling methods that quantify the exponent of decay functions. We broadly group these methods into Variance Scaling Methods (VSMs) and Probability Scaling Methods (PSMs). Some of these methods include Rescaled Range Analysis (R/S), Detrended Fluctuation Analysis (DFA), the Truncated Lévy Flight (TLF), and Diffusion Entropy Analysis (DEA).

Hurst's Rescaled Range Analysis (R/S) method is a variance scaling method that subdivides an integrated time series into contiguous segment sizes and examines the integrated range of variation (R). Then, a measure of variance, usually the standard deviation (S), is determined as a function of segment size. The power law defines the approximate relationship between the Rescaled Range Analysis (R/S) statistic and segment size [4].

The Detrended Fluctuation Analysis (DFA) by Peng [5] (1994) is a technique that quantifies the same power law exponent of the R/S method. Like Hurst's R/S method, DFA is based on random walk theory [6, 7]. Addressing difficulties in determining correct power law exponents of the R/S in non-stationary time series resulted in the introduction of the DFA. Unlike the R/S, the DFA uses a local detrending approach (usually linear regression) in the segments of the integrated series. It provides its power law exponents protection against the effects of nonstationarity and pollution of time series by external signals while eliminating spurious detection of long memory. It makes detecting the intrinsic self-similarities in seemingly volatile time series possible. Empirical evidence has shown that the DFA performs well compared to other VSMs, including the R/S, when estimating power law exponents. Even though the DFA performs better than the R/S, it sometimes tends to overestimate/underestimate the Hurst exponent.

In [8], the authors show that the estimation of the Hurst exponent is subject to the choice of scale that can lead to an overestimation/underestimation of the Hurst exponent. In order to address this DFA-related problem, we proposed Cantor Detrended Fluctuation Analysis (CDFA). CDFA uses Cantor set theory to define non-overlapping equal segments of the DFA methods to solve the problem. The Cantor set is constructed by removing the middle part of a three-part series and repeating this process with the remaining short segments. It is the prototype of a fractal [9]. There are videos and animations showing the fractal properties of the Cantor set [10]. However, we will focus on this phenomenon using the DFA power exponent as a fractal measure to prove the self-similarity property of Cantor sets. We further check the relationship using real-time series and how the Cantor scale ($1/3^n, n \geq 0$) solves the overestimation problem. The obtained Hurst exponent is then

compared to the conventional DFA and TLF to examine its performance. We implement parallel CDFA optimization scheme (PCDFA) to address the computational complexity of potentially large amounts of data (i.e., big data) at a high frequency, such as minutes per minute. We investigate the impact of the CDFA Hurst exponent on runtime. We use high- and low-frequency DJI and BTC data to evaluate the PCDFA algorithm for parallel computing metrics such as speed-up ratio, efficiency, and overhead.

Usually, characterizing stochastic processes requires the study of determining asymptotic probability density distributions (*pdf*) and temporal correlations. Brownian motion models the evolution of a particle's position over time by assuming that the movement of particles follows a diffusive process with Gaussian distribution. This model did not accurately describe real-world time series because the kurtosis of the associated pdf is greater than that of the Gaussian distribution [11]. The Truncated Lévy Flight (TLF) model originated to address the difficulties of the Brownian motion for working in long-range correlation scales. The TLF is a random walk with a heavy-tailed probability distribution with a step length. Researchers have extended Lévy flight to include cases where random walks occur on discrete grids rather than in contiguous space. [12, 13]. For instance, animals (including birds and humans) follow paths modeled by Lévy's flight when searching for food [14, 15, 16]. For a typical step size distribution that satisfies the power condition, the distance from the start of a random walk approaches a stable distribution after many steps. It allows you to model many processes using Lévy flights. In particular, this is a result of the generalized central limit theorem. Therefore, we use Lévy's flight to model long memory processes as demonstrated by real-time series data. Scaling exponent ($0 < \alpha \leq 2$) of the TLF uses Gaussian and non-Gaussian distributions to measure the memory behavior of time series following a diffusion process [1, 3, 17].

An alternative method to VSMs is the Diffusion Entropy Analysis (DEA) model developed by Scafetta et al. (2002), which falls under probability scaling methods (PSMs). DEA can be used to determine whether a time series' characteristics follow a Gaussian or Lévy distribution. It can also determine if a time series has long-range correlations. Scafetta

argued that DEA is the only way to scale time series of complex processes [18] properly.

This method transforms the series into a diffusion process and determines the probability that a particle belongs to a given optimal bin size. The monofractal approach of the DEA uses the Shannon Entropy of Scafetta et al. [19]. The multifractal method uses a mix of the Shannon and Rényi Entropy [20, 21] for different q weights to estimate scaling exponents. An important result of this study is the proposal to study bull and bear markets in stocks using a scale with multiple time lags to determine the Conditional Diffusion Entropy (CDE) and Multi Scale Conditional Diffusion Entropy (MS-CDE). We use the study of optimal bin widths in empirical histograms by Jizba et al. [22] to estimate the underlying probability density function.

Understanding the stability of financial systems, as a result of the US sub-prime crisis, has grown in importance for asset and derivative pricing, asset allocation, and risk management among practitioners, researchers, and regulators in the domains of economics, mathematics, and physics in recent years [23, 24, 25, 26, 27, 28, 29]. These scaling approaches characterize the dependence of observations separated in time series dominated by stochastic properties. Researchers have applied some of these models to various domains, including DNA sequencing, [30, 31, 32] neural oscillations, [33] speech pathology detection, [34] and heartbeat fluctuation during different sleep stages [35], etc. We use established concepts and methods in complex systems and econophysics to investigate unusual, chaotic, and erratic behavior in economic systems, of which entropy is one. Numerous examples in the literature show abrupt transitions from a steady state to a fundamentally different one, building on previous research on financial crises as complex dynamic systems. In [36], the authors discuss whether entropy reflects stock market uncertainty and disorder.

In this study, we use DFA and DEA to investigate long memory behavior in financial and geophysical time series. We report the DFA and DEA scaling parameters, H and δ respectively, and say that the time series exhibit long memory behavior if $(H, \delta) \in (0.5, 1]$, anti-persistence if $(H, \delta) \in [0, 0.5)$, and $H, \delta = 0.5$ for completely random noise to draw conclusions. We generate a mathematical relationship between the TLF parameter α that

characterizes data and the resulting parameters H and δ of DFA and DEA that characterize the self-similar feature, respectively. Then, we analytically prove the relationship between α , H and δ . We also use a biological time series to determine the Hurst exponents of DFA (H) and CDFA (H^c) and compare them with the corresponding TLF α . Here we observe the overestimation problem of H and how H^c solves it. Next, we investigate the PC DFA for a large financial time series. This shared-memory multiprocessing computational technique tells us that the execution time of the sequential CDFA code increases significantly as the amount of data increases. On the other hand, the PC DFA run time increases by a smaller margin as the data size increases but provides the same H^c scaling exponent, making it more efficient when dealing with big datasets.

Since its detection in December 2019, the novel coronavirus resulting from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a respiratory illness, has claimed thousands of lives and devastated global economies together with the US. The World Health Organization (WHO) classified it as a public health emergency. Since then, numerous precautionary measures have been set to protect lives (e.g. mask mandate, social distancing etc.) and policies to get the economy back to normal (e.g. stimulus payments). Loads of research are currently taking place to overcome the virus and its impact. Another application we investigate in this study uses the Q-CDFA to analyze the highly erratic behavior of daily Covid-19 cases in the United States before and after November 2020 when drug manufacturers announced that they had developed and tested a vaccine that was over 90% effective. We use the multifractal spectrum of the Q-CDFA to characterize the path and predict the memory behavior of the time series on different time scales.

In [37], based on Tsallis and the Shannon entropy, the authors investigated the volatility of seven stock market indices. Compared to other methods, such as convexity, variance, and vector autoregression (VaR), the authors in [38] found that the information entropy method can better quantify the risk associated with bonds and other financial securities. Therefore, the need to identify potentially significant factors to reduce the negative consequences on economic systems has received attention recently, even though global financial crisis are

caused by events generated in the financial industry sectors. In light of that, we employ the proposed CDE and MS-CDE techniques to predict bull and bear markets. A market that is expanding and where economic conditions are typically favorable is known as a bull market. Bear market develops when the economy contracts and most stocks and equities lose value. This experiment uses daily sampled Dow Jones Industrial Average (DJI) market data from 2013 to 2021. For conditional entropy, $C_q(t) = 1$ implies random behavior in financial markets, $C_q(t) > 1$ in a bull market and is less than 1 in a bear market.

The outline of this study is as follows. We look at the variance and probability scaling methods such as the DFA, Q-DFA, DEA, and Q-DEA in Chapter 2. I also introduce the concept of conditional diffusion entropy (CDE) and Multi Scale CDE (MS-CDE). TLF model is also presented in this section to determine the relationship of its parameter to DFA and DEA's scaling parameters. I introduce the CDFA with proofs, illustrations, and required steps for implementation. Finally, this section ends with the CDFA parallel processing framework known as the PCDFA. Chapter 3 presents all the time series data sets used for applications ranging from finance to geophysical and medical data. I review the results and discussion of all applications undertaken in this study in Chapter 4. Chapter 5 concludes the study and outlines future research work.

Chapter 2

Methodology

There are several methods to analyze self-similar behavior in time series. There are two (2) main classes of models, namely variance scaling methods (VSMs) and probability-based scaling methods (PSMs). VSMs use different measures of dispersion, like range, standard deviation, variance, etc., to assess self-similar behavior. On the other hand, PSMs use probability estimation theory to model disorder and chaos to determine the fractal pattern in time series.

2.1 Variance Scaling Method (VSM)

2.1.1 Detrended Fluctuation Analysis (DFA)

The Monofractal DFA method, also called the DFA technique, is an innovative strategy for uncovering long-range correlations in non-stationary time series. The technique was advanced by Peng et al. in 1993 and has numerous applications in fields such as cloud formation analysis, DNA examination, cardiac dynamics, climate study, solid-state physics, and financial time series analysis. The following section demonstrates the steps for calculating DFA exponents numerically.

Let N be the length of the time series $\{X_i\}_{i=1}^N$. The logarithmic ratio of the time series is obtained. The length of the new time series $M(t)$ will be $N - 1$.

$$M(t) = \log \left(\frac{X_{t+1}}{X_t} \right) \quad (2.1)$$

for $t = 1, 2, \dots, N - 1$. The absolute value of $M(t)$ is integrated:

$$X(t) = \sum_{i=1}^t |M(i)| \quad (2.2)$$

Next, the $N - 1$ length integrated time series is divided into m segments, each with a length of n , without overlapping. Since the data is divided into segments of equal length, there may be some remaining values. To consider these values, the same process is repeated but starting from the end, resulting in a total of $2(N - 1)$ segments. These are then averaged. For each segment, a line of best fit is calculated to represent the trend, resulting in $y_n(i)$. Finally, the root mean square errors (RMSE) is determined using the following formula.

$$F(n) = \sqrt{\frac{1}{2(N - 1)} \sum_{i=1}^{2(N-1)} [X(i) - X_n(i)]^2} \quad (2.3)$$

The calculation is repeated for different box sizes to identify the relationship between box size n and $F(n)$. If there is a linear relationship between $F(n)$ and n (i.e., box size) when plotted on a logarithmic scale, it indicates that the fluctuations can be characterized by a scaling exponent, H , which is the slope of the line that relates $\log F(n)$ to $\log n$. This results in the following mathematical relationship:

$$F(n) \propto n^H \quad (2.4)$$

For data series with no fractal pattern, $H = 0.5$. For data series with long-memory, $0.5 < H < 1$ and $0 < H < 0.5$ for short-memory behavior.

2.1.2 Q-Order Detrended Fluctuation Analysis (Q-DFA)

As described by Kantelhardt in [39], the profile of the cumulative series X_t is divided into N_s segments of equal length s . To include all data points, the same procedure is repeated in the reverse direction, resulting in a total of $2N_s$ segments. A least squares regression is performed on each segment of length u to estimate the local trends, as shown in equation

(2.5). The average of each segment collects the q th order fluctuation \mathbb{F}_q , as described in equation (2.6).

$$\mathbb{F}_S^2(u) = \frac{1}{s} \sum_{i=1}^s [X_u(i) - \bar{X}_u(i)]^2 \quad (2.5)$$

$$\mathbb{F}_q(S) = \left\{ \frac{1}{2N_s} \sum_{u=1}^{2N_s} [\mathbb{F}_S^2(u)]^{q/2} \right\}^{1/q} \quad (2.6)$$

Next, the scaling behavior of fluctuations is obtained by determining the slope $H(q)$ of log-log plots of $\mathbb{F}_q^2(S)$ versus segment length s for each value of q [3].

$$\mathbb{F}_q(S) \propto S^{H(q)} \implies \log [\mathbb{F}_q(S)] - H(q) \log (S) = 0 \quad (2.7)$$

Based on the generalized Hurst Exponent (gHE), $H(q)$, we derive the mass exponent (as per equation (2.8)), the singularity exponent (equation (2.9)), and the singularity spectrum (equation (2.10)).

$$\tau(q) = qH(q) - 1 \quad (2.8)$$

$$\alpha = H(q) + qH'(q) \quad (2.9)$$

$$f(\alpha) = q[\alpha - H(q)] + 1 \quad (2.10)$$

It is worth mentioning that when $q = 2$, the Q-DFA methodology is equivalent to the traditional DFA. For time series data with no fractal characteristics, $H(q)$ would equal 0.5. If the scaling exponent $H(q)$ falls within the range (0.5, 1], this suggests the presence of long-memory in the data. Conversely, a value of $H(q)$ in the interval [0, 0.5) implies that the data has short-memory properties.

2.2 Probability Scaling Methods (PSM)

2.2.1 Diffusion Entropy Analysis (DEA)

The DEA can be utilized to identify the scaling characteristics of both low and high-frequency time series. Using DEA allows for the characterization of a time series as either a Gaussian or Lévy distribution and determining the presence of long-range correlations in the series.

A function $\Phi(r_1, r_2, \dots)$ is considered scale-invariant if it meets the requirement stated.

$$\Phi(r_1, r_2, \dots) = \gamma^a \Phi(\gamma^b r_1, \gamma^c r_2, \dots). \quad (2.11)$$

The equation (2.11) demonstrates that we can obtain the original function by scaling all coordinates r with a suitable set of exponents a, b, c, \dots . In the context of a time series, the sequence of numbers resulting from the composition can be understood as producing a diffusion process. Its corresponding probability distribution function $\rho(x, t)$ can be analyzed, where x represents the variable that captures the diffusion process variations. If the time series is stationary, the scale property is expressed as follows:

$$\rho(x, t) = \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right), \quad (2.12)$$

where the δ is the scaling exponent.

Use the following steps to find the scaling exponent [19]. Let N be the length of a time series $\{X_i\}_{i=1}^N$

- Convert the series $\{X_i\}_{i=1}^N$ into a diffusion process. Consider the series such that it can be represented as:

$$\{X_i, X_{i+1}, X_{i+2}, \dots, X_{i+t-1}\}$$

where $i = 1, 2, \dots, N - t + 1$ and $t \in [1, N]$ is the time scale. For any given diffusion time t and initial state $\psi_i^0 = 0$, the matrix ψ_i^j defined as

$$\psi_i^j = X_{i+j} - X_{i+j-1}, \quad j = 0, 1, \dots, N - t \quad (2.13)$$

can be denoted as sub-sequences.

Using the stochastic process

$$\eta_t^j = \sum_{i=1}^t \psi_i^j, \quad (2.14)$$

we create a diffusion trajectory for each sub-sequences where η_t^j defines the new position of the j th particle in the diffusion process.

- Determine the diffusion entropy. To begin, divide the x-axis into bins of size $B(t)$ and assume that $N_i(t)$ indicates the number of particles falling in each bin at time t , where $i = 1, 2, \dots, B(t)$.
- Pick the ideal bin size B . There is no "optimal" number of bins, and different bin sizes B can identify different data features. When the density of the underlying data points is low, wider bins are used to reduce sampling-related noise; when the density is high, narrower bins are used to increase the density estimation's precision. As a result, modifying the bin size within a histogram may be advantageous. We use Freedman-Diaconis' rule [40] for determining bin size B , which is given by

$$B = 2 \frac{IQR}{\sqrt[3]{n}}, \quad (2.15)$$

which is based on the interquartile range (IQR). The 3.5σ of Scott's rule in $B = \frac{3.5\hat{\sigma}}{\sqrt[3]{n}}$ [41], where $\hat{\sigma}$ is the sample standard deviation, is substituted with $2 IQRs$. Scott's rule is most effective with data that has a Gaussian distribution. The Freedman-Diaconis rule is less sensitive to anomalies and outliers in data than the standard deviation. As a result, the FD rule is more robust.

- The relative frequency can then be used to approximate the probability (pdf) that a particle will fall into a bin at time t as

$$p(i, t) = \frac{N_i(t)}{N - t + 1}. \quad (2.16)$$

At time t , the Shannon diffusion entropy will be determined as:

$$S(t) = - \sum_{i=1}^{B(t)} p(i, t) \ln[p(i, t)]. \quad (2.17)$$

or

$$S(t) = - \int_{-\infty}^{\infty} \rho(x, t) \ln [\rho(x, t)] dx \quad (2.18)$$

Consider the equation (2.18), assuming that $\rho(x, t)$ satisfies the scale condition (2.12).

Then, by substituting $\rho(x, t)$ in (2.18), we get

$$S(t) = - \int_{-\infty}^{\infty} \rho(x, t) \ln [\rho(x, t)] dx \quad (2.19)$$

by assuming that

$$\rho(x, t) = \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right). \quad (2.20)$$

and $F(y)$ retains its form, allowing the statistics of the processes to be time-independent.

Substituting (2.20) into (2.19) gives

$$\begin{aligned} S(t) &= - \int_{-\infty}^{\infty} \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right) \ln \left[\frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right) \right] dx \\ &= - \frac{1}{t^\delta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\delta}\right) \left[\ln\left(\frac{1}{t^\delta}\right) + \ln F\left(\frac{x}{t^\delta}\right) \right] dx \\ &= - \frac{1}{t^\delta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\delta}\right) \ln\left(\frac{1}{t^\delta}\right) dx - \frac{1}{t^\delta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\delta}\right) \ln \left[F\left(\frac{x}{t^\delta}\right) \right] dx \\ &= - \frac{1}{t^\delta} \ln\left(\frac{1}{t^\delta}\right) \int_{-\infty}^{\infty} F\left(\frac{x}{t^\delta}\right) dx - \frac{1}{t^\delta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\delta}\right) \left[\ln F\left(\frac{x}{t^\delta}\right) \right] dx \end{aligned}$$

Let $y = \frac{x}{t^\delta} \Rightarrow dx = t^\delta dy$. Substituting for dx , we get

$$S(t) = - \frac{1}{t^\delta} \ln\left(\frac{1}{t^\delta}\right) \int_{-\infty}^{\infty} F(y) t^\delta dy - \frac{1}{t^\delta} \int_{-\infty}^{\infty} F(y) \ln [F(y)] t^\delta dy$$

We get the following after canceling out certain common terms:

$$\begin{aligned}
&= -\ln\left(\frac{1}{t^\delta}\right) \int_{-\infty}^{\infty} F(y) dy - \int_{-\infty}^{\infty} F(y) \ln[F(y)] dy \\
&= -[\ln(1) - \ln(t^\delta)] \int_{-\infty}^{\infty} F(y) dy - \int_{-\infty}^{\infty} F(y) \ln[F(y)] dy \\
&= \delta \ln(t) \int_{-\infty}^{\infty} F(y) dy - \int_{-\infty}^{\infty} F(y) \ln[F(y)] dy
\end{aligned}$$

The derived equation yields the linear-log relationship between diffusion entropy $S(t)$ and t denoted by [1, 3]

$$S(t) = \delta \ln(t) + A \tag{2.21}$$

where δ is the DEA scaling exponent and the constant

$$A = - \int_{-\infty}^{\infty} F(y) \ln[F(y)] dy \tag{2.22}$$

and the Cumulative Density Function (CDF) equals 1. Equation (2.21) illustrates the growth of entropy in a linear manner with $\ln(t)$, and the scaling coefficient δ represents the slope of the linear function.

2.2.2 Conditional Diffusion Entropy (CDE)

At time t , normalizing diffusion entropy in equation (2.17) can be written as

$$\bar{S}(t) = \frac{S(t)}{t}. \tag{2.23}$$

The concept of conditional diffusion entropy, $C(t)$, offers a method for identifying bull and bear markets. A bull market is characterized by growth and favorable economic conditions, while a bear market occurs during an economic downturn where stocks generally

decrease in value. As demonstrated in [42, 43], $C(t)$ can be defined as a measure of this differentiation.

$$C(t) = 1 + \alpha [1 - \bar{S}(t)] \quad (2.24)$$

where $\bar{S}(t)$ is the normalized entropy as shown in equation (2.23) and

$$\alpha = \begin{cases} -1, & p < q \\ 0, & p = q \\ 1, & p > q \end{cases} \quad (2.25)$$

In equation (2.13), the amount of positive and negative values in the series are designated as p and q , respectively. A bear market is represented by $C(t) < 1$, while a bull market is indicated by $C(t) > 1$. In the case of an indifferent or completely random market, $C(t) = 1$.

2.3 Q-Order Diffusion Entropy Analysis (Q-DEA)

The property of multifractality in time series is continuous, and techniques that deal with the finite size of histograms and discretization must employ an interpolation method, leading to a potential for bias. To find the optimal bin size B for various values of q , the Rényi entropy calculates both p_i and p_i^q , their q -th power, for different values of q . The Scott's rule for determining the q -order bin size B_q is expressed as follows:

$$B_q = (24\sqrt{\pi})^{1/3} \frac{\sqrt{q}}{\sqrt[6]{2q-1}} \left(\frac{\sum_{k=1}^m \sigma_{s_k}^{2(1-q)} / N_{s_k}}{\sum_{k=1}^m \sigma_{s_k}^{-(1+2q)}} \right)^{1/3},$$

where $q = 0, 1, 2, 3, 4$ and scale $s_k = 2^k, k = 1, 2, 3, \dots, m = \text{floor}(\log N)$. In practice, the theoretical standard deviation σ is replaced with the empirical standard deviation $\hat{\sigma}$ to get

$$\hat{B}_q = (24\sqrt{\pi})^{1/3} \frac{\sqrt{q}}{\sqrt[6]{2q-1}} \left(\frac{\sum_{k=1}^m \hat{\sigma}_{s_k}^{2(1-q)} / N_{s_k}}{\sum_{k=1}^m \hat{\sigma}_{s_k}^{-(1+2q)}} \right)^{1/3} \quad (2.26)$$

$$\equiv (24\sqrt{\pi})^{1/3} \frac{\sqrt{q}}{\sqrt[6]{2q-1}} N_{q,m}^{\hat{\sigma}}$$

where

$$N_{q,m}^{\hat{\sigma}} = \left(\frac{\sum_{k=1}^m \hat{\sigma}_{s_k}^{2(1-q)} / N_{s_k}}{\sum_{k=1}^m \hat{\sigma}_{s_k}^{-(1+2q)}} \right)^{1/3}$$

For Freedman-Diaconis' rule, the estimated standard deviation is replaced by interquartile ranges (IQR), to obtain

$$\hat{B}_q = 2.6 \frac{\sqrt{q}}{\sqrt[6]{2q-1}} N_{q,m}^{IQR}. \quad (2.27)$$

Please see [22] for more details.

To study the multifractal scaling characteristics of a time series, we employ the q -order entropy, a set of Shannon and Rényi entropies given by

$$S_q(t) = \begin{cases} - \sum_{i=1}^{B_q(t)} p(i,t) \ln[p(i,t)], & q = 1 \\ \frac{1}{1-q} \ln \sum_{i=1}^{B_q(t)} p^q(i,t), & q \neq 1, q \in \mathbb{R}^+ \end{cases}. \quad (2.28)$$

where $q \in \mathbb{R}^+$ denotes the weight of different probabilities of a particle falling in a bin. We restrict $q \geq 0$ due to the fact that information extraction is compromised for $q < 0$ [44]. This method is usually known as the multifractal diffusion entropy analysis (MFDEA). The DEA relate to the q -order DEA where $q = 1$. We express the linear-log relationship between the q -order diffusion entropy $S_q(t)$ and time t as

$$S_q(t) = A + \delta_q \ln(t). \quad (2.29)$$

2.4 Multi Scale Conditional Diffusion Entropy (MS-CDE)

The Multi Scale Conditional Diffusion Entropy (q -order MS-CDE) is an extension of the CDE that distinguishes between bear and bull markets at different probabilities (q -weights)

of a particle being placed in a bin. It is defined as follows.

$$C_q(t) = I + \alpha [I - \bar{S}_q(t)], \quad (2.30)$$

where I is a vector of ones and $\bar{S}_q(t)$ is the q -order vector of normalized entropy as defined in equation (2.28) and

$$\alpha = \begin{cases} -1, & p < q \\ 0, & p = q \\ 1, & p > q \end{cases} . \quad (2.31)$$

2.5 Truncated Lévy Flight (TLF)

As reported in [45], a stable Lévy process, also known as the Lévy flight model, has the property of independent increments but is still considered a long memory process. Lévy and Khintchine [46] solved the issue of defining a universal form for stable distributions and concluded that the most complete representation is through characteristic functions $\varphi(q)$, which are defined in the equation:

$$\ln(\varphi(q)) = i\mu q - \gamma |q|^\alpha \left[1 - i\beta \frac{q}{|q|} \tan\left(\frac{\pi\alpha}{2}\right) \right] \quad (2.32)$$

if $\alpha \neq 1$, and

$$\ln(\varphi(q)) = \mu q - \gamma |q| \left[1 + i\beta \frac{q}{|q|} \frac{2}{\pi} \log(q) \right] \quad (2.33)$$

If $\alpha = 1$, where α is referred to as the scaling exponent or the characteristic parameter and it has a value between 0 and 2 ($0 < \alpha \leq 2$), γ is a positive value known as the scale factor, μ is a real value known as the location parameter, and β , which ranges from -1 to 1, is known as the skewness parameter. The mean is represented by μ and the variance, σ^2 , is equal to γ^2 .

We study the symmetric stable Lévy distribution, where $\beta = 0$ and $\mu = 0$, and a limited number of variables of α and β are known to determine its analytical form. The

characteristic function for this particular scenario is:

$$\varphi(q) = \exp(-\gamma |q|^\alpha) \quad (2.34)$$

The stable distribution of index α and scale factor γ is determined by the Fourier transform, which is also known as the characteristic function of the distribution.

$$P_L(x) = \frac{1}{\pi} \int_0^\infty \exp(-\gamma |q|^\alpha) \cos(qx) dq \quad (2.35)$$

One way to tackle the issue of infinite second moment in stable Lévy processes with $\alpha < 2$ is to consider using a Truncated Lévy Flight (TLF) process with finite variance. The TLF process adheres to scale relationships and was introduced by Mantegna and Stanley in 1994. The TLF distribution is defined as follows:

$$T(x) = cP(x)\chi_{(-l,l)}(x), \quad (2.36)$$

with $P(x)$ a symmetric Lévy distribution. The TLF distribution is characterized by its unstable nature and limited variability. The speed of convergence may be impacted by the value of the cut-off length l . If l is low, the convergence will be quicker, but the cut in the tails will be more pronounced. In 1995, Koponen proposed using a decreasing exponential cut-off function with a parameter l in order to smooth out the tails and make them continuous.

The characteristic function of this distribution is defined as:

$$\varphi(q) = \exp \left[c_0 - c_1 \frac{(q^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l |q|)) \right] \quad (2.37)$$

with scale factors:

$$c_1 = \frac{2\pi \cos(\pi\alpha/2)}{\alpha\Gamma(\alpha) \sin(\pi\alpha)} At \quad (2.38)$$

and

$$c_0 = \frac{l^{-\alpha}}{\cos(\pi\alpha/2)} c_1 = \frac{2\pi}{\alpha\Gamma(\alpha) \sin(\pi\alpha)} Al^{-\alpha} t \quad (2.39)$$

By dividing time into intervals of size Δt , we obtain a total time of $T = N\Delta t$, requiring us to add up N randomly distributed variables, each of which is independent and has the

same distribution, at each time step. The distribution of the result is likely to resemble the stable Lévy distribution for small values of N , and standardizing the model may improve it if the variance is

$$\sigma^2 = -\frac{\partial^2 \varphi(q)}{\partial q^2} \Big|_{q=0} . \quad (2.40)$$

and

$$-\frac{\partial^2 \varphi(q/\sigma)}{\partial q^2} \Big|_{q=0} = -\frac{1}{\sigma^2} \frac{\partial^2 \varphi(q)}{\partial q^2} \Big|_{q=0} = 1 \quad (2.41)$$

Then, the standardized model becomes

$$\ln \varphi_s(q) = \ln \varphi \left(\frac{q}{\sigma} \right) = c_0 - c_1 \frac{((q/\sigma)^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos \left(\alpha \arctan \left(l \frac{|q|}{\sigma} \right) \right) \quad (2.42)$$

$$= \frac{2\pi A l^{-\alpha} t}{\alpha \Gamma(\alpha) \sin(\pi\alpha)} \left[1 - ((q/\sigma)^2 + 1)^{\alpha/2} \cos \left(\alpha \arctan \left(\frac{ql}{\sigma} \right) \right) \right]. \quad (2.43)$$

In this study, the equation presented represents a normalized Lévy model. The numerical simulations were carried out by adjusting the values of A , l (arbitrary scale parameter), and α (characteristic exponent) to obtain the best fit for the cumulative function. For additional details, refer to the references [17, 45, 65, 66].

Property: Stable distributions

For a stochastic process X , we can claim that it is stable, or that it has a stable distribution, if we express its probability density function as

$$Law(X_1 + X_2 + \dots + X_n) = Law(C_n X + D_n) \quad (2.44)$$

for any $n \geq 2$ where $C_n \in \mathbb{R}^+$, $D_n \in \mathbb{R}$, and $Law(X)$ denotes pdf of X . For instance, $Law(X) = N(\mu, \sigma^2)$ will denote the Gaussian random variable. Note that, X_1, X_2, \dots, X_n are

independent random copies of X , which implies that $Law(X_i) = Law(X)$ for $i = 1, 2, \dots, n$. Samorodnitsky and Taqqu (1994) have shown that X is strictly stable when $D_n = 0$ and

$$C_n = n^{1/\alpha}, \quad 0 < \alpha \leq 2 \quad (2.45)$$

We rewrite equation (2.45) as

$$\log C_n = \frac{1}{\alpha} \log n \quad (2.46)$$

using logarithms.

2.6 Relationship Between VSM and TLF

2.6.1 DFA and TLF

The equation (2.4) shows that there is a linear relationship in log-log space between the root mean square fluctuation $F(n)$ and the box size n .

$$F(n) \propto n^H \quad (2.47)$$

$$\log F(n) = \log K + H \log(n), \quad (2.48)$$

where K is a positive constant of proportionality.

$$\log F(n) - \log K = H \log(n) \quad (2.49)$$

Equations (2.49) - (2.46) gives:

$$H = 1/\alpha$$

if

$$\lim_{n \rightarrow \infty} \left| \frac{\log F(n) - \log K - \log C_n}{\log n} \right| \rightarrow 0, \quad (2.50)$$

for strictly stable variable X .

2.6.2 Q-DFA and TLF

The scaling behavior of fluctuations from the slopes $H(q)$ of the q – order DFA generates the relationship [3]

$$\mathbb{F}_q(S) \propto S^{H(q)} \quad (2.51)$$

$$\log [\mathbb{F}_q(S)] = \log K + H(q) \log (S), \quad (2.52)$$

K is a positive constant of proportionality.

$$\log [\mathbb{F}_q(S)] - \log K = H(q) \log (S) \quad (2.53)$$

Equations (2.53) - (2.46) gives:

$$H(q) = 1/\alpha$$

if

$$\lim_{n \rightarrow \infty} \left| \frac{\log F_q(S) - \log K - \log C_n}{\log S} \right| \rightarrow 0, \quad (2.54)$$

for strictly stable variable X by assuming that n equals the segment size S and $q = 2$.

2.7 Relationship Between PSM and TLF

2.7.1 DEA and TLF

We can calculate the Shannon entropy [45, 65] using the pdf in equation (2.12) as

$$S(t) = - \int_{-\infty}^{\infty} \rho(x, t) \ln [\rho(x, t)] dx. \quad (2.55)$$

Note that, we already derived the linear-log relationship between diffusion entropy $S(t)$ and t as [1, 3]

$$S(t) = \delta \ln(t) + A \quad (2.56)$$

in equation (2.21), where $A = S(1)$, Shannon entropy at time $t = 1$. Equation (2.56) becomes

$$S(t) - S(1) = \delta \ln(t) \quad (2.57)$$

From **property 1**, for some $C_n \in \mathbb{R}^+$,

Equations (2.57) - (2.46) gives

$$|S(t) - S(1) - \ln(C_n)| = 0, \text{ where } t \geq 2 \quad (2.58)$$

such that

$$\delta \approx \frac{1}{\alpha}.$$

2.7.2 Q-DEA and TLF

Earlier, we presented the linear-log relationship between the time t and the q -order diffusion entropy $S_q(t)$ in the context of the Q-DEA, where $q = 1$, which is related to the DEA.

$$S_q(t) = A + \delta_q \ln(t), \quad (2.59)$$

where $A = S_q(1)$, q -order diffusion entropy at $t = 1$.

$$S_q(t) - S_q(1) = \delta_q \ln(t) \quad (2.60)$$

Using **property 1**, for $C_n \in \mathbb{R}^+$,

Equations (2.60) - (2.46) gives

$$|S_q(t) - S_q(1) - \ln(C_n)| = 0, \text{ where } t \geq 2 \quad (2.61)$$

and $q = 1$ such that

$$\delta_q \approx \frac{1}{\alpha}.$$

2.8 Cantor Detrended Fluctuation Analysis (CDFA)

Here, we show that the Hurst exponents' subspace $[H_{min}, H_{max}]$ is homeomorphic to the Cantor set's $[0,1]$. We also show an illustration of the Cantor set and the CDFA method.

Theorem 1. *A map $f : [H_{min}, H_{max}] \rightarrow [0, 1]$ between the topological spaces of Hurst exponents of noise-like time series and the Cantor set is a homeomorphism if it has the following properties:*

1. f is a bijection ;
2. f is continuous;
3. inverse function f^{-1} is continuous.

We say two topological spaces are homeomorphic if they admit a homeomorphism between them: they are fundamentally the same topological space.

Proof of Theorem: Let $H_{min} \leq H \leq H_{max}$ and $0 \leq y = f(H) \leq 1$, then the map $f : [H_{min}, H_{max}] \rightarrow [0, 1]$ gives

$$H_{min} - H_{min} \leq H - H_{min} \leq H_{max} - H_{min} \tag{2.62}$$

$$0 \leq \frac{H - H_{min}}{H_{max} - H_{min}} \leq 1. \tag{2.63}$$

Thus,

$$y = f(H) = \frac{H - H_{min}}{H_{max} - H_{min}}. \tag{2.64}$$

We must now prove that the map f is homeomorphic to the Cantor set.

1. The map $f(H)$ is said to be bijective iff $f(a) = f(b)$ for all a, b implies that $a = b$.

From

$$\begin{aligned} f(a) &= \frac{a - H_{min}}{H_{max} - H_{min}} \quad \text{and} \quad f(b) = \frac{b - H_{min}}{H_{max} - H_{min}}, \\ f(a) &= f(b) \\ \implies a - H_{min} &= b - H_{min} \\ \implies a &= b. \end{aligned}$$

Therefore, the map $f(H)$ is a bijection.

2. The map $f(H)$ is continuous at some value c in its domain if

- $f(c)$ is defined,
- limit of f as H approaches c exists, and
- the function value of f at c equals the limit of f as H approaches c .

$f(c)$ is defined as

$$f(c) = \frac{c - H_{min}}{H_{max} - H_{min}}. \quad (2.65)$$

The limit of f as H approaches c gives

$$\lim_{H \rightarrow c^+} f(H) = \lim_{H \rightarrow c^-} f(H) = \frac{c - H_{min}}{H_{max} - H_{min}}. \quad (2.66)$$

The left and right-sided limits are equal from equation (2.66). Therefore,

$$\lim_{H \rightarrow c} f(H) = \frac{c - H_{min}}{H_{max} - H_{min}}. \quad (2.67)$$

Equations (2.65) and (2.67) are the same. Hence, it follows that

$$\lim_{H \rightarrow c} f(H) = f(c) = \frac{c - H_{min}}{H_{max} - H_{min}}.$$

As a result, for a differentiable fractal, the map f is continuous at some value $H = c$.

3. The inverse function of f (i.e. $f^{-1}(H)$) exists.

$$y = f(H) = \frac{H - H_{min}}{H_{max} - H_{min}} \quad (2.68)$$

$$(H_{max} - H_{min})y = H - H_{min} \quad (2.69)$$

$$H = H_{min} + (H_{max} - H_{min})y \quad (2.70)$$

Interchanging H and y gives

$$y = f^{-1}(H) = H_{min} + (H_{max} - H_{min})H, \quad (2.71)$$

the inverse function of $f(H)$.

The inverse map f^{-1} is continuous at some value s in its domain if $f^{-1}(s)$ is defined, the limit of f^{-1} as H approaches s exists and the function value of f^{-1} at s equals the limit of f^{-1} as H approaches s . $f^{-1}(s)$ is defined as

$$f^{-1}(s) = (1 - s)H_{min} + sH_{max}. \quad (2.72)$$

The limit of f^{-1} as H approaches s equals

$$\lim_{H \rightarrow s^+} f^{-1}(H) = \lim_{H \rightarrow s^-} f^{-1}(H) = H_{min} + (H_{max} - H_{min})s. \quad (2.73)$$

$$\Rightarrow \lim_{H \rightarrow s} f^{-1}(H) = H_{min} + (H_{max} - H_{min})s. \quad (2.74)$$

Equation (2.72) = (2.74) implies that,

$$\lim_{H \rightarrow s} f^{-1}(H) = f^{-1}(s) = (1 - s)H_{min} + sH_{max}.$$

The inverse map f^{-1} exists and is continuous at some value $H = s$.

Therefore, the map $f(H)$ is a homeomorphism. For noise-like time series, $H \in [H_{min}, H_{max}]$ is homeomorphic to the Cantor set's $[0, 1]$. \square

2.8.1 The Cantor set

Take a time series and remove middle thirds up to many levels, what you observe is similar to the Cantor set. This phenomenon is depicted in Figure 2.1 [47]. It demonstrates that the segments seem the same at different scales in successive Cantor magnifications ranging from C_0 to C_6 . C_0 represents the initial time series with no missing parts and C_6 represents the leftover time series after deleting the middle thirds for the sixth time. We limit our scope to C_0 through C_3 , for the sake of experimentation.

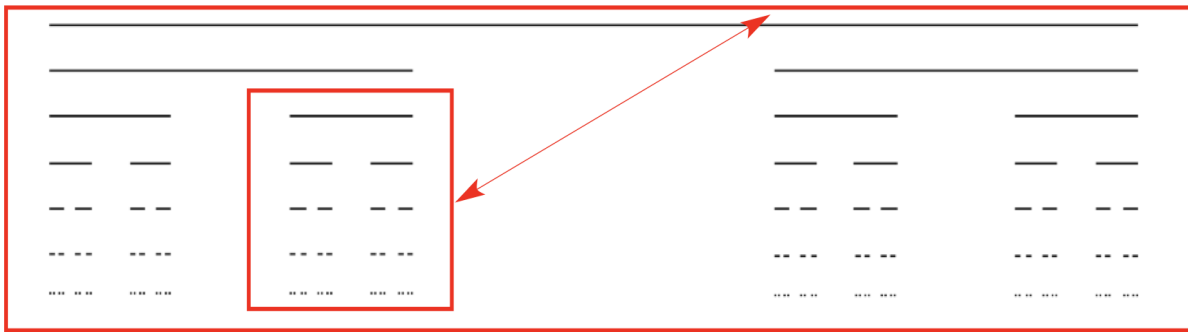


Figure 2.1: Fractal behavior of a Ternary Cantor Set.

2.8.2 Definition

The interval subset of the Cantor set is defined recursively as:

1. $C_0 = [0, 1]$;
2. $C_1 = \left(\frac{1}{3}, \frac{2}{3}\right)$;
3. $C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right)$ for $n \geq 2$.

The ternary Cantor set is defined as $C = [0, 1] \setminus (\cup_{n=1}^{\infty} C_n)$. The interval we begin with is indicated by level C_0 . For C_1 , $[0, 1]$ is divided into three sub-intervals, with the middle sub-interval $\left(\frac{1}{3}, \frac{2}{3}\right)$ deleted. For C_2 , each of the remaining intervals from C_1 is broken into

three sub-intervals, with the middle sub-intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ removed. This technique can be repeated indefinitely by deleting the open middle third sub-interval of each interval created in the preceding level. We rescale the integrated series ψ_t by dividing each data point by the maximum data point due to concerns with the dimension of the Cantor sets (i.e. dimension of 0.631 which is less than 1),

$$\Psi_t = \frac{\psi_t}{\max(\psi_t)}.$$

s.t. $\Psi_t \in [0, 1]$.

2.8.3 Algorithm of the CDFA

We present a Cantor version of the DFA method to generalize the DFA's segment division step. The CDFA algorithm is comprised of four (4) major steps.

1. Find the integrated series shifted by the mean $\langle \Psi \rangle$ given the time series Ψ_t of length N .

$$Y_j = \sum_{i=1}^j (\Psi_i - \langle \Psi \rangle).$$

2. The series Y_j is then divided into equal non-overlapping segments of varying sizes Δs . Δs is based on the Cantor set theory scale ($\Delta s = 3^n, n \geq 0$). The number of non-overlapping segments are as follows:

$$N_{\Delta s} \equiv \text{int} \left(\frac{N}{\Delta s} \right) = \text{int} \left(\frac{N}{3^n} \right).$$

The Cantor set scaling function is calculated for each segment to emphasize both slow and fast-evolving fluctuations that control the time series structure.

3. The integrated series' Root Mean Squared Fluctuation (RMSF) is calculated at several scales:

$$F(\Delta s) \equiv \left\{ \frac{1}{2N_{\Delta s}} \sum_{j=1}^{2N_{\Delta s}} [Y_j - Y_j^{\Delta s}]^2 \right\}^{1/2}.$$

j denotes the sample size of segments $N_{\Delta s}$. We compute RMSF from $j = 1$ to $2N_{\Delta s}$ rather than $N_{\Delta s}$. We add from beginning to end and from end to beginning, then compute an average of the values to consider every data point. Conversely, the large segments interweave several local periods with both minor and major fluctuations and so average out their disparities in magnitude.

4. The power-law notation is computed for multiple scales using the least square regression fit of $F(\Delta s)$ vs the Cantor scales Δs on a log-log scale

$$F(\Delta s) \propto (\Delta s)^{H^c}$$

$$\log(F(\Delta s)) = H^c \log(\Delta s) + \log(C).$$

$H^c :=$ CDFA Hurst exponent that measures memory behavior in the noise-like time series.

2.9 Parallel Framework for CDFA (PCDFA)

This section introduces the computation optimization strategy in which CDFA loop iterations are processed in a parallel pool of four processors. The CDFA algorithm's core workhorse is the use of least squares to calculate fluctuations in non-overlapping segments. As a result, fluctuations are computed across available processors (procs) utilizing parallel for-loops before being gathered by a single processor for further processing in determining the Hurst exponent [48]. We call it the PCDFA and the algorithm is depicted in the chart below.

2.9.1 Hardware & Software Environment

Experiments are conducted in a system with:

- Processor: 1.4 GHz Quad – Core Intel Core i5

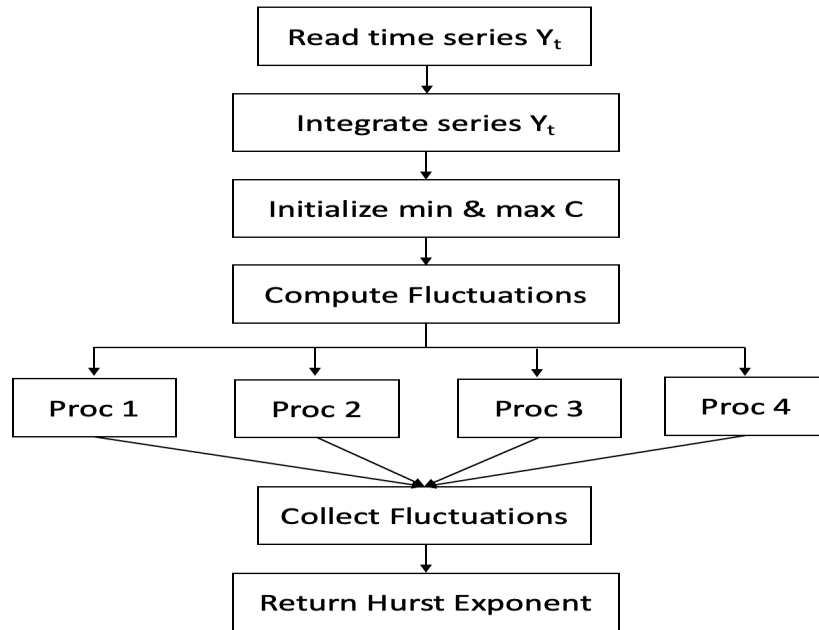


Figure 2.2: A flow chart of PCDFA for 4 available processors

- Memory: 16 GB 2133 MHz LPDDR3
- Graphics: Intel Iris Plus Graphics 645 1536 MB
- Software: MATLAB R2021b 64 – bit, Parallel Computing Toolbox
- Operating System: macOS Big Sur Version 11.6.6

Chapter 3

Data

We adopt the data provided in each subsection below for our research. This includes financial market data (stock and cryptocurrency time series), geophysical data (volcanic eruption series), and medical data (bio-medical and COVID-19 time series).

3.1 Financial Time Series

Financial market data used to analyze the relationship between DFA and TLF Hurst index and between DEA and TLF were obtained from YAHOO FINANCE. All data points collected are daily closing prices. Following is the name of the countries from which data was collected and the start and end dates of that data collection. Hong Kong (HSI) from 2 January 1991 to 25 October 2001, and the USA (S&P500) from 2 January 1991 to 25 October 2001. The evolution of prices in the time series of financial markets is shown in Figures [3.1](#) and [3.2](#).

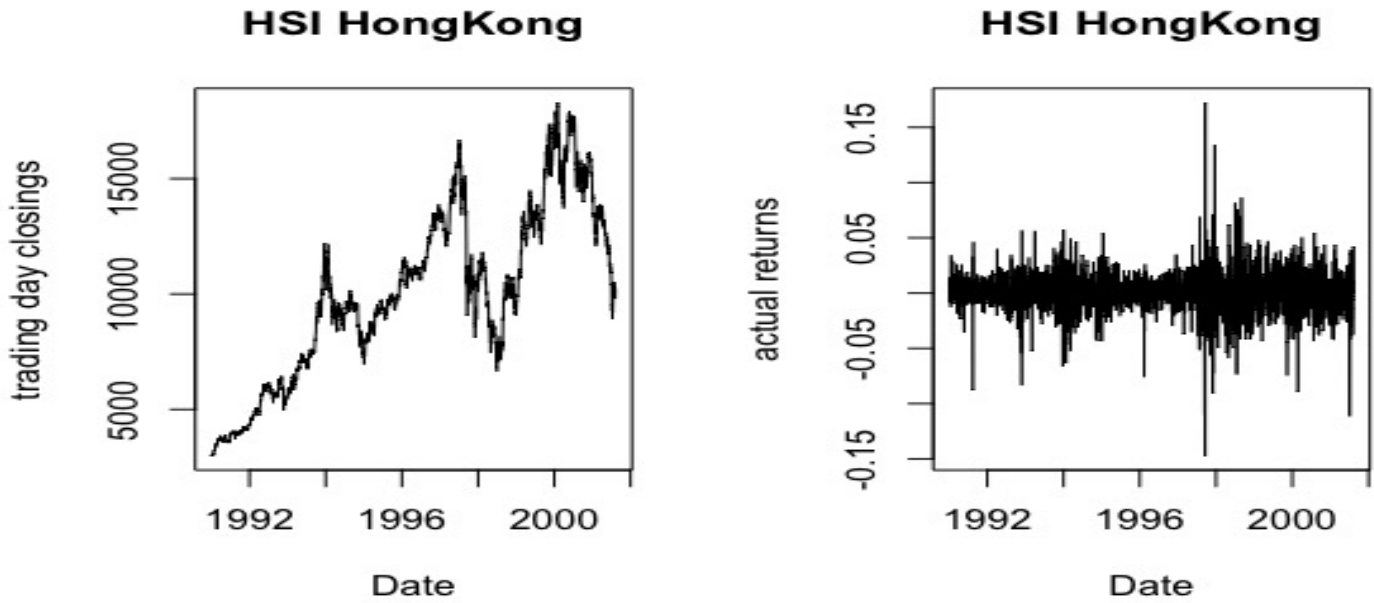


Figure 3.1: Plots of trading days closings and actual returns of HSI

We also use the daily close values of the Dow Jones Industrial Average (DJI) from the NY Stock Exchange and Bitcoin (BTC) from the cryptocurrency market for investigating the PCDFa technique. The table below shows the start and end dates of the data used.

Table 3.1: Financial Market Data for Analysis.

Data	Start Date	End Date
<i>DJI</i>	09/30/2013	04/19/2021
<i>BTC</i>	09/30/2013	04/19/2021

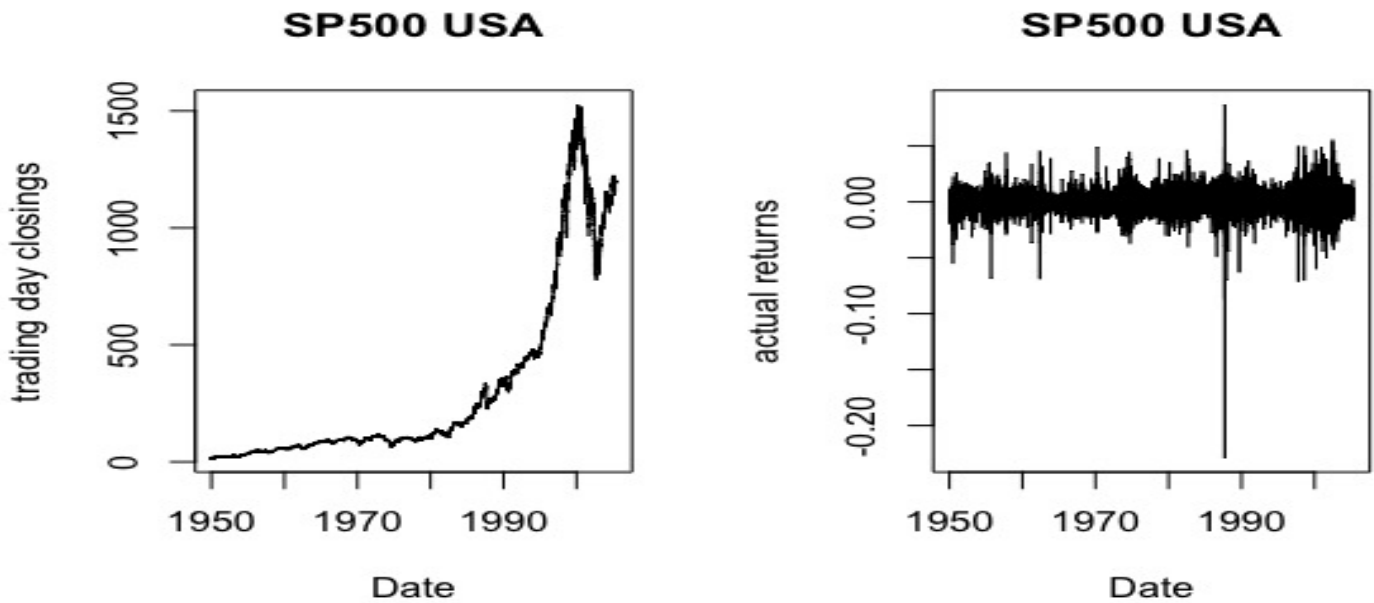


Figure 3.2: Plots of trading days closings and actual returns of SP500

3.2 Geophysical Time Series

The seismic network of the Bezymyanny Volcano Campaign (PIRE) collected volcanic data at different times of the eruption from two other seismic stations, BEZB and BELO. The data used in this article were requested 10 days before and 5 days after the eruption. Volcanic eruptions 1 and 2 originate from BEZB, and volcanic eruptions 7 and 8 originate from BELO. Time series graphs of volcanic eruptions are shown in Figures 3.5 and 3.6.

Dow Jones Industrial Average (DJI)

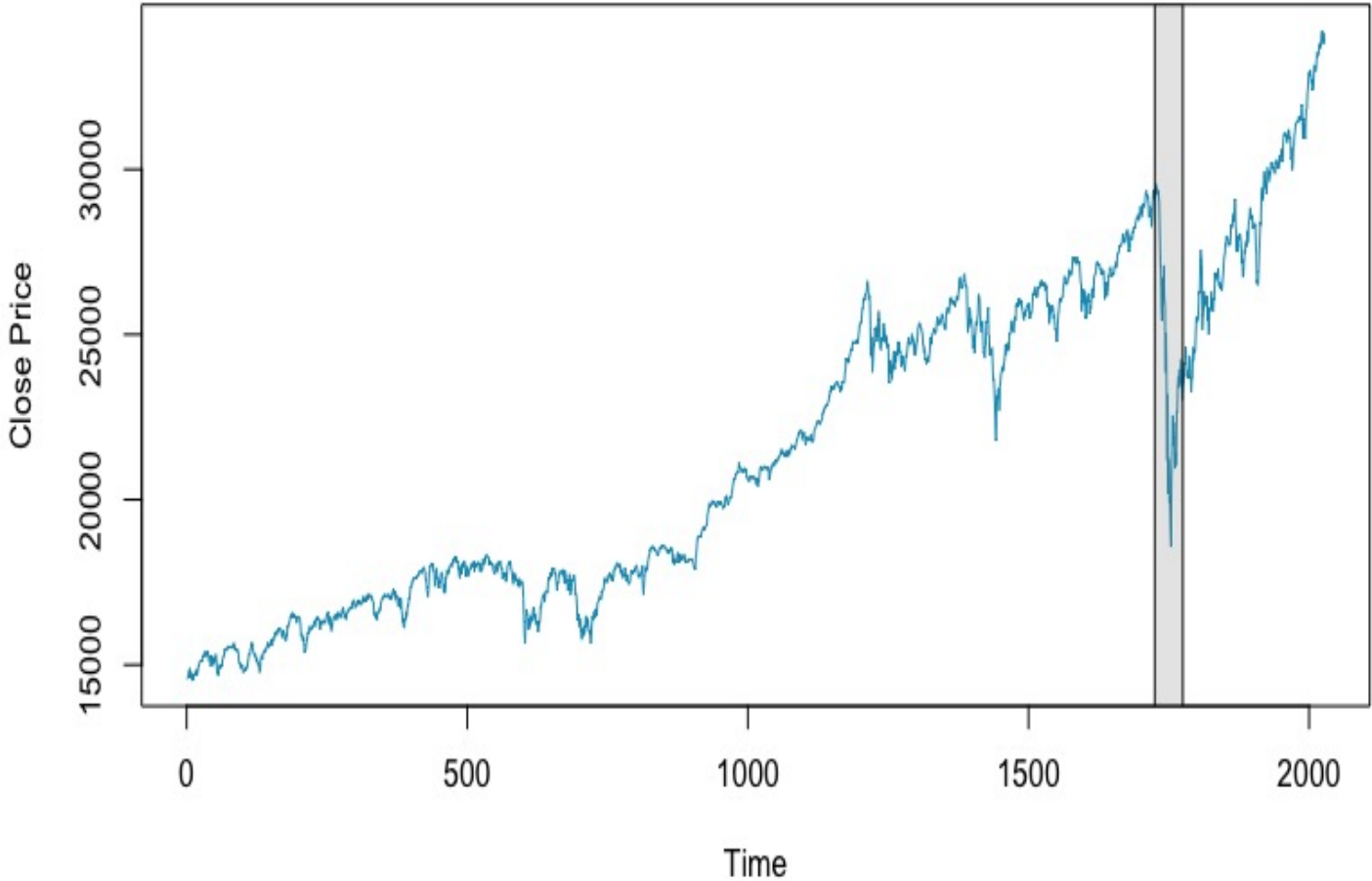


Figure 3.3: Daily Close Prices Plot of DJI

3.3 Biomedical Time Series

The white noise (top panel), monofractal (middle panel), and multifractal (bottom panel) time series used in the experiments for CDFA and PCDFa are noise-like biomedical time series with 8000 scaled sample data points each. White noise time series has a time-independent structure with Hurst exponent close to $H = 0.5$, whereas monofractal and

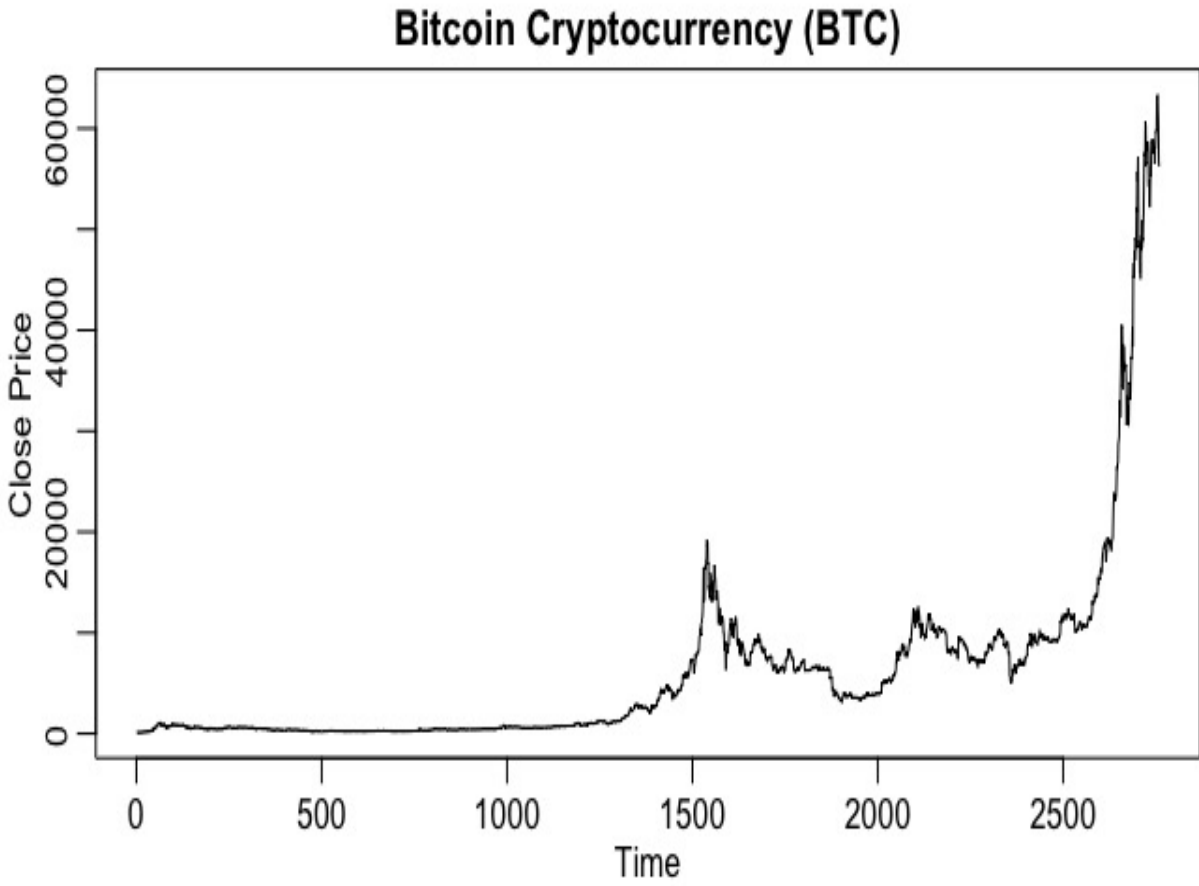


Figure 3.4: Daily Close Prices Plot of BTC

multifractal time series have persistent behavior such that $0.5 < H \leq 1$.

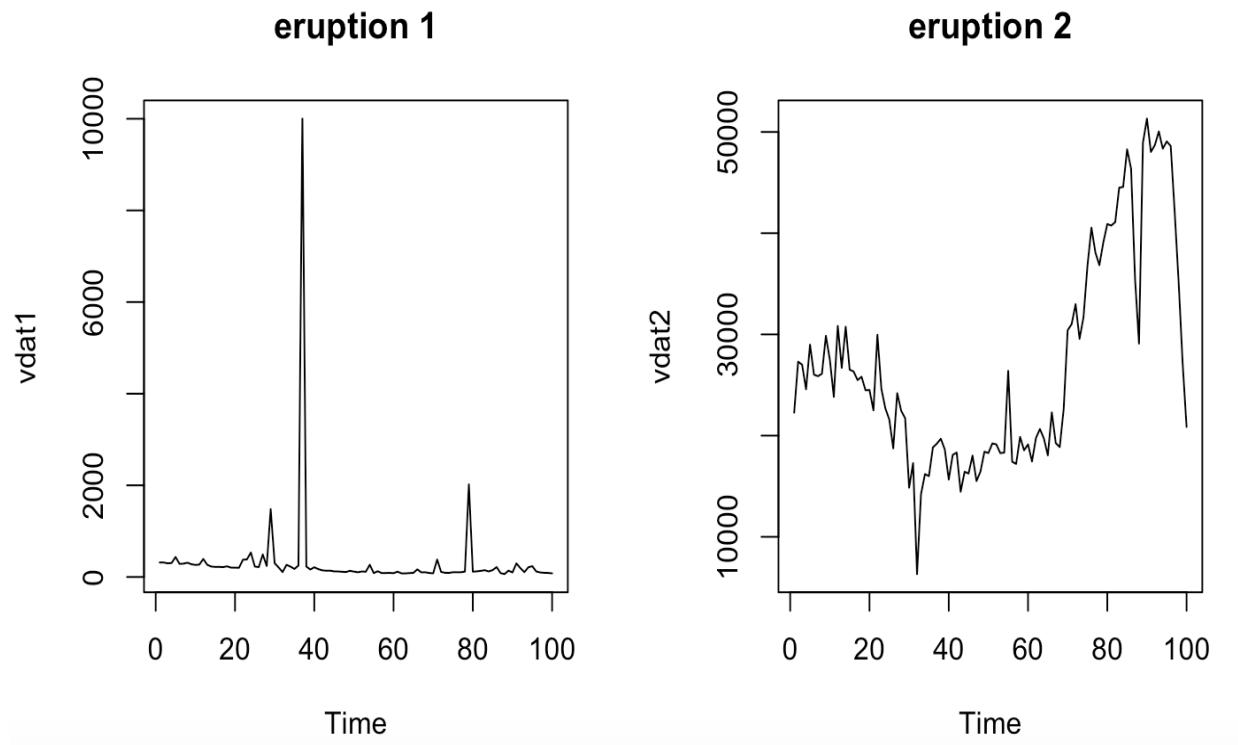


Figure 3.5: Volcanic Eruption 1 and 2 Plots

3.4 Covid-19 Time Series

The time series used here were collected daily from January 21, 2020 to April 19, 2021, via the New York Times (NYT) US Coronavirus Cases and Deaths GitHub repository and had 422 observations. Due to the extreme fluctuations in the time series, we analyze the following:

$$D_t = \log(C_t) - \log(C_{t-1}), \quad (3.1)$$

where D_t denotes the log-difference of daily cumulative cases (C_t) of Covid-19 on day t . Figure 3.8 provides a plot of the series.

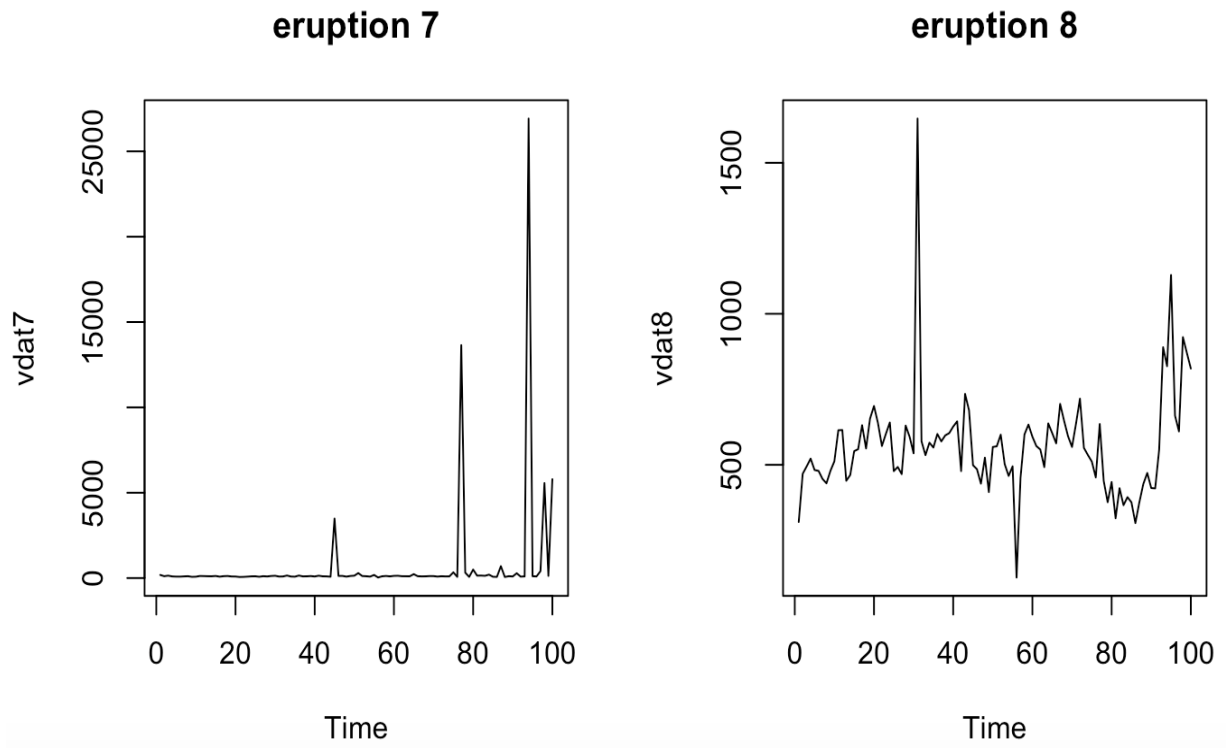


Figure 3.6: Volcanic Eruption 7 and 8 Plots

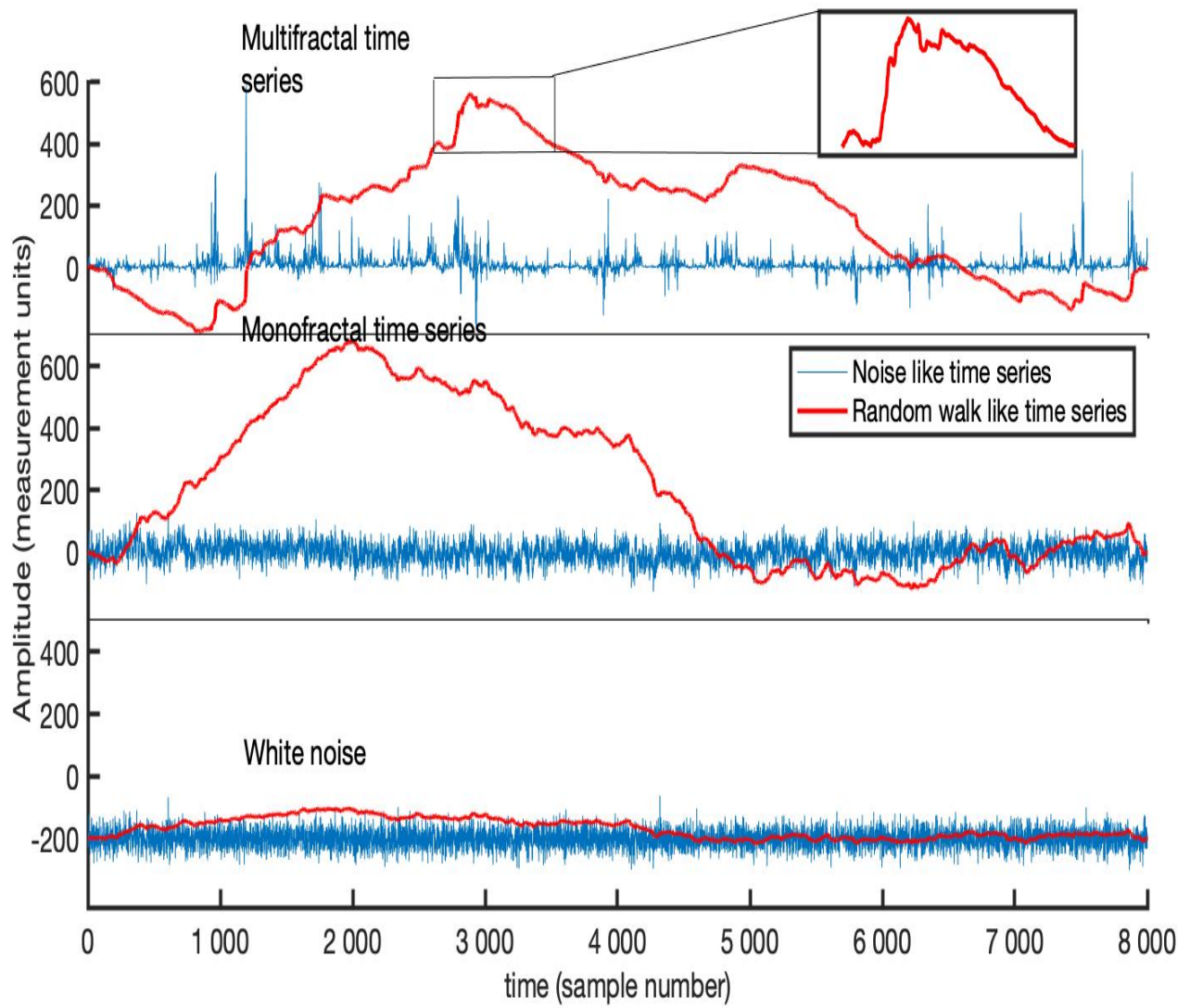


Figure 3.7: Biomedical time series plots

Daily Covid-19 Cases in the US: 01/21/2020 - 04/19/2021

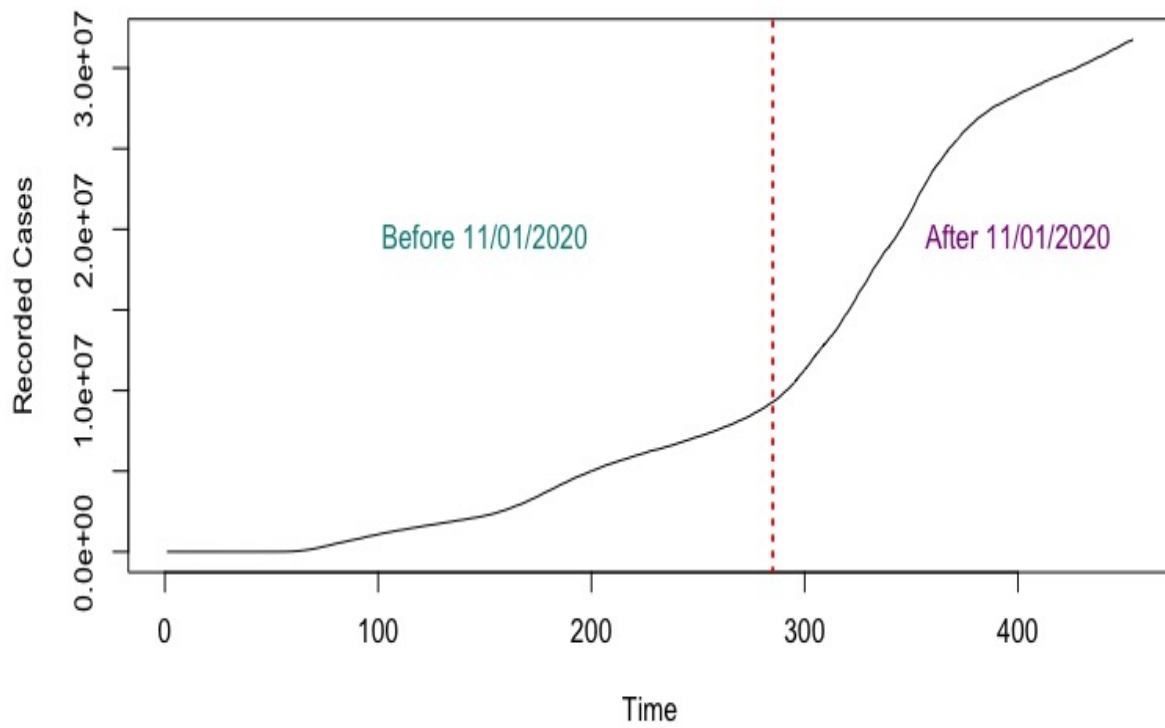


Figure 3.8: Evolution of daily US cumulative and log-difference of Covid-19 cases

Chapter 4

Applications

4.1 Relationships Between DFA, DEA, and TLF

In this section, we present numerical results by fitting all three models, namely DFA, DEA and TLF, to the data. Based on the numerical results, we propose a relationship between the self-similar models and the Lévy truncated flight model, which confirms the analytically derived relationship. Below is a table of financial market data and volcanic eruption data. Each table has 6 columns; the first column represents a specific financial market or some volcanic eruptions, the second through fourth columns represent the DFA and DEA scale indicators and the best Lévy parameter α . The last three columns are essentially the product of their scaling factor and the Lévy parameter α that characterizes the data.

Table 4.1: Scaling exponents of various financial markets

Stock Index	DFA (H)	DEA (δ)	Lévy (α)	$H.\alpha$	$\delta.\alpha$
HSI	0.70	0.60	1.40	0.98	0.84
SP500	0.66	0.65	1.40	0.92	0.91

The empirical data presented in Tables 4.1 and 4.2 show an approximately inverse relationship between the DEA parameter δ and the Lévy parameter α . Meanwhile, for a very inconsistent time series from volcanic eruption 1, it is not close to 1.

$$\delta \approx \frac{1}{\alpha} \tag{4.1}$$

Also, there is an inverse relationship between the DFA parameter H and the resulting Lévy

Table 4.2: Scaling exponents of Volcanic Data from the two seismic stations

Seismic Station	Eruption	DFA (H)	DEA(δ)	Lévy (α)	$H.\alpha$	$\delta.\alpha$
BEZB	1	0.74	0.68	1.12	0.83	0.76
	2	0.92	0.68	1.34	1.23	0.91
BELO	7	0.81	0.68	1.40	1.13	0.95
	8	0.75	0.68	1.34	1.01	0.91

parameter α .

$$H \approx \frac{1}{\alpha} \quad (4.2)$$

NOTE: The DFA is susceptible to overestimating the Hursts exponent (H), as seen in Table 4.2. Observe that the product of H and α sometimes is greater than 1. It was this problem that required the CDFA method.

Plots obtained from the numerical simulation of the HSI, S&P500, and Volcanic Eruption time series data are shown in Figures 5.1 - 5.12 in Appendix A.

4.2 CDFA Applied to Biomedical Signals

Here is the resulting table of Hurst exponents for implementing of the DFA algorithm for the respective time series.

Table 4.3: DFA's Hurst Exponents for White noise time series

<i>Levels</i>	<i>Hurst Exponents</i>
C_0	$H_1 = 0.50$
C_1	$H_1 = 0.50, H_2 = 0.45$
C_2	$H_1 = 0.54, H_2 = 0.45, H_3 = 0.52, H_4 = 0.42$
C_3	$H_1 = 0.50, H_2 = 0.54, H_3 = 0.4, H_4 = 0.49, H_5 = 0.59, H_6 = 0.43,$ $H_7 = 0.42, H_8 = 0.57$

In table 4.3, we observe that the Hurst exponent of the white noise series is close to $H = 0.5$ for all levels C_0, C_1, C_2 and C_3 . This confirms the phenomena appearing in Cantor's fractal properties set in the white noise time series.

Table 4.4: DFA's Hurst Exponents for Monofractal time series

<i>Levels</i>	<i>Hurst Exponents</i>
C_0	$H_1 = 0.79$
C_1	$H_1 = 0.80, H_2 = 0.68$
C_2	$H_1 = 0.81, H_2 = 0.69, H_3 = 0.74, H_4 = 0.68$
C_3	$H_1 = 0.65, H_2 = 0.80, H_3 = 0.63, H_4 = 0.72, H_5 = 0.78, H_6 = 0.68,$ $H_7 = 0.67, H_8 = 0.79$

Table 4.4 presents Hurst indices between 0.5 and 1 ($0.5 < H \leq 1$) for long-memory monofractal time series for levels C_0, C_1, C_2 and C_3 . The phenomena seen in the monofractal time series in the table above are similar to the fractal properties of the Cantor set.

The Hurst exponent of the multifractal time series is in the range $0.5 < H \leq 1$ for all levels C_0, C_1, C_2 and C_3 of Table 4.5 describing the fractal phenomena described by the

Table 4.5: DFA's Hurst Exponents for Multifractal time series

<i>Levels</i>	<i>Hurst Exponents</i>
C_0	$H_1 = 0.86$
C_1	$H_1 = 0.86, H_2 = 0.75$
C_2	$H_1 = 0.75, H_2 = 0.88, H_3 = 0.70, H_4 = 0.78$
C_3	$H_1 = 0.82, H_2 = 0.69, H_3 = 0.77, H_4 = 0.97, H_5 = 0.69, H_6 = 0.70,$ $H_7 = 0.91, H_8 = 0.90$

Cantor set. The results from Tables 4.3 - 4.5 confirm that successive increases in noise-like time series show similar patterns at progressively smaller scales. This phenomenon is widely known as self-similarity in fractals.

Table 4.6: Comparison of Scaling Exponents of DFA(H) & CDFA(H^c) & TLF (α) on Noise-like Time series

<i>Time Series</i>	H	H^c	<i>Difference</i>	α	$H\alpha$	$H^c\alpha$
White noise	0.5	0.4997	0.0003	1.97	0.985	0.9844
Monofractal	0.79	0.781	0.009	1.28	1.0112	0.9997
Multifractal	0.86	0.851	0.009	1.17	1.0062	0.9976

When examining monofractal and multifractal time series using the Hurst exponents of DFA, and CDFA of white noise, we observe the difference in exponents as shown in Table 4.6. The Hurst exponent of white noise time series changes slightly, but that of the monofractal and multifractal time series changes by about 1%. The slight change in score is the result of dividing the time series by a multiple of 3 at each level using CDFA. This will help resolve overestimation and underestimation issues related to the DFA. Despite the

differences between the parameters, they still represent the same process modeled here (i.e. noise-like time series). The exponent of the white noise series is close to 0.5, while that of the monofractal and multifractal series is in the range $0.5 < H \leq 1$, which describes the behavior of long memories. The results also show that the segment size does not always need to be hard-coded into the DFA algorithm based on the length of the time series under consideration. In particular, for odd-length time series, the process can be automated using the fractal phenomenon of Cantor sets to obtain equal segment sizes and satisfactory Hurst exponents.

Also, the product of the Hurst exponents DFA and TLF, and CDFA and TLF exponent (α) suggests that H^c is a better estimate. It can be seen that $H^c\alpha \approx 1$, but for time series such as monofractal and multifractal noise, DFA produces $H\alpha > 1$. The latter does not satisfy the inverse relationship between the Hurst exponent and the TLF scaling exponent for Gaussian noise, as discussed in [3]. This also highlights the overestimation/underestimation of the Hurst index of conventional DFA approaches.

4.3 Q-CDFA Applied to Covid-19 Time Series

4.3.1 Before November 2020

Table 4.7: Q-DFA scaling exponents from Covid cases before Vaccines

$q - order$	$H(q)$	α	$f(\alpha)$
-5	0.94	1.20	-0.27
-4	0.89	1.20	-0.25
-3	0.81	1.18	-0.10
-2	0.69	1.04	0.31
-1	0.52	0.69	0.82
0	0.34	0.34	1.00
1	0.20	0.11	0.90
2	0.11	-0.02	0.73
3	0.04	-0.10	0.59
4	0.00	-0.13	0.48
5	0.00	-	-

4.3.2 After November 2020

Table 4.8: Q-DFA scaling exponents from Covid cases after Vaccines

$q - order$	$H(q)$	α	$f(\alpha)$
-5	0.81	0.95	0.31
-4	0.78	0.93	0.43
-3	0.75	0.89	0.57
-2	0.70	0.83	0.73
-1	0.63	0.74	0.89
0	0.52	0.52	1.00
1	0.36	0.22	0.86
2	0.22	0.02	0.60
3	0.12	-0.08	0.40
4	0.05	-0.13	0.27
5	0.01	-	-

Let us set the weights of the $q - order$ of the local fluctuation to lie in the $[-5, 5]$ interval. Tables 4.7 and 4.8 show the calculated values $H(q) = 0.11$ and $H(q) = 0.22$ at $q = 2$, indicating anti persistence behavior in time series paths. However, moderate changes in $H(q)$ are inversely proportional to changes in q over time. Before the introduction of the vaccines, the Hurst score fell more rapidly than after the introduction of the vaccines. This indicates that the incidence rate is slowing rather than worsening. Also, the peaks of the multifractal spectrum lie at $\alpha = 0.3$ and $\alpha = 0.52$, respectively, in a wide range from 0.2 to 1.25. This suggests that volatility clustering of series underlined with negative q tends to indicate long memory behavior, while positive q weights tend to represent short memory behavior. We observe that the case's multifractal spectrum is not symmetric until November 2020. It is asymmetrical and has a long tail. It represents a multifractal structure with high sensitivity to large and small fluctuations in the cases. In contrast, the multifractal spectrum of post-November 2020 cases has a close-to-symmetrical shape,

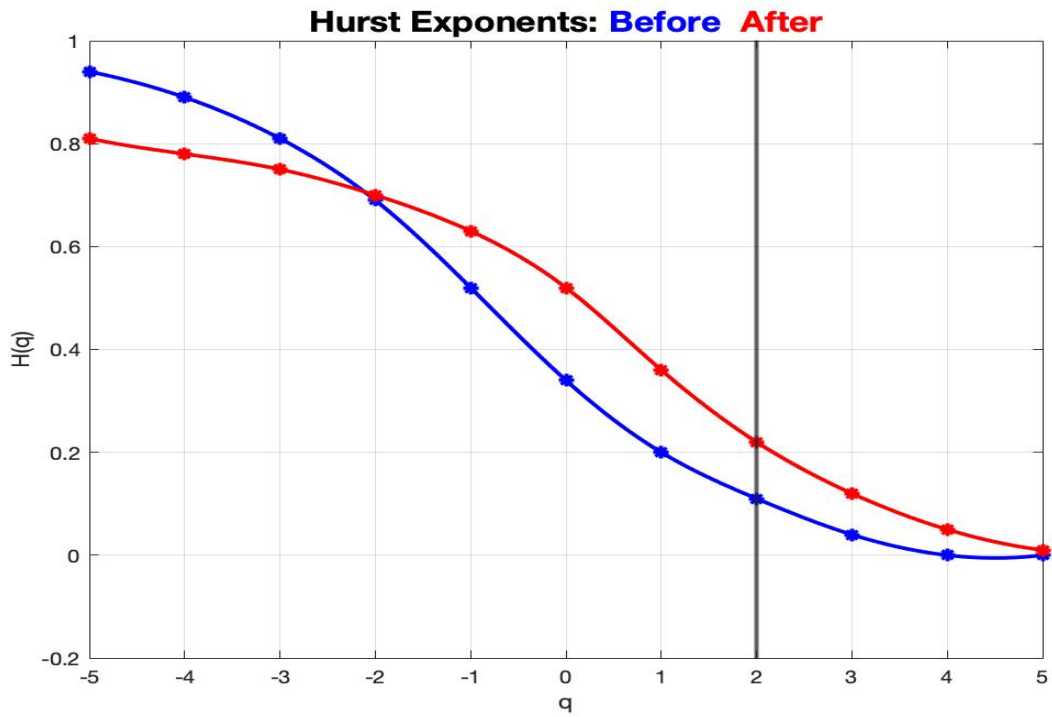


Figure 4.1: Plot of q -order Hurst exponents Before and After the Vaccines

indicating moderate sensitivity to both small and large fluctuations in the US Covid-19 case time series.

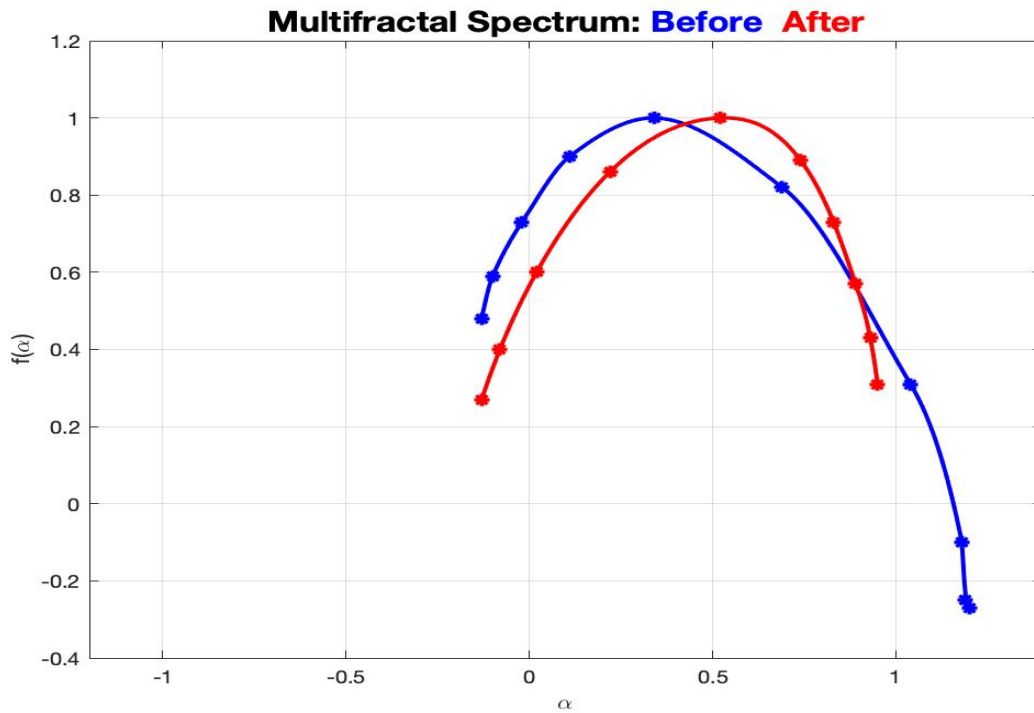


Figure 4.2: Plot of q-order singularity spectrum Before and After the Vaccines

4.4 PC DFA Applied to Financial Time Series

4.4.1 Results and Discussions

Table 4.9: PC DFA Time Analysis of Dow Jones Industrial Average (in seconds)

Data	Data Points (N)	1 Proc	2 Procs	3 Procs	4 Procs
<i>DJI daily</i>	1901	0.05	0.04	0.03	0.04
<i>DJI 1 hour</i>	15630	0.22	0.16	0.08	0.06
<i>DJI 30 mins</i>	28276	0.37	0.21	0.13	0.10
<i>DJI 5 mins</i>	153168	2.15	1.19	0.75	0.57

The execution time of the CDFA's sequential code grows by a larger margin as the data

Table 4.10: PCDFA Time Analysis of Bitcoin Data (in seconds)

Data	Data Points (N)	1 Proc	2 Procs	3 Procs	4 Procs
<i>BTC daily</i>	2759	0.07	0.04	0.03	0.03
<i>BTC 1 hour</i>	65994	0.89	0.51	0.33	0.26
<i>BTC 30 mins</i>	131892	1.93	0.99	0.69	0.51
<i>BTC 5 mins</i>	772339	10.24	5.23	3.63	2.65

amount increases. The execution time of the CDFA parallel code (PCDFA), on the other hand, grows by a lesser margin as the data size grows. Tables (4.9) and (4.10) demonstrate the PCDFA’s scalability with varying numbers of processors and data points. It is worth noting that as more processors are used, the CDFA’s execution time on each data set lowers. It is also obvious that PCDFA performs best for scalability on ”large data” with 772K data points. In other words, the more data there is, the greater the PCDFA’s scaling performance.

In the analysis of the parallel implementation of the CDFA, we assume the following variables:

- p : Number of processors
- T_1 : Execution time of the sequential algorithm
- T_p : Execution time of the parallel algorithm with p processors.

4.4.2 Speedup Ratio of PCDFFA

The speedup ratio of the PCDFFA is calculated as the ratio of the execution time of the sequential algorithm of the CDFA to the execution time of the parallel algorithm of the CDFA with p processors. Speedup ratio may be less than, greater than or equal to the

number of processors p .

$$\text{Speedup Ratio, } S_p = T_1/T_p$$

- $S_p < p$: Sub-linear Speedup: Common
- $S_p = p$: Linear Speedup (Amdahl's law): Ideal
- $S_p > p$: Super-linear Speedup: Wonderful

Table 4.11: PCDFA Speedup Ratios of Low- & High-Frequency DJI & BTC Data

Data	S_1	S_2	S_3	S_4
<i>DJI daily</i>	1	1.25	1.67	1.25
<i>DJI 1 hour</i>	1	1.38	2.75	3.67
<i>DJI 30 mins</i>	1	1.76	2.85	3.70
<i>DJI 5 mins</i>	1	1.81	2.87	3.77
<i>BTC daily</i>	1	1.75	2.33	2.34
<i>BTC 1 hour</i>	1	1.75	2.70	3.42
<i>BTC 30 mins</i>	1	1.94	2.80	3.78
<i>BTC 5 mins</i>	1	1.96	2.82	3.86

For the DJI, when processing 1901 data points of the daily close values, 2 processors generated 1.25 times speedup whereas 4 processors also generated the same 1.25 times speedup. However, for a high-frequency DJI 5-mins data with 153168 data points, 2 processors generated 1.81 times speedup, and 4 processors generated 3.77 times speedup.

A similar relation is observed for the BTC data. Processing 2759 data points of the daily close values using 2 processors generated 1.75 times speedup, whereas 4 processors generated 2.34 times speedup. For the high-frequency 5-mins BTC data with 772339 data

samples, 2 processors generated 1.96 times speedup whereas 4 processors generated 3.86 times speedup.

The computation shows that the speedup ratio of the two largest data with 153K and 772K samples, respectively, is "asymptotic" to the ideal speedup with increased processors. Figures (5.13) and (5.14) show that the speedup ratios from both DJI and BTC data go close but do not achieve the ideal speedup (linear speedup or Amdahl's law) [49]. We observe a sub-linear speedup ratio as the number of processors increases.

4.4.3 Parallel Efficiency of PCDFA

The efficiency of the PCDFA is calculated as the ratio of the speedup, S_p to the number of processors p . Efficiency measures the fraction of time for which a processor is usefully utilized in computing Hurst's exponent of the CDFA.

$$\text{Parallel Efficiency, } E_p = S_p/p = T_1/pT_p, \quad 0 \leq E_p \leq 1. \tag{4.3}$$

Table 4.12: PCDFA Parallel Efficiency Analysis of DJI & BTC

Data	E_1	E_2	E_3	E_4
<i>DJI daily</i>	1	0.63	0.56	0.31
<i>DJI 1 hour</i>	1	0.69	0.92	0.92
<i>DJI 30 mins</i>	1	0.88	0.95	0.93
<i>DJI 5 mins</i>	1	0.91	0.96	0.94
<i>BTC daily</i>	1	0.88	0.78	0.59
<i>BTC 1 hour</i>	1	0.88	0.90	0.86
<i>BTC 30 mins</i>	1	0.97	0.93	0.95
<i>BTC 5 mins</i>	1	0.98	0.94	0.97

For the DJI daily recorded data, 2 processors produce 63% efficiency while 4 processors generate 31% efficiency. However, for the high-frequency 5-min-by-min DJI data, 2 processors produce 91% efficiency while 4 processors generate 94% efficiency.

The daily BTC data show a similar relationship where 2 processors generate 88% efficiency to 59% efficiency resulting from 4 processors. For the high-frequency 5-min-by-min BTC data, we generate 98% efficiency from 2 processors compared to 97% efficiency from 4 processors.

We can see from the results of the computation in Table (4.12), Figures (5.15), and (5.16) that the efficiency of the PCDFA method increases when there is enough computation to undertake (i.e. big data). Small data renders some of the processors idle, thus, resulting in low efficiency as more and more processors are employed. Parallel optimization of the CDFA is essential and efficient when there is more computation. Also, because of Amdahl's law, we observed sub-linear speedup ratios. This results in less than 100% efficiency every time.

4.4.4 Parallel Overhead of PCDFA

In parallel computation, several factors appear as overhead including

- inter-process interaction: communication time between processors.
- idling: periods when not all the available processors are performing useful tasks.
- excess computation: excess activities the serial program does not perform.

For an efficient parallel program, these overheads are minimized to zero, but this is not always possible.

$$\textit{Parallel Overhead} : T_0 = pT_p - T_1 \tag{4.4}$$

Table 4.13: PCDFA Parallel Overhead Analysis of DJI & BTC

Data	T_0^1	T_0^2	T_0^3	T_0^4
<i>DJI daily</i>	0	0.03	0.04	0.11
<i>DJI 1 hour</i>	0	0.10	0.02	0.02
<i>DJI 30 mins</i>	0	0.05	0.02	0.03
<i>DJI 5 mins</i>	0	0.23	0.10	0.13
<i>BTC daily</i>	0	0.01	0.02	0.05
<i>BTC 1 hour</i>	0	0.13	0.10	0.15
<i>BTC 30 mins</i>	0	0.05	0.14	0.11
<i>BTC 5 mins</i>	0	0.22	0.65	0.36

We observe from the overheads in Table (4.13) that, parallel overheads increase for the smallest data while it decreases for the most part when dealing with large data sets. Empirically, the overheads obtained are minimized close to zero as the number of processors increases which depict a good performance of the PCDFA.

4.5 Q-DEA for Dow Jones Industrial Average (DJI)

Figure 4.3 below shows a plot of diffusion entropy versus scale for q weighted in the range 0 to 4 for DJI during April 2013 to April 2021. Observe that at some diffusion time t , entropy decreases with the increase in q weights, whereas entropy increases with the increase in q weights at some diffusion time. This is because the Rényi entropy changes more promptly at large q weights as compared to small q weights.

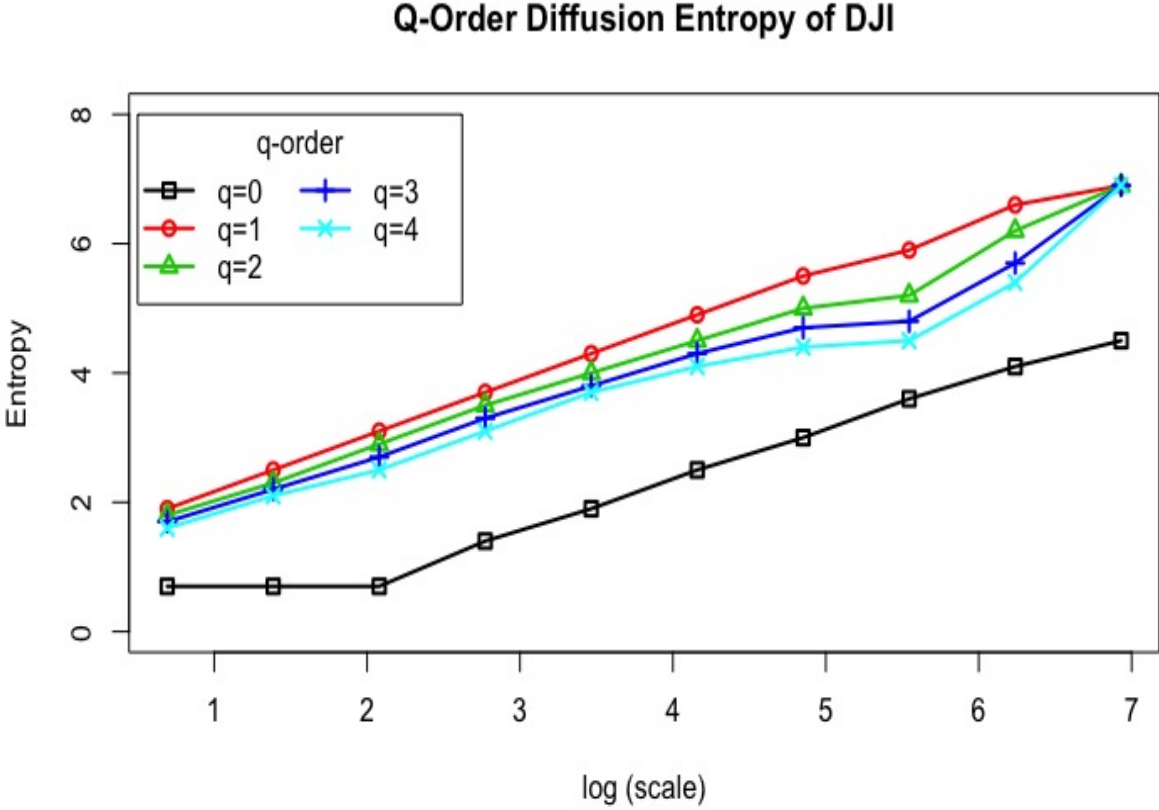


Figure 4.3: Q-order Diffusion Entropy as a function of diffusion time constructed based on different q values of the Rényi family of entropies using DJI

4.6 CDE and MS-CDE of DJI from April 2013 to April 2021

Here, we present results by looking into the stability of the stock market throughout the COVID-19 pandemic stock market crash using the multi-scale conditional diffusion entropy (MS-CDE). The S&P 500 index dropped 34%, 1145 points, from its peak of 3386 on February 19, 2020 to 2237 on March 23, 2020 . This crash was part of a worldwide recession caused by the COVID-19 lockdowns.

Figures 5.17 - 5.26 show the monthly conditional diffusion entropy $C_q(t)$ at different

time scales from April 2013 to April 2021 where $C_q(t) = 1$ for random behavior in the financial market, $C_q(t) > 1$ during a bull market and is less than 1 during a bear market.

Dow Jones Industrial Average (DJI)

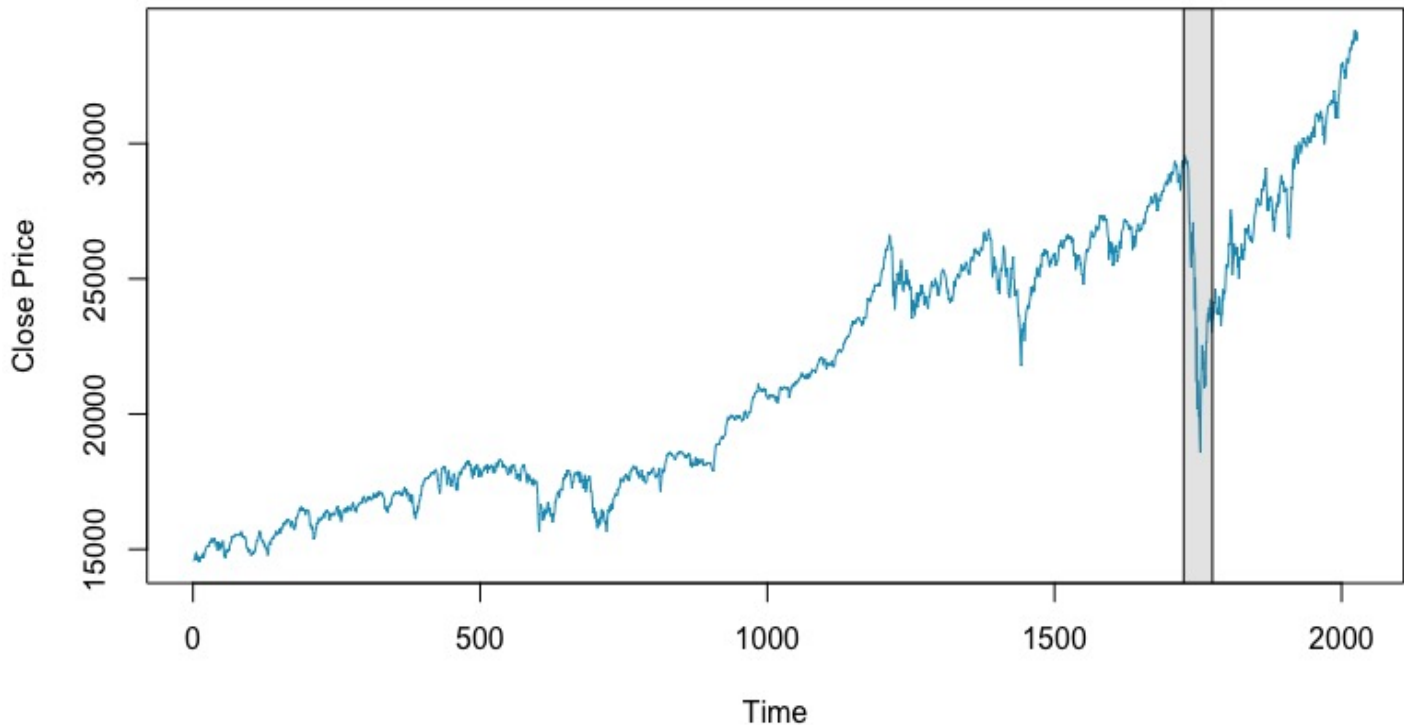


Figure 4.4: Daily Close Prices Plot of DJI. The gray region corresponds to the Covid-19 US Stock Market Crash of February/March 2020

Figure 4.4 shows instances of bull and bear markets in the DJI index. The highlighted portion represents the period from late February 2020 to early April 2020, depicting the 2020 coronavirus stock market crash. During this period, the COVID-19 pandemic spread globally during the week of February 24-28, causing a significant decline in global stock markets. The DJI dropped 11-12%, marking the most significant weekly decline since the 2007/2008 financial crisis. On March 12, a day after the announcement of a travel ban from Europe, the DJI fell sharply by 10%. After it became clear that a recession was inevitable,

the DJI dropped another 12.93% on March 16. Stock market indices briefly recovered to their levels at the end of February 2020 by early June 2020.

Figure 4.4 shows the fluctuations in the series over time. Many values of the conditional diffusion entropy at a time scale $t = 2$ are larger than 1 during bull market periods. In contrast, they are less than 1 during periods of the financial crisis (or bear market), as shown by the grey highlighted part of Figure 5.17. Observe that, as the time scale increases from $t = 2$ to $t = 256$, the conditional diffusion entropy approach is a random behavior where most of the $C_q(t) \approx 1$. During this period, the market is indifferent. Beyond the time scale of $t = 256$, conditional diffusion entropy is less than 1, indicating that the bear market is dominant. Therefore, the conditional diffusion entropy value coincides with the state of the stock market. Still, at different time lags, the conditional diffusion entropy depicts the other states of the stock market. A shorter time lag scale like $t = 2$ to $t = 256$ coincides more with the state of the stock market than larger time lag scales like $t = 256$ to $t = 1024$. This is because, for a short time lag, extreme fluctuations in the conditional diffusion entropy are observed.

At the same time, smoother (or averaged-out) fluctuations result in lower values of the conditional diffusion entropy. Thus, care should be taken when using conditional diffusion entropy to analyze the state of the stock market. Preferably, the conditional diffusion entropy of a time series is more informative if we investigate using a multi-time lag scale.

Chapter 5

Conclusion

This study compared and quantified the scaling parameters and long-term memory behavior of financial and geophysical time series using the DFA, Q-DFA, DEA, and Q-DEA methods. We also characterize the relevant time series with the TLF process to obtain the most robust Lévy distribution parameters. The stock market and volcanic eruption time series exhibit long memory behavior because $0.5 < H, \delta < 1$. Note that the choice of parameter α for the Lévy process is stochastic. Corresponding time variables $T = 1, 4, 8, 16$ affect it. Therefore, we select the appropriate TLF parameters for the period. Note that the numerical simulations adjust the value of A , the arbitrary scale parameter l , and the characteristic exponent α of the TLF model to best fit the TLF process cumulative function. The empirical data in Tables 4.1 and 4.2 suggest an inverse relationship between the DFA H parameter and the resulting TLF α parameter. Similarly, there is an inverse relationship between the δ DEA and the α TLF parameters. We support this proof of the relationship.

The DFA sometimes overestimates or underestimates the Hurst exponents based on the segment sizes. We generate better estimates of Hurst exponents by automating segment sizes in the DFA algorithm using the number base three(3) theory of the Cantor Set ($1/3^n, n \geq 0$). Here, we divide the time series into multiples of 3 at each level. We call this approach the Cantor Detrended Fluctuation Analysis (CDFA). It curbs the overestimation problem of the H by determining segment sizes based on the fractal phenomena depicted by the Cantor set while correctly predicting the memory behavior of the time series using H^c . We show this in Table 4.6, where we compare H and H^c exponents with the scaling exponent (α) of the TLF. We can apply the CDFA approach to time series with odd

lengths, time series whose lengths are not divisible by 2, time series whose lengths do not permit equal segment sizes after subdividing, etc. These series exist in several industries like finance, geophysics, etc. We can use the CDFA as a control experiment for the ordinary DFA. We have also proved that the Hurst exponent is homeomorphic to the Cantor Set. Using real-time series in Tables 4.3 - 4.5, we verify the proof by describing the fractal phenomenon represented by the Cantor set.

We also exploit the Q-DFA to analyze the highly irregular behavior of daily cases of COVID-19 in the United States before and after drugmakers announced in November 2020 that they had developed and tested 90 percent effective vaccines. We use the multifractal spectrum of the Q-DFA to characterize the path and predict the memory behavior of the time series on different time scales. Empirical results from the generalized Hurst exponent (gHE) indicate moderate variations over time after 11/1/2020 compared to the period before that. Also, the multifractal spectrum estimation before November 2020 has a long tail. It means high sensitivity to small fluctuations in the recorded cases. We record moderate sensitivity to small and large fluctuations after 11/1/2020. This results in the closeness to the symmetry of the multifractal spectrum. It implies that after 11/1/2020, daily covid-19 cases started becoming less fractal. Over time, the amount of daily change becomes steady.

We then investigated the effect of low and high-frequency data on estimating CDFA's Hurst index. We propose a PCDFA method to detect memory behavior in time series. We also calculate the speed-up ratios, efficiency, and overheads to evaluate PCDFA performance. The speed-up ratio from the implementation provided sub-linear speed-ups as the number of processors increased. The efficiency of the PCDFA method is directly related to data size or computational cost. We minimize inter-process interactions, idling, and implementation time overhead to zero by increasing the number of available processors.

Finally, we investigate the stability of the Dow Jones Industrial Average (DJI) from the US stock markets using a multi-scale normalized diffusion entropy and conditional diffusion entropy. We observe that conditional diffusion entropy (CDE) is used to analyze fluctua-

tions (or volatility) in the stock market data. It also depicts bear and bull markets, but we must be careful when using them. At different time lag scales, the conditional diffusion entropy can show conflicting information about the state of the stock market. Therefore, a multi-time lag scale (MS-CDE) is more informative for investigating the complex state of a financial market time series (or any time series), as shown in this paper. It is because they are often multifractal in nature.

For future work, we aim to investigate the memory behavior in time series using deep learning techniques from computer vision. Train a multi-layer convolutional neural network (CNN) by converting the time series to images and then applying cross-validation. To check the performance of this model, we will compute the root mean square errors and compare the results to that of VSMS and PSMs.

References

- [1] Mariani, M. C.; Asante, P. K.; Bhuiyan, M. A. M.; Beccar-Varela, M. P.; Jaroszewicz, S.; Tweneboah, O. K. Long-Range Correlations and Characterization of Financial and Volcanic Time Series. *Mathematics* **2020**, *8*, 441. <https://doi.org/10.3390/math8030441>.
- [2] Mariani, M. C.; Kubin, W.; Asante, P. K.; Tweneboah, O. K.; Beccar-Varela, M. P.; Multifractal Analysis of Daily US COVID-19 Cases. *10th Annual AHSE, STEM/STEAM and Education Conference. 2021*, ISSN 2333-4908. <https://huichawaii.org/wp-content/uploads/2021/07/Mariani-Maria-C.-2021-HUIC.pdf>
- [3] Mariani, M. C.; Kubin, W.; Asante, P. K.; Tweneboah, O. K.; Beccar-Varela, M. P.; Jaroszewicz, S.; Gonzalez-Huizar, H. Self-Similar Models: Relationship between the Diffusion Entropy Analysis, Detrended Fluctuation Analysis and Lévy Models. *Mathematics*. **2020**, *8*(7), 1046. <https://doi.org/10.3390/math8071046>
- [4] Hurst, H. E. Long term storage capacity of reservoirs. *Trans. Am. Soc. Eng.* **1951**, *116* (1), 770-799. <https://doi.org/10.1061/TACEAT.0006518>.
- [5] Peng, C. K. ; Buldyrev, S. V.; Havlin, S.; Simons, M. ; Stanley, H. E. ; Goldberger, A. L. Mosaic organization of DNA nucleotides. *Phys. Rev. E. American Physical Society.* **1994**, *49* (2), 1685-1689. <https://doi.org/10.1103/PhysRevE.49.1685>.
- [6] Shlesinger, M. F.; West, B. J.; Klafter, J. Lévy dynamics of enhanced diffusion: Application to turbulence. *Physical Review Letters.* **1987**, *58*, 11-16.
- [7] Ben-Avraham, D.; Havlin, S. Diffusion and Reactions in Fractals and Disordered Systems. *Cambridge University Press.* Cambridge, **2000**.

- [8] Mariani, M. C.; Kubin, W.; Asante, P. K.; Guthrie, J. A.; Tweneboah, O. K. Relationship between Continuum of Hurst Exponents of Noise-like Time Series and the Cantor Set. *Entropy* **2021**, 23, 1505. <https://doi.org/10.3390/e23111505>.
- [9] Guthrie, J. A. ; Nymann, J. E. The topological structure of the set of subsums of an infinite series. *In Colloquium Mathematicum*. Institute of Mathematics: Polish Academy of Sciences. **1988**, 55, 323–327. http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.desklight-a3fbdce0-a55c-413d-a53f-e1228d61fd02/c/cm55_2_15.pdf.
- [10] Cantor set. https://en.wikipedia.org/wiki/Cantor_set
- [11] Mantegna, R. N. ; Stanley, H. E. An introduction to Econophysics: Correlations and Complexity in Finance. *Physics Today*. **2000**, 53 (12), 148. <https://doi.org/10.1063/1.1341926>.
- [12] Kleinberg, J. M. Navigation in a small world. *Nature*. **2000**, 406(6798), 845.
- [13] Li, G.; Reis, S. D. S.; Moreira, A. A.; Havlin, S.; Stanley, H. E.; Andrade Jr, J. S. Towards Design Principles for Optimal Transport Networks. *PRL*. **2010**, 104, 1.
- [14] Viswanathan, G. M.; Buldyrev, S. V.; Havlin, S.; da Luz, M. G. E.; Raposo, E. P.; Stanley, H. E. Optimizing the success of random searches. *Nature*. **1999**, 401 (6756), 910-914.
- [15] Reynolds, G. Navigating Our World Like Birds and some authors have claimed that the motion of bees. *The New York Times*. **2014**.
- [16] Sims, D. W.; Reynolds, A. M.; Humphries, N. E.; Southall, E. J.; Wearmouth, V. J.; Metcalfe, B.; Twitchett, R. J. Hierarchical random walks in trace fossils and the origin of optimal search behavior. *Proceedings of the National Academy of Sciences*. **2014**, 111 (30), 1107-1108.

- [17] Beccar-Varela, M.; Gonzalez-Huizar, H.; Mariani, M. C.; Serpa, L. F.; Tweneboah, O. K. Chile2015: Lévy Flight and Long-Range Correlation Analysis of Earthquake Magnitudes in Chile. *Pure Appl. Geophys.* **2016** *173*, 2257-2266. <https://doi.org/10.1007/s00024-016-1334-8>.
- [18] Scafetta, N. An entropic approach to the analysis of time series. *University of North Texas*, **2003**, 5906-5906.
- [19] Scafetta, N.; Grigolini, P. Scaling detection in time series: diffusion entropy analysis. *Phys Rev E Stat Nonlin Soft Matter Phys.* **2002**, *66* (3), 036130. <https://doi.org/10.48550/arXiv.cond-mat/0202008>.
- [20] Huang, J.; Shang, P.; Zhao, X. Multifractal diffusion entropy analysis on stock volatility in financial markets. *Physica A.* **2012**, *391* (22), 5739-5745.
- [21] Morozov, A. Y. Comment on 'multifractal diffusion entropy analysis on stock volatility in financial markets' [Physica A. 391 (2012) 5739-5745]. *Physica A.* **2012**, *392* (10), 2442.
- [22] Jizba, P., Korbel, J. Multifractal diffusion entropy analysis: Optimal bin width of probability histograms. *Physica A: Statistical Mechanics and its Applications.* **2014**, *413*, 438-458. <https://doi.org/10.1016/j.physa.2014.07.008>.
- [23] Yamasaki, K.; Muchnik, L.; Havlin, S.; Bunde, A.; Stanley, H. E. Scaling and memory in volatility return intervals in financial markets. *Proc. Natl. Acad. Sci.* **2005**, *102*, 9424-9428.
- [24] Wang, F.; Yamasaki, K.; Havlin, S.; Stanley, H. E. Multifactor analysis of multiscaling in volatility return intervals. *Phys. Rev. E.* **2009**, *79*, 016103.
- [25] Ponzi, A.; Lillo, F.; Mantegna, R. N. Market reaction to a bid-ask spread change: a power-law relaxation dynamics. *Phys. Rev. E.* **2009**, *80*, 016112.

- [26] Siokis, F. M. Stock market dynamics: before and after stock market crashes. *Phys. A.* **2012**, 391, 1315-1322.
- [27] Andreev, B.; Sermpinis, G.; Stasinakis, C. Modelling Financial Markets during Times of Extreme Volatility: Evidence from the GameStop Short Squeeze. *Forecasting.* **2022**, 4, 654-673. <https://doi.org/10.3390/forecast4030035>.
- [28] Caraiani, P. Modeling the Comovement of Entropy between Financial Markets. *Entropy.* **2018**, 20, 417. <https://doi.org/10.3390/e20060417>.
- [29] Mariani, M.C.; Kubin, W.; Asante, P.K.; Guthrie, J.A.; Tweneboah, O.K. Relationship between Continuum of Hurst Exponents of Noise-like Time Series and the Cantor Set. *Entropy.* **2021**, 23, 1505. <https://doi.org/10.3390/e23111505>.
- [30] Buldyrev, S. V.; Goldberger, A. L.; Havlin, S. ; Mantegna, R. N.; Malsa, M. E. ; Peng, C. K. ; Simons, M. ; Stanley, H. E. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Phys. Rev. E.* **1995**, 51 (5), 5084-5091. <https://doi.org/10.1103/PhysRevE.51.5084>.
- [31] Bunde, A.; Havlin, S. Fractals and Disordered Systems. *Springer*. Berlin, Heidelberg, New York, **1996**.
- [32] Kantelhardt, J. W.; Koscielny-Bunde,E.; Rego, H. H. A.; Havlin, S.; Bunde, A. Detecting Long-range Correlations with Detrended Fluctuation Analysis. *Physica A.* **2001**, 295 (3-4), 441-454. [https://doi.org/10.1016/s0378-4371\(01\)00144-3](https://doi.org/10.1016/s0378-4371(01)00144-3).
- [33] Hardstone, R.; Poil, S. S.; Schiavone, G.; Jansen, R.; Nikulin, V. V.; Mansvelder, H. D.; Linkenkaer-Hansen, K. Detrended Fluctuation Analysis: A Scale-Free View on Neuronal Oscillations. *Frontiers in Physiology.* **2012**, 3, 450. <https://doi.org/10.3389/fphys.2012.00450>.
- [34] Little, M.; McSharry, P.; Moroz, I.; Roberts, S. Nonlinear, Biophysically-Informed Speech Pathology Detection. *IEEE International Conference on*

- Acoustics Speed and Signal Processing Proceedings*. **2006**, 2, 1080-1083.
<https://doi.org/10.1109/ICASSP.2006.1660534>.
- [35] Bunde, A.; Havlin, S.; Kantelhardt, J. W. ; Penzel, T.; Peter, J. H.; Voigt, K. Correlated and Uncorrelated Regions in Heart-Rate Fluctuations during Sleep. *Phys. Rev. Lett.* **2000**, 85 (17), 3736-3739. <https://doi.org/10.1103/PhysRevLett.85.3736>.
- [36] McCauley, J. L. Thermodynamic analogies in economics and finance: instability of markets. *Phys. A*. **2003**, 329 (1), 199-212.
- [37] Bentes, S. R.; Menezes, R. Entropy: A new measure of stock market volatility? *J. Phys. Conf. Ser.* **2012**, 394 (1), 012033.
- [38] Zhou, R.; Zhang, J.; Xiong, M.; Yang, F.; Yu, M. Using information entropy to measure bond risk: an empirical investigation. *J. Inf. Comput. Sci.* **2015**, 12(3), 1089-1100.
- [39] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E. , Havlin, S., Bunde, A. and Stanley, H. E. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*. **2002**.
[https://doi.org/10.1016/S0378-4371\(02\)01383-3](https://doi.org/10.1016/S0378-4371(02)01383-3).
- [40] Freedman, D.; Diaconis, P. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. **1981**, 57(4), 453-476. <https://doi.org/10.1007/BF01025868>.
- [41] Scott, D. W. On optimal and data-based histograms. *Biometrika*. **1979**, 66 (3), 605-610. <https://doi.org/10.1093/biomet/66.3.605>.
- [42] Oh, G., Kim, H.Y., Ahn, S. W., Kwak, W. Analyzing the financial crisis using the entropy density function. *Physica A: Statistical Mechanics and its Applications*. **2015**, 419, 464-469. <https://doi.org/10.1016/j.physa.2014.10.065>.

- [43] Li, S., Zhuang, Y., He, J. Stock market stability: Diffusion Entropy Analysis. *Physica A: Statistical Mechanics and its Applications*. **2016**, 450, 462-465. <https://doi.org/10.1016/j.physa.2016.01.037>.
- [44] Jizba, P., Arimitsu, T. The world according to Rényi : thermodynamics of multifractal systems. *Annals of Physics*. **2004**, 312, 17-59. <https://doi.org/10.48550/arXiv.cond-mat/0108184>.
- [45] Barany, E.; Beccar Varela, M. P.; Florescu, I.; Sengupta, I. Detecting market crashes by analyzing long-memory effects using high-frequency data. *Quantitative Finance* **2012**, 12, 4, 623-634.
- [46] Khintchine, A. Y.; Lévy, P. Sur les lois stables. *C.R. Acad. Sci. Paris*. **1936**, 374-376.
- [47] https://wiki.math.ntnu.no/_media/tma4225/2015h/cantor_set_function.pdf
- [48] Khan, M.; Ashton, P. M.; Li, M.; Taylor, G. A.; Pisica, I.; Liu, J. Parallel Detrended Fluctuation Analysis for Fast Event Detection on Massive PMU Data. *IEEE Transactions on Smart Grid*. **2015**, 6(1), 360-368. <https://doi.org/10.1109/TSG.2014.2340446>.
- [49] Amdahl, G. M.; Validity of the single processor approach to achieving large scale computing capabilities. *Proc. of AFIPS Spring Joint Computer Conference*. **1967**.
- [50] Gabaix, X.; Gopikrishnan, P.; Plerou, V. ; Stanley, H. E. A theory of power-law distributions in financial market fluctuations. *Nature* **2003**, 423, 267-270. <https://doi.org/10.1038/nature01624>.
- [51] Schinckus, C. From DNA to Economics: Analogy in Econobiology. *Review of Contemporary Philosophy* **2018**, 17 31-42. <https://doi.org/10.22381/RCP1720183>.
- [52] Scafetta, N. An entropic approach to the analysis of time series. *University of North Texas. ProQuest Dissertations Publishing* **2001**. <https://doi.org/3073550>.

- [53] Ivanova, K.; Ausloos, M. Application of the detrended fluctuation analysis (DFA) method for describing cloud breaking. *Physica A. Statistical Mechanics and its Applications*. **1999**, *274* (1-2), 349-354. [https://doi.org/10.1016/S0378-4371\(99\)00312-X](https://doi.org/10.1016/S0378-4371(99)00312-X).
- [54] de Moura, E. P.; Vieira, A. P.; Irmao, M. A. S.; Silva, A. A. Applications of detrended-fluctuation analysis to gearbox fault diagnosis. *Mechanical Systems and Signal Processing*. **2009**, *23* (3), 682-689. <https://doi.org/10.1016/j.ymssp.2008.06.001>.
- [55] Govindan, R. B.; Wilson, J. D.; PreiBl, H.; Eswaran, H.; Campbell, J. Q.; Lowery, C. L. Detrended fluctuation analysis of short datasets: An application to fetal cardiac data. *Physica D. Nonlinear Phenomena*. **2007**, *226* (1), 23-31. <https://doi.org/10.1016/j.physd.2006.10.019>.
- [56] Li, E.; Mu, X.; Zhao, G.; Gao, P. Multifractal Detrended Fluctuation Analysis of Streamflow in the Yellow River Basin, China. *Water*. **2015**, *7* (4), 1670-1686. <https://doi.org/10.3390/w7041670>.
- [57] Benes, E.; Fodor, M.; Kovacs, S.; Gere, A. Application of Detrended Fluctuation Analysis and Yield Stability Index to Evaluate Near Infrared Spectra of Green and Roasted Coffee Samples. *Processes*. **2020**, *8*, 913. <https://doi.org/10.3390/pr8080913>.
- [58] Weron, R. Measuring long-range dependence in electricity prices. In: Takayasu H. (eds) Empirical Science of Financial Fluctuations. *Springer, Tokyo*. **2002**. https://doi.org/10.1007/978-4-431-66993-7_12.
- [59] Kristoufek, L. Long-range dependence in returns and volatility of Central European Stock Indices. *IES Working Paper 3/2010*. IES FSV. Charles University. **2010**.
- [60] Ihlen, E. Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Frontiers in Physiology*. **2012**, *3*, 141. <https://doi.org/10.3389/fphys.2012.00141>.
- [61] Koponen, I. Analytic approach to the problem of convergence of truncated Lévy flights

towards the Gaussian stochastic process. *Phys. Rev. E.* **1995**, em 52 (1), 1197-1199.
<https://doi.org/10.1103/PhysRevE.52.1197>.

[62] <https://tasks.illustrativemathematics.org/content-standards/tasks/929>

[63] Clausius, R. Available online: <http://www.ueltschi.org/teaching/chapShannon.pdf>
(accessed on 15 October 2019). **1865**.

[64] Clausius, R. Available online: <http://galileo.phys.virginia.edu/classes/152.mf1i.spring02/Entropy.pdf>
(accessed on 10 October 2019). **1865**.

[65] Mariani, M. C.; Florescu, I.; Beccar Varela, M. P.; Ncheuguim, E. Long correlations and Lévy models applied to the study of memory effects in high frequency (tick) data. **2008**.

[66] Beccar-Varela, M. P.; Gonzalez-Huizar, H.; Mariani, M. C.; Tweneboah, O. K. Lévy Flights and Wavelets Analysis of Volcano-Seismic Data. *Pure and Applied Geophysics, Online first*. **2019**.

[67] Lévy, P. Calcul des probabilités. Paris, **1925**.

[68] Scafetta, N.; Hamilton, P.; Grigolini, P. *Fractals*. **2001**, 9, 193-208.

[69] Mariani, M. C.; Bhuiyan, M. A. M.; Tweneboah, O. K.; Gonzalez-Huizar, H.; Florescu, I. Volatility models applied to geophysics and high frequency financial market data. *Physica A*. **2018**, 503, 1, 304-321.

[70] M. C. Mariani and H. Gonzalez-Huizar and M. A. M. Bhuiyan and O. K. Tweneboah, Using Dynamic Fourier Analysis to Discriminate Between Seismic Signals from Natural Earthquakes and Mining Explosions. *AIMS Geosciences*, 2017, Vol. 3, No. 3, pp. 438-449.

- [71] Mariani, M. C.; Gonzalez-Huizar, H.; Bhuiyan, M. A. M.; Tweneboah, O. K. Forecasting the Volatility of Geophysical Time Series with Stochastic Volatility Models. *World Academy of Science, Engineering and Technology*. **2017**, 11, 10, 393-399.
- [72] Fuller, W. A. Introduction to Statistical Time Series. *New York: John Wiley and Sons*. **1976**.
- [73] Tang, X.; Yang, X. ; Wu, F. ; Multifractal detrended fluctuation analysis parallel optimization strategy based on openMP for image processing. *Neural Comput & Applic.* **2020** 32, 5599-5608. <https://doi.org/10.1007/s00521-019-04164-2>.
- [74] Romero-Laorden, D.; Villazón-Terrazas, J.; Martínez-Graullera O.; Ibáñez A.; Parrilla M.; Santos P. M. Analysis of parallel computing strategies to accelerate ultrasound imaging processes. *IEEE Trans Parallel Distrib Syst.* **2016**, 27 (12), 3429-3440.
- [75] <https://tasks.illustrativemathematics.org/content-standards/tasks/929>

- BLUE points depict the Normalized Data
- RED curve depicts the Cumulative Distribution Function
- GREEN curve depicts the Gaussian(Normal) distribution fit.

Relationships Between DFA, DEA and TLF using Financial Time Series

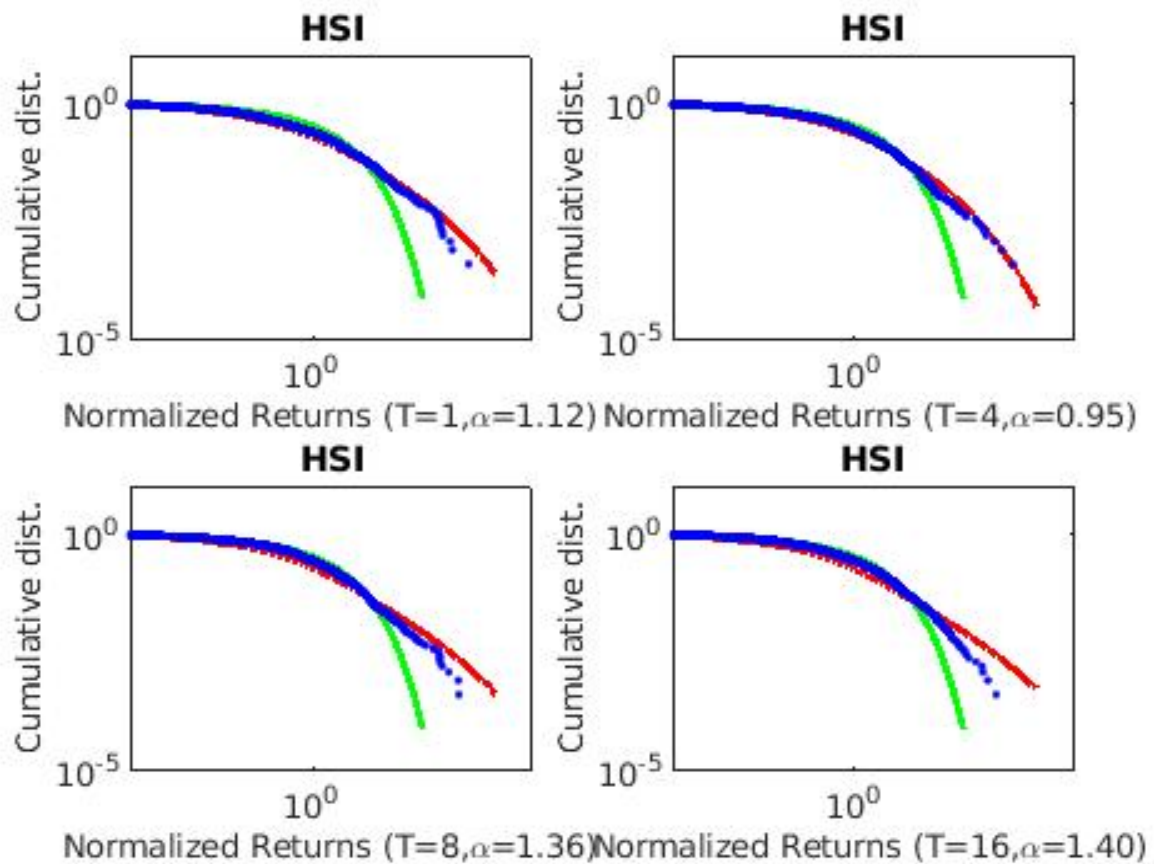


Figure 5.1: TLF model fits for HSI

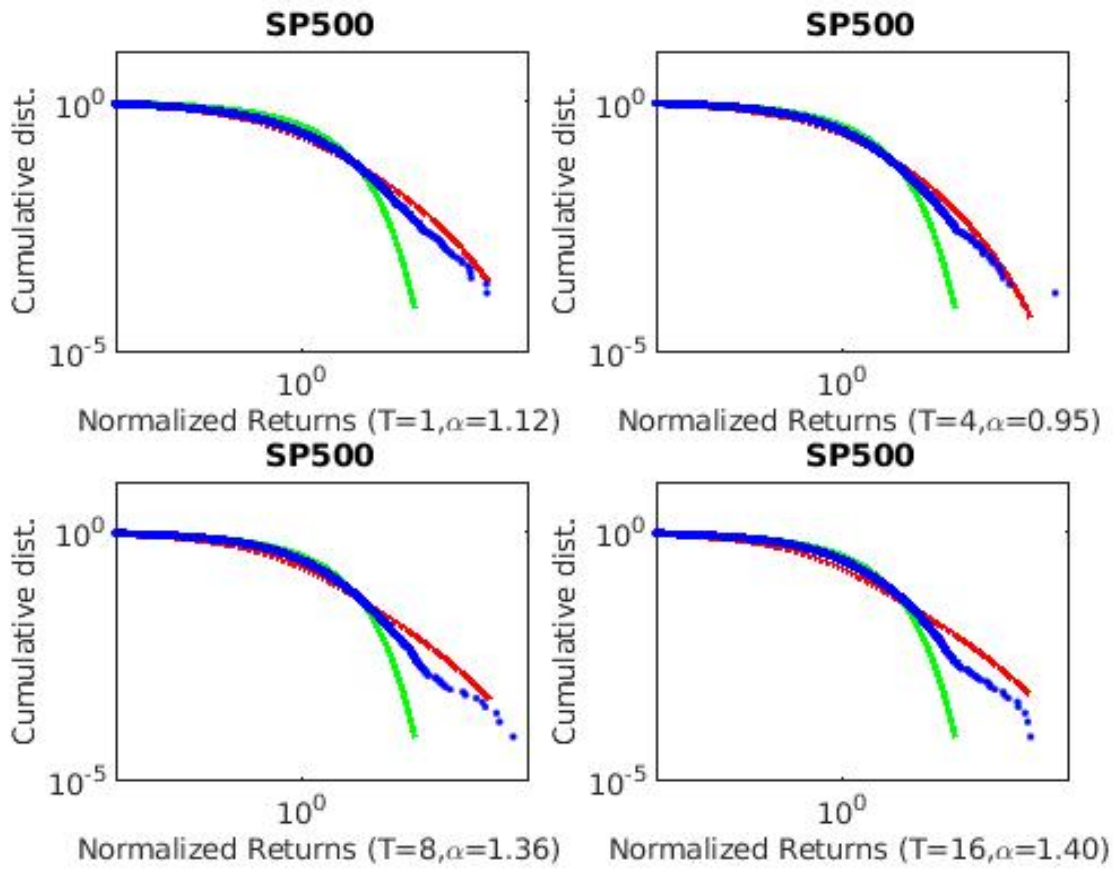


Figure 5.2: TLF model fits for SP500 USA

Relationships Between DFA, DEA and TLF using Geophysical Time Series

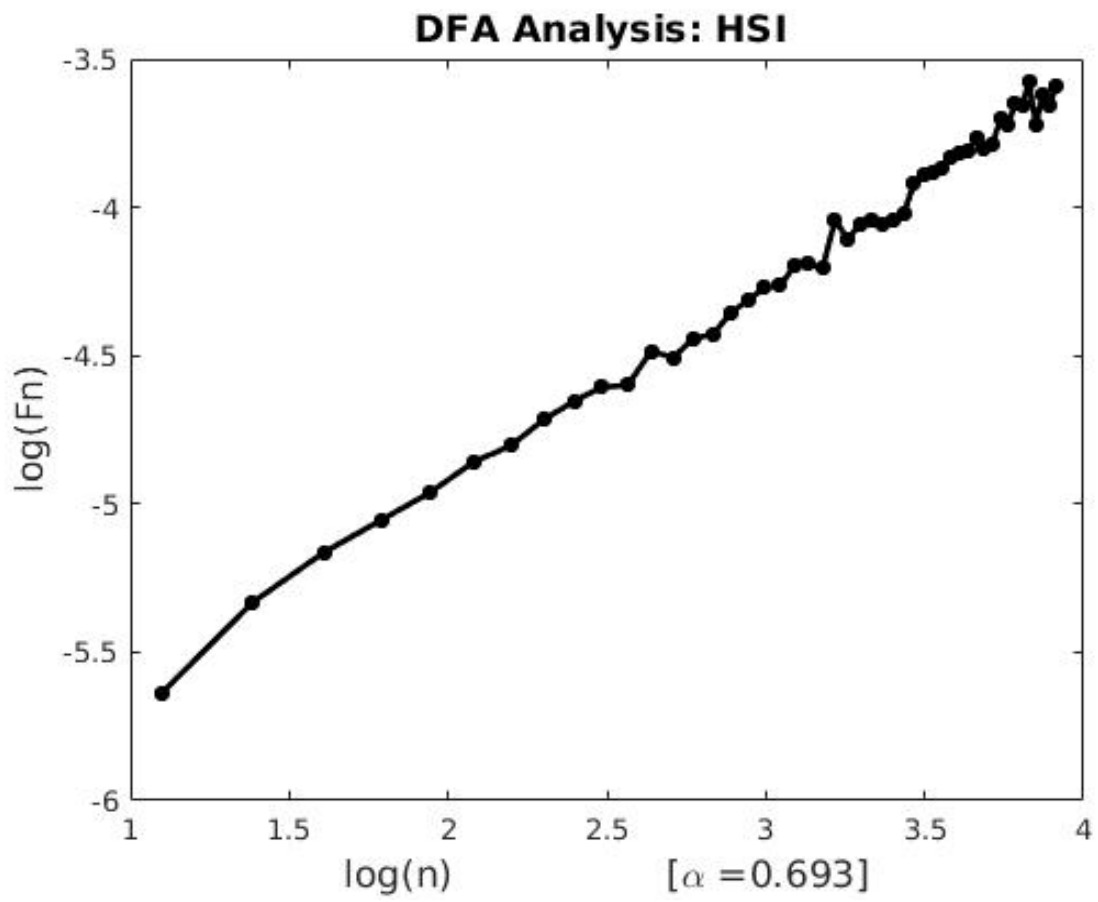


Figure 5.3: DFA Plot for HSI

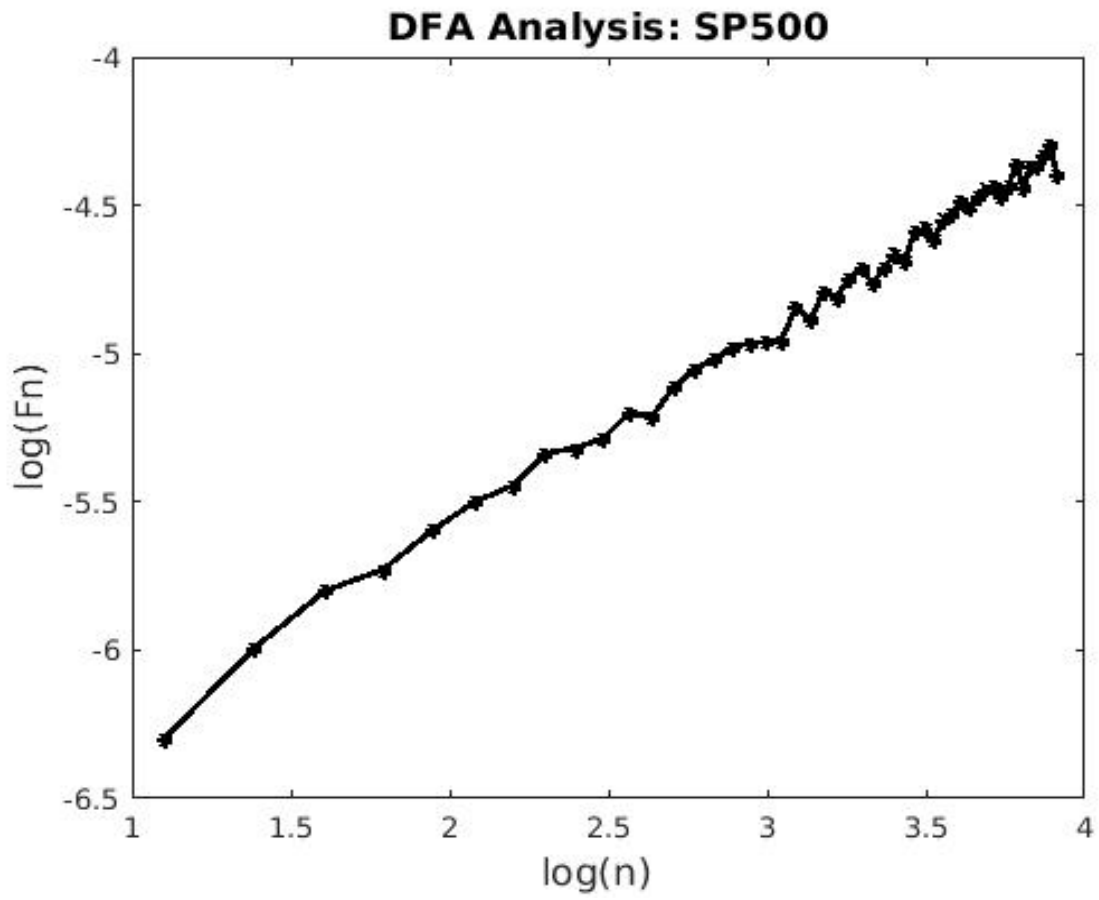


Figure 5.4: DFA for SP500 USA

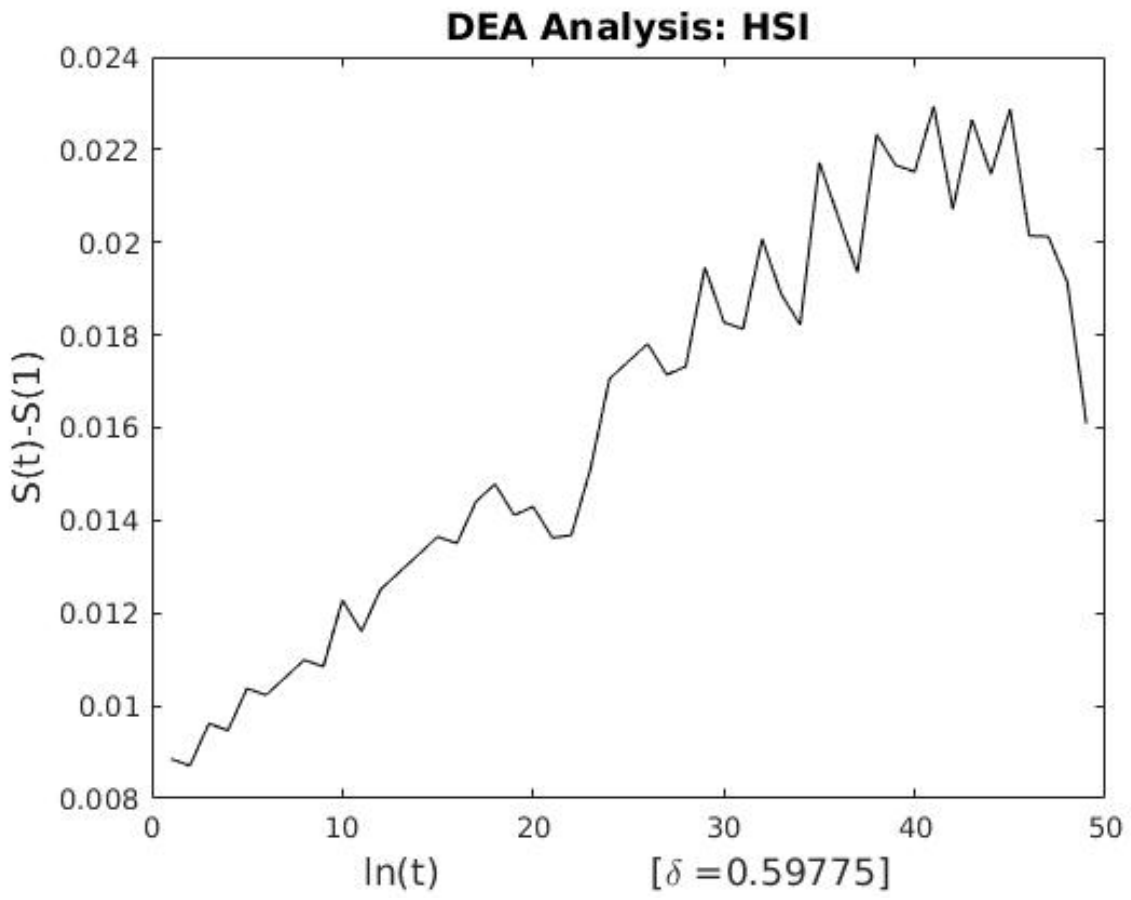


Figure 5.5: DEA for HSI

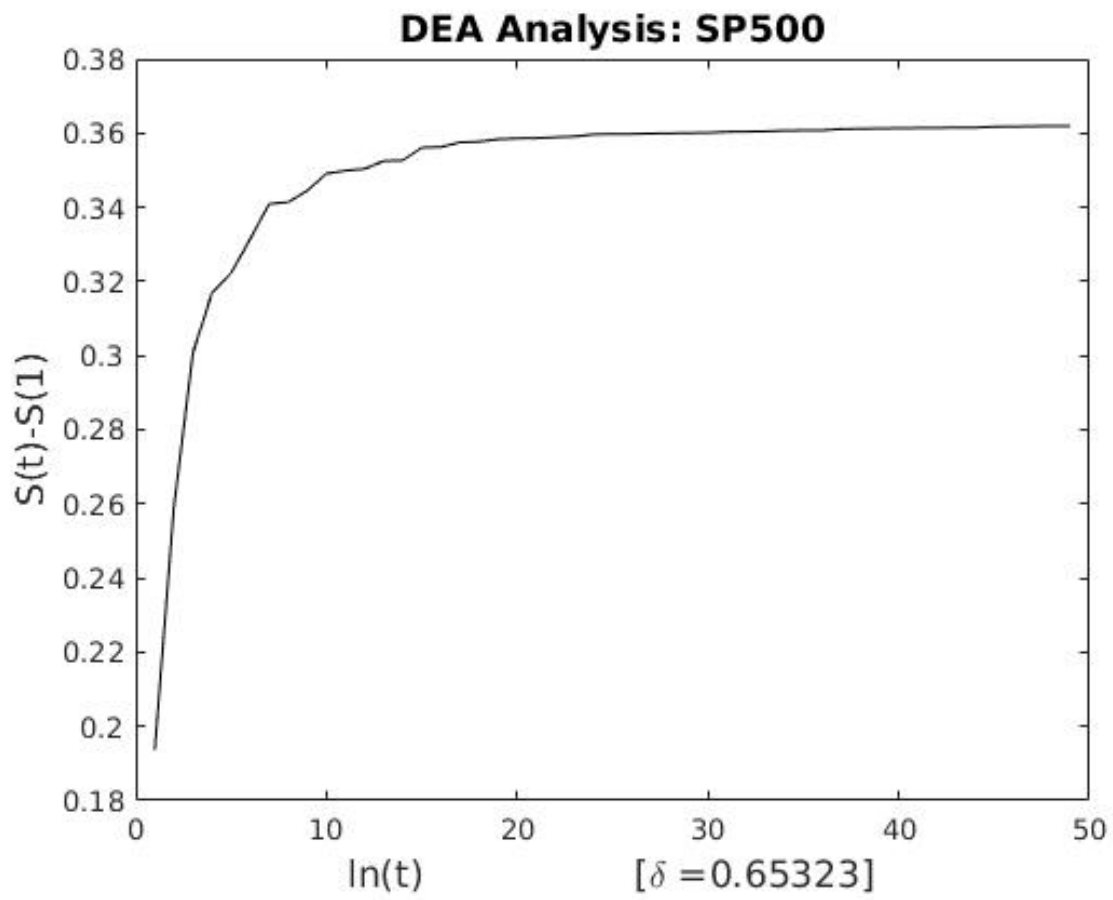


Figure 5.6: DEA for SP500 USA

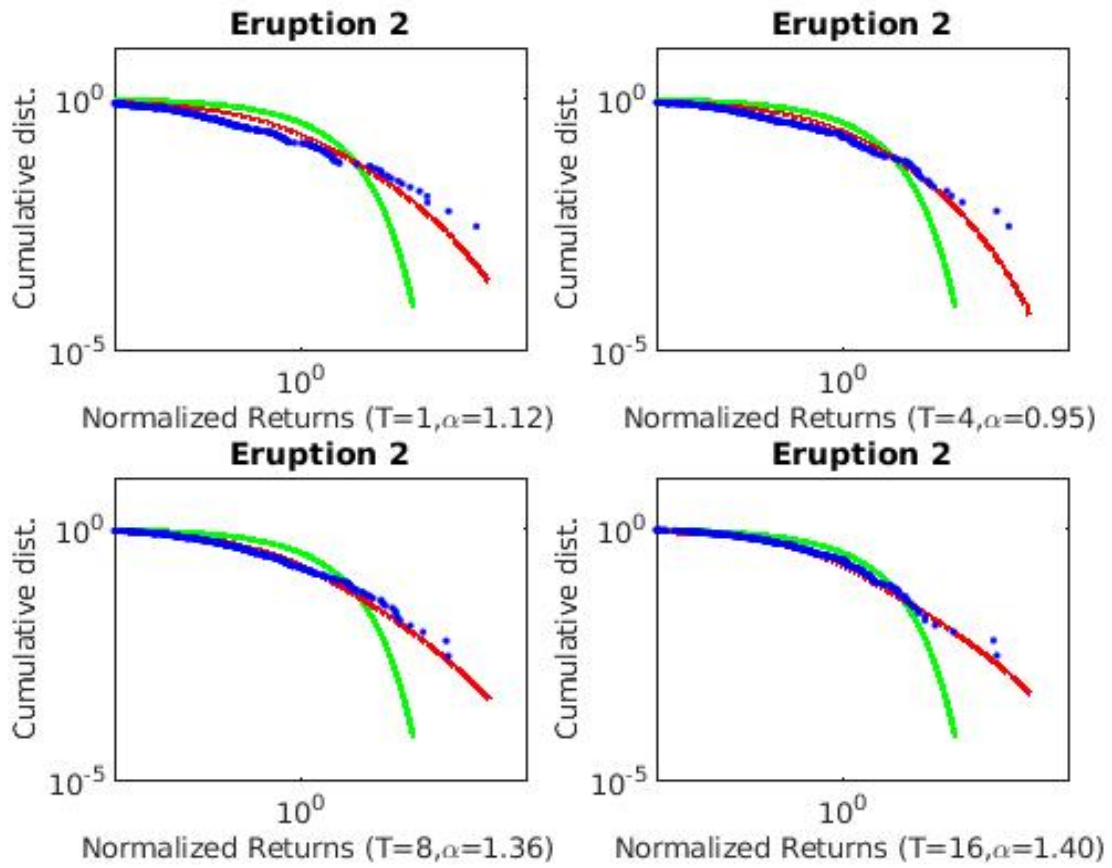


Figure 5.7: TLF model fit for Volcanic Eruption 2

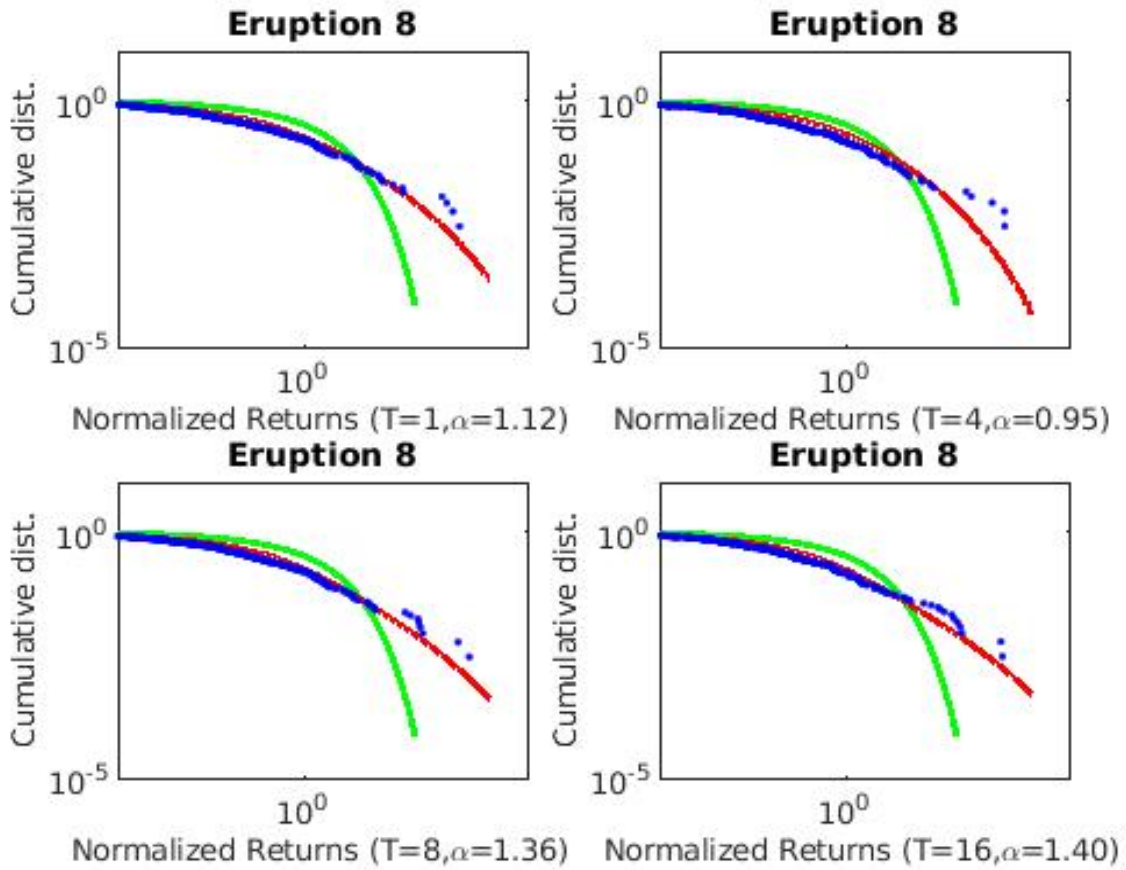


Figure 5.8: TLF model fit for Volcanic Eruption 8

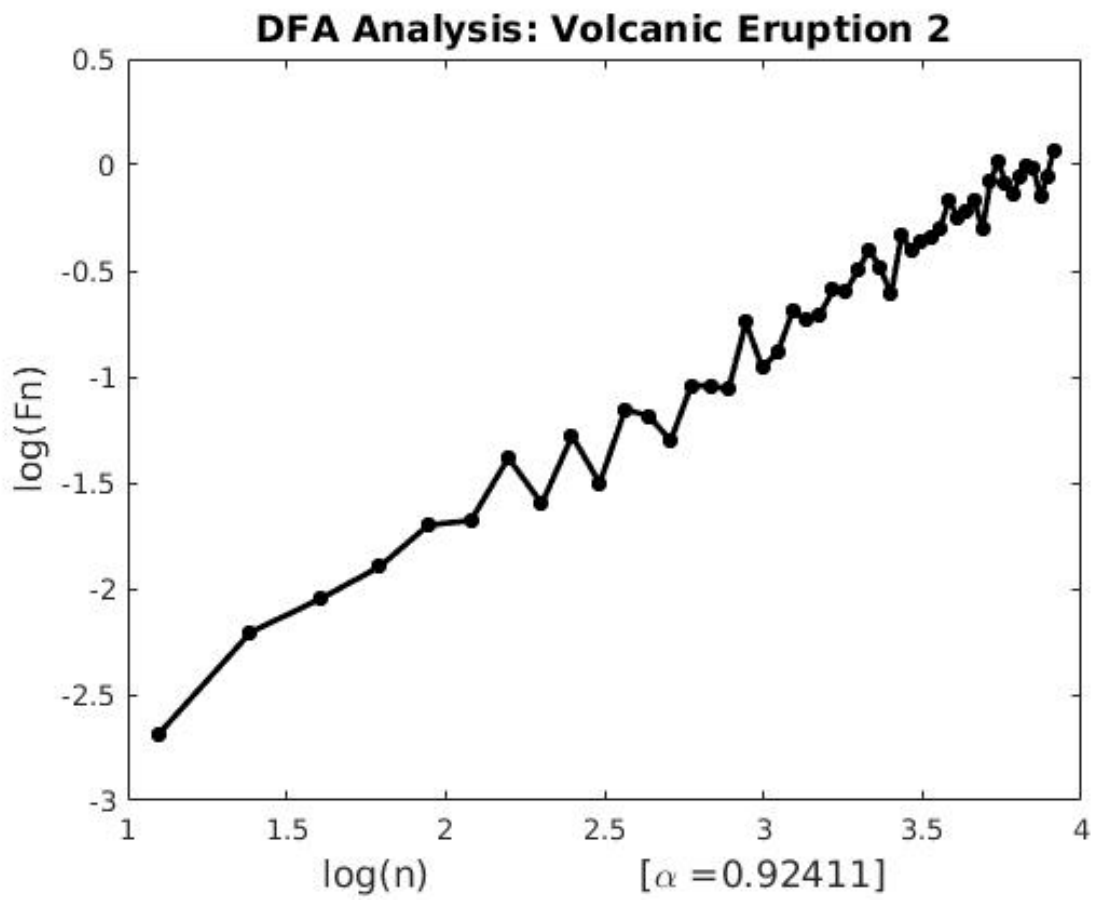


Figure 5.9: DFA for Volcanic Eruption 2

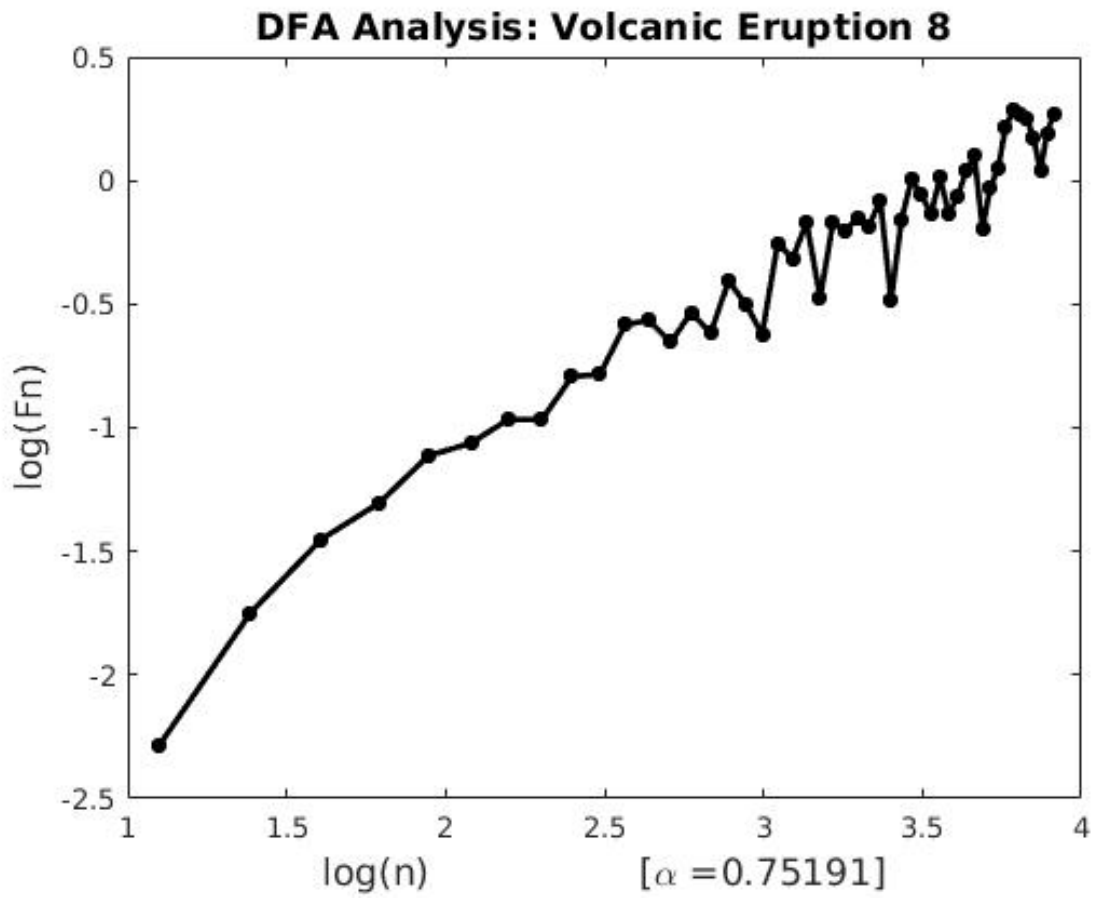


Figure 5.10: DFA for Volcanic Eruption 8

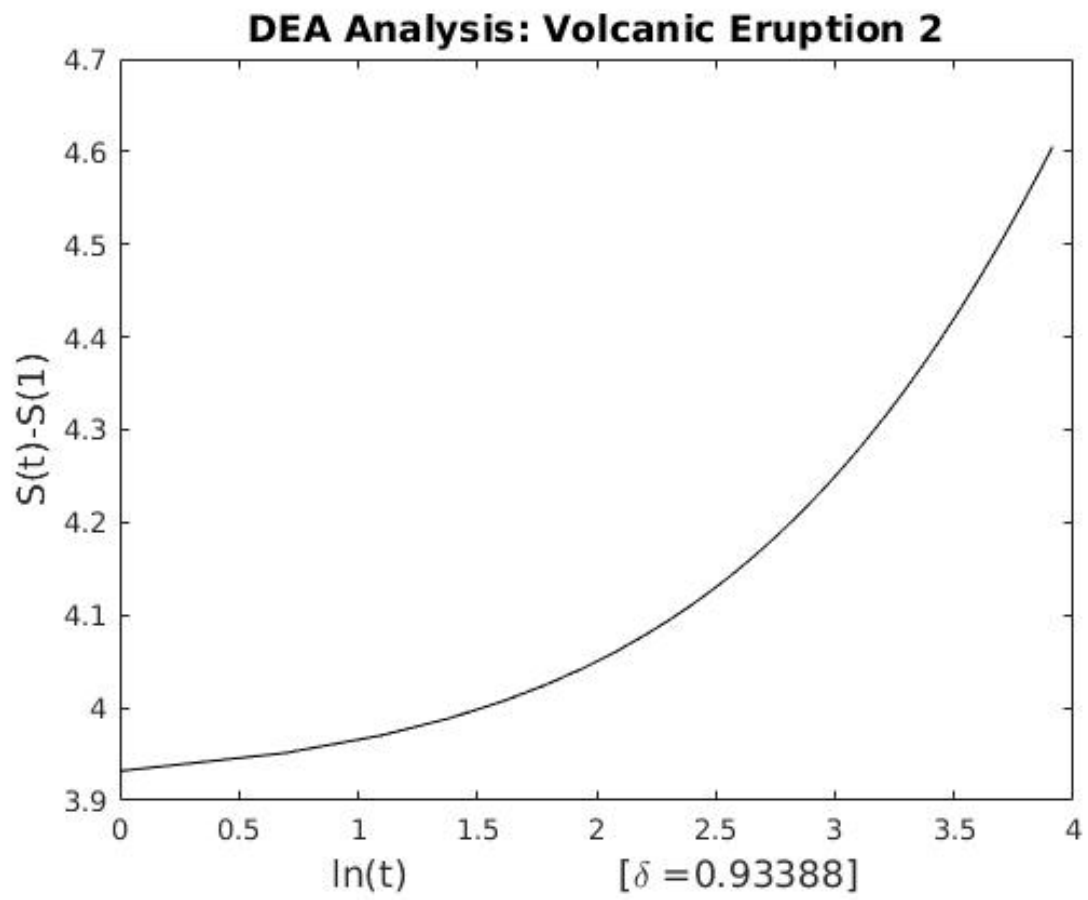


Figure 5.11: DEA for Volcanic Eruption 2

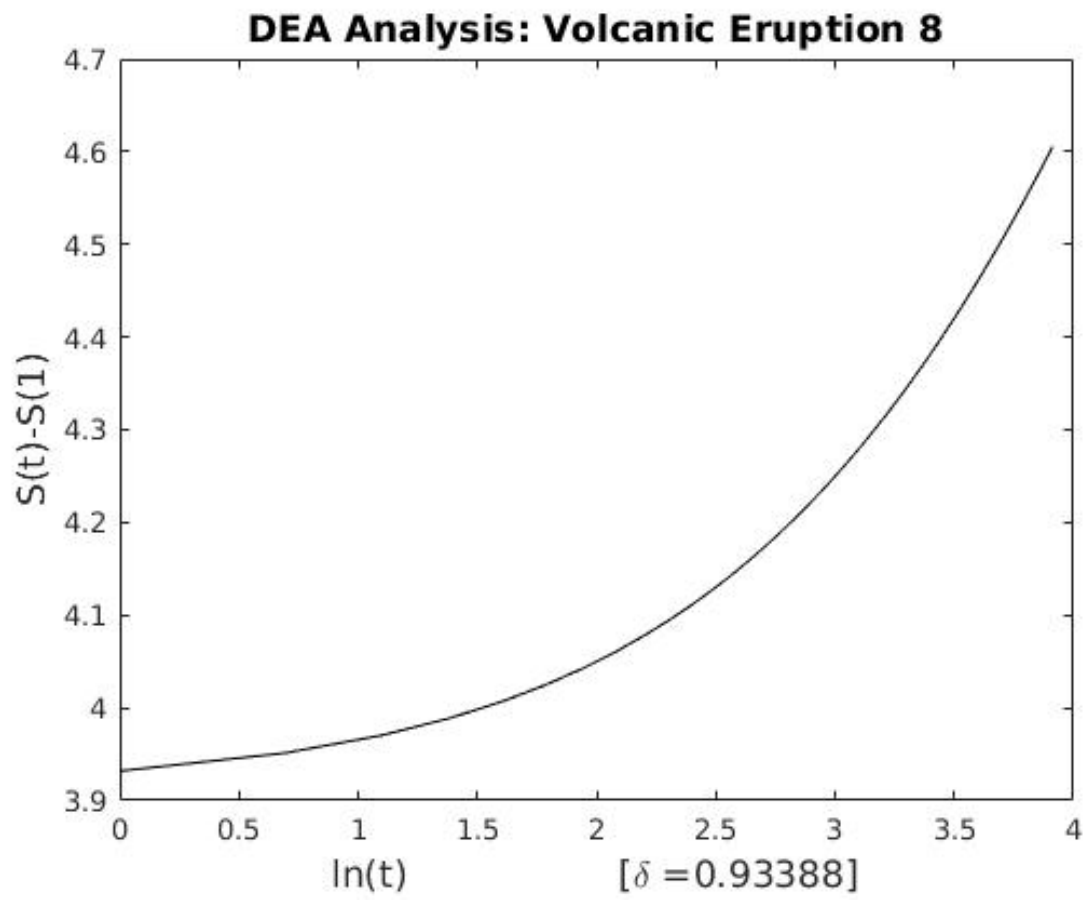


Figure 5.12: DEA for Volcanic Eruption 8

PCDFA Applied to Financial Time Series

Speedup Ratio of PCDFA

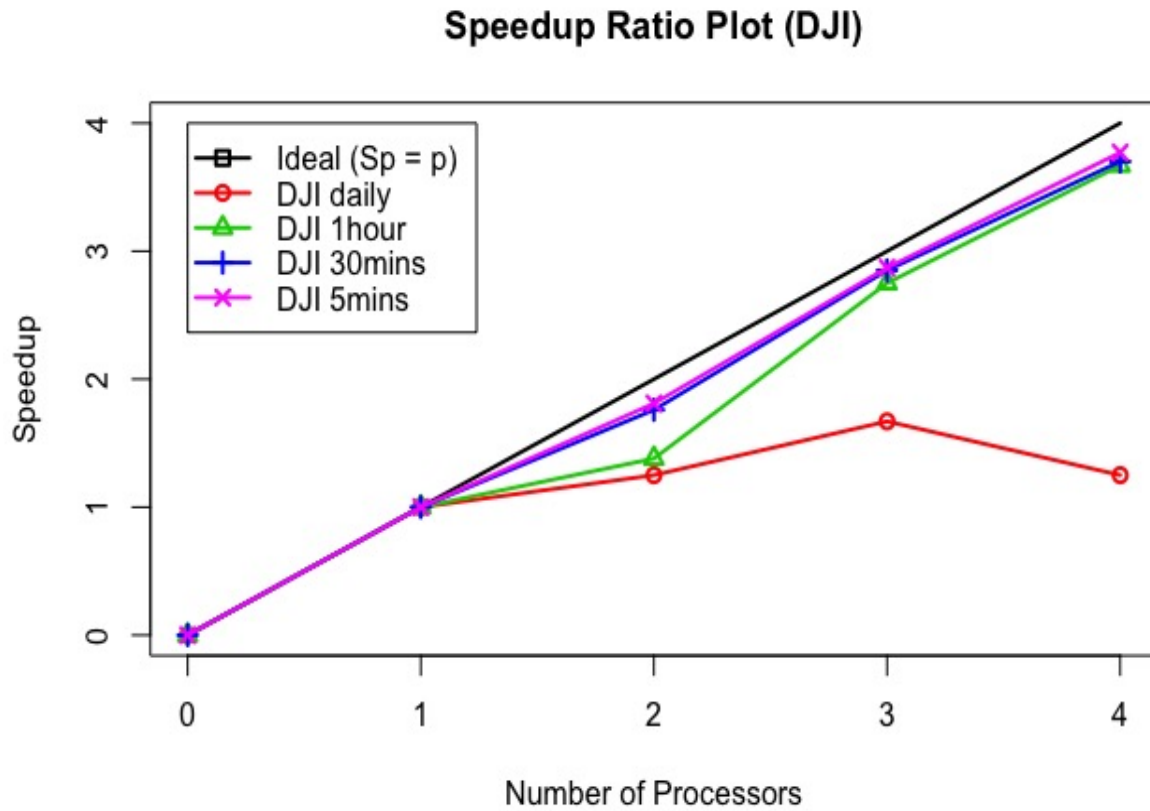


Figure 5.13: Speedup Ratio plot of PCDFA for low-to-high frequency DJI data

Speedup Ratio Plot (BTC)

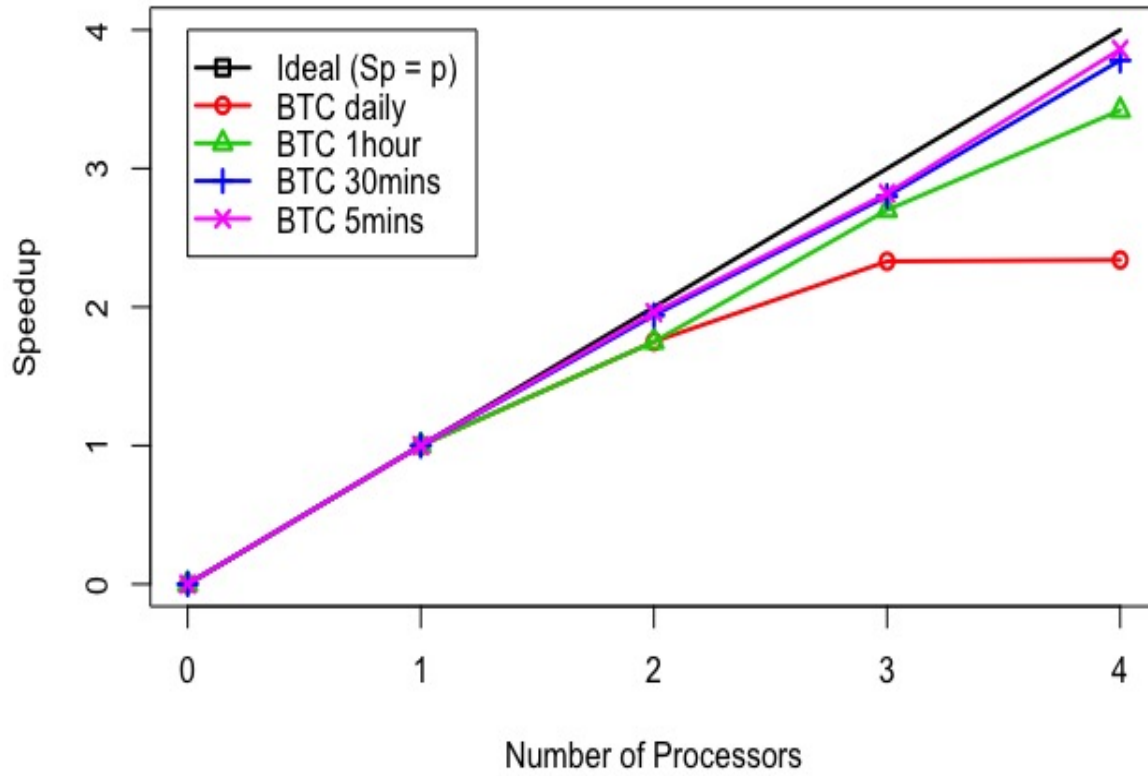


Figure 5.14: Speedup Ratio plot of PCDFA for low-to-high frequency BTC data

Parallel Efficiency of PCDFA

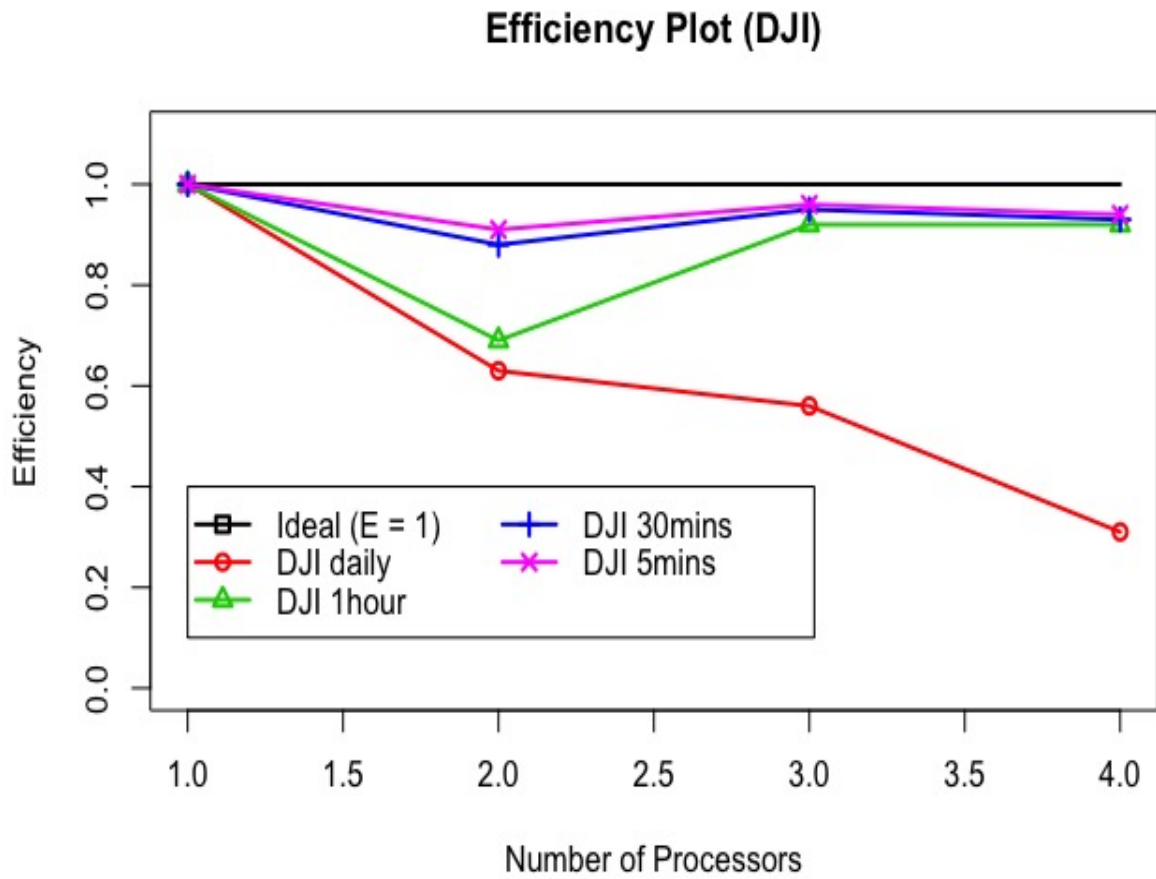


Figure 5.15: Efficiency plot of PC DFA for low-to-high frequency DJI data

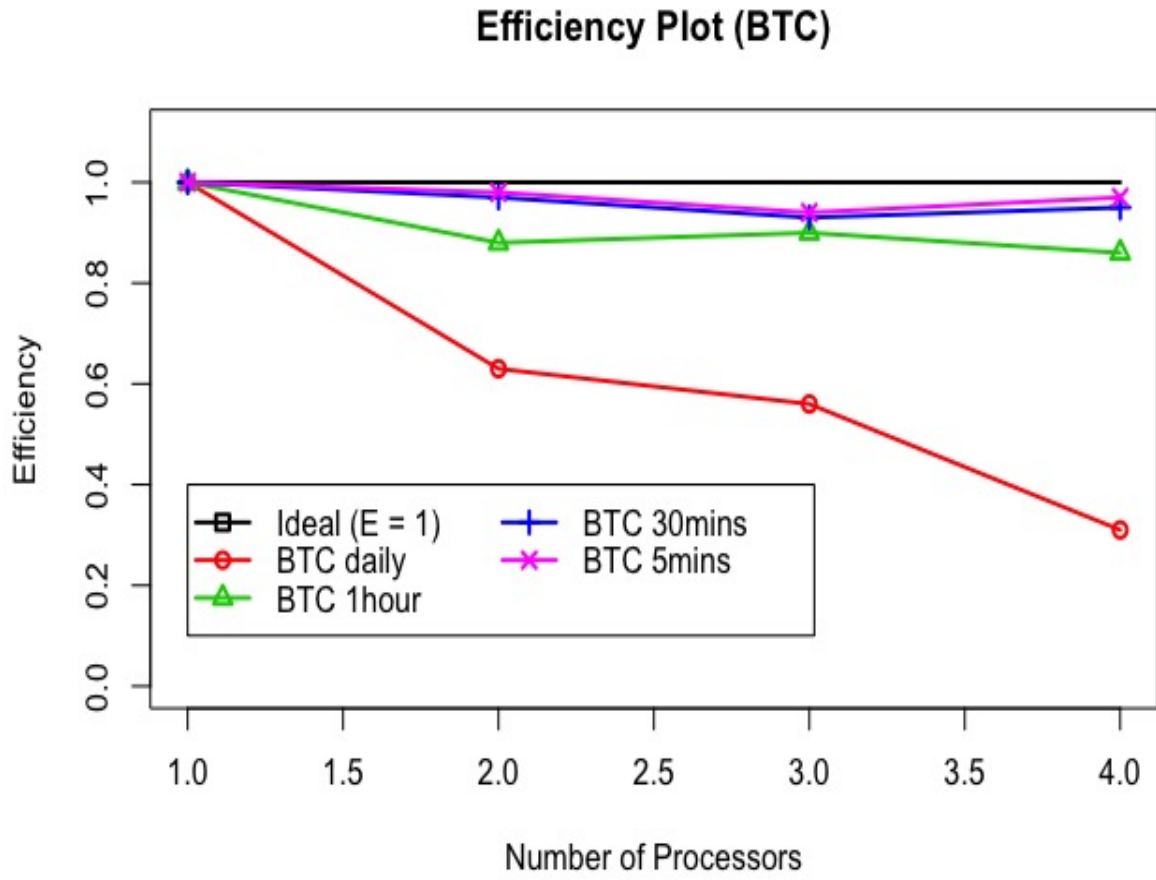


Figure 5.16: Efficiency plot of PCDFFA for low-to-high frequency BTC data

CDE and MS-CDE Applied to DJI

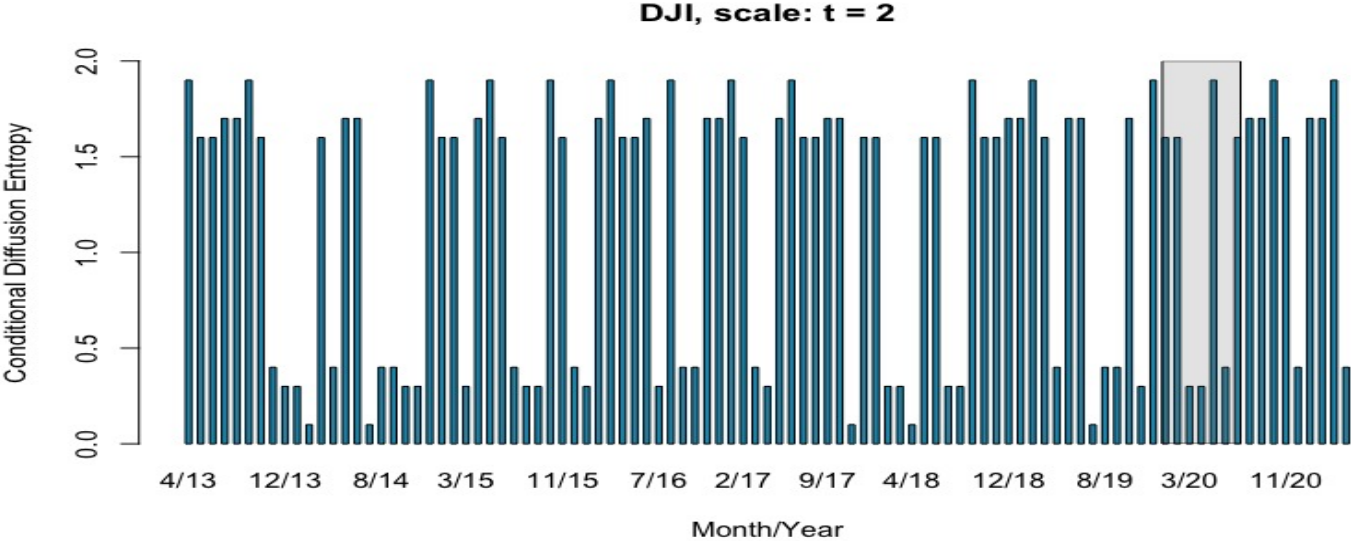


Figure 5.17: Monthly Conditional Entropy of DJI at Time Scale $t=2$

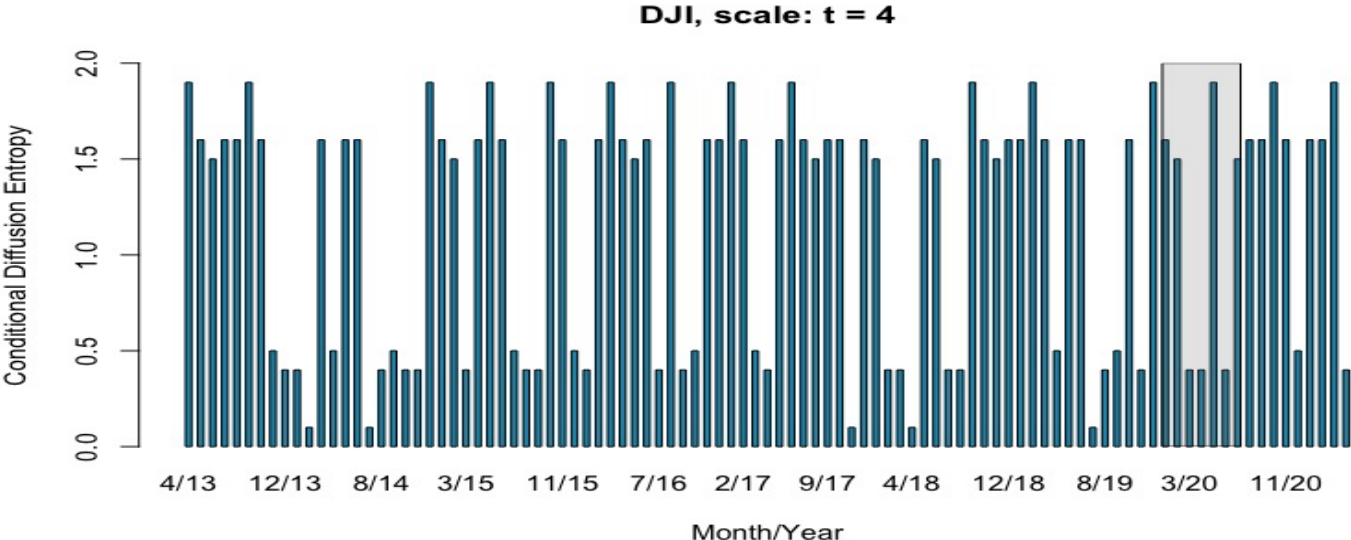


Figure 5.18: Monthly Conditional Entropy of DJI at Time Scale $t=4$

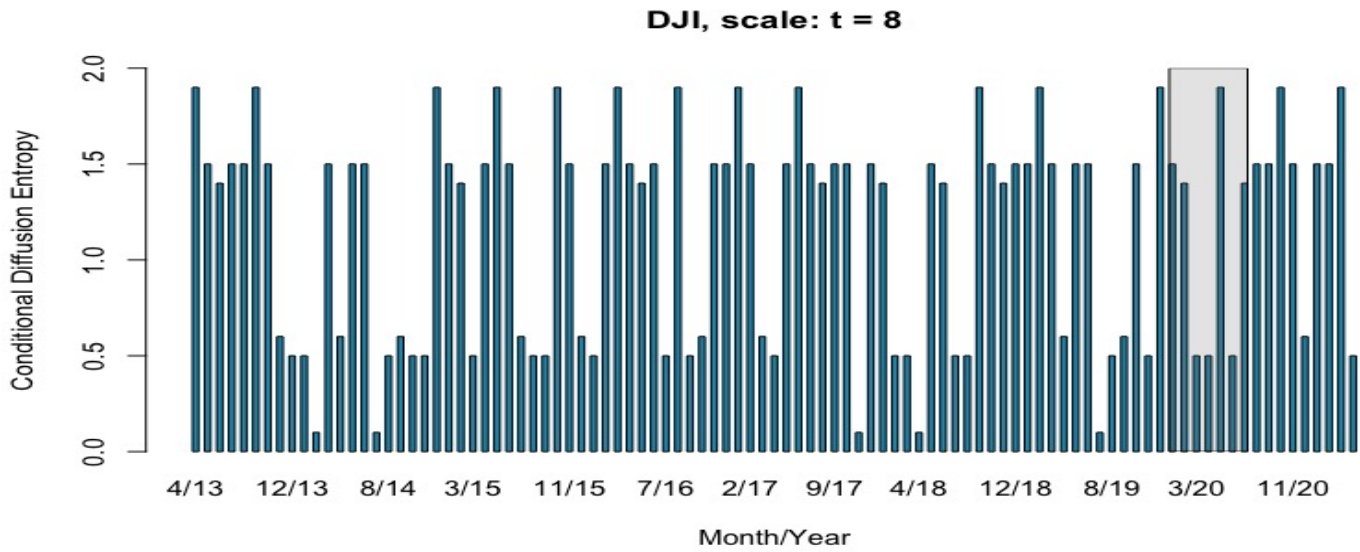


Figure 5.19: Monthly Conditional Entropy of DJI at Time Scale $t=8$

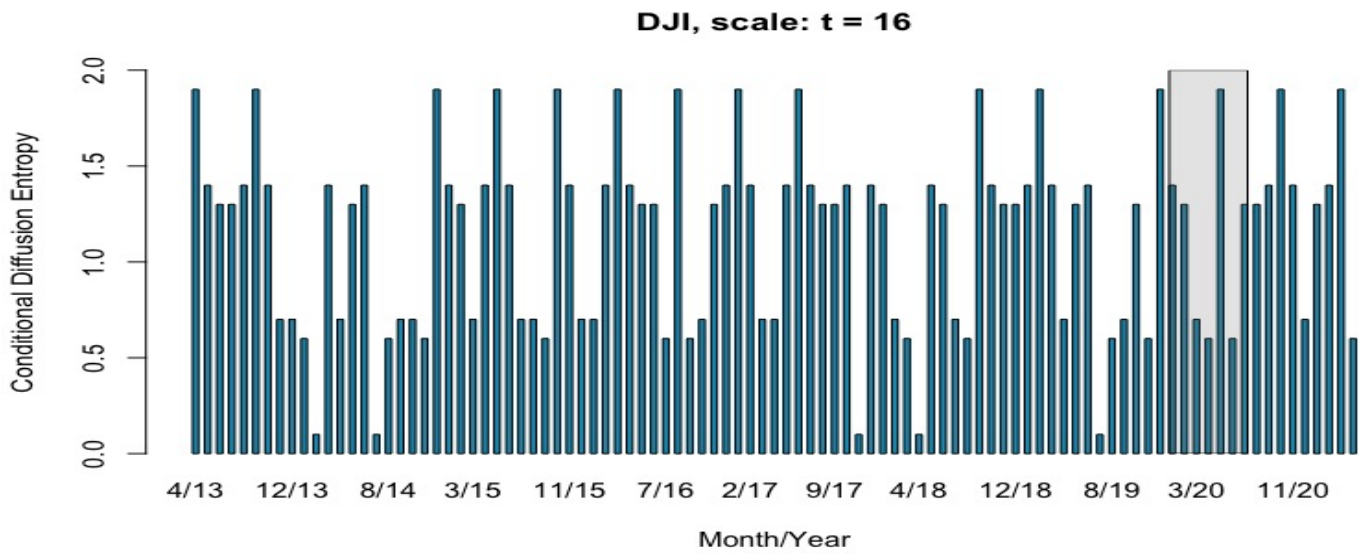


Figure 5.20: Monthly Conditional Entropy of DJI at Time Scale $t=16$

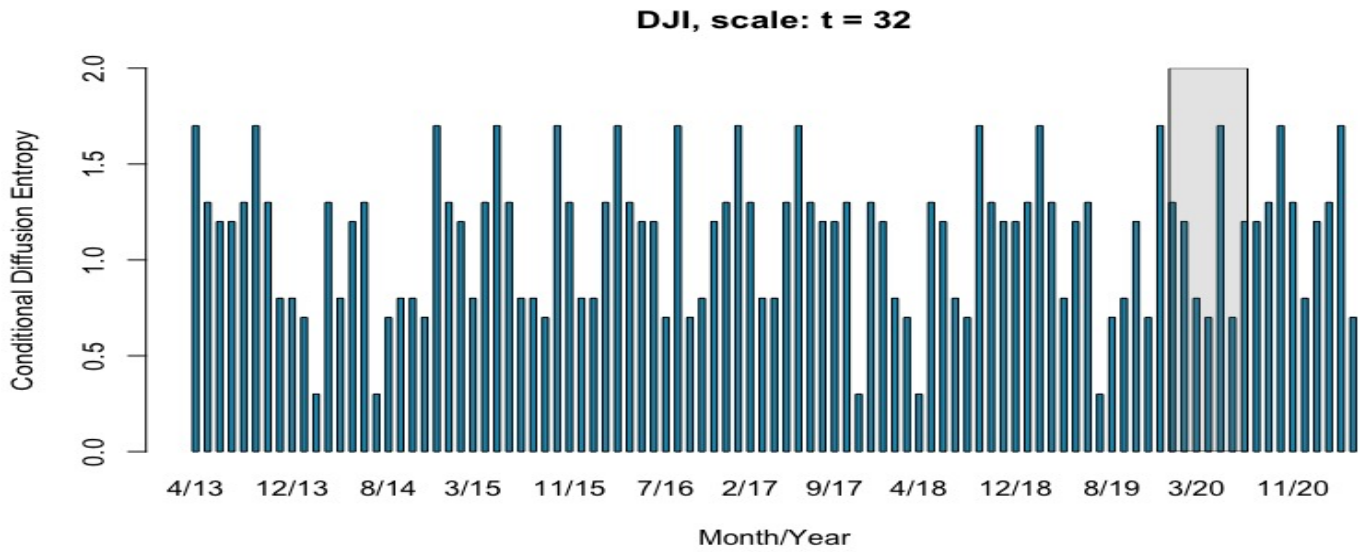


Figure 5.21: Monthly Conditional Entropy of DJI at Time Scale $t=32$

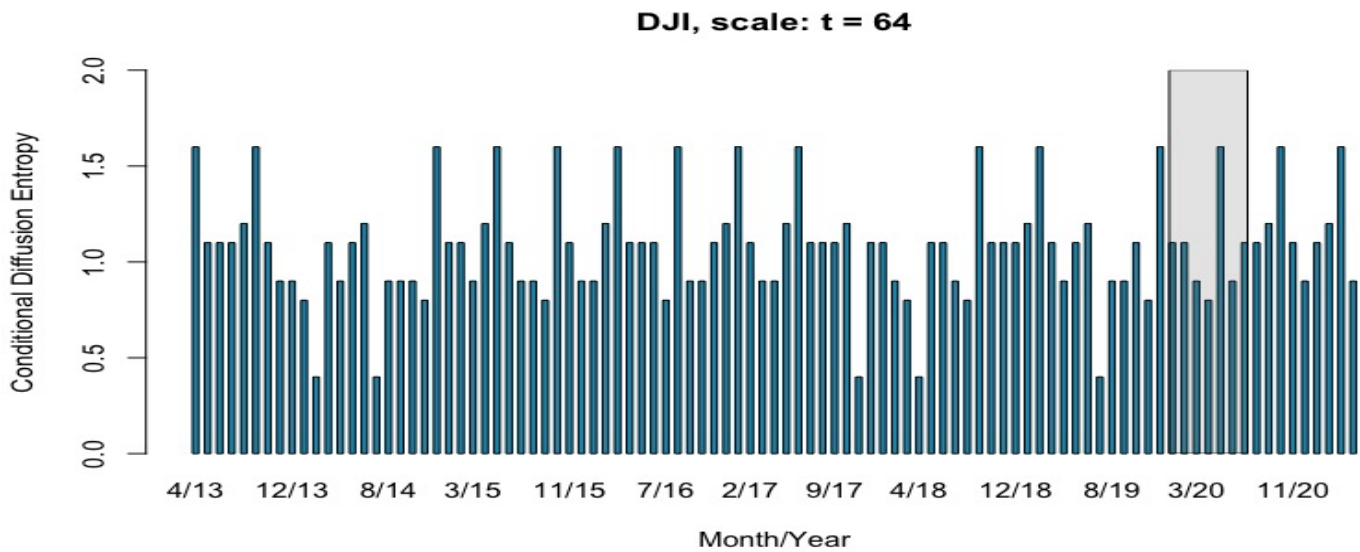


Figure 5.22: Monthly Conditional Entropy of DJI at Time Scale $t=64$

DJI, scale: t = 128

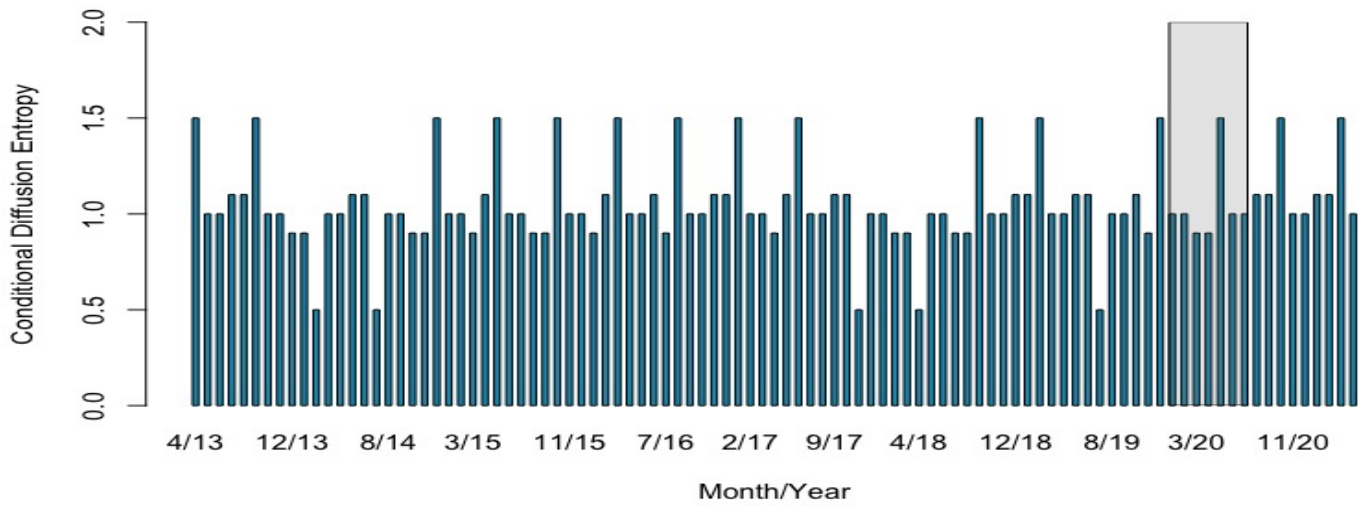


Figure 5.23: Monthly Conditional Entropy of DJI at Time Scale t=128

DJI, scale: t = 256

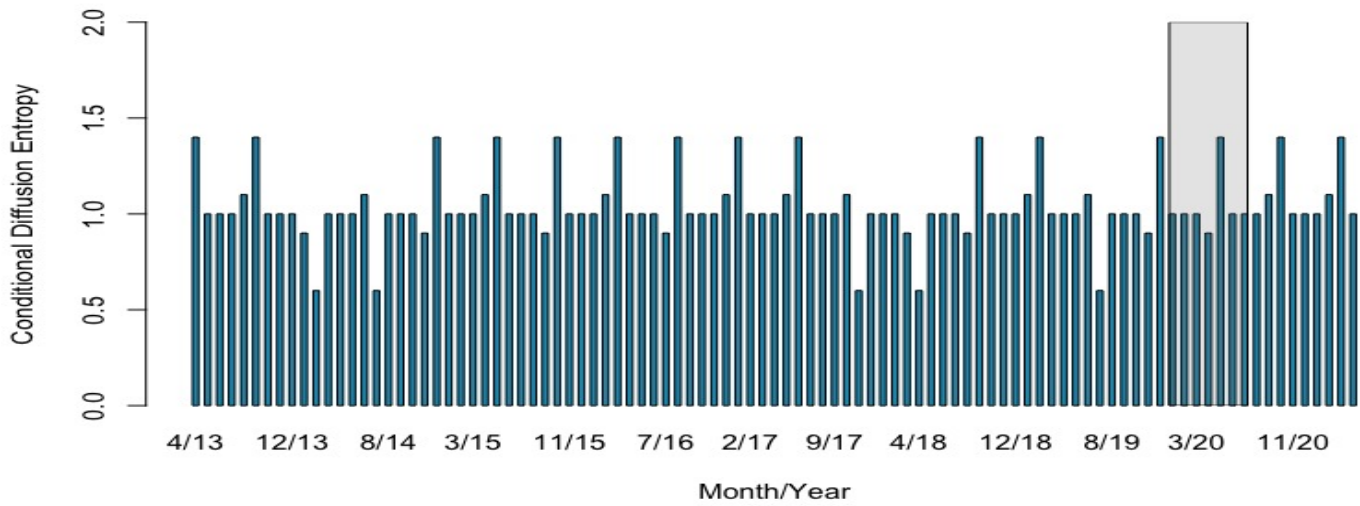


Figure 5.24: Monthly Conditional Entropy of DJI at Time Scale t=256

DJI, scale: t = 512

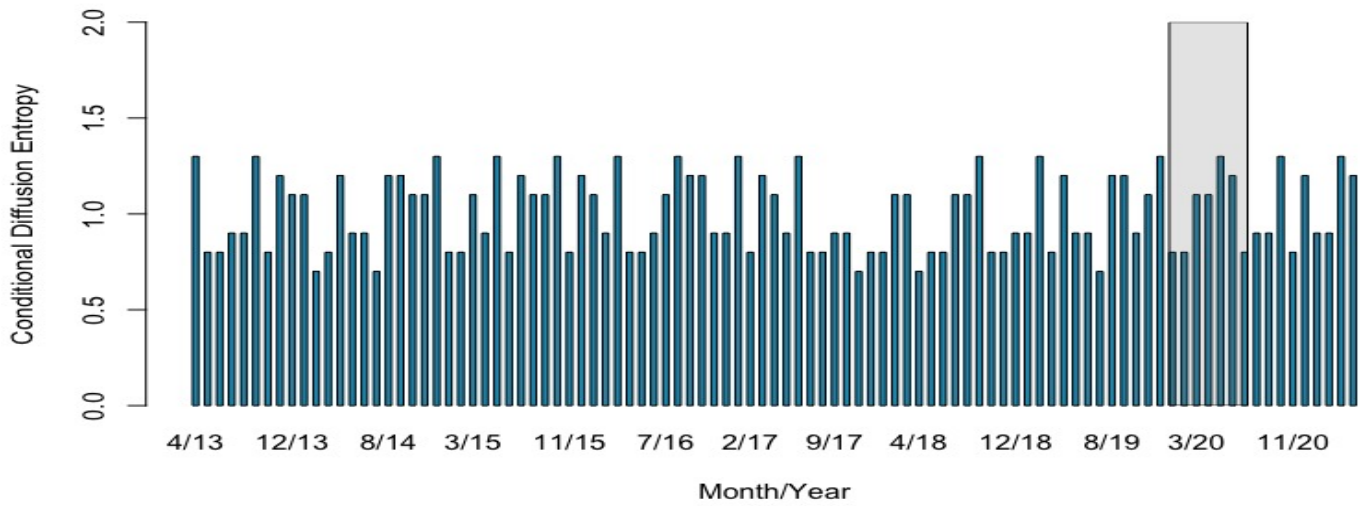


Figure 5.25: Monthly Conditional Entropy of DJI at Time Scale $t=512$

DJI, scale: t = 1024

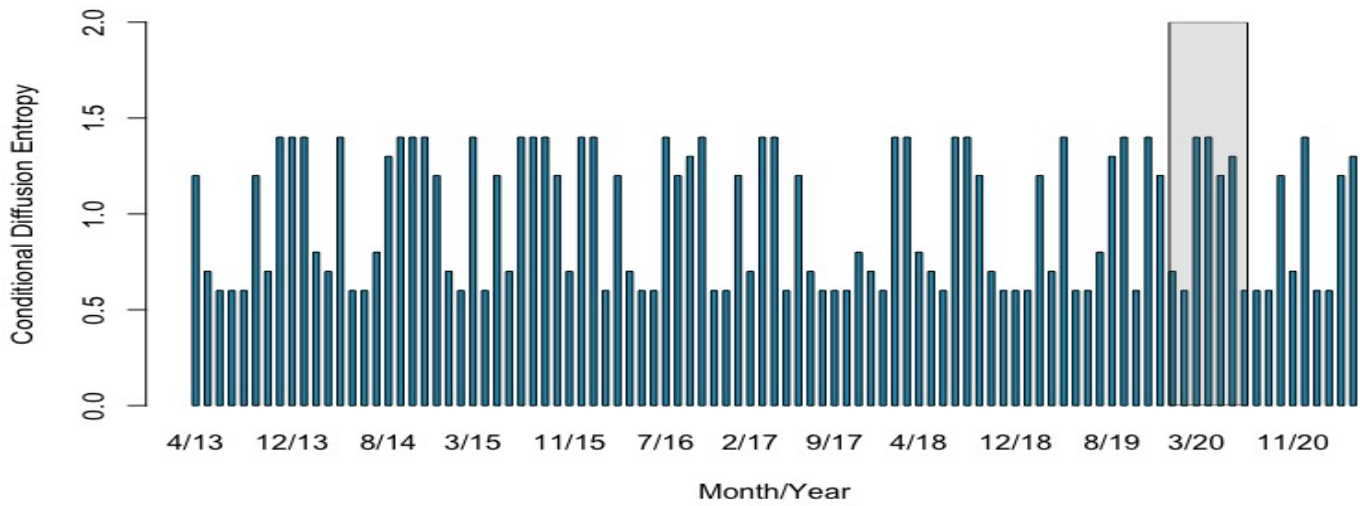


Figure 5.26: Monthly Conditional Entropy of DJI at Time Scale $t=1024$

Curriculum Vitae

William Kubin was born on June 4, 1990. He started schooling at Moree Star Of Hope International Primary and Junior Secondary School. He studied Business (Financial Accounting option) at The Academy of Christ The King Senior Secondary School at Cape Coast. He graduated The University of Cape Coast in Cape Coast - Ghana in 2015 where he pursued a Bachelor's degree in Mathematics-with-Economics and worked as a Teaching Assistant for an academic year for Algebra, Trigonometry, and Differential Equations.

He entered The University of Akron in the fall of 2016. He graduated with a Master's degree in Applied Mathematics where he was a Graduate Assistant for two (2) academic years during his studies.

In the fall of 2018, he entered the Graduate School of The University of Texas at El Paso to pursue a PhD in Computational Science.

Email: wkubin@miners.utep.edu