2023-05-01

Performance Classification Of Ornstein-Uhlenbeck-Type Models Using Fractal Analysis Of Time Series Data.

Peter Kwadwo Asante
University of Texas at El Paso

Recommended Citation

PERFORMANCE CLASSIFICATION OF ORNSTEIN-UHLENBECK-TYPE MODELS

USING FRACTAL ANALYSIS OF TIME SERIES DATA.

PETER KWADWO ASANTE

Doctoral Program in Computational Science

APPROVED:

_____
Maria C. Mariani, Ph.D., Chair

_____
Osei Kofi Tweneboah, Ph.D.

_____
Thompson Sarkodie-Gyan, Ph.D.

_____
Granville Sewell, Ph.D.

_____
Elsa Villa, Ph.D.

_____
Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

This Dissertation is dedicated to

my wife, son, parents and siblings

with love

PERFORMANCE CLASSIFICATION OF ORNSTEIN-UHLENBECK-TYPE MODELS

USING FRACTAL ANALYSIS OF TIME SERIES DATA.

by

PETER KWADWO ASANTE, B.S,M.S.

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

I would like to express my heartfelt gratitude to the Almighty God for his love, mercy, and abundant grace that has sustained me throughout my educational journey.

My advisor and chair of the Mathematical Sciences Department at The University of Texas at El Paso, Dr. Maria Christina Mariani, deserves special recognition for her invaluable guidance, encouragement, enduring patience, and keen interest that led to the successful completion of my dissertation.

I am also grateful to the other members of my committee, Dr. Granville Sewell of the Mathematical Sciences Department and Computational Science Program, Dr. Thompson Sarkodie-Gyan of the Electrical & Computer Engineering Department, and Dr. Elsa Villa of the Department of Education, all at The University of Texas at El Paso, as well as Dr. Osei Kofi Tweneboah of the Department of Data Science at Ramapo College. Their insightful suggestions, comments, and additional guidance were instrumental in completing this work. I also extend my appreciation to Dr. Maria Pia Beccar-Varela for her contributions and support throughout this journey.

Furthermore, I am thankful to the University of Texas at El Paso, the Graduate School, and the staff and professors at the Computational Sciences Program and Mathematical Sciences Department for their unwavering support and dedication to providing me with the resources I needed to complete my degree.

I would also like to express my warmest gratitude to my dear wife, son, parents, siblings, and in-laws for their prayers, love, encouragement, and constant support. I would not have come this far without this support system and I am grateful to God for them. May the blessings of God be upon them.

Finally, I extend my sincere thanks to all colleagues, friends, and loved ones who contributed in various ways to make this work a success.

Abstract

This dissertation aims to assess the performance of Ornstein-Uhlenbeck models by examining the fractal characteristics of time series data from various sources, including finance, volcanic and earthquake events, US COVID-19 reported cases and deaths, and two simulated time series with differing properties. The time series data is categorized as either a Gaussian or a Lévy process (Lévy walk or Lévy flight) by using three scaling methods: Rescaled range analysis, Detrended fluctuation analysis, and Diffusion entropy analysis. The outcomes of this analysis indicate that the financial indices are classified as Lévy walks, while the volcanic, earthquake and COVID-19 data are classified as Lévy flights. The two simulated Brownian motions are classified as Gaussian processes, as expected. The simulation results of the time series using Ornstein-Uhlenbeck models emphasize the need for selecting an appropriate background driving process, combining solutions of Ornstein-Uhlenbeck-type SDEs, and considering the correlations between time series events to improve the performance of the Ornstein-Uhlenbeck type models.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Analyzing time series data is a crucial aspect of research, and the insights gained from these data sets have significantly enhanced various products across various industries and improved the quality of life. One of the key results that researchers aim to identify is whether the time series shows persistence (long-term correlations), randomness, or anti-persistence (short-term memory effect). Persistence, also referred to as long-term memory effect or long-term dependence, refers to a gradual decline of the temporal or spatial correlation function, which is defined as:

$$\gamma_{xy}(\delta) = \langle X(t)Y(t+\delta) \rangle. \tag{1.1}$$

Long-range correlation can impact the statistical characteristics of a time series, such as its power spectral density and autocorrelation function. Therefore, it is crucial to identify and measure the presence of long-range correlation in a time series to ensure proper modeling and analysis. Time series data that exhibit long-range correlations suggest that the evolution of the system is impacted by its previous state over extended periods ([6]-[9],[25]). Verifying the existence of long-range correlation using the formula in equation 1.1 can be challenging due to sensitivity to noise and other factors, leading to the development of various scaling methods ([6]-[9],[10]-[15],[16]-[19]). These methods are commonly used to detect persistence or anti-persistence in financial and geophysical time series data. Examples of scaling methods include rescaled range analysis (R/S), detrended fluctuation analysis (DFA), relative dispersion analysis (RDA), and the newer diffusion entropy analysis (DEA) introduced by Scafetta ([10]-[12]). Scafetta utilized DEA to detect scaling

behavior in DNA sequences. R/S, DFA, and RDA are examples of variance (dispersive) scaling methods and have a scaling index known as the Hurst index, named after Hurst who first studied it in the field of hydrology. DEA, on the other hand, is a PDF scaling method.

Variance scaling methods, however, face challenges with time series data exhibiting unusual behavior. For instance, R/S analysis is generally ineffective for non-stationary time series data, while DFA tends to overestimate the scale factor. As a result, these variance scaling methods have two disadvantages: they can detect scale invariance, but determining the exact value of the exponent is difficult, and they are not suitable for processes with infinite variances, such as Lévy flight [10]. This is why DEA, as a PDF scaling method, is crucial for determining the scaling exponent of time series data. DEA uses the probability density function of the diffusion process obtained from the time series data to find the scaling parameter δ. The advantage of DEA over variance scaling methods is that it can effectively identify the presence of scaling in time series data, whether stationary or nonstationary, without changing the data through detrending and it can be used even for processes with infinite variance [[10]-[12], [20]]. This resolves the shortcomings of variance-based methods.

Research in the fractal analysis of time series, like research in time series modeling and forecasting, has significantly impacted various fields, such as geophysics, health sciences, financial markets, traffic analysis, and bioengineering. The results obtained in these studies have helped in improving the performance of time series models [[10]-[13], [30]]. It is essential to understand the characterization of time series data, as it can prevent incorrect assumptions from affecting the results. If the data follows a Gaussian distribution, some traditional forecasting methods based on this assumption can yield accurate results. On the other hand, if the data follows a Lévy distribution (Lévy walk or Lévy flight), these traditional methods with Gaussian assumptions may provide incorrect forecasts. Further-

more, if the time series follows a Lévy flight distribution, it has an infinite variance, which contradicts the assumption of finite variance in some forecasting models. Thus, knowing the characteristics of time series data can aid in selecting the appropriate forecasting model. The Lévy process is named after the French mathematician Paul Lévy, who made significant contributions to the study of stochastic processes. Lévy processes are widely used in finance, physics, and other fields to model random fluctuations in various phenomena.

There are many prediction methods ranging from linear models to qualitative models. However, some of these models fail to capture the stochastic nature of complex data sets (i.e. data sets with unusual characteristics), which can lead to poor predictions. Hence, many stochastic models have been developed to model these complex systems, including the Monte Carlo simulations, Cox-Ingersoll-Ross (CIR) models, and the Black-Scholes model. The Ornstein-Uhlenbeck stochastic model happens to be another important stochastic model with applications in a variety of fields, from healthcare [[59],[60]], physics and finance [[7],[38],[81]] to name a few. The Ornstein-Uhlenbeck model is a stochastic differential equation that describes the mean-reverting behavior of time series data. It assumes that the underlying process has a drift term that pulls it back toward its mean and a diffusion term that adds noise to the process. This makes the Ornstein-Uhlenbeck model particularly useful for modeling time series data that exhibit mean-reverting behavior, such as stock prices, interest rates, and exchange rates. The model can also be used to estimate the parameters of the underlying process, such as the mean, volatility, and speed of mean reversion.

The Ornstein-Uhlenbeck model, introduced in [47] has been used in many areas of application, including but not limited to areas such as health [64], nanotechnology/ thermodynamics [68], geophysics [7] and Finance [[14] [61] [81]]. Unlike its original formulation, which used Brownian motion as a background driving process [[47] [80]], there have been many extensions or modifications to capture the behavior of data sets that cannot be adequately

3

modeled with Brownian motion [[44],[80]]. Experimental results have shown non-Brownian behavior in many complex real-world systems [[1],[2], [14],[69]]. In fact, according to [69], the statistic of type Lévy is a common phenomenon observed in many different fields including physics, seismology, engineering, etc. Lévy motions constitute one of the fundamental families of random motions with independent and stationary increments. This means that the distribution of the increments is the same for any time interval and that the increments are statistically independent of each other. In Economics, the Ornstein-Uhlenbeck Model is known as the Vasiček model [46]. The Vasicek model is a mathematical model that describes changes in interest rates, by specifying that the instantaneous interest rate follows the SDE

$$d_{x_t} = \alpha(\beta - x_t)dt + \gamma d\lambda_t$$

where λ_t is the Weiner process under the risk-neutral framework modeling the market risk factor [46]. Overall, the Ornstein-Uhlenbeck model is a versatile and widely used tool for modeling complex systems in a variety of fields.

The prediction of stock market trends is an ongoing area of research as it is of great significance for individual investors and economic policymakers. Accurate forecasting of earthquakes, rockslides, and volcanic eruptions is also crucial as these natural disasters can result in massive loss of life and property if not forecasted correctly. For instance, the 1989-1990 eruption of Mount Redoubt in Alaska caused \$240 million in damages and losses [83]. Similarly, the 1964 rockslide in Mt Toc, Italy resulted in the loss of 2000 lives and \$200 million worth of property [75]. The 2011 Tohoku earthquake in Japan resulted in the loss of 18,000 lives. This highlights the importance of research in forecasting such events [1] [2] [72] [73] [74] [75] [76].

In January 2020, the first case of COVID-19 infection was reported in the United States of America, since then the entire year of 2020 and parts of 2021 were devoted to battling

the spread of the COVID-19 virus. In the wake of this unprecedented pandemic, many researchers around the world sought to model the spread of the COVID-19 virus to help officials understand the severity of the situation as well as enact preventive measures to control the spread [[62] [65] [66] [67]]. In reading the literature on these models, it was observed that most of these models were in the class of deterministic compartmental models. As stated earlier, data from these events usually exhibit stochastic behaviors; hence to capture such stochastic trends in the data, we implement a modification of the Ornstein-Uhlenbeck stochastic model. Introducing stochasticity in predicting the spread of COVID-19 would have the advantage of causing the disease to die out in scenarios where deterministic models may predict disease persistence.

In reference to various types of literature, they show occurrences of this phenomenon when the action of one event causes a specific action of another event. It is not surprising that some dependencies appear in the stock market. When studying market trends, we often find positive correlations in the movements of stock portfolios such as the Dow Jones, Nasdaq, Russell, and S&P500. In [75], the authors presented a scenario where the volcanic eruption preceded an earthquake from up to 120 miles away. Needless to say, stock markets across the globe were hit hard by the COVID-19 pandemic.

In [[1] [2]], we showed that both traditional and emerging financial markets are Lévy processes and can be explicitly characterized as Lévy walk processes. In [14], the authors showed that the Ornstein-Uhlenbeck SDE modeled financial and geophysical data better when using the gamma process as the background driving process (BDP) instead of the standard Brownian motion. Also, in [6] the authors used a truncated Lévy model to identify market crashes by analyzing the long-term impact of high-frequency financial markets. Therefore, there is empirical evidence that real-world stochastic systems exhibit non-Gaussian behavior in most cases through the above results and other studies. These empirical results are testaments to the fact that extending the classic Ornstein-Uhlenbeck

process to Ornstein-Uhlenbeck processes with non-Gaussian BDP may be the best course of action for most real-world data, as in [38], the author showed that the Gamma-driven Ornstein-Uhlenbeck model could predict the time for a major earthquake by looking at minor earthquakes up to a certain point in time. Again, in [14] the authors applied a superposed Gamma Ornstein-Uhlenbeck model to both geophysical and financial data, showing that it produced small error margins when compared to the real data, and finally [44] and [80] presented further extensions of the Ornstein-Uhlenbeck model to non-Gaussian processes. Following current research works in extending the Ornstein-Uhlenbeck model, we propose a three-component superposed Inverse-Gaussian(a,b) and a three-component superposed $\Gamma(a, b)$ Ornstein-Uhlenbeck (OU) model in this study. We compare both models with the ordinary OU models and the two-component superposed Inverse-Gaussian(a, b) and $\Gamma(a, b)$ OU models.

This work analyzes the long-term memory of time series data from finance, geophysics, and health sciences using the rescaled range analysis (R/S), the detrended fluctuation analysis (DFA), and the diffusion entropy analysis (DEA). We identify the features of time series data (i.e. if they follow a Gaussian or Lévy distribution) by examining the correlation between the scaling exponent of the DEA to that of the R/S and DFA. After characterizing the various datasets, we model them using the classic Ornstein-Uhlenbeck model, and the two- and three-component superposed Ornstein-Uhlenbeck models. Also, a system of coupled Ornstein-Uhlenbeck stochastic differential equations (SDEs) is modeled with the datasets to establish and validate intra-dependencies (i.e. dependencies in similar events, e.g. stock markets) or interdependencies (i.e. dependencies in different events, e.g. stock markets and earthquakes) to investigate how the predictive performance of the Ornstein-Uhlenbeck model is affected by such dependencies. For our model analysis, if a dataset is characterized by the Lévy walk processes or the Lévy flight processes, we consider the $\Gamma(a, b)$ process or the IG(a,b) process as the background driving process (BDP). By characterizing the data and modeling the overlaps and associated OU-SDEs with different

BDPs, we can observe different scenarios where model performance is affected. For our model performance, we compute the root mean squared errors (RMSE), the mean absolute percentage errors (MAPE), the mean absolute errors (MAE), and the average relative percentage errors (ARPE).

The dissertation is organized into three parts following the conclusion: the first focuses on characterizing data; the second introduces the Ornstein-Uhlenbeck model and its variations; and the final part examines the performance of the Ornstein-Uhlenbeck-type models in various applications. In Chapter 2, the Lévy walk and Lévy flight, as well as the scaling methods used to characterize the time series, are discussed. Chapter 3 uses the scaling methods discussed in Chapter 2 to characterize time series data from different sources as either Gaussian or Lévy processes. The classic Ornstein-Uhlenbeck model (OU model) with modifications and the proposed parameter estimation method is presented in Chapter 4. Chapter 5 presents some applications, and finally, a conclusion is reached in Chapter 6. The results from this dissertation have been published in four peer-reviewed scientific journals.

Part I

Data Characterization

Chapter 2

Methods for Analyzing
Long-Memory-Effects In Time Series

This chapter reviews some existing scaling methods and some relevant research areas in which these methods have been applied. Also, some similarities and differences between the Lévy walk and the Lévy flight in bounded domains will be discussed.

2.1   Lévy Processes

This section begins with a definition of the Lévy process, then the similarities and differences between the Lévy flight and the Lévy walk are outlined. As this chapter explores the characterization of time series data using scaling methods, it is observed that the self-similarity of both the variance scaling and probability density function (PDF) scaling methods discussed in this work lead to the characterization of time series as either Gaussian or Lévy processes.

Definition 1. *A Lévy process is a stochastic process* $\{X_t : t \geq 0\}$ *on* $\mathbf{R}^n$ *if the following conditions are satisfied.*

1 *For any choice of* $n \geq 1$ *and* $0 \leq t_0 < t_1 < ... < t_n$, *the random variables* X_{t_0}, $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}$, ..., $X_{t_n} - X_{t_{n-1}}$ *are independent. The process has independent increments.*

2 $X_0 = 0$

3 The distribution of $X_{s+t} - X_s$ does not depend on s therefore the process has the stationary increments.

4 It is stochastically continuous.

5 There is a $\Omega_0 \in \mathcal{F}(\sigma - algebra)$ with $\mathbb{P}[\Omega_0] = 1$ such that for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous on $t \geq 0$ and has left limits on $t > 0$

Lévy flights and Lévy walks are two examples of random walks; they share some common characteristics but also have significant distinctions [34]. They are two well known stochastic models that exhibit anomalous diffusion. By analyzing two models of stochastic motion in bounded domains, Bartlomiej et al., [34] showed that both Lévy flights and Lévy walks assume a random walker performs long jumps distributed according to a heavy-tailed power law density. However, Lévy walks have continuous paths and finite velocities, while Lévy flights have discontinuous paths and infinite propagation velocities. Lévy flights can serve as an approximation to Lévy walks with an improper prediction of the moments of the jump length distribution.

2.2 Applications of Lévy Processes.

This section highlights recent research areas where Lévy walks and Lévy flights have been utilized.

2.2.1 Volcanic Eruptions [9]

This study focuses on analyzing volcano-seismic data using Lévy flights and wavelet techniques.

Overview

In this work, the authors estimate the scaling parameter α of the normalized truncated Lévy walk, obtained from analyzing volcanic data collected from a seismic station. The

scaling index α of the truncated Lévy walk gave insight into the long-memory effects of the volcanic eruptions. For $\alpha < 2.0$, the release of energy has a long-memory dependence in its evolution.

Conclusion

The authors conclude that the α index for each volcanic eruption is less than 2.0, proving that the evolution of the published data exhibits long-range-dependence. These findings suggest that present information has a strong connection with past information at different scales, enabling the forecasting of future volcanic activity in the area.

2.2.2   An Application in Transportation [7]

Overview

It is not surprising that there is ongoing research on transportation and how to improve it, as it has become an integral part of society. Whether walking, taking public transit, or driving a private car, you find the best way possible to arrive on time with less stress and delays. Therefore, pavement upkeep is a concern that various government organizations and consulting companies have to tackle.

This work focuses on the ability to maintain pavement structures in an acceptable condition from a structural and functional point of view involving many factors that are often ambiguous and change over time [37]. In this work, the authors applied the normalized truncated Lévy walk method (TLW) to flexible pavement performance to predict the evolution of service levels provided by pavement structure and traffic conditions.

Results and Conclusion

Numerical results obtained by the authors in this work show that the TLW provides another way to represent the trends of pavement usefulness. They concluded that the parameters of the TLW function and the time gap could be adjusted to represent the stochastic character

11

of the interplay of factors driving road surface deterioration, maintenance policies, and regular maintenance programs [37]. Such information can describe the lifetime response of pavement structures using larger datasets collected over long periods.

2.3 Methods for the Hurst Exponent

Past research has shown evidence of long-term-memory in many time series data collected from various domains. The ability to detect these phenomena results from applying scaling techniques to these time series. Hurst was a pioneer in this field of long memory characterization, so it is not surprising that Mandelbrot, another famous figure in the field, named the scaling parameter after him.

This section covers the historical background of long-term-dependence, the development of scaling methods, and fascinating applications utilizing time series data from multiple fields.

2.3.1 History of Long-Term-Dependence

After the Industrial Revolution in the 19th century, there was a need to build large reservoirs by damming river basins. The ideal solution to this issue was a dam that never exceeded its capacity or ran dry. Rippl (1883) provided an effective solution to this problem, but was compromised by the requirement to know or assume the future variability of river flow. Hazen (1914) found a breakthrough that led to the birth of stochastic hydrology using the simplest model, which turned out to be a Gaussian process.

In 1965, Hurst presented techniques for investigating fractal characteristics in his book "Long-Term Storage: An Experimental Study". He developed this method while studying reservoirs on the Nile, in an attempt to design reservoirs that would never overflow or empty, given the observed water flow. In 1969, Mandelbrot and Wallis later popularized the term "ideal dam" for such a reservoir. Hurst's work generated a lot of interest and controversy. After almost a decade of debate, Mandelbrot published his first stationary model, called

Fractional Gaussian Noise (FNR), which could explain the Hurst phenomenon. Later, the ARFIMA(p,d,q) model of Hoskin and Granger [35] incorporated long-range-dependence via a fractional difference parameter "d" into the existing ARMA(p,q) model.

2.4 Applications of Scaling Methods

This section presents some applications of the scaling method in various fields. Recent and early studies are presented, and findings and conclusions are briefly discussed.

2.4.1 Hydrology [21]

Overview

Hurst (1951) used a method similar to Rippl's in this study. In his method, Hurst analyzed a specific statistic about a river's cumulative runoff over time in R, called its "adjusted range."

Conclusion

Hurst studied 690 different time series covering 75 different geophysical events across varying quantities, such as river levels, precipitation, temperature, atmospheric pressure, tree rings, silt thickness and sunspots. In each case, Hurst found the statistic is $R/S(n) \propto n^k$.

He then evaluated the variable k and found that it was approximately normal with a mean of 0.72 and a standard deviation of 0.006. However, his approach to estimating k was based on inadequate analysis, namely the assumption of a known proportionality constant, which is the asymptotic law $R/S(n) = (n/2)^k$. This was later resolved when Mandelbrot abandoned Hurst's fixed-point assumption and performed a two-parameter logarithmic regression to obtain the gradient parameter k [35].

Hurst's work showed that the exponent $k = 0.5$ under the Gaussian assumption of independence, which suggested that current hydrological models did not agree with the empirical data. The exponential value k obtained by Hurst was greater than 0.5, suggesting a "long-range-dependence" of the time series, and caused controversy in the hydrological community because of its importance for Hazen's model assuming *i.i.d.* Gaussian process. The discrepancies found in Hurst's work regarding the theory available at the time and his empirical results would later become known as the "Hurst phenomenon," [35].

2.4.2   Stock Market [6]

Overview

Crashes in the stock market have been recorded in the past years where market behavior was previously unpredictable, such as the infamous 2008 global financial downturn. This volatility in the stock market has increased the interest in research aimed at predicting potential crushes from current and past data.

Assuming normal market conditions, we expect all participating agents to have different views. This means that the pricing process does not exhibit long-memory-effects. However, in normal market conditions, when working with high-frequency data, the pricing process is known to be far from the log-normal specification [36]

In this work, by Barany et al., scaling techniques were used to detect market crashes by analyzing long-term-dependence using high-frequency data collected every minute for several stock indices. They also considered the relationship between the Lévy parameter α and the scaling parameter H characterizing self-similarity. Data used in this work were stock market indices from different industries, including entertainment, technology, retail, oil, and finance.

Conclusion

The results of this work showed that most of the stock market data analyzed exhibited long-memory-effects. Of particular interest is the fact that the data showed long-memory-effects on normal trading days without major events. However, when the stock market crashes, the parameter estimating long-term dependence seems to increase, approaching random behavior during the crash. Further, the results show that after a crash, the process exhibited long-memory-effects, thus, returning to normal behavior [6]. This was a very interesting finding by the authors of this study, making it possible to predict the potential for market crashes by examining the behavior of the crash parameters.

2.4.3 DNA Sequencing [10] [11]

Overview

Recent advances in experimental methods of molecular genomics have made vast amounts of genomic data available [12]. In this work, Scafetta et al. wanted to solve the problem of statistical analysis of time series generated by complex dynamics using the Diffusion Entropy Analysis (DEA). DEA is a relatively recent scaling method introduced by Scafetta that, unlike the existing scaling methods, uses the Shannon entropy of the diffusion process derived from a time series to estimate the correct scaling factor.

In this work, Scafetta et al. studied time series data derived from DNA sequences, considering both coding and non-coding sequences. This is because the early articles focused on the controversial question of whether the propert of long-range-dependence is common in both coding and non-coding sequences. This discrepancy may be due to the inability to determine the correct scaling of the time series, which exhibits anomalous behavior because of the limitations of existing methods [12]. Therefore, the authors of this article specifically wanted to show that DEA can detect correct time series scaling independent of anomalous behavior, in contrast to traditional scaling methods.

Conclusion

Using DEA with traditional methods show that traditional methods are prone to erroneous conclusions and that both coding and non-coding DNA sequences produce Lévy statistics in the long term. The results obtained in this study proved to be groundbreaking, fundamentally clarifying the cause of the controversy over coding and non-coding DNA sequences and providing a way to characterize time series. This is what drives one of the main goals of this dissertation, aiming to determine whether, for those time series characterized by a Lévy process, we can further classify the characterization as a Lévy flight or a Lévy walk. As stated earlier, there is a fundamental difference between these two forms of random walk.

2.5 Characterizing Time Series as Gaussian or Lévy Processes

In the previous section, we have seen Lévy walks and Lévy flights and their usefulness in some interesting applications of scaling methods to complex real-world-systems. It is clear how important the Lévy process is in various studies on time series analysis of data and its prediction.

Variance scaling methods are based on the Gaussian assumption and therefore have the potential to lead to erroneous conclusions when used to derive scaling measures for a time series characterized by a Lévy distribution. This is what led Scaffeta to introduce the Diffusion Entropy Analysis (DEA).

In this section, we analyze the long-memory effects of time series and characterize them as Gaussian or Lévy processes. In addition, for the time series that follow Lévy distributions, we classify them further as Lévy walks or Lévy flights. Differentiating the Lévy process will aid in selecting appropriate time series models, which will capture the characteristics of the Lévy walk and the Lévy flight.

2.6 Variance Scaling Methods

Let us now briefly introduce the Rescaled Range Analysis (R/S) [[15],[19],[89],[90]] and the Detrended Fluctuation Analysis (DFA) [[5],[13],[24]]. These two methods are well-known methods with many research papers about them. We refer the reader to the referenced articles for further information on the R/S and the DFA. We will then discuss the Diffusion Entropy Analysis in more detail, including our proposed methods using the self-similarity of the DEA, the R/S, and the DFA to characterize time series as either Gaussian or Lévy processes.

2.6.1 Rescaled Range Analysis

Hurst introduced the rescaled-range (R/S) analysis as part of his study of long-term water level fluctuations in the Nile [21]. Since then, it has become popular and applied in various fields, including transportation analysis, biotechnology, physics, geology, biology, and geophysics.

Mandelbrot coined the name H for the parameter obtained using this method in honor of Hurst and the Hölder. The H parameter (dependency index) represents the relative trend of the time series and is always between 0 and 1, and equal to $\frac{1}{2}$ for a random process. In our work, the case where $0.5 < H < 1$ is of particular interest since it is an indicator of long-memory-effects. However, because of its sensitivity to outliers in the series, the rescaled range analysis is not suitable for analyzing long-range auto-correlations of non-stationary series.

Algorithm [[89] [90]]

Given a time series $X = X_1, X_2, ..., X_n$. The rescaled range is calculated as follows:

i Calculate the mean:

$$m = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

ii Create a mean adjusted series

$$Y_t = X_t - m,$$

for $t = 1, 2, 3, ..., n$.

iii Calculate the cumulative derivative series Z as follows:

$$Z_t = \sum_{i=1}^{t} Y_i,$$

for $t = 1, 2, 3, ..., n$.

iv Create a range series R as follows:

$$R_t = max(z_1, z_2, ..., z_n) - min(z_1, z_2, ..., z_n),$$

for $t = 1, 2, 3, ..., n$.

v Create a standard deviation series S as follows:

$$S_t = \sqrt{\frac{1}{t} \sum_{i=1}^{t} (x_i - m(t))^2},$$

for $t = 1, 2, 3, ..., n$ and where $m(t)$ is the mean of the series of length t.

vi Calculate the rescaled range series $(\frac{R}{S})_t = \frac{R_t}{S_t}$ for $t = 1, 2, 3, ..., n$.

Hurst Exponent from Rescaled Range Analysis

Now to calculate the Hurst exponent from the rescaled range we follow the steps below:

i Average the rescaled range values for each range region

ii Plot $\log(\frac{R}{S})$ and $\log(X)$, where X is the average size of each range of region.

iii The calculated slope of the plotted points gives the Hurst exponent.

For this work, the *polyfit* function in matlab is used to fit the data points to a linear equation from which the slope can be extracted.

2.6.2 Detrended Fluctuation Analysis

Peng et al. proposed a non-trend variance analysis, named the detrended fluctuation analysis (DFA) while examining DNA nucleotides. His goal was to study the self-similarity and long-term-dependence of time series obtained from DNA structures. Since its development, DFA has become a popular method for determining fractal properties and detecting long-range-correlations in non-stationary time series. It has been used, for example, in biology, meteorology, geophysics, and economics [[22]-[29]].

An advantage of the DFA is its ability to distinguish internal time series auto-correlations from those imposed by unusual external trends or non-stationary trends. In other words, this method puts aside non-stationary trends and focuses on the internal structure of the correlation between market fluctuations on different time scales.

Applying the DFA method, the scaling factor α is obtained from the estimate of the slope of the function $F(s)$, which measures the standard deviation from the best linear approximation of the trend signal over the segment length s. The variable function versus s works like a power law. Therefore, it is possible to calculate the value of the exponent α from the slope of the function in a plot of $F(s)$ versus s on a logarithmic scale plot (log-log plot). The DFA exponent α and the Hurst parameter H are related as follows:

$$
H = \begin{cases} \alpha & \text{if } 0 < \alpha < 1 \\ \alpha - 1 & \text{if } \alpha \geqq 1 \end{cases} \tag{2.1}
$$

Algorithm

Suppose N is the length of time series $(y_1, y_2, y_3, ..., y_N)$. To calculate the α exponent of the DFA we proceed as follows.

   i Generate a new time series $M(t)$ of length $N - 1$ from the logarithmic ratio of the time series using:

$$M(t) = \log\left(\frac{y_{t+1}}{y_t}\right), \quad t = 1, 2, ..., N - 1$$

   ii integrate the absolute value of $M(t)$:

$$y(t) = \sum_{i=1}^{t} | M(i) |$$

   iii The integrated time series of length N is divided into m boxes of equal length n with no intersection between them.

   iv Repeat $[i] - [iii]$ but beginning from the end of the series, thus obtaining $2N/n$ boxes. This repetition is done in order to capture leftover values after the 1st 3 steps are performed since the lenght of the series may not be divisible into m boxes without leftover values.

   v Fit a least square line to each box, representing the trend in each box, thus obtaining $y_n(t)$.

   vi Compute the Root Mean Square Fluctuation (RMSF) as follows:

$$F(n) = \sqrt{\frac{1}{2N} \sum_{t=1}^{2N} [y(t) - y_n(t)]^2}$$

After repeating the above steps over all box sizes, we obtain a relation between the box

size n and $F(n)$ as $F(n) \propto n^\alpha$, where α, the slope of the line relating to the log-log plot of $F(n)$ and n is the Hurst exponent.

2.6.3 Diffusion Entropy Analysis

The DEA, based on Shannon's direct estimate of entropy [[10]-[12],[17],[18]], is a PDF scaling method that perceives numbers in time series as the trajectory of the diffusion process [16].

The scaling property of a stationary time series takes the form:

$$p(x,t) = \frac{1}{t^\delta} F\left(\frac{x}{t^\delta}\right). \tag{2.2}$$

where x denotes the diffusion variable, $p(x,t)$ is its probability density function (PDF) at time t, and $0 < \delta < 1$ is the scaling exponent.

The scaling property for the non-stationary time series takes the form

$$p(x,t) = \frac{1}{t^{\delta(t)}} F\left(\frac{x}{t^{\delta(t)}}\right). \tag{2.3}$$

As derived in ([10]-[11]), a diffusion process generated by Lévy walk is characterized by the following relation:

$$\delta = \frac{1}{3 - 2(H,\alpha)} \tag{2.4}$$

In the equation (2.4), (H,α) refers to the scaling factor derived from the two variance scaling methods used in this work. If $\delta = (H,\alpha) = \frac{1}{2}$, the time series can be characterized by Fractional Brownian Motion (FBM), since the variance methods are based subtly on the Gaussian assumption [[21] [25]]. However, if the equation (2.4) holds for $(H,\alpha) \neq \frac{1}{2}$, the time series can be characterized by a Lévy statistic, specifically a Lévy walk. Now, if $\delta \neq (H,\alpha) \neq \frac{1}{2}$ and the equation (2.4) do not hold, the noise can be characterized as Lévy flight.

2.6.4 Estimation Procedure

This subsection describes techniques for estimating the δ scaling factor. First, let us present a summary of Shannon's entropy used to estimate δ.

The Shannon Entropy

A few years after Rudolph Clausius formulated the laws of thermodynamics, he introduced the term entropy. Later in 1872, Ludwig Boltzmann developed an expression for the entropy of the system in a particular state as proportional to the logarithm of the probability of that state. Entropy is an indicator of the lack of information about the scale of an event occurring with probability p [30].

Other types of entropy are Kolmogorov-Sinai entropy, Rényi entropy, and Tsallis entropy [[10] - [12], [30]]. Shannon's entropy measures information about a probability distribution as:

$$S(t) = -\sum_{1}^{N} p_i \log p_i \tag{2.5}$$

For continuous probability distributions, the sum is replaced by the integral. The above equation is used to derive the logarithmic equation used to determine the DEA δ scaling. Below is the process for estimating δ.

- The time series data is first transformed into a diffusion process.

- Shannon's entropy of the diffusion process is calculated. A log-linear equation or log-quadratic equation is derived from the Shannon entropy by substituting equation 2.2 and 2.3 respectively. Simplifying the result from the substitutions, we have the following relation for stationary time series:

$$S(t) = A + \delta ln(t) \tag{2.6}$$

For the non-stationary series, the relation is as follows:

$$S(t) = A + \delta(t)\tau \tag{2.7}$$

where $\delta(t) = \delta_0 + \eta \log(t)$ and $\tau = \log(t)$ with $\eta \log(t) < 1 - \delta_0$. After some simplifications, equation 2.7 becomes

$$S(t) = A + (\delta_0 - K)\log(t) + (1 - \delta_0)(\log(t))^2 \tag{2.8}$$

where $\delta_0 \equiv \delta$ from equation 2.7. Therefore, the scaling exponent δ (or δ_0) can be determined by fitting a log-quadratic model to a non-stationary series and a log-linear model to a stationary series. For $t = 1$ in both equations 2.6 and 2.7, the constant A is the expression $C(1)$.

So δ (or δ_0) can be estimated by the slope of the linear-logarithmic equation or derived by the coefficients of the quadratic-logarithmic equation. For more information on the algorithm used to transform a series into a diffusion process, see [10].

Chapter 3

Time Series Data

This chapter focuses on the characterization process of the time series datasets used in Chapter 5 for our model application. The data presented are from three main areas: stock market datasets, geophysical datasets, and health datasets.

3.1 Stock Market Data

For the stock market data, we obtained five stock indices. We got the Dow Jones, Standard and Poor 500, NASDAQ, and Russel from the US stock market, and from the Japanese stock market, we got the Nikkei225. We obtained the stock market data from Yahoo Finance. For our analysis, we used the reported daily closing values. For our datasets, we considered the period from February 19, 2020, to April 16, 2021. Additionally, for NASDAQ and Nikkei225, we collected data from January 1, 2011, to January 1, 2012. Before using the data, we performed an exploratory data analysis using the R software to remove missing values (NAs) due to market closures on holidays and weekends.

3.2 Volcanic and Earthquake Time Series

Seismic stations that are part of the Seismic Network of the Bezymyanny Volcano Campaign PIRE recorded the volcanic data. For the volcanic dataset, we requested data from ten days before and five days after each eruption. We obtained the volcanic datasets from seismic stations BEZB and BELO. Eruptions one and two originate from BEZB, and eruptions three-eight originate from BELO. For our analysis, we used eruptions two, four, and eight.

Earthquake data used is from the 2011 earthquake in Japan. The data collected captures the seismic intensity before and after the magnitude nine earthquake, termed EQ1 and EQ2, respectively.

3.3   USA COVID-19 Data

We downloaded US COVID-19 data from the New York Times COVID-19 data website [85]. This data comprises cumulative daily reports of cases and deaths in the United States, individual states, and counties. For this work, we only use data covering the entire United States. Because the data is cumulative, we transform it into daily reports of new cases and deaths by subtracting current values from previous values. The period of interest used for analysis is from 19 February 2020 to 16 April 2021.

3.4   Gaussian Processes

Here we are simulating a Wiener process and a fractional Brownian motion (FBM) using R software. To do this, we generate a Gaussian process by injecting the Wiener kernel into the function. We then model a fractional Brownian motion using the R software, with the value of the Hurst exponent set to 0.5. For our simulated fractional Brownian Motion, we use the *fbm* function from the *somebm* package in R. We thus model two different Gaussian processes using two different approaches, with the first approach injecting a Wiener kernel into the R function and the second setting the Hurst exponent at 0.5. As noted in Chapter 2, a Hurst exponent of 0.5 means the time series is a random process and can be characterized by a fractional Brownian motion provided the DEA exponent $\delta = 0.5$. Also, we use the self-similar relation between the scaling methods developed in Chapter 2 and apply the two simulated datasets as controls in the fractal analysis of the time series data. With the self-similar relation between the different scaling methods, we characterize the distribution of the time series data as Gaussian or Lévy. We will subsequently use the selected distribution

of the time series to investigate the performance of our Lévy-driven Ornstein-Uhlenbeck-type models.

3.5 Stationarity of the Time Series

In this section, the Augmented Dickey-Fuller Test (ADF) is used to determine the stationarity of the time series data. For comparison, we implemented the method in R and Python.

3.5.1 Augmented Dickey-Fuller

The ADF test is a type of statistical test called *unit root test*. The test's null hypothesis is that a time series is not stationary (with some time-dependent structure) if we can express it as a unit root. The alternative (rejecting the null hypothesis) is that the time series is stationary.

3.5.2 Stock Indices

After implementing the ADF test to the stock market data, we obtained the following p-values at $\alpha = 0.05$.

Table 3.1: Results from the Stationarity Test on Stock Indices: p-values [1]

Market	p-value
S&P 500	0.4729
Russel	0.9378
Dow Jones	0.2575
Nasdaq	0.9453
Nikkei225	0.6252

3.5.3 Earthquake and Volcanic time series

After implementing the ADF test to the Volcanic and Earthquake time series, we obtain the following results for p-values at $\alpha = 0.05$.

Table 3.2: Results from the Stationarity Test on the Geophysical Data: p-values [1]

Data	p-value
V2	0.6747
V4	0.095
V8	0.4059
EQ1	0.01
EQ2	0.01

3.5.4 US COVID-19 Data

Table 3.3: ADF test applied to the US COVID-19 data

US COVID-19 Deaths	0.7055
US COVID-19 Cases	0.9633

3.5.5 Gaussian Processes

Table 3.4: ADF test applied to the simulated Gaussian processes

Wiener	0.4753
FBM	0.925

3.5.6 Remarks

After running the ADF-test on the datasets at an α level of 0.05, the stock market data, the Covid-19 data, the simulated Gaussian processes, and the volcanic data are all observed to be non-stationary with $p-values > 0.05$. However, the two earthquake data are stationary

with $p - values < 0.05$. We will use the stationarity of the time series in obtaining the δ exponent of the diffusion entropy analysis (DEA), as stated in Chapter 2.

3.6   Scaling Exponents

In this section, the variance scaling methods (Rescaled Range Analysis and Detrended Fluctuation Analysis), and the pdf scaling method (Diffusion Entropy Analysis), are used to compute the scaling exponents of the time series datasets discussed above. Each data set is then characterized as either Gaussian or Lévy (Lévy walk or Lévy flight) processes, using the method described in Chapter 2. $\delta_{Levy}(R/S) = \frac{1}{3-2H}$ and $\delta_{Levy}(DFA) = \frac{1}{3-2\alpha}$.

Table 3.5: Scaling exponents for Financial time series

Market	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
S&P500	0.63	0.65	0.58	0.57	0.59
Russel	0.65	0.7	0.60	0.59	0.63
Dow Jones	0.64	0.70	0.60	0.58	0.63
NASDAQ	0.6	0.6	0.56	0.56	0.56
Nikkei	0.66	0.70	0.62	0.60	0.63

Table 3.6: Scaling exponents of Volcanic and Earthquake Data [1]

Data	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
V2	0.51	0.92	0.934	0.5093	0.8682
V4	0.39	0.66	0.934	0.4509	0.5957
V8	0.504	0.75	0.934	0.5018	0.6684
EQ1	0.4221	0.066	0.15422	0.4638	0.3487
EQ2	0.3149	0.6518	0.77046	0.4219	0.5895

Table 3.7: Scaling exponents of the US COVID-19 data

Data	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
USA Cases	0.8209	0.4950	0.3526	0.7363	0.4975
USA Deaths	0.7068	0.2354	0.3627	0.4396	0.3954

Table 3.8: Scaling exponents for Gaussian Processes

Data	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
Weiner	0.5055	0.498	0.5015	0.5	0.5
FBM	0.4928	0.4992	0.4986	0.5	0.5

3.6.1 Discussion

In Table 3.5, we observe that the scaling exponents suggest a long-term memory effect for the stock indices under consideration. To characterize the stock indices, using the expression $\delta = \frac{1}{3-2(H,\alpha)}$, we compare the value of the DEA scaling exponent to that of both the δ_{Levy}(R/S) and the δ_{Levy}(DFA) and conclude that the relation $\delta = \frac{1}{3-2(H,\alpha)}$ holds. Hence the stock indices can be characterized by a Lévy process, particularly a Lévy walk process. The volcanic data in Table 3.6 exhibits long-memory effects when we observe the scaling of the DEA and DFA. However, due to the non-stationarity of the volcanic time series, the R/S method fails to detect the memory behavior correctly since it concludes that V_2 and V_8 are random while V_4 is anti-persistent. The magnitudes measured before the magnitude nine event in Japan in 2011 (EQ1) show evidence of short-memory behavior. This significant observation shows that the earthquake activity in the region before the magnitude nine earthquake event was abnormal. On the other hand, the magnitudes recorded after the magnitude nine event depict long memory, which indicates a return to normal activity. In terms of characterization, the scaling exponents derived for the three volcanic eruptions and two earthquake datasets do not satisfy the relation $\delta = \frac{1}{3-2(H,\alpha)}$ or $\delta = (H, \alpha) = 0.5$, so they are both classified as Lévy processes, specifically Lévy flight processes. In Table 3.7, both the COVID-19 cases and COVID-19 deaths exhibit short-memory behavior, indicating the possibility of them dying out. In terms of characterization, the COVID-19 cases and COVID-19 deaths do not satisfy the relation $\delta = \frac{1}{3-2(H,\alpha)}$ or $\delta = (H, \alpha) = 0.5$, hence, both are classified as Lévy processes, specifically Lévy flight. Finally, the simulated FBM in Table 3.8 satisfies the relation $\delta = (H, \alpha) = 0.5$, confirming that both simulated datasets are Gaussian processes.

Part II

The Ornstein-Uhlenbeck Model and its Variations.

Chapter 4

Stochastic Processes and The Ornstein-Uhlenbeck Model

This chapter introduces stochastic processes, the Ornstein-Uhlenbeck model, the two-component and three-component superposed Ornstein-Uhlenbeck models, and the coupled Ornstein-Uhlenbeck model used for analysis in this dissertation. In addition, the Gamma and Inverse-Gaussian processes are also briefly discussed.

4.1 Stochastic Process

When we talk of stochastic processes, we refer in simple terms to systems that change over time, i.e., any process describing the evolution in time of a random phenomenon. As evidenced by the literature, stochastic processes occur in a wide range of natural and physical systems, including health sciences, geophysics, finance, and social sciences [[91], [92], [93], [94], [95], [96]]. Common examples of stochastic processes are the simple random walk and the martingale. Let's define a stochastic process below with some relevant terminologies.

Definition 2. *(Stochastic Process) A stochastic, or random, process is a mathematical object usually defined as a family of random variables.*

4.1.1 Simple Random Walk

One common stochastic process is the simple random walk. This will be touched on briefly in this section.

Definition 3. *(Random Walk) Let $\mathbb{Z}$ be the set of integers. We define a random walk on the integers $\mathbb{Z}$ with step function Λ and initial state $x \in \mathbb{Z}$ as a sequence of random variables*

$$X_n = x + \sum_{i=1}^{n} \epsilon_i$$

where ϵ_is are i.i.d. random variables with common distribution Λ.

Definition 4. *(Rademacher distribution) The Rademacher distribution is a discrete probability distribution where a random variate X has a 50% chance of being +1 and a 50% chance of being -1.*

Definition 5. *(Random Walk) A simple random walk is a stochastic process with Rademacher-$\frac{1}{2}$ increments, i.e.,*

$$P(\epsilon_i = 1) = P(\epsilon_i) = -1 = \frac{1}{2}$$

So a simple random walk is defined as a series of fair coin flips. One well-known discrete random process modeled by a simple random walk is the evolution of player wealth, an example of which is the player ruin problem.

Theorem 1. *A simple random walk visits every state $i \in \mathbb{Z}$ infinitely often with probability 1.*

4.1.2 Martingale

This section briefly discusses the martingale, mathematical probability theory, and stochastic processes. The definition of a martingale takes inspiration from a basic concept of gambling. However, its diverse applications have made it a more advanced tool in modern abstract mathematics and have influenced and contributed to other social and scientific fields. To formally define a martingale, we first define two basic concepts: continuous or discrete filtering and an adaptive process [97]].

Definition 6. *(Continuous Filtration) A continuous filtration of a set Ω is a collection $\mathbb{F}_t$ of σ-algebra subsets of Ω such that $\mathbb{F}_s \subset \mathbb{F}_t$ for all $s < t$. For $n \in \mathbb{N}$, a discrete filtration $\mathbb{F}_n$ is similarly defined on a random variable $\{\mathbb{X}_n\}$*

Definition 7. *($\mathcal{F}_t$-measurable) An $\mathcal{F}_t$-measurable random variable is a random variable whose value is known at time t.*

Definition 8. *(Adaptive process) A sequence $\{X_t\}$ of random variables is an adaptive process relative to $\{\mathbb{F}_t\}$, if the random variable $\{X_t\}$ is $\{\mathbb{F}_t\}$-measurable for each t.*

We now define a martingale using a discrete-time-adaptive process and a continuous-time-adaptive process.

Definition 9. *(Martingale) A discrete-time-adapted process $\{\mathbb{X}_n\}$ of integrable random variables is a martingale relative to $\{\mathbb{F}_n\}$ if $\mathbb{E}(\mathbb{X}_n|\mathbb{F}_n) = \mathbb{X}_n$ almost surely for all $n \in \mathbb{N}$*

Definition 10. *(Martingale) A continuous-time-adapted process $\{\mathbb{X}_t\}$ of integrable random variables is a martingale relative to $\{\mathbb{F}_t\}$ if $\mathbb{E}(\mathbb{X}_t|\mathbb{F}_s) = \mathbb{X}_s$ almost surely for all $s < t$*

4.1.3 Standard Brownian Motion

The standard Brownian motion is an example of a continuous-time stochastic process. In 1828, botanist Robert Brown observed the erratic motion of pollen suspended in water, later called Brownian motion. As pointed out by [100], standard Brownian motion is a subclass of the continuous martingale, Markov process, Gaussian process, and Itô diffusion process. Below, we define Brownian motion, represented by [99], and show the situation in which this is standard Brownian motion.

Definition 11. *(Brownian Motion) Let $\{B(t) \geq 0\}$ be a continuous time stochastic process, and let $B(0)$ be the initial distribution, μ the drift vector, and Σ the diffusion matrix. Then the Brownian motion with drift μ and diffusion matrix Σ satisfies the conditions below:*

i For all times $0 \leq t_1 \leq ... \leq t_n$ the random variables $B(t_n) - B(t_{n-1})$, $B(t_{n-1}) - B(t_{n-2}), ..., B(t_2) - B(t_1)$ are independent.

ii The distribution of the increment $B(t + h) - B(t)$ does not depend on t

iii The process $\{B(t) : t \geq 0\}$ has almost surely continuous paths.

iv For every $t \geq 0$ and $h \geq 0$ the increment $B(t + h) - B(t)$ is multivariate normally distributed with mean $h\mu$ and covariance $h\Sigma\Sigma^T$

If the drift vector is $\mu = 0$ and the diffusion matrix Σ is the identity matrix then we simply have the Brownian motion. If in addition, $B(0) = 0$ then we have a standard Brownian motion.
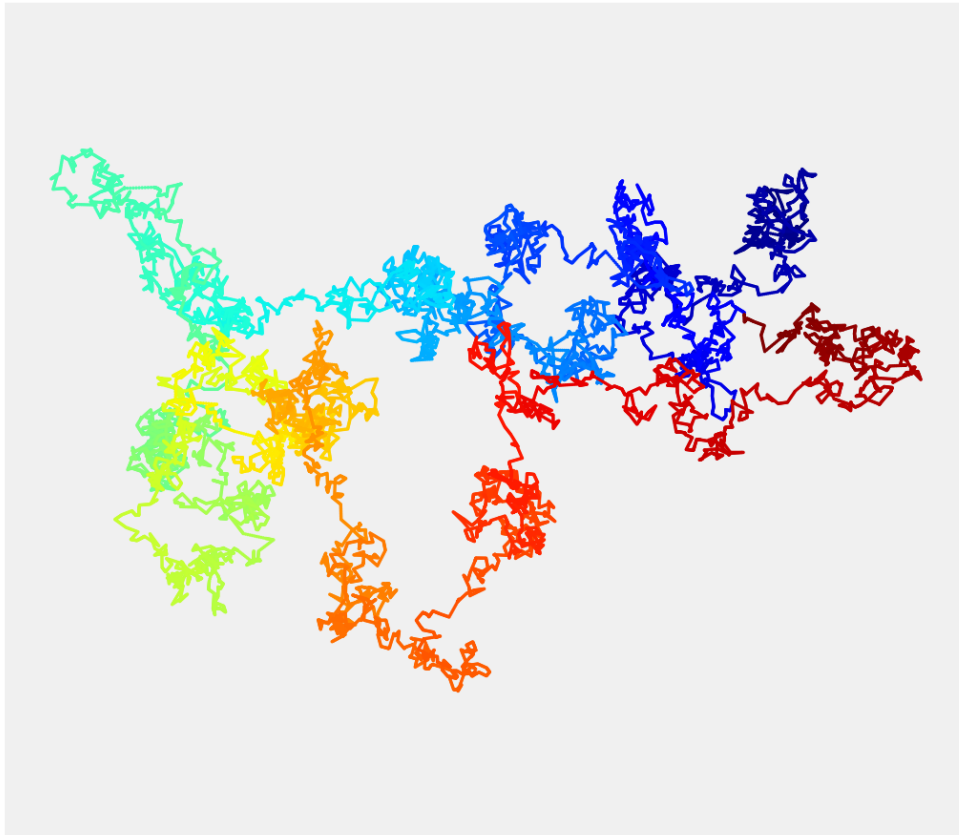
Figure 4.1: The figure of a Brownian motion from [101] using the Python Jupyter Notebook.

4.2 Stochastic Differential Equations

4.2.1 SDEs

The theory of Differential equations is the origin of classical calculus and motivated the creation of calculus and integral calculus. A differential equation is an equation that relates an unknown function and its derivative. In general, differential equations are functional relations.

$$f(t, x(t), x'(t), x''(t), \dots) = 0, \quad 0 \le t \le T \tag{4.1}$$

involving the time t, an unknown function $x(t)$ and its derivative. The solution of the DE is to find a function $x(t)$ which satisfies Equation (4.1).

Now consider the deterministic differential equation:

$$dx(t) = a(t, x(t))dt, \quad x(0) = x_0. \tag{4.2}$$

The easiest way to introduce randomness into this equation is to randomize the initial conditions. The solution $x(t)$ then devolves into a random process $(X_t, t \in [0, T])$ defined as:

$$dX_t = a(t, X_t)dt, \quad X_0(\omega) = Y(\omega). \tag{4.3}$$

The equation (4.3) is called a random differential equation. Random differential equations are deterministic equations with perturbed initial conditions. Equation (4.3) is not a complete stochastic differential equation.

Definition 12. *(Stochastic Differential Equation) A stochastic differential equation (SDE) is defined as a deterministic differential equation that is perturbed by random noise.*

In general, an SDE is formulated as:

$$dX(t, \omega) = f(t, X(t, \omega))dt + g(t, X(t, \omega))dW(t, \omega) \qquad (4.4)$$

where ω denotes that $X = X(t, \omega)$ is a random variable and possesses the initial condition $X(0, \omega) = X_0$ with probability one.

As an example we have

$$dY(t, \omega) = \mu(t)dt + \sigma(t)dW(t, \omega),$$

The white noise process is denoted by $dW(t, \omega)$. White noise can be viewed as a derivative of Brownian motion. This means that white noise, defined as the derivative of $dW(t, \omega)/dt$ of $W(t)$, does not exist in the conventional sense. Recall that the Brownian motion is a continuous stationary random process $W(t)$ with independent increments and a Gaussian random variable with mean 0 and variance t for each t and $W(t)$.

4.2.2 Itô Calculus

To solve the Ornstein-Uhlenbeck equation, we compute the integral of the stochastic process. That is, the solution of the SDE involves the integrand and integrator, which are stochastic processes. Therefore, classical methods of calculus cannot be used to solve SDEs. The Japanese mathematician Kiyoshi Itô advanced a solution to SDE by developing Itô's theory of calculus, which extended calculus methods to random processes. In this section, the Itô integral is presented, and the existence and uniqueness of theorems for SDE are studied. Before explaining the properties of the Itô integral, we first define some relevant terms.

Definition 13. *(Probability Space): A probability space is a triplet $(\Omega, \mathcal{F}, P)$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω and P is a probability measure $P : \mathcal{F} \to [0, 1]$.*

Itô Integrals

To proceed with the theory of the Itô integral, let's first find a mathematical interpretation of the "noise" term in the SDE. Consider the SDE below:

$$\frac{dN}{dt} = (b(t) + \text{ "noise" })N(t) \tag{4.5}$$

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{ "noise"} \tag{4.6}$$

where $b(t, X_t)$ and $\sigma(t, X_t)$ are some given deterministic functions. Considering a 1-dimensional case of the "noise" term leads to a description of the "noise" term by some stochastic process W_t, thus

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot W_t \tag{4.7}$$

Below are the properties of the stochastic process W_t :

(i) For $t_1 \neq t_2$, the stochastic processes W_{t_1} and W_{t_2} are independent.

(ii) The stochastic process $\{W_t\}$ is stationary, i.e. the joint distribution of $\{W_{t_1+t}, \ldots, W_{t_k+t}\}$ does not depend on t.

(iii) $E[W_t] = 0$ for all t.

It turns out that no suitable stochastic process exists that meets properties (i) and (ii), implying that such a W_t cannot have continuous paths. However, we can represent W_t as a generalized stochastic process called the *white noise process*. Here, generalized means that the process can be constructed as a probability measure on the space of tempered distributions on $[0, \infty)$, and not as a probability measure on the much smaller space $\mathbb{R}^{[0,\infty)}$.

If we let $0 = t_0 < t_1 < \ldots < t_m = t$ we can discretize (4.7) as follows:

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)W_k\Delta t_k \tag{4.8}$$

where $X_j = X(t_j)$, $W_k = W_{t_k}$ and $\Delta t_k = t_{k+1} - t_k$

Replacing $W_k \Delta t_k$ by $\Delta V_k = V_{t_{k+1}} - V_{t_k}$ in (4.8) where $\{V_t\}_{t \geq 0}$ is a suitable stochastic process. The properties (i), (ii), and (iii) on W_t suggest that V_t should be stationary. The only process with continuous paths is the Brownian motion B_t. Therefore we substitute V_t with B_t in (4.8) to obtain:

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j) \delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j) \Delta B_j \tag{4.9}$$

Assuming that the limit of the right hand side of (4.9) exist when $\Delta t_j \to 0$, then applying the usual integration notation we obtain:

$$X_t = X_0 + \int b(s, X_s) ds + \int \sigma(s, X_s) dB_s \tag{4.10}$$

where the first integral on the right-hand side is a Riemann integral, and the second one is an Itô stochastic integral. We would adopt as a convention that (4.9) really means that $X_t = X_t(\omega)$ is a stochastic process satisfying (4.10):

We will proceed to prove the existence of

$$\int_0^t f(s, \omega) dB_s(\omega), \tag{4.11}$$

where $B_t(\omega)$ is a 1- dimensional Brownian motion starting at the origin, for a wide class of functions $f : [0, \infty] \times \Omega \to \mathbb{R}$.

Definition 14. *(Itô Integral) Let $f \in V(S, T)$. Then the Itô integral of f is defined by*

$$\int_s^T f(t, \omega) dB_t(\omega) = \lim_{x \to \infty} \int_s^T \phi_n(t, \omega) dB_t(\omega), \tag{4.12}$$

where ϕ_n is a sequence of elementary functions such that

$$\mathbb{E}\left[\int_s^T (f(t, \omega) - \phi_n(t, \omega))^2 dt \right] \to 0 \quad as \quad n \to \infty. \tag{4.13}$$

38

From Definition 14, we get the following,

Corollary 1 (The Itô Isometry).

$$\mathbb{E}\left[\left(\int_s^T f(t,\omega)dB_t\right)^2\right] = \mathbb{E}\left[\left(\int_s^T f^2(t,\omega)dt\right)\right] \quad \forall \quad f \in V(S,T). \tag{4.14}$$

Corollary 2 (The Itô Isometry). *If $f(t,\omega) \in V(S,T)$ and $f_n(t,\omega) \in V(S,T)$ for $n = 1,2,\ldots$ and $\mathbb{E}\left[\int_s^T f_n(t,\omega) - f(t,\omega)dt\right] \to 0$ as $n \to \infty$, then*

$$\int_s^T f_n(t,\omega)dB_t(w) \to \int_s^T f(t,\omega)dB_t(w) \quad in \quad L^2(P) \quad as \quad n \to \infty. \tag{4.15}$$

Theorem 2 (Integration by parts). *Suppose $f(s,\omega) = f(s)$ only depends on s and that f is continuous and of bounded variation in $[0,t]$. Then*

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_s df_s. \tag{4.16}$$

Properties of the Itô integral

Theorem 3. *Let $f, g \in V(S,T)$ and let $0 \le S < U < T$. Then*

1. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$.

2. $\int_S^T (cf + g)dB_t = \dot{c} \int_S^T f dB_t + \int_S^T g dB_t$, for $c \in \mathbb{R}$.

3. $\mathbb{E}\left[\left(\int_S^T f dB_t\right)\right] = 0$.

4. $\int_S^T f dB_t$ is $\mathcal{F}_T-$ measurable.

Another essential fact of the Itô integral is that it is a martingale.

4.2.3   Existence and Uniqueness Theorems for SDEs

The existence and uniqueness theorems are important results in the theory of stochastic differential equations (SDEs). The theorems provide conditions under which an SDE has a

unique solution, and they are essential for studying the properties of stochastic processes.

Given the general form of an SDE below:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t$$

where X_t is the stochastic process, W_t is a Wiener process, f and g are deterministic functions, and dt and dW_t are the infinitesimal increments of time and Brownian motion, respectively. We state the existence and uniqueness theorems for SDEs as follows:

Theorem 4. *(Existence Theorem): If f and g are Lipschitz continuous and satisfy certain growth conditions, then there exists a solution to the SDE.*

Theorem 5. *(Uniqueness Theorem): If f and g are globally Lipschitz continuous, then there exists a unique solution to the SDE.*

Below, we list some examples of SDEs satisfying the existence and uniqueness theorems:

(i) Geometric Brownian Motion: $dS_t = \mu S_t dt + \sigma S_t dW_t$, where S_t is the stock price, μ is the drift (the expected rate of return), σ is the volatility (the standard deviation of the rate of return), and W_t is the Wiener process.

(ii) Ornstein-Uhlenbeck Process: $dX_t = \theta(\mu - X_t)dt + \sigma dW_t$, where X_t is the process, θ is the speed of mean reversion, μ is the mean, σ is the volatility, and W_t is the Wiener process.

(iii) Heston Model: $dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^1$, $dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t^2$, where S_t is the stock price, μ is the drift, v_t is the variance, κ is the rate of mean reversion, θ is the long-term variance, σ is the volatility of volatility, and W_t^1 and W_t^2 are two correlated Wiener processes.

4.2.4   Proof of Existence and Uniqueness Theorems

To prove the existence and uniqueness theorems of SDEs, we first define Lipschitz continuity Gronwall's inequality, and Gronwall-Bellman's inequality.

Lipschitz continuity

Lipschitz continuity is a type of mathematical continuity that characterizes functions that do not vary too much over small distances. It is a useful concept in many areas of mathematics, including analysis, differential equations, and optimization. It provides a way to quantify the smoothness of a function and to establish bounds on the behavior of solutions to equations involving the function.

Definition 15. *(Lipschitz continuity): A function $f(x)$ is said to be Lipschitz continuous if there exists a constant $K \geq 0$ such that for any two points x_1 and x_2 in the domain of f:*

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2| \tag{4.17}$$

In other words, the absolute difference between the values of f at any two points in its domain is no larger than K times the distance between the points. This means that as the distance between two points in the domain of f gets smaller, the difference between their function values also gets smaller at a rate that is at most proportional to the distance. The constant K is called the Lipschitz constant of f. In the case of global Lipschitz continuity, the above holds for all points in the domain of f. The constant K in this case is called the global Lipschitz constant of f.

Gronwall's inequality

Gronwall's inequality is a mathematical inequality that provides a bound on the growth of a nonnegative function based on its initial value and the integral of another function. The inequality is named after the Swedish mathematician T.H. Gronwall. Gronwall's inequality has many applications in mathematics and physics, including the analysis of differential equations, the study of growth and decay processes, and the derivation of various estimates and bounds. It is often used in conjunction with other mathematical techniques, such as the method of successive approximations and the Picard-Lindelöf theorem, to establish the existence and uniqueness of solutions to various types of equations.

Definition 16. *(Gronwall's inequality): The simplest form of Gronwall's inequality states that if $y(t)$ is a nonnegative function that satisfies the inequality:*

$$y(t) \leq a + \int_0^t b(s)y(s)ds \qquad (4.18)$$

for some nonnegative constants a and $b(t)$, then:

$$y(t) \leq ae^{\int_0^t b(s)ds} \qquad (4.19)$$

In other words, the function $y(t)$ is bounded above by an exponential function of the integral of $b(t)$ up to time t.

Gronwall-Bellman's Inequality

The Gronwall-Bellman inequality is a generalization of Gronwall's inequality that provides a bound on the growth of a function that depends on another function and a set of inequalities. It is named after the mathematicians T.H. Gronwall and Richard E. Bellman.

Definition 17. *(Gronwall-Bellman's Inequality): The inequality can be stated as follows: Let $y(t)$ and $z(t)$ be non-negative functions that satisfy the inequalities:*

$$y(t) \leq a + \int_0^t b(s)y(s)ds + \int_0^t c(s)z(s)ds$$

and

$$z(t) \leq d + \int_0^t e(s)z(s)ds$$

for some non-negative constants $a, b(t), c(t), d,$ and $e(t)$. Then, for any $t \geq 0$, we have:

$$y(t) \leq A(t)\left(a + \int_0^t c(s)dA(s)ds\right)$$

where $A(t) = \exp\left(\int_0^t b(s) + e(s)dA(s)ds\right).$

In other words, the function $y(t)$ is bounded above by an exponential function of the integrals of $b(t)$ and $e(t)$ up to time t, multiplied by a constant that depends on a and the integral of $c(t)$ times the exponential function of the integral of $e(t)$ up to time t.

Proving the existence and uniqueness theorems for stochastic differential equations (SDEs) requires some technical mathematical machinery, hence we present below the main ideas of the proof.

Proof. (Existence Theorem): Suppose that $f(x,t)$ and $g(x,t)$ are Lipschitz continuous in x with a constant K and satisfy certain growth conditions. We want to show that there exists a solution to the SDE:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t$$

First, we consider the truncated SDE:

$$dX_t^k = f(X_t^k, t)dt + g(X_t^k, t)dW_t$$

where X_t^k satisfies $|X_t^k| \leq k$. By the Lipschitz continuity of f and g, we can show that there exists a solution to this SDE on a finite time interval $[0, T]$ with probability 1. Next, we use the Gronwall-Bellman inequality to show that the solution is globally bounded with probability 1. Finally, we take the limit as $k \to \infty$ to show that there exists a solution to the original SDE. $\square$

Proof. (Uniqueness Theorem): Suppose that $f(x,t)$ and $g(x,t)$ are globally Lipschitz continuous in x with a constant K. We want to show that there exists a unique solution to the SDE:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t$$

To prove uniqueness, we use the Itô's formula to derive an expression for the difference between two solutions $X_t^{(1)}$ and $X_t^{(2)}$:

$$d(X_t^{(1)} - X_t^{(2)}) = [f(X_t^{(1)}, t) - f(X_t^{(2)}, t)]dt + [g(X_t^{(1)}, t) - g(X_t^{(2)}, t)]dW_t$$

Using the Lipschitz continuity of f and g, we can show that there exists a constant $C > 0$ such that:

$$d|X_t^{(1)} - X_t^{(2)}|^2 \leq 2C|X_t^{(1)} - X_t^{(2)}|^2 dt$$

By applying Gronwall's inequality, we obtain:

$$|X_t^{(1)} - X_t^{(2)}|^2 \leq e^{2Ct}|X_0^{(1)} - X_0^{(2)}|$$

which shows that the difference between the solutions is exponentially bounded by the initial difference. Therefore, if $X_0^{(1)} = X_0^{(2)}$, then $X_t^{(1)} = X_t^{(2)}$ for all t. This proves the uniqueness of the solution.

$\square$

In summary, the existence and uniqueness theorems for SDEs provide important results for studying the properties of stochastic processes. The theorems require certain regularity conditions on the coefficients of the SDE, and the proofs rely on mathematical tools such as Gronwall's inequality and Itô's formula.

4.3   Ornstein-Uhlenbeck Model (OU)

With the information we have, we are now at a good point to introduce the main model for this dissertation, namely the Ornstein-Uhlenbeck model. Since the Ornstein-Uhlenbeck model was proposed in 1930, various researchers in different fields have utilized it, be it in finance, geophysics, survival models, etc. [[7] [14] [32] [38] [46] [47] [49] [71]]. According to

[49] the Ornstein-Uhlenbeck process is a natural model to consider in a biological context since it stabilizes around some equilibrium point. When the OU model was proposed by Ornstein and Uhlenbeck, it was an alternative to the Brownian Motion, and thus in its original presentation, the background driving process (BDP) was a standard Brownian Motion.

The Gaussian OU process can be defined as the solution to the stochastic differential equation:

$$dX_t = \lambda(m - X_t)dt + \alpha dB_t, \ \ t > 0, \tag{4.20}$$

where λ, m and α are real constants and B_t is a standard Brownian Motion on $\mathbb{R}$. The initial value X_0 is a random variable independent of $(B_t)_{t\geq0}$. It has been shown that the stochastic integral:

$$X_t = m(1 - e^{-\lambda t}) + \alpha e^{-\lambda t} \int_0^t e^{\lambda s}dB_s + X_0 e^{-\lambda t}, \ \ t \geq 0, \tag{4.21}$$

satisfies 4.20 for any λ, m, α and choice of X_0. The solution X as defined in equation 4.21 is the unique, strong Markov solution to 4.20 [44].

As stated earlier, empirical results have shown that many stock indices deviate from normalcy and hence modeling with the ordinary Gaussian OU may result in poor forecasts a.s. Thus, a modification of the Gaussian OU model through replacing the Weiner process in 4.20 with a Lévy process has been developed and applied in various literature. To define the Lévy OU process let (ϵ, η) be a bivariate Lévy process and define:

$$X_t = m(1 - e^{-\epsilon_t}) + e^{-\epsilon_t} \int_0^t e^{\epsilon_s-}d\eta_s + X_0 e^{-\epsilon_t}, \ \ t \geq 0, \tag{4.22}$$

where X_0 is independent of $(\epsilon_t, \eta_t)_{t\geq0}$ and assumed F_0-measurable. Equation 4.22 is the generalized OU model (GOU) and seems to have been first considered by Carmona, Petit, and Yor (1997) as well as being implicit in Haan and Karandukar (1989) [[44], [50], [51]].

Now setting $m = 0$, $\alpha = 1$ and replacing B_t with a Lévy process $Z_{\lambda t}$ we get the SDE:

$$dX_t = -\lambda X_t dt + dZ_{\lambda t}, \quad X_0 > 0, \quad \lambda \in \mathbb{R}^+. \tag{4.23}$$

The solution for equation 4.23 is given by the OU process:

$$X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dZ_{\lambda s}, . \tag{4.24}$$

4.3.1 Superposed Ornstein-Uhlenbeck Model

Using the idea of superposition of solutions to ODEs, the OU process in 4.24 is redefined as a sum of m independent Ornstein-Uhlenbeck processes [58] as :

$$X_t = \Sigma_{i=1}^m w_i e^{-\lambda_i t} X_0 + \int_0^t \Sigma_{i=1}^m w_i e^{-\lambda_i(t-s)} dZ_{\lambda_i s}, \tag{4.25}$$

where a $\Sigma_{i=1}^m w_i = 1$.

For this work, we consider a two-component and a three-component model of 4.25 which results in

$$X_t = w_1 X_{t_1} + w_2 X_{t_2},$$

and

$$X_t = w_1 X_{t_1} + w_2 X_{t_2} + w_3 X_{t_3},$$

with $\sum_i w_i \approx 1$.

4.4 Parameter Estimation

This section presents methods for estimating the λ parameters and weights for both the two-component and three-component OU models. This is achieved through the application

46

of an iterative approach that uses the data to estimate λ values for various lags of the auto-correlation function. The auto-correlation function is then used together with the estimated λ values to derive a matrix system, which is used to estimate the weights (w_i). The Ljung-Box test statistic and the ARIMA model will also be discussed briefly in this section.

4.4.1 Ljung-Box Statistic

The Ljung-Box test is commonly used in time series analysis to check the adequacy of a model and to determine the appropriate order of an autoregressive integrated moving average (ARIMA) model, as well as in the diagnosis of the residuals of the model. We state the null and alternate hypothesis below:

H_0 : The residuals are independently distributed.

H_A : The residuals are not independently distributed; they exhibit serial correlation.

The independence of residuals means that the error terms of the dependent variable at different time points are not related to each other. This assumption is usually checked by examining the autocorrelation function (ACF) of the residuals. If the residuals are independent, the ACF will be close to zero for all lags. If the ACF shows significant values at one or more lags, it indicates that the residuals are not independent, and some kind of correlation exists. Thus, if the p-value is less than a specified significance level (e.g., 0.05), then the null hypothesis of independently distributed residuals is rejected, and it is concluded that there is evidence of residual autocorrelation. In this case, further diagnostic checks may be needed to improve the model fit. If the p-value is greater than the significance level, then the null hypothesis is not rejected, and it is concluded that the residuals are independently distributed. This is a desirable property for a time series model, as it indicates that the model has captured all of the relevant information in the data and is not leaving any systematic patterns unexplained.

4.4.2 Autoregressive Integrated Moving Average (ARIMA)

To ensure that the matrix A is invertible, there is the need to choose appropriate lags for each λ_i. Thus, we fit an ARIMA model to the data and perform a Ljung-box statistic of the fitted ARIMA model to select the most significant lag based on the p-value at a 5% significant level. The ARIMA model assesses the significance of one dependent variable to other changing variables. We iterate the ARIMA model over some lags and perform the Ljung-Box on each fit to examine the null hypothesis of independence in our time series. We then select the lags that give a p-value > 0.05 as significant lags. The three lags chosen are then used in computing the λ_is. This process is performed with R-software. See [103] and [104] for more information on the ARIMA model and Ljung-Box statistic.

4.4.3 Two-Component OU Model

To solve for the parameters λ_1, λ_2, w_1 and w_2 we proceed as follows.

Consider the autocorrelation function at lags k and $k + h$:

$$
\begin{aligned}
\rho(k) =&w_1 e^{-\lambda_1|k|} + w_2 e^{-\lambda_2|k|} \\
\rho(k + h) =&w_1 e^{-\lambda_1|k+h|} + w_2 e^{-\lambda_2|k+h|}
\end{aligned}
\tag{4.26}
$$

Assume $\lambda_1 = \lambda_2$, then from the first expression in equation 4.26

$$
\begin{aligned}
\rho(k) =&(w_1 + w_2)e^{-\lambda_1|k|} \\
\lambda_1 =& -\frac{\log(\rho(k))}{|k|}
\end{aligned}
\tag{4.27}
$$

and from the second expression in equation 4.26,

$$
\rho(k + h) =(w_1 + w_2)e^{-\lambda_2|k+h|} \tag{4.28}
$$

$$
\lambda_2 = -\frac{\log(\rho(k + h))}{|k + h|} \tag{4.29}
$$

To solve for the weights, we derive a matrix equation from equation 4.26 as follows :

$$A = \begin{pmatrix} e^{-\lambda_2|k|} & e^{-\lambda_2|k|} \\ e^{-\lambda_2|k|} & e^{-\lambda_2|k+h|} \end{pmatrix} \tag{4.30}$$

$$b = \begin{pmatrix} \rho(k) \\ \rho(k+h) \end{pmatrix} \tag{4.31}$$

$$W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \tag{4.32}$$

From equation 4.30, 4.31 and 4.32, we can solve for the weights as : $W = A^{-1}b$, with $\sum W \approx 1$.

Where h is a shift from lag k. We fit an ARIMA model to the data, then we perform a Ljung-box statistic of the fitted ARIMA model to choose the most significant lag based on the p-value at a 5% significant level.

4.4.4   Three-Component OU Model

For the three-component case we need to estimate λ_1, λ_2, λ_3, w_1, w_2 and w_3. We proceed as follows:

Consider the auto-correlation function at lags k, $k + h_1$ and $k + h_2$:

$$
\begin{aligned}
\rho(k) &= w_1 e^{-\lambda_1|k|} + w_2 e^{-\lambda_2|k|} + w_3 e^{-\lambda_3|k|} \\
\rho(k+h_1) &= w_1 e^{-\lambda_1|k+h_1|} + w_2 e^{-\lambda_2|k+h_1|} + w_3 e^{-\lambda_3|k+h_1|} \\
\rho(k+h_2) &= w_1 e^{-\lambda_1|k+h_1|} + w_2 e^{-\lambda_2|k+h_1|} + w_3 e^{-\lambda_3|k+h_2|}
\end{aligned} \tag{4.33}
$$

Assume $\lambda_1 = \lambda_2 = \lambda_3$ then from the first expression in equation 4.33

$$\rho(k) = (w_1 + w_2 + w_3)e^{-\lambda_1|k|} \tag{4.34}$$

$$\lambda_1 = -\frac{\log(\rho(k))}{|k|} \tag{4.35}$$

from the second expression in equation 4.33

$$\rho(k + h_1) = (w_1 + w_2 + w_3)e^{-\lambda_2|k+h_1|} \tag{4.36}$$

$$\lambda_2 = -\frac{\log(\rho(k + h_1))}{|k + h_1|} \tag{4.37}$$

and from the third expression in equation 4.33

$$\rho(k + h_2) = (w_1 + w_2 + w_3)e^{-\lambda_3|k+h_2|} \tag{4.38}$$

$$\lambda_3 = -\frac{\log(\rho(k + h_2))}{|k + h_2|} \tag{4.39}$$

whereas in the case of the two-component model, h_1 and h_2 are shifts from lag k. To solve for the weights, we derive a matrix equation from 4.33 as follows :

$$A = \begin{pmatrix} e^{-\lambda_1|k|} & e^{-\lambda_2|k|} & e^{-\lambda_3|k|} \\ e^{-\lambda_1|k+h_1|} & e^{-\lambda_2|k+h_1|} & e^{-\lambda_3|k+h_1|} \\ e^{-\lambda_1|k+h_2|} & e^{-\lambda_2|k+h_2|} & e^{-\lambda_3|k+h_2|} \end{pmatrix} \tag{4.40}$$

$$b = \begin{pmatrix} \rho(k) \\ \rho(k + h_1) \\ \rho(k + h_2) \end{pmatrix} \tag{4.41}$$

50

$$W = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \quad (4.42)$$

From equation 4.40, 4.41 and 4.42, we can solve for the weights as : $W = A^{-1}b$ with $\sum W \approx 1$.

4.4.5 SDE Approximation: Euler-Maruyama and Euler-Milstein Methods

The Euler-Maruyama and Euler-Milstein methods are two widely used numerical techniques for solving stochastic differential equations (SDEs), which are differential equations that involve random noise.

The Euler-Maruyama method is named after Leonhard Euler, a Swiss mathematician, and Gisiro Maruyama, a Japanese mathematician, who independently developed the method in the 1950s. The method is an extension of the classical Euler method for ordinary differential equations and is based on approximating the solution of the SDE at each time step using a Taylor expansion. The Euler-Maruyama method is relatively simple and computationally efficient but can suffer from numerical instability and produce inaccurate results for certain types of SDEs.

The Euler-Milstein method is named after the Russian mathematician Grigori Milstein, who developed the method in the 1970s. The method is a modified version of the Euler-Maruyama method that includes an additional correction term to account for the second-order effects of the random noise. The method is more accurate than the Euler-Maruyama method for certain types of SDEs, particularly those with strong non-linearities or high volatility. However, it is also more computationally expensive, since it requires the evaluation of additional partial derivatives.

Since their development, the Euler-Maruyama and Euler-Milstein methods have become two of the most widely used numerical techniques for solving SDEs, and have found applications in a wide range of fields, including finance, physics, chemistry, and biology. The methods are also the basis for many more advanced numerical techniques for solving SDEs, such as the stochastic Runge-Kutta methods and the multi-level Monte Carlo methods.

4.4.6   Euler-Maruyama

The Euler-Maruyama method is a numerical technique used to approximate solutions to stochastic differential equations. It is an extension of the classical Euler method for ordinary differential equations. The Euler-Maruyama method is given by the following formula:

$$Y_{n+1} = Y_n + f(X_n, Y_n)\Delta t + g(X_n, Y_n)\Delta W_n + \frac{1}{2}g(X_n, Y_n)f'(X_n, Y_n)(\Delta W_n)^2 \qquad (4.43)$$

where Y_n is the numerical approximation to the solution of the stochastic differential equation at time t_n, $f(X_n, Y_n)$ and $g(X_n, Y_n)$ are the drift and diffusion coefficients of the stochastic differential equation, respectively, Δt is the time step size, ΔW_n is a random variable with mean 0 and variance Δt, and $f'(X_n, Y_n)$ is the partial derivative of f with respect to Y evaluated at (X_n, Y_n).

The term ΔW_n represents the increment of a Wiener process, which is a continuous-time stochastic process with independent and normally distributed increments. The term $\frac{1}{2}g(X_n, Y_n)f'(X_n, Y_n)(\Delta W_n)^2$ is a correction term that accounts for the fact that the Wiener process has non-zero quadratic variation.

The Euler-Maruyama method is a simple and computationally efficient method for approximating solutions to stochastic differential equations. However, it can suffer from numerical instability and can produce inaccurate results for certain types of stochastic differential equations.

Algorithm

Consider the SDE in equation 4.43 with initial condition $Y_0 = y_0$ and all parameters as previously described. Suppose that we wish to solve this SDE at some interval of time $[0, T]$. The steps below describe the Euler-Maruyama algorithm:

(i) Set the initial condition Y_0.

(ii) Set the time step size Δt and the number of time steps N.

(iii) For $n = 0, 1, 2, ..., N-1$: Generate a random variable ΔW_n with mean 0 and variance Δt. Compute $Y_{n+1} = Y_n + f(X_n, Y_n)\Delta t + g(X_n, Y_n)\Delta W_n + \frac{1}{2}g(X_n, Y_n)f'(X_n, Y_n)(\Delta W_n)^2$.

(iv) Return the sequence of approximations $Y_0, Y_1, Y_2, ..., Y_N$.

4.4.7 Euler-Milstein

The idea in the Euler-Milestein scheme is to consider expansions on the coefficients b and σ of the SDE. This method is applied when the coefficients of the process are functions of only the main process i.e. do not depend on time. The Milstein scheme is more accurate than the Euler-Maruyama method for certain types of stochastic differential equations, particularly those with strong non-linearities or high volatility.

The general form of a one-dimensional SDE is:

$$dY_t = f(Y_t, t)dt + g(Y_t, t)dW_t \tag{4.44}$$

where Y_t is the unknown stochastic process, $f(Y_t, t)$ is the drift coefficient, $g(Y_t, t)$ is the diffusion coefficient, dW_t is the increment of a Wiener process (a standard Brownian motion), and t is time. The problem is to find the solution Y_t of the SDE for $t \in [0, T]$, given some initial condition Y_0.

Algorithm

The Euler-Milstein method approximates the solution of the SDE at each time step $t_n = n\Delta t$, where Δt is the time step size and n is an integer. The method involves the following steps:

(i) Set the initial condition $Y_0 = y_0$.

(ii) For $n = 0, 1, 2, ..., N - 1$, do the following:

(iii) Compute $Y_{n+1} = Y_n + f(Y_n, t_n)\Delta t + g(Y_n, t_n)\Delta W_n + \frac{1}{2}g(Y_n, t_n)f'(Y_n, t_n)(\Delta W_n)^2 + \frac{1}{2}g(Y_n, t_n)^2(f''(Y_n, t_n)(\Delta W_n)^2 - f'(Y_n, t_n)\Delta t)$

(iv) Where $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ is a random variable with mean 0 and variance Δt, $f'(Y_n, t_n) = \frac{\partial f}{\partial Y}(Y_n, t_n)$, and $f''(Y_n, t_n) = \frac{\partial^2 f}{\partial Y^2}(Y_n, t_n)$ are the first and second partial derivatives of the drift coefficient with respect to Y, evaluated at (Y_n, t_n).

(v) Return the sequence of approximations $Y_0, Y_1, Y_2, ..., Y_N$.

In each step of the method, the solution Y_{n+1} is obtained by adding to Y_n the deterministic term $f(Y_n, t_n)\Delta t$, the random term $g(Y_n, t_n)\Delta W_n$, and two correction terms that account for the second-order effects of the random noise. The first correction term is proportional to $(\Delta W_n)^2$, while the second correction term is proportional to $(\Delta W_n)^2 - \Delta t$.

The first correction term is a modified version of the correction term used in the Euler-Maruyama method and accounts for the curvature of the drift coefficient. The second correction term accounts for the fact that the diffusion coefficient is not constant, but depends on the value of Y_t and t. This term is typically small for small values of Δt but can be significant for large values of Δt or SDEs with high volatility.

4.5 Coupled Ornstein-Uhlenbeck Model

Assume that $Z_1(\lambda t)$ and $Z_2\lambda 2$ are two stochastic processes relative to two different time series data collected within a specified period. And suppose the data sets have some

correlation, i.e., their correlation coefficient is non-zero. We further assume the stochastic processes to be Lévy driven and simulate the model using either a $\Gamma(a, b)$ process or an IG(a,b) process for comparison purposes. Then we can model a coupled system of Ornstein-Uhlenbeck SDE, as shown below in equations 4.45 and 4.46

$$dX(t) = -\lambda X(t)dt + \sigma_{11}dZ_1(\lambda_1 t)_{t\geq 0} + \sigma_{12}dZ_2(\lambda_2 t)_{t\geq 0} \qquad (4.45)$$

$$dY(t) = -\lambda Y(t)dt + \sigma_{21}dZ_1(\lambda_1 t)_{t\geq 0} + \sigma_{22}dZ_2(\lambda_2 t)_{t\geq 0}. \qquad (4.46)$$

Where λ_1 and λ_2 are the intensity parameters, σ_{11} and σ_{22} determine volatility, and σ_{12} and σ_{21} describes the correlation between the data sets. Now we observe that when $\sigma_{12} = \sigma_{21} = 0$ we end up with a decoupled system and thus conclude that the two occurrences do not have any correlation and when $\sigma_{12} = \sigma_{21} = \sigma_{22} = \sigma_{11} = 0$ we end up with a decoupled deterministic system and each equation can be solved independently. Z_1 and Z_2 are the background driving processes. For the system, we assume both Z_1 and Z_2 are either $\Gamma(a, b)$ processes or IG(a,b) processes.

In matrix form, the coupled system of Ornstein-Uhlenbeck equations can be written as:

$$dX(t) = AX(t)dt + \sum_{i=1}^{2} B_i(t)dZ(\lambda t)$$

Where

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$A = \begin{pmatrix} \lambda_2 & 0 \\ 0 & -\lambda_2 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{21} \end{pmatrix} \tag{4.47}$$

$$B_2 = \begin{pmatrix} \sigma_{21} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$$

and

$$Z(\lambda t) = \begin{pmatrix} Z_1(\lambda t) \\ Z_2(\lambda t) \end{pmatrix}$$

The solution to this system has been obtained in [14] with a clear step-by-step proof, hence the proof is omitted in this work. The solution is given as:

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-s)}B_1 dZ(\lambda s) + \int_0^t e^{A(t-s)}B_2 dZ(\lambda s) \tag{4.48}$$

4.6   Background Driving Lévy Process

4.6.1   Gamma Process

Definition 18. *The Gamma process is a stochastic process $X = X_t, t \geq 0$ with parameters a and b which satisfies the following conditions:*

- *$X_0 = 0$.*

- *The process has independent increments.*

- *For $s < t$ the random variable $X_t - X_s$ has a $\Gamma(a(t-s), b)$ distribution.*

A random variable X has a gamma distribution $\Gamma(a, b)$ with rate and shape parameters, $a > 0$ and $b > 0$ respectively if its density function is given by:

$$f_x(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad \forall x > 0 \tag{4.49}$$

4.6.2   Inverse Gaussian Process

Definition 19. *The IG process $Y(t); t \geq 0$ is defined as the stochastic process satisfying the following properties:*

- *$Y(t)$ has independent increments*

- *$Y(t) - Y(s)$ follow an inverse gaussian distribution $IG(\Lambda(t) - \Lambda(s), \eta[\Lambda(t) - \Lambda(s)]^2)$ for all $t > s \geq 0$,*

where $\Lambda(t)$ is a monotone increasing function and $IG(a,b)$, $a, b > 0$, denotes the IG distribution with probability density function,

$$f_{IG}(y; a, b) = \sqrt{\frac{b}{2\pi y^3}} exp\left[-\frac{b(y-a)^2}{2a^2 y}\right], \quad y > 0 \tag{4.50}$$

The inverse Gaussian distribution is infinitely divisible, thus we redefine $IG(a,b)$ as a stochastic process X with parameters a,b to be the process that starts at zero and has independent and stationary increments such that

$$\mathbb{E} = \phi(z; at, b) \tag{4.51}$$
$$= exp(-at(\sqrt{-2iz + b^2} - b)).$$

Part III

Performance Analysis of Ornstein-Uhlenbeck Type Models.

Chapter 5

Applications

In this chapter, we present some applications which will help us analyze the performance of the classic and modified Ornstein-Uhlenbeck-type models. Before the applications are presented, we discuss below the error analysis methods used in ascertaining the performance of our models.

5.1   Error Analysis

In this section, we briefly discuss the errors generated from the model results. Four different error calculations are made to ascertain the accuracy of predictions using the OU system to model multiple data sets. We calculate the root mean squared errors (RMSE), the mean absolute percentage errors (MAPE), the mean absolute errors (MAE), and the average relative percentage errors (ARPE). These four error metrics are commonly used in evaluating the performance of forecasting models. The RMSE measures the average magnitude of the error by taking the square root of the mean of the squared errors. The MAPE gives an estimate of the average percentage error, which is useful in cases where the relative error is more important than the absolute error. The MAE provides a simple measure of the average magnitude of the errors, and the ARPE measures the average relative error as a percentage of the actual values. By using these four metrics, we can get a comprehensive understanding of the performance of our models and determine which ones provide the best predictions. Formulas used in computing the respective error estimates are explained below.

5.1.1 Error Formulas

Suppose y is the true value, p is the predicted value, and n is the number of data points, then:

Root Mean Squared Error:

$$\sqrt{\sum_{i=1}^{n} \frac{y_i - p_i}{n}} \tag{5.1}$$

Mean Absolute Percentage Error:

$$\frac{1}{n} \sum_{i=1}^{n} \left| \frac{(y_i - p_i)^2}{n} \right| \tag{5.2}$$

Mean Absolute Error:

$$\sum_{i=1}^{n} \left| \frac{y_i - p_i}{n} \right| \tag{5.3}$$

Average Relative Percentage Error:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - p_i|}{n} \tag{5.4}$$

5.2 Analyzing the performance of Ornstein-Uhlenbeck-Type models via superposition of solutions

5.2.1 Introduction

In this application, the performance of the classic Ornstein-Uhlenbeck model is compared with a two- and three-component superposed Ornstein-Uhlenbeck model. As mentioned earlier, the Ornstein-Uhlenbeck stochastic model is one of the important stochastic mod-

els with applications across multiple fields ranging from health [[59],[60]], geophysics and finance [[7],[38], [81]] to mention a few.

As the experimental results have shown, extending the classical Ornstein-Uhlenbeck model to the Ornstein-Uhlenbeck model with a non-Gaussian background driving process may yield the best solution for most real-world data. As in [38], the author has shown that the gamma-driven Ornstein-Uhlenbeck model can be used to predict the duration of a large earthquake by considering small earthquakes at a single instant in time. Again, in [14], the authors applied the gamma-driven Ornstein-Uhlenbeck model to geophysical and stock market data, showing that it produces small deviations from the actual data.

In [1] the authors show that both established and emerging stock indices are Lévy processes, specifically characterized as Lévy walk processes. In [14] the authors show that the Ornstein-Uhlenbeck SDE better models stock market and geophysical data by using the Gamma process as the background driving process (BDP) instead of the Weiner process. Furthermore, citing Barany et al. in [6], the authors used a truncated Lévy model to detect market downturns by examining the long-term impact of high-frequency stock markets. Thus, there is empirical evidence that many stock markets exhibit non-Gaussian behavior, with the above results and results from several other types of studies.

Anticipation of possible volatility in stock markets has made it more useful to predict stocks as accurately as possible. For the reasons above, there are many forecasting methods, from linear models to qualitative models. However, some of these models cannot capture the stochastic nature of these stock market indices and may give incorrect predictions. Because of this problem, many probabilistic models have been developed to predict stock markets. Some of these models are the Monte Carlo simulation, the Cox-Ingersoll-Ross (CIR) model, and the Black-Scholes model.

According to recent literature on Ornstein-Uhlenbeck model modification, a full numerical estimate of the rate parameter (λ) and weight (W) is developed, with the logarithm of the sum of the estimated weights equal to 0. In addition, by proposing a superposition of the Ornstein-Uhlenbeck model, driven by either an inverse-Gaussian (a,b) or a $\Gamma(a, b)$ process, a comparison is drawn on the accuracy and prediction of the models through various simulations. The model simulation uses four stock indices characterized as Lévy processes according to the characterization method discussed in Chapter 3 [1].

5.2.2 Data and Results

The data sets used in this application are closing prices of BVSP, MERV, MXX, and NASDAQ. All four closing prices are obtained through Yahoo Finance. Below, we show the time series graph of the datasets.
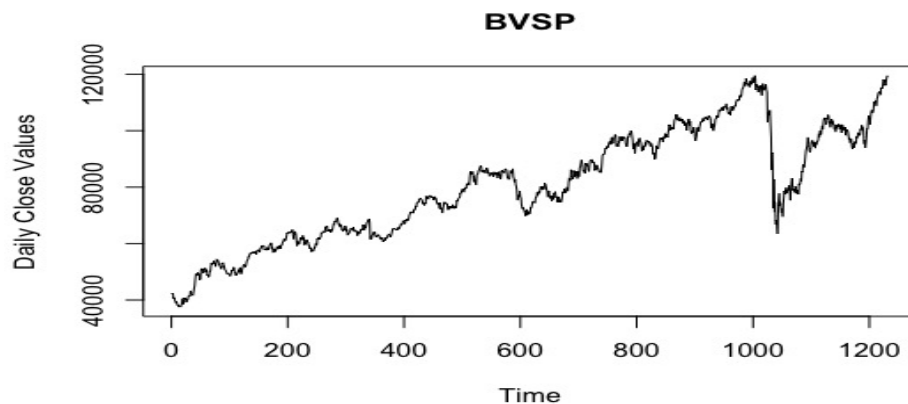
Figure 5.1: A time series plot showing daily closing values of BVSP

5.2.3 Results: Model Assessment

This subsection presents the model results in tabular form. This includes the root-mean-square error (RMSE) and estimated parameters. It also shows that the sum of the weights estimated from the superposed OU model is approximately 1. Each dataset had over 1000

62
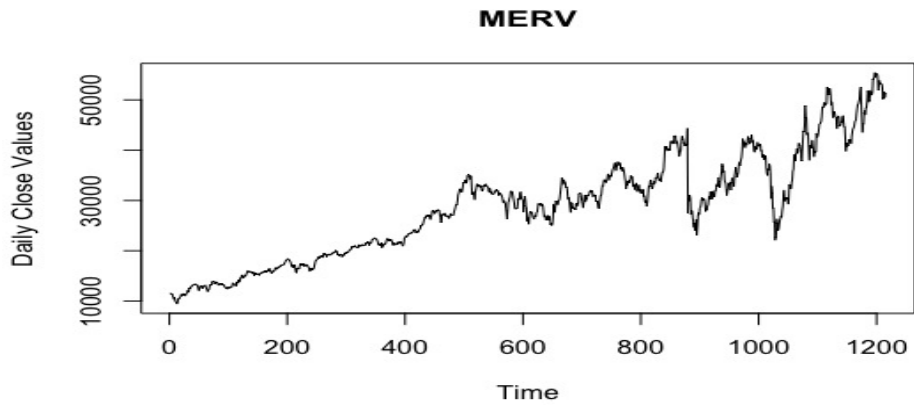
Figure 5.2: A time series plot showing daily closing values of MERV
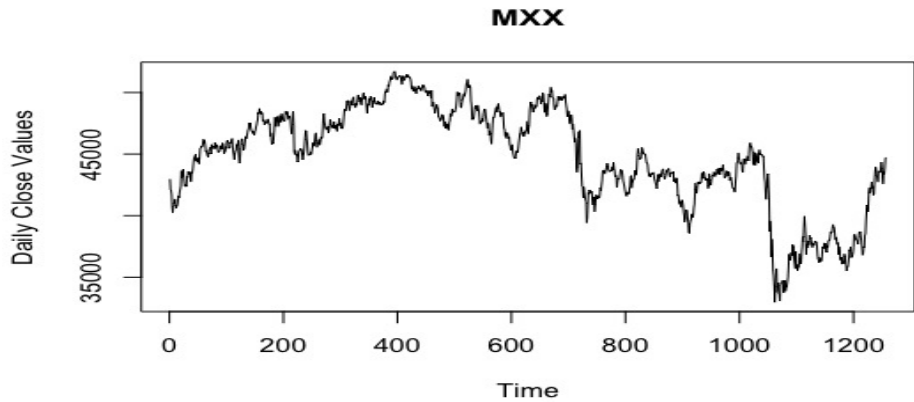
Figure 5.3: A time series plot showing daily closing values of MXX

entries, and we used approximately two-thirds of the entries in each dataset to estimate the parameters. The remaining entries are predicted using the estimated parameters to determine the prediction accuracy.

5.2.4 Ordinary OU model

Tables 5.1 and 5.2 show results from the Ordinary OU model with the Gamma and Inverse-Gaussian processes as background driving processes, respectively.

63
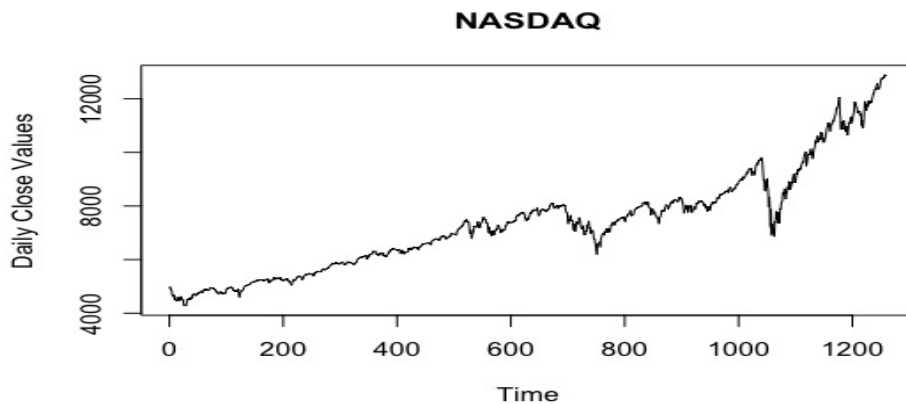
Figure 5.4: A time series plot showing daily closing values of NASDAQ

Table 5.1: Table for Ordinary $\Gamma(a,b)$

Data	k	λ	RMSE
MXX	11	0.01358	0.90640
BVSP	11	0.006258	0.9466
NASDAQ	11	0.00452	0.7051
MERV	11	0.00488	0.8290

Table 5.2: Table for Ordinary IG(a,b)

Data	k	λ	RMSE
MXX	11	0.01358	0.04342
BVSP	11	0.006258	0.1032
NASDAQ	11	0.00452	0.1006
MERV	11	0.00488	0.2410

5.2.5 Two-Component OU model

Tables 5.3 and 5.4 show results from the two-component OU model with the Gamma and Inverse-Gaussian processes as background driving processes, respectively.

5.2.6 Three-Component OU model

Tables 5.5 and 5.6 show results from the three-component OU model with the Gamma and Inverse-Gaussian processes as background driving processes, respectively.

Table 5.3: Table for two-component superposed $\Gamma(a, b)$

Data	$w_1 + w_2$	λ_1	λ_2	RMSE
MXX	1.0105	0.01358	0.013655	0.367
BVSP	1.0069	0.00626	0.00631	0.52017
NASDAQ	1.0004	0.00452	0.00453	0.254
MERV	1.0001	0.00488	0.004882	0.3543

Table 5.4: Table for two-component IG(a,b)

Data	$w_1 + w_2$	λ_1	λ_2	RMSE
MXX	1.0105	0.01358	0.013655	0.00545
BVSP	1.0069	0.00626	0.00631	0.0203
NASDAQ	1.0004	0.00452	0.00453	0.02033
MERV	1.0001	0.00488	0.004882	0.0354

Table 5.5: Table for three-component superposed Γ(a,b)

| Data	$w_1 + w_2 + w_3$	λ_1	λ_2	λ_3	RMSE
MXX	1.00006	0.01358	0.013655	0.01697	0.1056
BVSP	1.005	0.006258	0.00631	0.00717	0.38034
NASDAQ	0.99989	0.00452	0.004526	0.00465	0.253
MERV	0.9974	0.00488	0.004882	0.00495	0.3468

Table 5.6: Table for three-component superposed IG(a,b)

Data	$w_1 + w_2 + w_3$	λ_1	λ_2	λ_3	RMSE
MXX	1.00006	0.01358	0.013655	0.01697	0.00534
BVSP	1.005	0.006258	0.00631	0.00717	0.0203
NASDAQ	0.99989	0.00452	0.004526	0.00465	0.02033
MERV	0.9974	0.00488	0.004882	0.00495	0.0354

5.2.7   Results: Model Testing

In this subsection, the estimated parameters are used to predict the remaining closing prices for each dataset. For each data, we forecast over 400 closing prices, which represents over one year of data predicted.

Ordinary OU model

Tables 5.7 and 5.8 show results after forecasting with the Ordinary OU model using the Gamma process and the Inverse-Gaussian process as background driving processes, respectively.

Table 5.7: Forecast results for Ordinary Γ(a,b)

Data	k	λ	RMSE
MXX	11	0.01358	0.902
BVSP	11	0.006258	0.9578
NASDAQ	11	0.00452	0.7227
MERV	11	0.00488	0.8613

Table 5.8: Forecast results for Ordinary IG(a,b)

Data	k	λ	RMSE
MXX	11	0.01358	0.07062
BVSP	11	0.006258	0.2366
NASDAQ	11	0.00452	0.1150
MERV	11	0.00488	0.3038

5.2.8   Two-Component OU model

Tables 5.9 and 5.10 show results after forecasting with the two-component OU model using the Gamma process and the Inverse-Gaussian process as background driving processes, respectively.

Table 5.9: Forecast results for two-component superposed $\Gamma(a, b)$

Data	$w_1 + w_2$	λ_1	λ_2	RMSE
MXX	1.0105	0.01358	0.013655	0.4664
BVSP	1.0069	0.00626	0.00631	0.6175
NASDAQ	1.0004	0.00452	0.00453	0.2673
MERV	1.0001	0.00488	0.004882	0.4152

Table 5.10: Table for two-component IG(a,b)

Data	$w_1 + w_2$	λ_1	λ_2	RMSE
MXX	1.0105	0.01358	0.013655	0.009164
BVSP	1.0069	0.00626	0.00631	0.02374
NASDAQ	1.0004	0.00452	0.00453	0.0286
MERV	1.0001	0.00488	0.004882	0.0404

5.2.9 Three-Component OU model

Tables 5.11 and 5.12 show results after forecasting with the three-component OU model using the Gamma process and the Inverse-Gaussian process as background driving processes, respectively.

Table 5.11: Table for three-component superposed Γ(a,b)

Data	$w_1 + w_2 + w_3$	λ_1	λ_2	λ_3	RMSE
MXX	1.00006	0.01358	0.013655	0.01697	0.1506
BVSP	1.005	0.006258	0.00631	0.00717	0.5170
NASDAQ	0.99989	0.00452	0.004526	0.00465	0.2577
MERV	0.9974	0.00488	0.004882	0.00495	0.4109

Table 5.12: Table for three-component superposed IG(a,b)

Data	$w_1 + w_2 + w_3$	λ_1	λ_2	λ_3	RMSE
MXX	1.00006	0.01358	0.013655	0.01697	0.00906
BVSP	1.005	0.006258	0.00631	0.00717	0.02037
NASDAQ	0.99989	0.00452	0.004526	0.00465	0.0286
MERV	0.9974	0.00488	0.004882	0.00495	0.0403

5.2.10 Stock Market Analysis

From the above table, we can see that by superposing the Ornstein-Uhlenbeck model solution, we can significantly improve the predictive performance of the model by comparing the results of the ordinary OU model (Tables 5.1 and 5.2) to both the two-component (Tables 5.3 and 5.4) and three-component (Tables 5.5 and 5.6) OU models. The $\Gamma(a, b)$

OU model show significant improvement from the ordinary OU model to the superposed two-component and three-component models. Also, the IG(a,b) OU model performs better overall compared to the $\Gamma(a, b)$ OU model.

Again, we see a significant improvement from the ordinary IG(a,b) model (Table 5.2) to the two-component IG(a,b) OU model (Table 5.4). However, as can be seen from the RMSEs, the three-component IG(a,b) OU model (Table 5.6) shows only a small improvement when compared to the two-component IG(a,b) OU model. This is seen with the NASDAQ index in tables 5.4 and 5.6. Observing the predicted values in tables 5.7 to 5.12 shows similar trends in the performance of our OU models as described earlier. The RMSEs observed in table 5.12 show the three-component IG(a,b) OU model performs better in comparison to the other models.

Sample Paths

In this subsection, we have generated some sample paths with their respective means for each stock index using the log returns. The means computed for each sample path are also presented in table 5.13.

Table 5.13: Sample Path Means for each index

Color	MXX	BVSP	NASDAQ	MERV
Blue	11.270	11.089	9.122	10.410
Red	11.126	11.407	8.787	9.612
Yellow	11.223	11.279	9.323	9.996

Figures 5.5 to 5.8 show 3 sample paths (trajectories) for the stock indices considered in this work. Each sample path is a possible solution for the SDE model and signifies the stochastic behavior of the stock market datasets. Thus, any one of the 3 sample paths drawn for each data set could be a possible forecast of the price action of the index, confirming the stochasticity of these stock market datasets and the difficulty in forecasting them. From the values of the means in each of the sample paths shown in table 5.13, we observe that
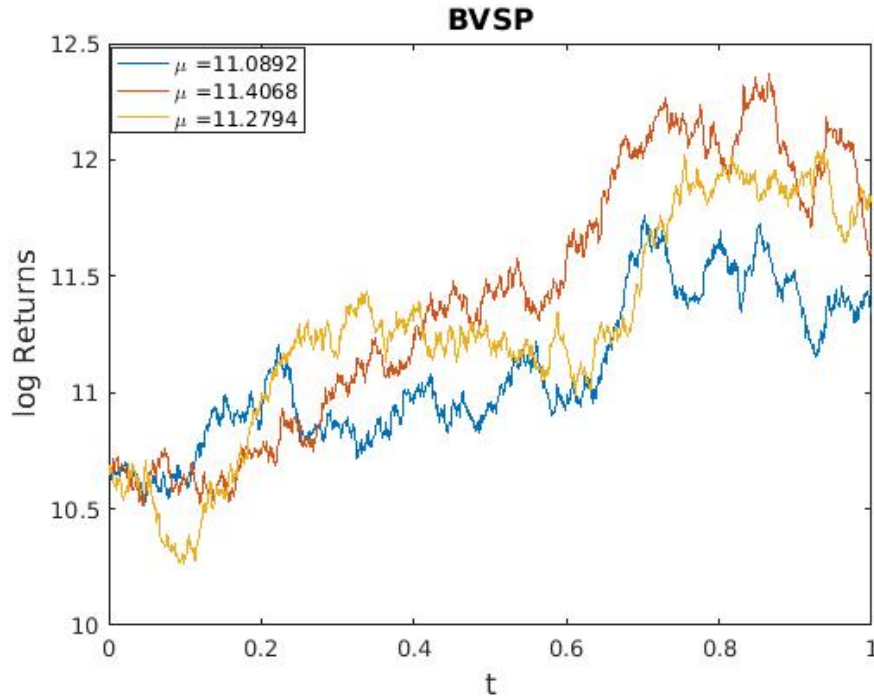
Figure 5.5: Three sample paths generated for the daily closing values of BVSP

the means are close. Therefore, we do not expect a drastic deviation of our model solution from the other possible solutions that are randomly generated.

5.2.11 Conclusion

This application shows that the superposition of Ornstein-Uhlenbeck models can significantly improve predictive performance. There are also trade-offs when increasing the number of components. That said, there is a point where additional components may result in decreased performance rather than increased performance. Hence, there is no gain in increasing the components beyond some limits.
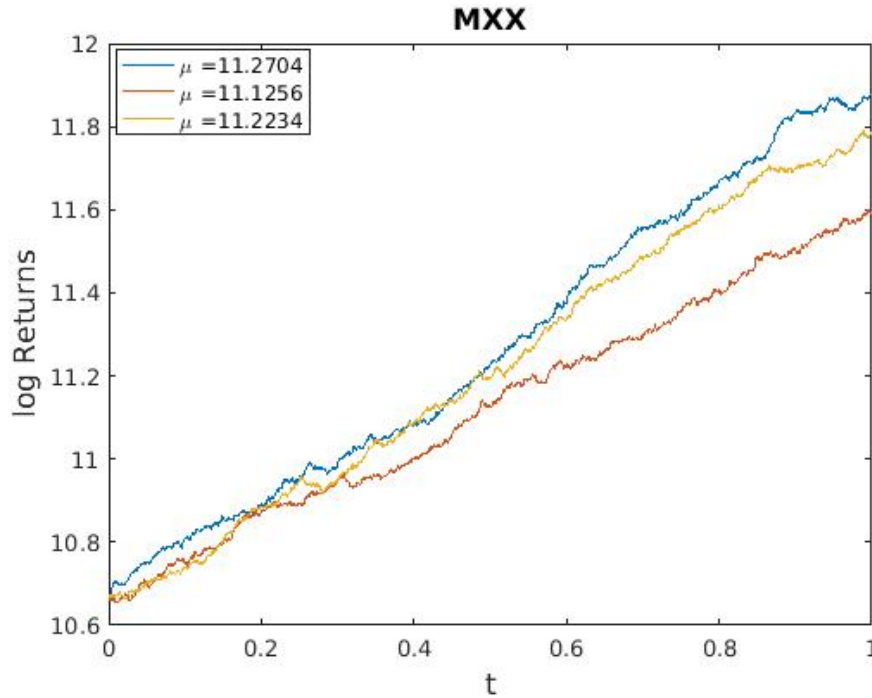
Figure 5.6: Three sample paths generated for the daily closing values of MXX

5.3 Analyzing time series events with dependencies.

5.3.1 Introduction

Naturally, some dependencies arise between some time series events. For example, when looking at market trends, we see several cases where the movements of various stock portfolios are positively correlated. In a study (referenced in [75]), the author found instances where volcanic eruptions occurred before earthquakes up to 120 miles away. In this application, we use a coupled system of stochastic Ornstein-Uhlenbeck differential equations (SDEs) driven by Lévy processes to model three different application domains. Applications featured include applications for stock market data, volcanic eruption data, and COVID-19 data. For the stock indices, we use the Dow Jones, Nasdaq, S&P500, and Russell. Volcanic eruption data comes from the Bezymianny seismic station, and COVID-19 data comes from the New York Times database. These datasets were described in Chapter
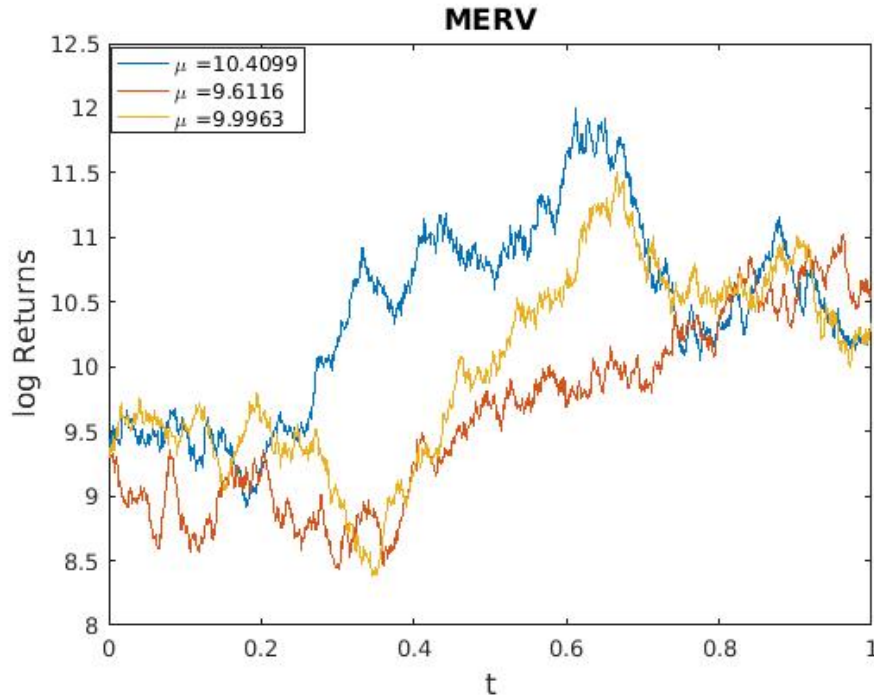
Figure 5.7: Three sample paths generated for the daily closing values of MERV

3.

The data sets being analyzed all have a significant impact on people and their assets. Stock market crashes, volcanic eruptions, and disease outbreaks can damage economies and negatively impact the quality of life. In a study (referenced in [41]), the authors showed the importance of complexity science in modeling these events, which can help save lives and property. Using these three data sets, this application model four different coupled OU systems. The first three are called "intra-dependent field applications," meaning they involve modeling two data sets from the same field. The last application, referred to as an "interdependent field application," involves combining data from two fields (in this case, stock market data with COVID-19 data).

Many studies have shown that events in one field can trigger specific responses in an-
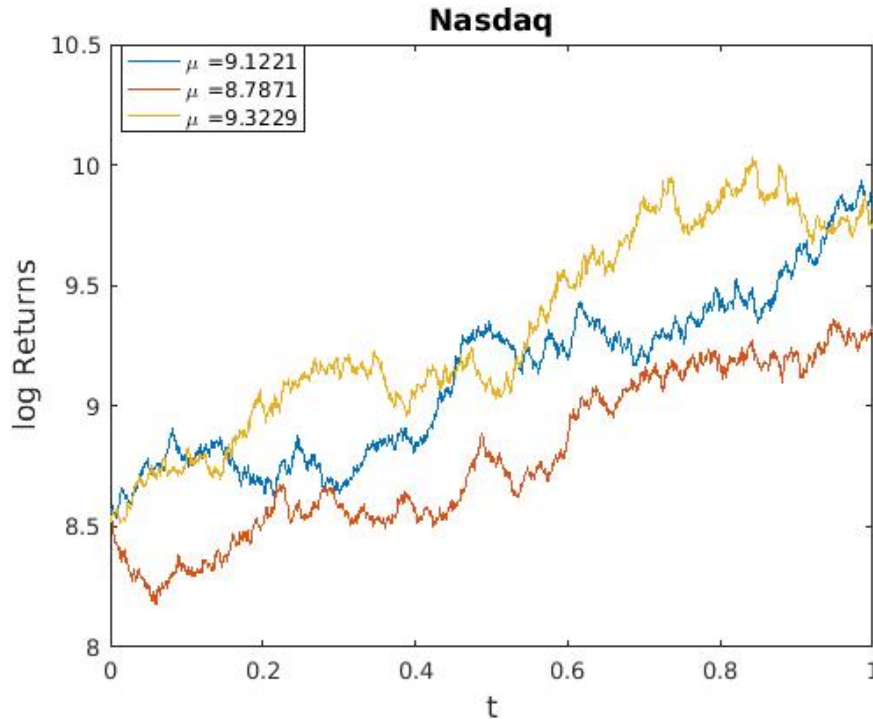
Figure 5.8: Three sample paths generated for the daily closing values of NASDAQ

other. The COVID-19 pandemic had a significant impact on the stock market. The coupled OU system calculates the correlation between time series events, allowing us to model different data sets. Two Lévy processes, the $\Gamma(a, b)$ process and the IG(a,b) process, are used as the background driving process to compare the performance of the models and determine the best option.

5.3.2 Derivation of Parameters

Matlab and R software are used to find the parameters σ_{11} and σ_{22} (volatility parameters) and σ_{12} and σ_{21} (correlation parameters) from the model presented in Chapter 4 (equations 4.45,4.46). Matlab software is used to compute the correlations between the data sets. The R software (using the astsa and stochvol packages) estimates the volatility at a 95% confidence interval. The estimation results are presented in the tables below, with corresponding graphs showing the estimated volatilities at 5%, 50%, and 95% posterior

quantiles.

Correlation Matrix Results for the Data Sets

Table 5.14: Correlation Matrix for Volcanic Eruptions

	Eruption 2	Eruption 4	Eruption 8
Eruption 2	1	0.0626	-0.0754
Eruption 4	0.0626	1	0.0127
Eruption 8	-0.0754	0.0127	1

Table 5.15: Correlation Matrix for Stock Markets

Stock Markets	Dow Jones	S&P500	NASDAQ	Russell
Dow Jones	1	0.9942	0.9618	0.9590
S&P500	0.9942	1	0.9840	0.9596
NASDAQ	0.9618	0.9840	1	0.9258
Russell	0.9590	0.9596	0.9258	1

Table 5.16: Correlation Matrix USA COVID-19 Cases and Deaths

	USA COVID-19 Cases	USA COVID-19 Deaths
USA COVID-19 Cases	1	0.6755
USA COVID-19 Deaths	0.6755	1

Table 5.17: Correlation Matrix for Stock Markets and USA COVID-19 Cases and Deaths

Stock Markets	USA COVID-19 Cases	USA COVID-19 Deaths
Dow Jones	0.5576	0.3829
S&P500	0.5861	0.4202
NASDAQ	0.6219	0.4539
Russell	0.5741	0.4858

Volatility Parameter

This section presents the results obtained for the volatility parameters using the stockvol package in R (referenced in [42]). The figures show the estimated volatilities with 5%, 50%,

and 95% posterior quantiles, and the values used in the study were chosen from the 95% posterior quantile.

Table 5.18: Volatility Values Estimated Using R stochvol package

Data	Volatility (σ_{ii})
Dow Jones	-5.123718
S&P500	-5.255099
NASDAQ	-5.444095
Russell	-5.120774
Eruption 2	-7.906
Eruption 4	-7.967
Eruption 8	-7.84
USA COVID-19 Cases	0.006128
USA COVID-19 Deaths	-1.116847
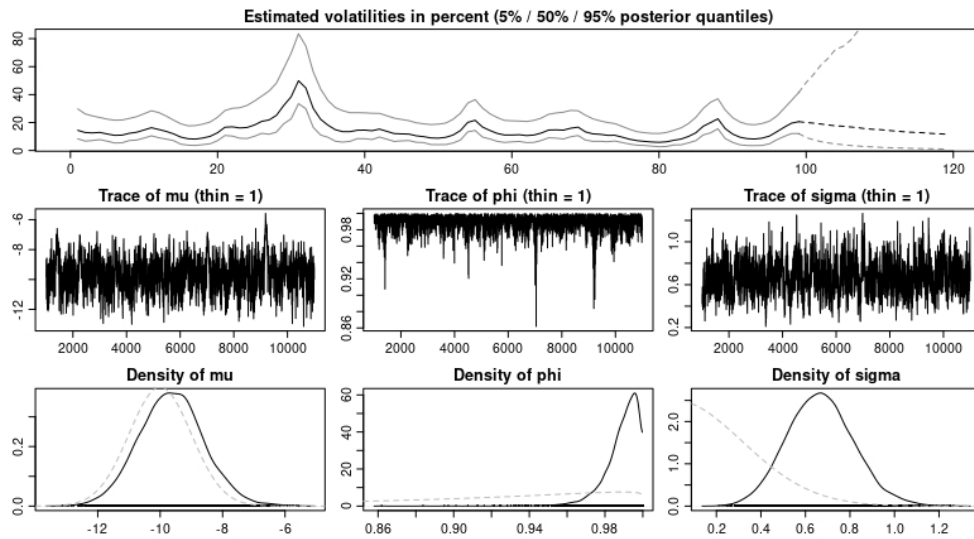
Figure 5.9: Graph showing Estimated Volatility For Eruption 2

5.3.3 Model Applications

In this section, the four different applications of the coupled OU system will be discussed. In addition, the error estimates for predictions made using the model on data related to

Figure 5.10: Graph showing Estimated Volatility For Eruption 4
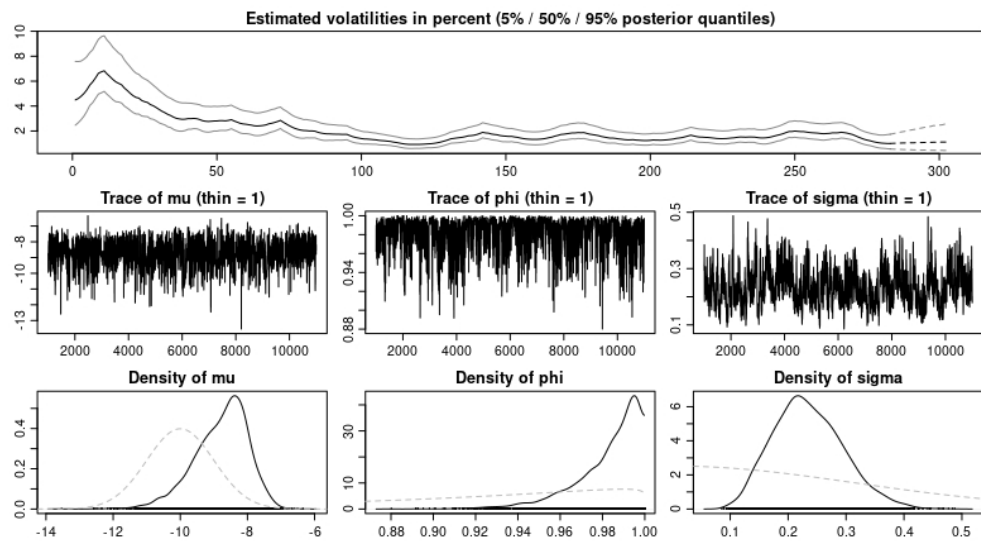
Figure 5.11: Graph showing Estimated Volatility for The Russell.

volcanic eruptions, stock market data, US COVID-19 data, and a combination of the stock market and COVID-19 data are shown.
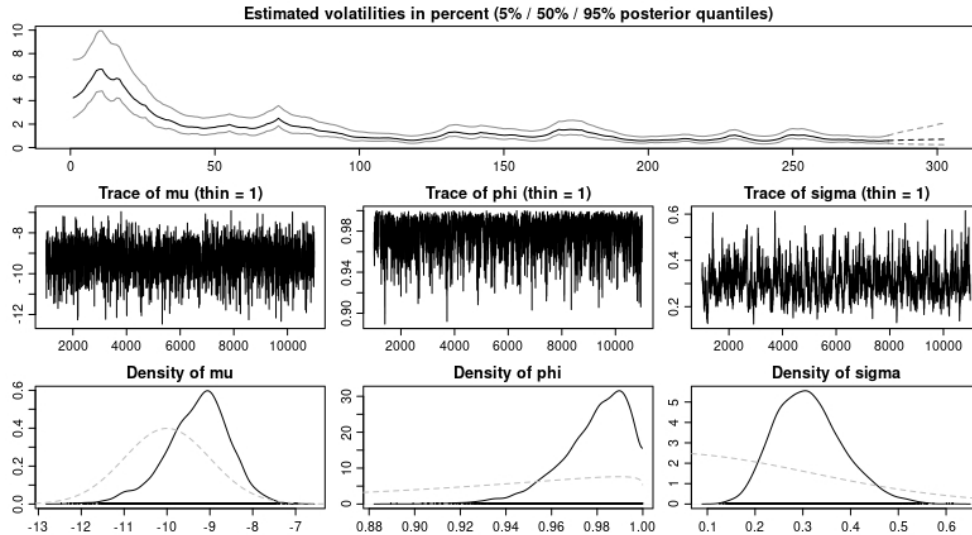
Figure 5.12: Graph showing Estimated Volatility for The Dow Jones
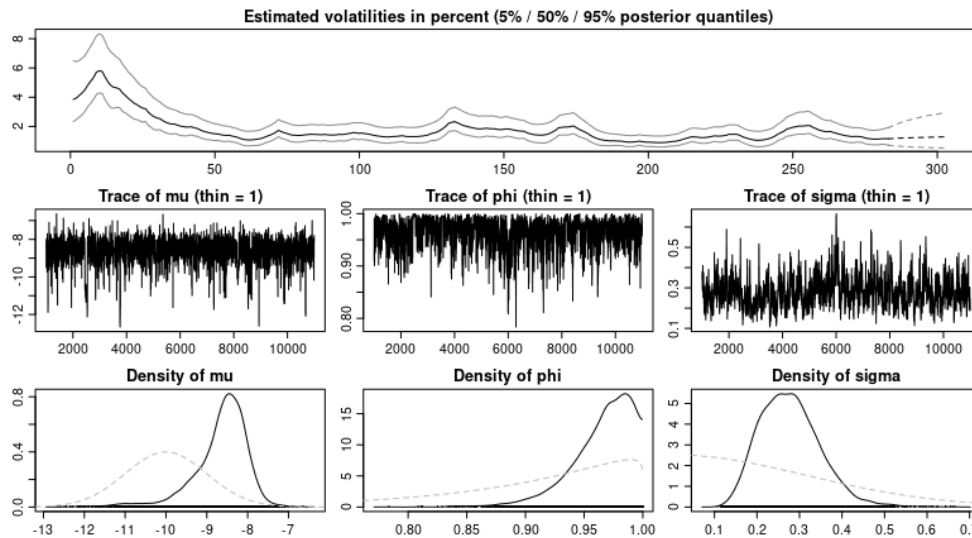
Figure 5.13: Graph showing Estimated Volatility for the NASDAQ.
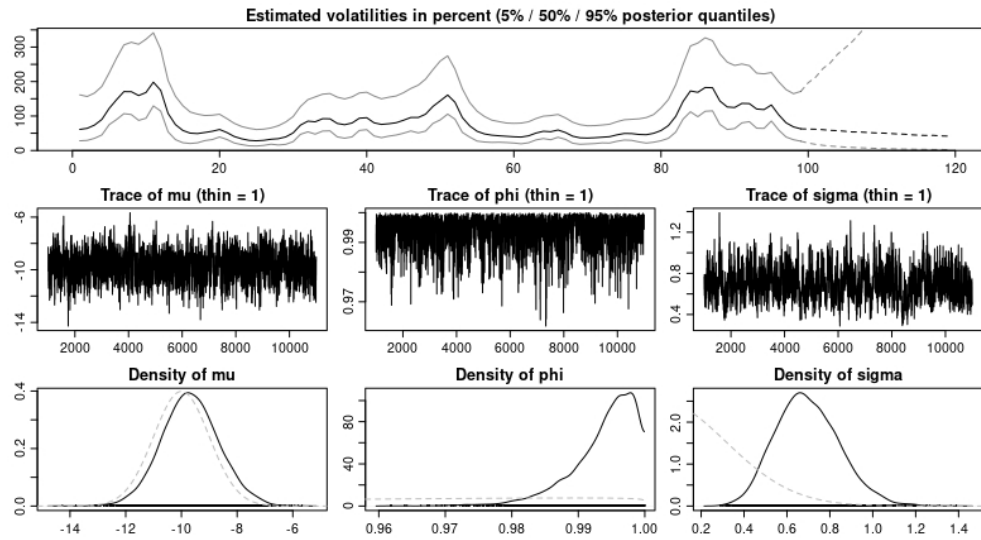
5.3.4 Application to Volcanic Data

This section presents the results of applying the model to data on volcanic eruptions. The data was modeled using both the IG(a,b) and $\Gamma(a, b)$ coupled Ornstein-Uhlenbeck (OU) systems. The analysis was performed on three of the eight eruptions, specifically Eruption 2, Eruption 4, and Eruption 8. The results are presented in Table 5.19 and Table 5.20 for

76
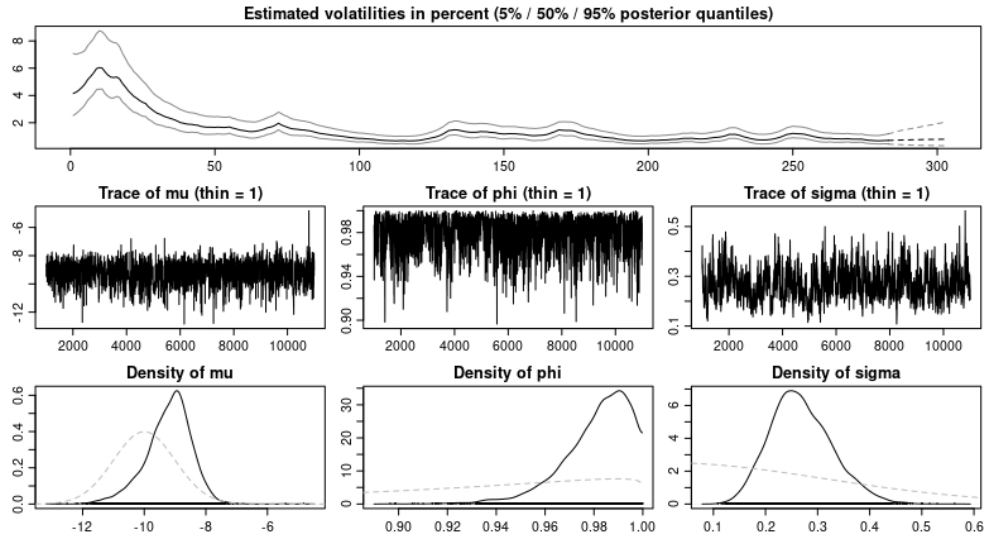
Figure 5.14: Graph showing Estimated Volatility for The S&P500
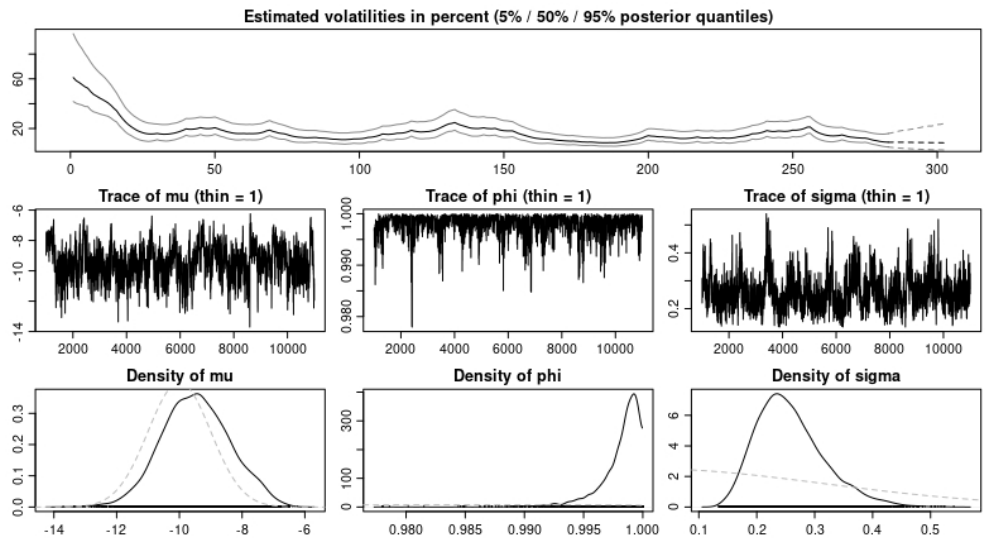
Figure 5.15: Graph showing Estimated Volatility for US COVID-19 Cases.

both Lévy processes.

Remark: The errors in the IG(a,b) OU model, shown in Table 5.20, are smaller compared to those of the $\Gamma(a, b)$ OU model in Table 5.19. This suggests that for the volcanic data, the coupled OU system with the IG(a,b) BDP performs better than the coupled OU system with the $\Gamma(a, b)$ BDP.
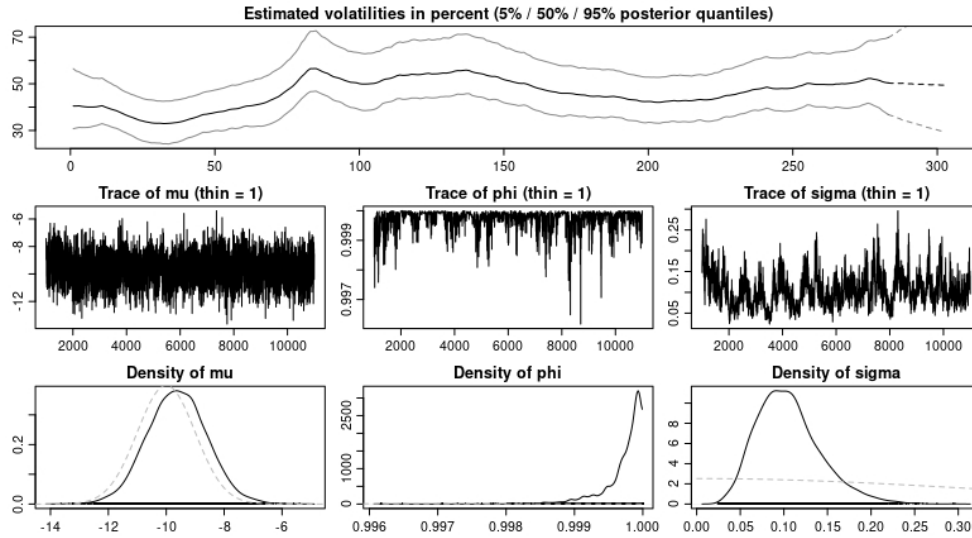
Figure 5.16: Graph Showing Estimated Volatility For US COVID-19 Deaths

Table 5.19: Results from System of Gamma(a,b) OU model. The results reported are the best results from modeling a combination of Eruptions 2, Eruptions 4, and Eruptions 8

| Eruptions	RMSE	MAPE	MAE	ARPE
Eruption 2	0.7654	0.0049	34685.5	1.3275
Eruption 4	1.3218	6.5506	156.69	13.33
Eruption 8	0.8088	0.8212	2419.4	4.6206

Table 5.20: Results from System of IG(a,b) OU model. The results reported are the best results from modeling a combination of Eruptions 2, Eruptions 4, and Eruptions 8

Eruptions	RMSE	MAPE	MAE	ARPE
Eruption 2	0.3927	0.0013	8518.93	0.3411
Eruption 4	1.8168	4.547	18.597	0.926
Eruption 8	0.5839	0.0646	216.49	0.3633

5.3.5 Application to US Stock Markets

This section presents the results of the model applied to stock indices from the US stock market, such as the Dow Jones, NASDAQ, S&P500, and Russell. As seen in the parameter estimation section, these data sets were highly correlated with correlation coefficients close to 1. The analysis was based on the daily closing values from 02/19/2020 to 04/16/2021, which exclude weekends as the stock market is closed during weekends. The results in

Table 5.21 and Table 5.22 display the error estimates from the predictions made using the model with either the IG(a,b) or $\Gamma(a, b)$ process as the BDP.

Table 5.21: Results from System of $\Gamma(a, b)$ OU model. The results reported are the best results from modeling a combination of the Dow Jones, the S&P500, the NASDAQ, and the Russell

Stock Markets	RMSE	MAPE	MAE	ARPE
Dow Jones	0.1798	0.0014	3950.69	0.1385
S&P500	0.1574	0.0111	446.16	0.1318
NASDAQ	0.2428	0.0046	2009.52	0.1838
Russell	0.2219	0.0338	320.85	0.1992

Table 5.22: Results from System of IG(a,b) OU model. The results reported are the best results from modeling a combination of the Dow Jones, the S&P500, the NASDAQ, and the Russell

Stock Markets	RMSE	MAPE	MAE	ARPE
Dow Jones	0.1259	0.001	2773.07	0.1027
S&P500	0.1360	0.0095	365.58	0.1125
NASDAQ	0.1762	0.0038	1547.44	0.1495
Russell	0.2235	0.0334	319.25	0.1967

Remark: This result indicates that the coupled OU system with the IG(a,b) as the BDP performs better in modeling the stock market data than the coupled OU system with $\Gamma(a, b)$ as the BDP.

5.3.6   Application to US COVID-19 Data

This section applies the model to the US COVID-19 data obtained from the New York Times between the dates 02/19/2020 and 04/16/2021. The model is used to predict both daily reported cases and daily reported deaths. Results in Table 5.23 and Table 5.24 show the error estimates obtained when the BDP in the model was either a $\Gamma(a, b)$ process or an IG(a,b) process.

Remark: The errors from the system of IG(a,b) OU model shown in table 5.24 are smaller compared to that of the system of Gamma(a, b) OU model in table 5.23. Thus,

Table 5.23: Results from System of $\Gamma(a, b)$ OU model.

US COVID-19 Data	RMSE	MAPE	MAE	ARPE
Daily Cases	1.648	0.3874	51573.23	106.74
Daily Deaths	3.1713	7.5384	28204.32	40.40

Table 5.24: Results from System of IG(a,b) OU model.

US COVID-19 Data	RMSE	MAPE	MAE	ARPE
Daily Cases	1.648	0.3874	51573.23	106.74
Daily Deaths	1.4193	3.1076	828.20	16.65

the model performs better when the BDP is the Inverse Gaussian process.

5.3.7 Applications to Coupled US COVID-19 and Stock Markets Data

This section models a combination of stock portfolios and the US COVID-19 data. At the height of the pandemic, as infections and deaths soared, investor panic, fearing an imminent market crash, drove stocks down. Section 3 Table 5.16 and Table 5.17 show a non-zero correlation between the stock market datasets and the daily reported cases and deaths. Each stock market data is modeled with the daily reported US COVID-19 cases and the daily reported US COVID-19 deaths. Here both the $\Gamma(a, b)$ and IG(a,b) are used as BDPs to compare the model performance based on the Lévy process used. Results from the error estimates are shown below in Table 5.25, Table 5.26, Table 5.27, and Table 5.28.

Table 5.25: Results from System of $\Gamma(a, b)$ OU model. The results reported are the best results from modeling a combination of USA COVID-19 Cases with the Stock Market Data

Data	RMSE	MAPE	MAE	ARPE
Daily Cases	1.6481	0.3874	51525.03	106.74
Dow Jones	0.2684	0.0027	6071.4	0.2052
S&P500	0.5678	0.0631	2516.53	0.7486
NASDAQ	0.182	0.0041	1551.14	0.1598
Russell	0.5211	0.1057	1133.73	0.6226

Table 5.26: Results from System of IG(a,b) OU model. The results reported are the best results from modeling a combination of USA COVID-19 Cases with the Stock Market Data

Data	RMSE	MAPE	MAE	ARPE
Daily Cases	1.6481	0.3874	51530.41	106.74
Dow Jones	0.1259	0.0015	2773.07	0.1027
S&P500	0.136	0.0095	365.56	0.1125
NASDAQ	0.1762	0.0038	1547.44	0.1495
Russell	0.2235	0.0334	319.25	0.1967

Remark: The errors from the system of IG(a,b) OU model shown in Table 5.26 are smaller compared to that of the system of Gamma(a, b) OU model in Table 5.25. Thus, the model performs better when the BDP of the coupled OU system is the Inverse Gaussian process.

Table 5.27: Results from System of $\Gamma(a, b)$ OU model. The results reported are the best results from modeling a combination of USA COVID-19 Deaths with the Stock Market Data

Data	RMSE	MAPE	MAE	ARPE
Daily Deaths	1.4192	3.1049	823.32	16.64
Dow Jones	0.1820	0.0014	4002.92	0.14
S&P500	0.1612	0.0011	456.29	0.1347
NASDAQ	0.267	0.0058	2535.05	0.2263
Russell	0.389	0.0537	537.49	0.3092

Table 5.28: Results from System of IG(a,b) OU model. The results reported are the best results from modeling a combination of USA COVID-19 Deaths with the Stock Market Data

Data	RMSE	MAPE	MAE	ARPE
Daily Deaths	1.4192	3.1076	828.20	16.65
Dow Jones	0.1259	0.0010	2773.07	0.103
S&P500	0.136	0.0095	365.58	0.1125
NASDAQ	0.1762	0.0038	1547.44	0.1495
Russell	0.2235	0.0334	319.25	0.1967

Remark: The errors from the system of IG(a,b) OU model shown in Table 5.28 are smaller compared to that of the system of Gamma(a, b) OU model in Table 5.27. Thus, the

model performs better when the BDP of the coupled OU system is the Inverse Gaussian process.
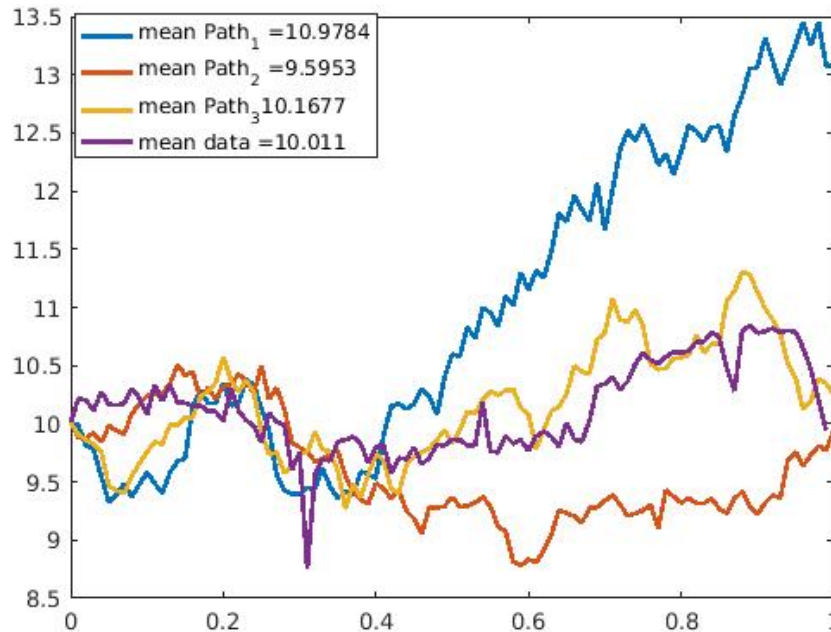

Figure 5.17: Graph showing three sample paths and time series plot for Eruption 2. We observe from the sample paths that sample-path 3 closely predicts the time series and see by comparison of the means, that sample-path 3 is closer in value to the mean of the time series data

5.3.8 Discussion

Running the model simulations on the four different applications and observing the values obtained from the error estimates shows that the model predictions fit the datasets well. Tables 5.19 and 5.20 show improvements in the MAPE, MAE, and ARPE error estimates for all three eruptions when the BDP is an IG(a,b) process. However, the RMSE of Eruption 4 shows better results using the $\Gamma(a, b)$ process. For the application to stock market data, in Tables 5.21 and 5.22 the coupled OU model with IG(a,b) as BDP gives better results than the $\Gamma(a, b)$ OU model.
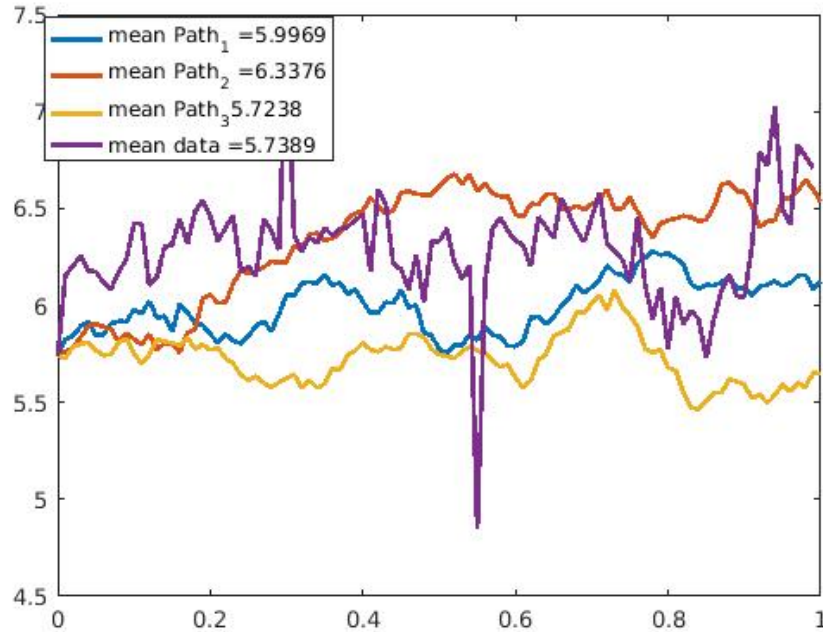
Figure 5.18: Graph showing three sample paths and time series plot for Eruption 8. We notice that Eruption 8 has 2 extreme values which may affect the model's performance. For this time series, we again see that sample-path 3 closely predicts it and by comparison of the means, sample path 3 is closer in value to the mean of the time series data

The results also highlight that the IG(a,b) as the BDP performs better than the $\Gamma(a,b)$ as the BDP for modeling daily reported deaths with the US COVID-19 data, and that both background driving Lévy processes give good error estimates when the stock market data is modeled with the US COVID-19 cases. The sample paths drawn in the figures suggest that the disease could potentially die out at some point after it has peaked. Comparing the sample paths to the original time series, the solution path that best models the data with a mean comparatively closer to that of the time series data are observed. Additionally, the error estimates in Tables 5.23 and 5.24 further support the validity of our model's predictions. However, it is important to note that these results are dependent on the assumptions and parameters used in the model and may not accurately reflect the actual future trend

Figure 5.19: Graph showing three sample paths and time series plot for the Russel index. For this time series, we see that sample-path 2 closely predicts it and by comparison of the means, sample path 2 is closer in value to the mean of the time series data

of the disease.

5.4 Analyzing the Background Driving Process of the Ornstein-Uhlenbeck Model.

5.4.1 Introduction

One of the major challenges when using the Ornstein-Uhlenbeck model is the ability to select an appropriate background driving process (BDP) for the random process under consideration [102]. As shown by [14], [38], and [102] the choice of BDP may significantly impact model performance. This application aims to investigate the performance of the
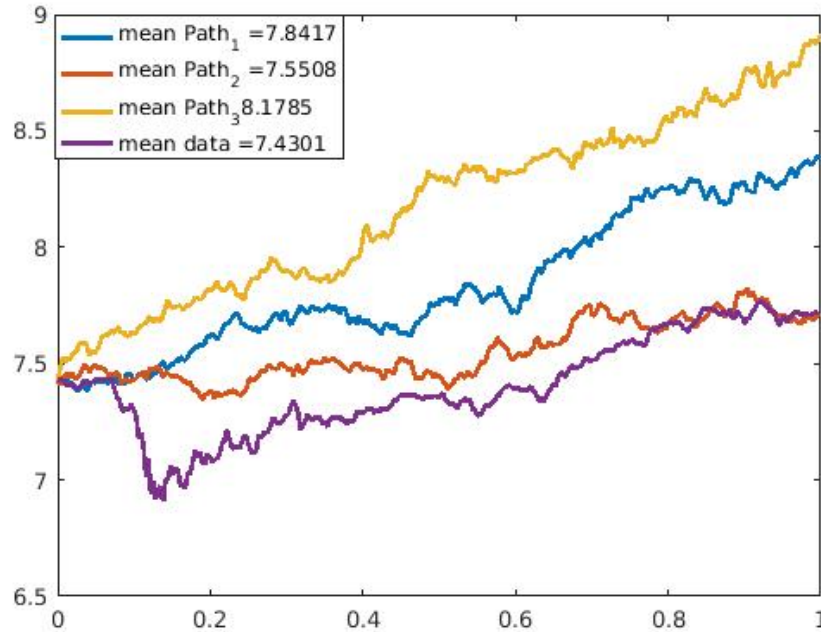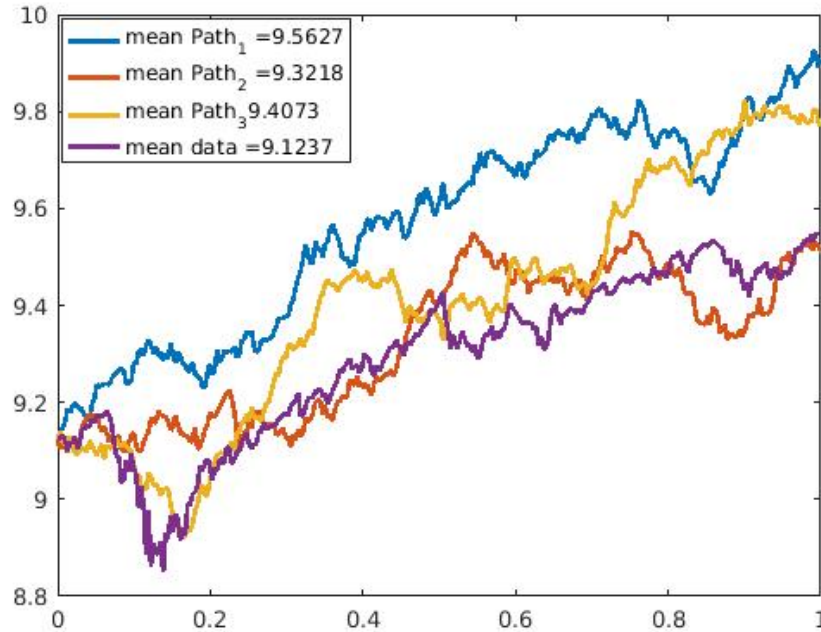
Figure 5.20: Graph showing three sample paths and time series plot for the NASDAQ index. For this time series, we see that sample-path 2 closely predicts it and by comparison of the means, sample path 2 is closer in value to the mean of the time series data

Ornstein-Uhlenbeck model by simulating three background driving processes: Gaussian, Gamma (Γ), and inverse-Gaussian processes. Since the data set has already been characterized, this application tests the effect of knowledge about time series characteristics on the performance of the Ornstein-Uhlenbeck model.

Thus, this application involves comparing the three different BDPs, the standard Brownian motion, the Gamma(a,b) process, and the Inverse-Gaussian(a,b) process, in modeling time series data. The choice of BDP has a significant impact on the performance of the OU model, as seen in the results of our simulation. The appropriate BDP depends on the characterization of the data as Gaussian, Lévy walk, or Lévy flight. From the results of our simulations, we observe that in cases where the data is Gaussian, the standard Brownian motion is a better choice as the BDP, while in cases where the data is a Lévy walk or a Lévy flight, a Lévy process such as the Gamma(a,b) or Inverse-Gaussian(a,b) would be a

Figure 5.21: Graph showing three sample paths and time series plot for the US COVID-19 cases. For this time series, we see that sample-path 1 closely predicts it and by comparison of the means, sample-path 1 is closer in value to the mean of the time series data

better choice as the BDP.

5.4.2 Data

For this application, simulation of the model will be performed using three different types of data, which include: daily closing values of Nasdaq obtained from Yahoo Finance, 2011 earthquake data from Japan following the 9.0 magnitude event (referred to as AfterM9), and a simulated fractional Brownian motion generated by the *fbm* function from the R package *somebm*.

5.4.3 Results

In this section, the results of using the proposed BDP approach in modeling with a three-component superposed Ornstein-Uhlenbeck equation are presented. Additionally, the re-
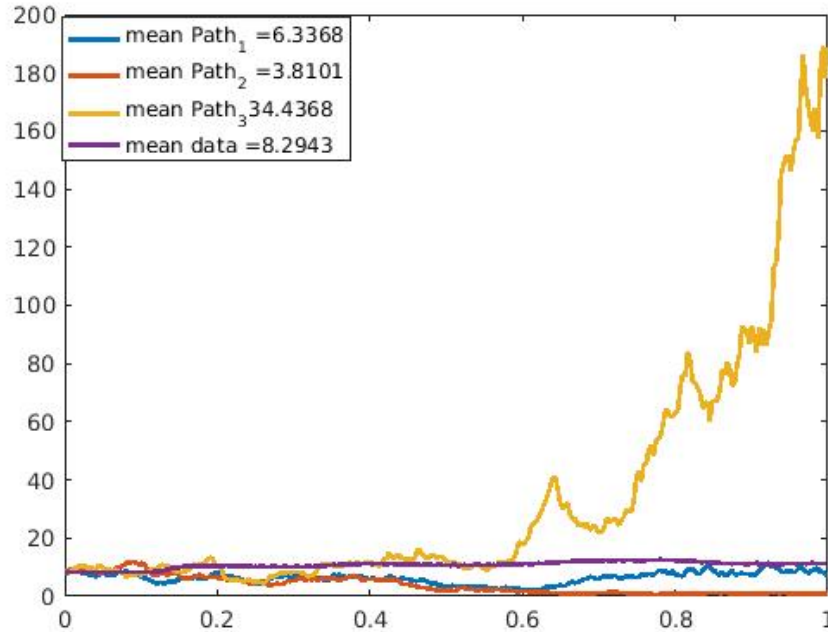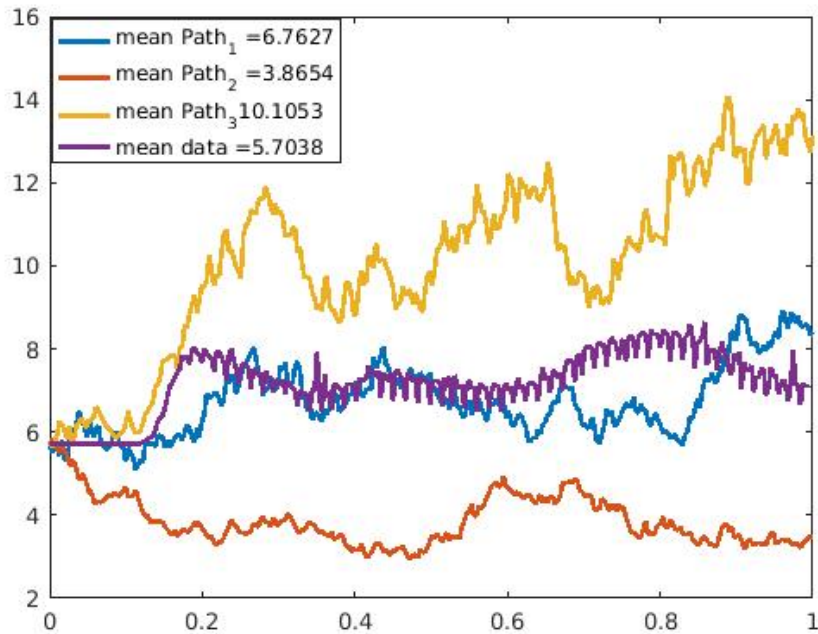
Figure 5.22: Graph showing three sample paths and time series plot for the US COVID-19 deaths. For this time series, we see that sample-path 1 closely predicts it and by comparison of the means, sample-path 1 is closer in value to the mean of the time series data
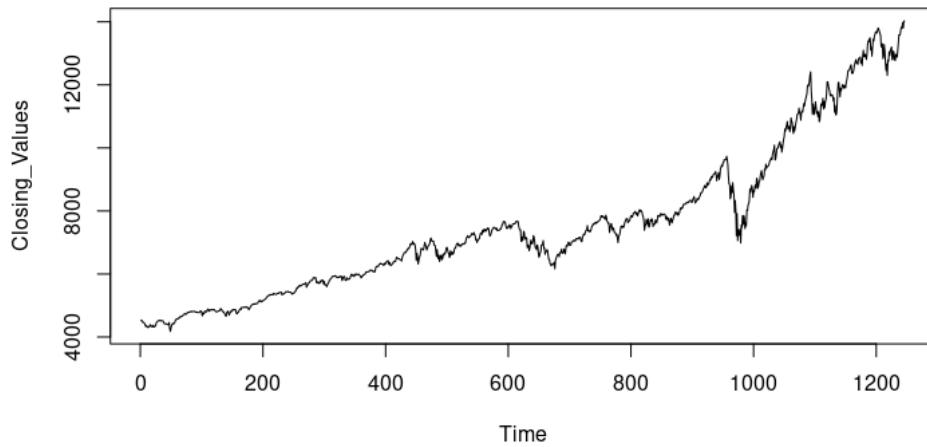
Figure 5.23: Time series plot of daily closing values of the Nasdaq index

87
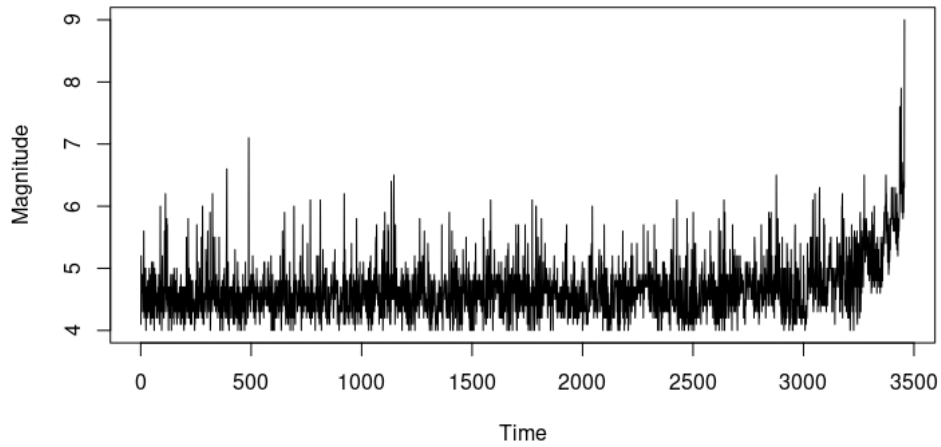
Figure 5.24: Time series plot of recorded earthquake magnitudes after magnitude nine event

Figure 5.25: Time series plot of simulated fractional Brownian motion

sults are compared with those obtained from modeling the data using the other two processes discussed, to assess the influence of the BDP on the performance of the model.

5.4.4 Model Simulation

This section presents the simulation of the three-component superposed Ornstein-Uhlenbeck model, including the software used for the simulation. The simulation process employed in this study is based on the solution of the proposed Ornstein-Uhlenbeck stochastic differential equation in equation 4.25. The model solutions of the Ornstein-Uhlenbeck stochastic differential equation are simulated using a Brownian motion, an Inverse-Gaussian (a,b) process, and a Gam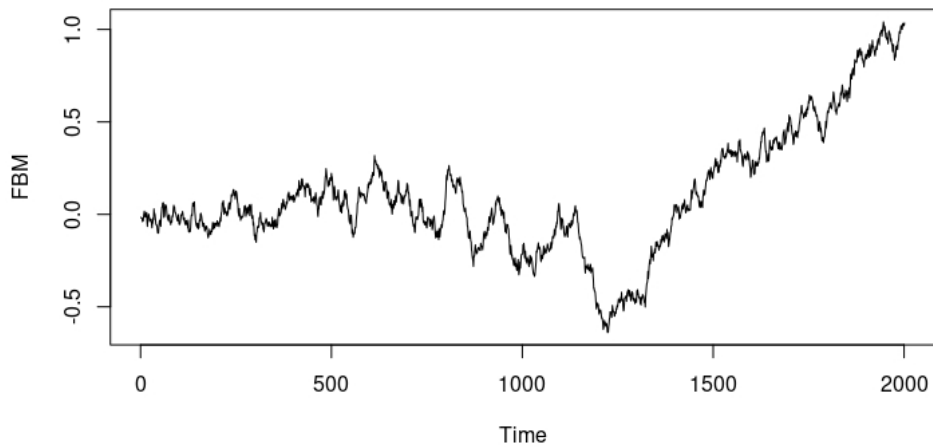ma (a,b) process. The performance of the model is then evaluated by computing the root mean squared errors. A table is generated to summarize the numerical results obtained from simulating each dataset. We use R and MATLAB software for the simulation of the model.

Table 5.29: RMSE's from model simulation with the three different background driving processes

Data	Inverse-Gaussian(a,b)	Gamma(a,b)	Brownian Motion
Nasdaq	0.0286	0.25774	1.5939
AfterM9	0.0488	0.8037	1.3796
FBM	1.0525	1.4037	0.478

5.4.5 Discussion

As seen from Table 5.29, for both stock market and earthquake data, using the standard Brownian motion as the BDP results in a comparatively lower performance of the three-component superposed Ornstein-Uhlenbeck model. On the other hand, using a Lévy-driven BDP for a time series characterized as a Lévy process, the three-component superposed Ornstein-Uhlenbeck model performs better. Similarly, if the time series data has Gaussian characteristics, the table shows that using the standard Brownian motion as the BDP leads to better model performance compared to using a Lévy-driven BDP (Inverse-Gaussian (a,b) or Gamma (a,b)). In the case of Lévy-driven BDP, the Inverse-Gaussian (a,b) performs better than the Gamma (a,b). Therefore, further examination of the time series data is

necessary to determine the best performance when modeling with a Lévy-driven three-component superposed Ornstein-Uhlenbeck equation.

Chapter 6

Conclusion

This dissertation aims to classify the performance of Ornstein-Uhlenbeck-type models using the fractal properties of specific time series data. Time series data from the stock market, volcano-seismic events, earthquake events, and disease spread events have been utilized. In addition, two Brownian motions are simulated to act as controls in the performance analysis of the models. By analyzing the scaling and dynamic behavior of the time series under observation, each has been characterized as either a Gaussian or a Lévy process (Lévy walk or Lévy flight). The first part of this work aims to characterize the time series data used in our Ornstein-Uhlenbeck-type models. We use the three scaling methods discussed in Chapter 2, namely the rescaled range analysis, the detrended fluctuation analysis, and the diffusion entropy analysis, in addition to the characterization relation:

$$\delta = \frac{1}{3 - 2(H, \alpha)}. \tag{6.1}$$

where δ is the PDF scaling factor of the diffusion entropy analysis and H and α are the Hurst exponents of the rescaled range analysis and the detrended fluctuation analysis, respectively. Through the relation in equation 6.1, we can conclude that the distribution of each time series is characterized by either a Gaussian or Lévy process.

The distribution of the time series is characterized as a Gaussian process if $\delta = (H, \alpha) = 0.5$, as a Lévy walk if $\delta = \frac{1}{3-2(H,\alpha)}$, and as a Lévy flight if neither $\delta = \frac{1}{3-2(H,\alpha)}$ nor $\delta = (H, \alpha) = 0.5$ holds. The technique used in part one of the work also helps determine the long- and short-memory effects of the time series. To obtain a good fit for the data, the stationary behavior of each time series is analyzed using unit root tests. Investigating

the stationarity of the time series in Tables 3.1 - 3.4 shows that the stock market data, the COVID-19 data, the simulated Gaussian processes, and the volcanic data are all non-stationary since their p-values are greater than $\alpha = 0.05$. However, the two earthquake datasets are stationary, with a p-value of $0.01 < \alpha = 0.5$.

Besides using the scaling exponents presented in Chapter 2 to characterize the time series, they also provide information on the time series' long- and short-term memory effects. A time series with long memory indicates that the system's evolution is influenced by the system's previous state over long periods. For the long memory effects of the time series using the scaling methods described in Chapter 2, we observe that for the three scaling methods used, the stock indices have scaling exponents greater than 0.5, as shown in 3.5, suggesting long-term memory. In Table 3.6, the volcanic data exhibits long-memory effects with the DEA and DFA, though the R/S method fails to detect the correct memory behavior. On the other hand, the magnitudes measured before the magnitude nine event in Japan in 2011 (EQ1) show evidence of short-memory behavior, while the magnitudes recorded after the magnitude nine event depict long memory, indicating that the seismic activity before the magnitude nine event was abnormal while the seismic activity after the magnitude nine event was normal. In Table 3.7, both the COVID-19 cases and COVID-19 deaths exhibit short-memory behavior, indicating the possibility of them dying out; this is one advantage of a stochastic disease model over a deterministic one.

In terms of characterization, the stock indices are characterized by a Lévy process, particularly a Lévy walk. Since the scaling exponents derived for the three volcanic eruptions and two earthquake datasets do not satisfy the relation $\delta = \frac{1}{3-2(H,\alpha)}$ or $\delta = (H, \alpha) = 0.5$, both are characterized as Lévy processes, specifically Lévy flight processes. Additionally, neither the COVID-19 cases nor COVID-19 deaths satisfy either the relation $\delta = \frac{1}{3-2(H,\alpha)}$ or $\delta = (H, \alpha) = 0.5$; hence, both are characterized as Lévy processes, specifically Lévy flight. Finally, in Table 3.8, the scaling exponents of the R/S, DFA, and the DEA satisfy the relation $\delta = H = \alpha \approx 0.5$, which confirms that both simulated Brownian motions are characterized by Gaussian processes based on our methodology. This confirms the correct-

ness of our characterization approach since the two simulated Brownian motions were used as controls and known beforehand to be Gaussian processes.

Thus, with a combination of the DEA, the DFA, and the R/S, the time series data in Chapter 3 have been characterized as either Gaussian or Lévy processes. The Lévy process helps detect stock market crashes, risky seismic events, or the increase in the spread or death cases during a pandemic. Since the high-frequency data follow an almost log-normal distribution, for any finite-variance Lévy process, randomizing time is equivalent to randomizing variance. Thus, the time-varying Lévy process generates stochastic volatility (SV) by randomizing time, which may improve the forecasting performance. The reason is that the SV model considers a stochastic component of the data volatility and estimates the time-varying parameters using filtering techniques to predict future volatility [33].

After characterizing the time series in Chapter 3, Chapter 4 presents the Ornstein-Uhlenbeck-type models used in this dissertation. Three modifications have been made to the classic Ornstein-Uhlenbeck model and used to analyze the time series data to determine the performance of the Ornstein-Uhlenbeck model against the modified models. A two-component and three-component Lévy-driven Ornstein-Uhlenbeck model is presented in the first two modifications. A parameter estimation procedure is developed for each superposed model to estimate the unknown parameters λ_i and w_i, with $\sum w_i \approx 1$ due to rounding errors. The third modification introduced is the Lévy-driven coupled Ornstein-Uhlenbeck model, used in applications to model events with dependencies. For our Lévy processes, the $\Gamma(a, b)$ and the $IG(a, b)$ have been used.

Chapter 5 presents three applications of the Ornstein-Uhlenbeck type models. In the first application, the performance of the Ornstein-Uhlenbeck model is investigated based on the superposition of solutions. To do this, the classic Ornstein-Uhlenbeck model is compared with the two- and three-component Lévy-driven Ornstein-Uhlenbeck models, respectively. For this application, the datasets used are the Dow Jones, Standard and Poor 500, NASDAQ, and Russell. After the model simulation is completed, it is observed that the superposition of the Ornstein-Uhlenbeck model improves its overall accuracy, as seen

from the two- and three-component OU models in comparison to the ordinary OU model. However, increasing the superposed components may not always lead to improvements, as shown in the two- and three-component IG(a,b) Ornstein-Uhlenbeck models. This could result from increased rounding errors introduced due to estimating more unknown parameters. As seen in this application, the performance of the Ornstein-Uhlenbeck model can be significantly improved by the superposition of solutions.

The second application analyzes the dependencies between time series events from similar or different fields. In this investigation, a coupled Lévy-driven Ornstein-Uhlenbeck model is used. The results of the second application show that the coupled Ornstein-Uhlenbeck model produces good predictions when the dependencies between the datasets are considered. The results also show that some events trigger other events, as seen with the stock market crash during the COVID-19 pandemic. Also, by modeling the US COVID-19 data using a stochastic SDE, we observe from the three sample paths drawn in Figures 5.21 and 5.22 that either the disease would potentially die out or the reported cases or deaths would be very few (almost no new related cases or deaths) at some point after it has peaked once or multiple times, thus showing there exist scenarios where the disease dies out.

Finally, in the third application, the Ornstein-Uhlenbeck model is applied to datasets from three different fields using three different background-driving processes to compare the model's predictive performance. The background driving processes used are either Gaussian or Lévy background driving processes. One simulated Gaussian process is used as a control in this application. From the results obtained, it is observed that for a time series characterized by a Lévy process, a Lévy-driven Ornstein-Uhlenbeck model performs better, while for a time series characterized by a Gaussian process, the standard Brownian motion as a background-driving process performs better.

Putting the three applications together, we arrive at three factors that could be considered when aiming to improve the performance of Ornstein-Uhlenbeck-type models. These are listed below.

1 The selection of an appropriate background driving process, based on the characterization of the time series improves the overall model performance.

2 Extending the Ornstein-Uhlenbeck SDEs solution to a superposition of two or three solutions improves the performance of the model.

3 For events that follow different events, modeling them with a coupled Ornstein-Uhlenbeck system that considers their dependencies, improves the model performance.

6.1   Future Work

1 An extension of this study could involve exploring other entropy measures apart from the Shannon entropy in the DEA methodology to enhance the accuracy of the δ scaling. Some examples of alternative entropies are the Kolmogorov entropy and the Rényi entropy.

2 This dissertation utilized a coupled system to analyze time series data assuming that both time series had the same distribution. An extension of this research could involve using a system of superposed Ornstein-Uhlenbeck models to model two distinct complex systems, where each time series is characterized by a different distribution.

6.1.1   Different Entropy Measures for DEA

In developing the DEA, Scafetta et al. [11] used the Shannon entropy to compute the probability of finding a given data point in a given cell. The scaling exponent derived from the DEA was shown to reflect the true scaling for the time series compared to the variance scaling methods. For the first proposed work, we can apply a different entropy measure in place of the Shannon entropy to explore how the accuracy of the scaling exponent would be affected. Below, we briefly discuss the Kolmogorov and Rényi entropy.

Kolmogorov Entropy

The Kolmogorov entropy is defined as the rate of change of the entropy due to finer partitioning given by each iteration or time step. Suppose we partition our diffusion process into non-overlapping boxes β_i and suppose after each iteration all we know is in which box the point x_i lies, Let $M^{-1}(\beta_i)$ be the pre-image of each β_i. The measurement that each x_0 lies in box i_0 and x_1 lies in box i_1 tells us that x_0 in fact lies in the region $\beta_{i_0} \cap M^{-1}(\beta i_1)$ this intersection gives a finer partition β_1 which can be expressed as $\beta_1 = \beta_i \cap M^{-1}(\beta_j)$ with which we get the entropy of that partition as $S(\beta_1) = -\Sigma p_{ij} \log p_{ij}$ where p_{ij} is the integral of the measure over the box $\beta_i \cap M^{-1}(\beta_j)$. The Kolmogorov entropy is thus measured as

$$K = \lim_{m \to \infty} \frac{1}{m} S(\beta_m).$$

As the definition shows, the Kolmogorov entropy measures the probability of iteratively refining each partition at each iteration.

Rényi Entropy

The Rényi entropy generalizes entropies like the Hartley entropy, the Collision entropy, and the Shannon entropy, which has been used by Scafetta et al. in the derivation of the scaling exponent of the diffusion entropy analysis (DEA). Since the Rényi entropy is a generalization of the Shannon entropy, we would like to explore how the derivation of the entropies for our diffusion process derived from the time series can be improved by using this generalized entropy method.

The Rényi entropy of order α, where $\alpha \geq 0$ and $\alpha \neq 1$ is defined as

$$H_\alpha(X) = \frac{1}{1 - \alpha} \log \Sigma p_i^\alpha$$

where X is a discrete random variable and $p_i = P(X = i)$, $i = 1, 2, 3, ...n$ are the corresponding probabilities.

To achieve our goal, we propose an algorithm similar to the one described for DEA in Chapter 3, but with the Shannon entropy replaced with either the Kolmogorov or Rényi entropy. We will then compare the exponent derived from both entropy measures with the scaling exponent derived from the R/S and the DFA and see how well both methods will perform in the characterization of the time series data.

6.1.2 System of Superposed Ornstein-Uhlenbeck Model involving Different Background-Driving Processes

Assume two stochastic processes recorded in the same region occur within a specified period. And suppose each process is characterized by a different background driving process, in particular, X_1 could be characterized by a gamma process while X_2 is described by a Gaussian process. We propose the coupled system of stochastic differential equations below with background processes Z_1 and Z_2 such that:

$$dX_1(t) = -\lambda X_1(t)dt + \sigma_{11}dZ_1(\lambda_1 t)_{t \geq 0} + \sigma_{12}dZ_2(\lambda_2 t)_{t \geq 0} \qquad (6.2)$$

$$dX_2(t) = -\lambda X_2(t)dt + \sigma_{21}dZ_1(\lambda_1 t)_{t \geq 0} + \sigma_{22}dZ_2(\lambda_2 t)_{t \geq 0}. \qquad (6.3)$$

Where λ_1 and λ_2 are the intensity parameters, σ_{11} and σ_{22} determine volatility , σ_{12} and σ_{21} describes the correlation of the system. The solution of X_1 and X_2 will take the form of the equation

$$X(t) = \Sigma_{i=1}^{m} w_i X_i e^{-\lambda_i t} + \int_0^t \Sigma_{i=1}^{m} e^{-\lambda_i (t-s)} dZ(\lambda_i s). \qquad (6.4)$$

As before, when $\sigma_{12} = \sigma_{21} = 0$ we end up with a decoupled system and thus conclude that the two occurrences do not have any correlation, and when $\sigma_{12} = \sigma_{21} = \sigma_{22} = \sigma_{11} = 0$ we end up with a decoupled deterministic system and each equation can be solved independently. We first characterize the data as Gaussian or Lévy using our proposed

characterization method. We then choose the appropriate background-driving process for the stochastic process being modeled.

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.

[2] Asante, Peter Kwadwo. *Lévy Processes: Characterizing Volcanic and Financial Time Series.* Ph.D. diss., The University of Texas at El Paso, 2020.

[3] Maria C. Mariani, Peter K. Asante, Osei K. Tweneboah and William Kubin (2022) *A 3-component superposed Ornstein-Uhlenbeck model applied to financial stock markets*, Research in Mathematics, 9:1, DOI: 10.1080/27658449.2021.2024339

[4] Mariani, Maria C., Peter K. Asante, William Kubin, and Osei K. Tweneboah. 2022. *Data Analysis Using a Coupled System of Ornstein–Uhlenbeck Equations Driven by Lévy Processes* Axioms 11, no. 4: 160. https://doi.org/10.3390/axioms11040160

[5] Mariani, Maria C., William Kubin, Peter K. Asante, Osei K. Tweneboah, Maria P. Beccar-Varela, Sebastian Jaroszewicz, and Hector Gonzalez-Huizar. 2020. *Self-Similar Models: Relationship between the Diffusion Entropy Analysis, Detrended Fluctuation Analysis, and Lévy Models* Mathematics 8, no. 7: 1046. https://doi.org/10.3390/math8071046

[6] Barany, E., Beccar Varela, M.P., Florescu, I., Sengupta, I. (2012), Detecting market crashes by analyzing long-memory effects using high-frequency data, *Quantitative Finance*, **12(4)**, 623-634, DOI 10.1080/14697688.2012.664937

[7] Mariani M.C. and Liu Y. (2007), Normalized truncated Lévy walks applied to the study of financial indices, *Physica A: Statistical Mechanics and its Applications*, 377, 590-598.

[8] Qi Jingchao and Yang Huijie (2011), Hurst exponents for short time series, *Physical Review E*, 84, 066114.

[9] Beccar Varela, M.P., Huizar, H.G., Mariani, M.C. and Tweneboah O.K. (2019), Lévy Flights and Wavelets Analysis of Volcano-Seismic Data, *Pure and Applied Geophysics*, 1420-9136.

[10] Scafetta, N. (2003), *An Entropic Approach to the Analysis of Time Series*, University of North Texas Libraries.

[11] Scafetta, N. and Grigolini, P. (2002), Scaling detection in time series: diffusion entropy analysis, *Physical Review E*, **66(3)**.

[12] Scafetta, N., Latora, V. and Grigolini, P. (2002), Lévy scaling: The diffusion entropy analysis applied to DNA sequences, *Physical review. E, Statistical, nonlinear, and soft matter physics*, 66. 031906. 10.1103/PhysRevE.66.031906.

[13] Vasile V. M., Luiza, B.I., Vamoş, C., Şoltuz, S.M. (2007), Detrended Fluctuation Analysis of Autoregressive Processes, *arXiv:0707.1437*.

[14] Mariani, M.C. and Tweneboah, O.K. (2016), Stochastic differential equations applied to the study of geophysical and financial time series, *Physica A: Statistical Mechanics and its Applications*, **443**, 170-178.

[15] Krištoufek, Rescaled range analysis, and Detrended Fluctuation Analysis: Finite sample properties and confidence intervals, *Czech Economic Review*, Charles University Prague, **4(3)**, 315-329.

[16] Huang, J., Shang, P., Zhao, X. (2012), Multifractal diffusion entropy analysis on stock volatility in financial markets, *Physica A: Statistical Mechanics and its Applications*, **391(22)**, 5739-5745.

[17] Haubold, H. J., Mathai, A. M., Saxena, R. K. (2014). Analysis of solar neutrino data from Super-Kamiokande I and II. *Entropy*, 16(3), 1414-1425.

[18] Brooks, C. (1995), A measure of persistence in daily pound exchange rates, *Applied Economics Letters*, **2(11)**, 428-431, ISSN 1466-4291.

[19] $https://en.wikipedia.org/wiki/Rescaled\_range$

[20] $http://fusionwiki.ciemat.es/wiki/Long-range\_correlation$

[21] Hurst, H. E. (1951), Long-term storage capacity of reservoirs, *Trans. Am. Soc. Eng.*, 116, 770-799.

[22] Peng, C.K., Sergey V. B., Shlomo H., Simons, M., Stanley, H.E. and Goldberger, A.L. (1994), Mosaic organization of DNA nucleotides, *Physical review. E*, 49, 1685.

[23] Buldyrev, S.V., Goldberger, A.L., Havlin, S., Mantegna, R.N., Matsa, M.E., Peng, C.K., Simons, M. and Stanley, H.E. (1995), Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis, *Physical review. E*, 51, 5084.

[24] Heneghan, C. and McDarby, G. (2000), Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes, *Physical review. E*, 62, 6103.

[25] https://www.physionet.org/tutorials/fmnc/node5.html

[26] Siwy, Z., Ausloos, M. and Ivanova, K. (2000), Correlation studies of open and closed state fluctuations in an ion channel: Analysis of ion current through a large-conductance locust potassium channel, *Physical review. E*, 65, 031907.

[27] Janosi, I.M. and Muller, R. (2005), Empirical mode decomposition and correlation properties of long daily ozone records., *Phys Rev E Stat Nonlin Soft Matter Phys.*, 71, 056126.

[28] Santhanam, M.S., Bandyopadhyay, J.N. and Angom, D. (2006), Quantum spectrum as a time series: Fluctuation measures, *Physical review. E*, 73, 015201.

[29] Talkner, P., Lutz, E. and Hänggi, P. (2007), Fluctuation theorems: Work is not an observable, *Physical review. E*, 75, 032903.

[30] *http : //www.ueltschi.org/teaching/chapShannon.pdf*

[31] *http : //galileo.phys.virginia.edu/classes/*152.*mf*1*i.spring*02*/Entropy.pdf*

[32] Mariani, M.C., Bhuiyan, M.A.Masum, Tweneboah, O.K., Gonzalez-Huizar, H. and Florescu, I. (2018), Volatility models applied to geophysics and high-frequency financial market data, *Physica A: Statistical Mechanics and its Applications*, 503, 304-321.

[33] Mariani, M.C., Bhuiyan, M.A.Masum, Tweneboah, O.K. (2018), Estimation of stochastic volatility by using Ornstein–Uhlenbeck type models, *Physica A: Statistical Mechanics and its Applications*, 491, 167-176.

[34] Bartlomiej Dybiec, Ewa Gudowska-Nowak, Eli Barkai, Alexandre A. Dubkov, (2017), Lévy flights versus Lévy walks in bounded domains. *PhysRevE*,052102(13),1-9.

[35] Graves Timothy, Gramacy B. Robert, Watkins W. Nicholas, Franzke L.E. Christian, (2016), A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA. *Entropy*,19(9),437

[36] Berestycki, H., Busca, J., Florent, I., (2004), Computing the implied volatility in stochastic volatility models. *Commun.Pure Appl.Math.*,57(10),1352-1373.

[37] Mariani, M. C., Bianchini, A., Bandini, P. (2012). Normalized truncated Levy walk applied to flexible pavement performance. *Transportation Research Part C: Emerging Technologies*, 24, 1-8.

[38] Semere Habtemicael, Indranil SenGupta *Ornstein-Uhlenbeck processes for geophysical data analysis.*, Physica A: Statistical Mechanics and its Applications,399, pages 147-156.

[39] Oravecz, Zita, Francis Tuerlinckx, and Joachim Vandekerckhove. *Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model.* Multivariate Behavioral Research 51, no. 1 (2016): 106-119.

[40] Oravecz, Zita, Francis Tuerlinckx, and Joachim Vandekerckhove. *A hierarchical Ornstein–Uhlenbeck model for continuous repeated measurement data.* Psychometrika 74, no. 3 (2009): 395-418.

[41] Helbing, Dirk, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Ulf Blanke, Olivia Woolley-Meza, Mehdi Moussaid, et al. *Saving human lives: What complexity science and information systems can contribute.* Journal of statistical physics 158, no. 3 (2015): 735-781.

[42] Kastner, Gregor. *Dealing with stochastic volatility in time series using the R package stochvol.* arXiv preprint arXiv:1906.12134 (2019).

[43] Jakub Obuchowski and Agnieszka Wylomanska *Ornstein-Uhlenbeck Process with Non-Gaussian Structure.*, Acta Physica Polonica Vol.44 (2013) 1123-1133.

[44] Maller, Ross A., Müller, G., Szimayer A. *Ornstein-Uhlehnbeck Process and Extensions.* Handbook of Financial Time Series (2009): 421-437.

[45] Lai, J., Ji, D., Yan, Z. *Extended Inverse Gaussian Distribution: Properties and Application.* Journal of Shanghai Jiaotong University (Science) (2019): 1-8

[46] O. Vasiček $https://en.wikipedia.org/wiki/Vasicek\_model$

[47] G.E. Uhlenbeck, L.S. Ornstein, (1930) Phys. Rev. 36, 823.

[48] Ernst Eberlein *Jump-type Lévy processes* Department of Mathematical Stochastics, University of Freiburg, Eckerstr.

[49] Aalen O.O., Gjessing H.K. *Survival Models Based on the Ornstein-Uhlenbeck Process* Lifetime Data Anal 10, 407-423 (2004). https://doi.org/10.1007/s10985-004-4775-9

[50] Carmona P., Petit F., Yor M.*On the distribution and asymptotic results for exponential functionals of Lévy processes. In: Yor M (ed) Exponential functionals and principal values related to Brownian motion.*(1997) Bibl. Rev. Mat. Iberoamericana, Madrid.

[51] De Haan L., Karandikar R.L. *Embedding a stochastic difference equation in a continuous-time process.* (1989) Stoch. Proc. Appl. 32:225–235

[52] Bernis, G., Brignone, R., Scotti, S., Sgarra, C., *A Gamma Ornstein-Uhlenbeck Model Driven by a Hawkes Process* (2020). SSRN:http://dx.doi.org/10.2139/ssrn.3370304

[53] Valdivieso L., Schoutens W., Tuerlinckx F. *Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type.* Statistical Inference for Stochastic Processes. 2009,12(1):1-9.

[54] Antonio Punzo *A new look at the inverse Gaussian distribution with applications to insurance and economic data.* (2019) Journal of Applied Statistics, 46:7, 1260-1287, DOI:10.1080/02664763.2018.1542668

[55] Barndorff-Nielsen O.E., Shephard N. *Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics.* Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63(2):167-241.

[56] Wikipedia contributors, *"Euler–Maruyama method,"* Wikipedia, The Free Encyclopedia, *https : //en.wikipedia.org/w/index.php?title = Euler%E2%80%93Maruyama_method&oldid =* 1005602209 (accessed October 20, 2020).

[57] $https://en.wikipedia.org/wiki/Stochastic_process$

[58] Nicolato E, Venardos E. *Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type.* Mathematical Finance: An International Journal of Mathematics, Statistics, and Financial Economics. 2003 Oct;13(4):445-66.

[59] Hashemi, Fariba. *Dynamics of firm size in the healthcare industry.* Health 4, no. 3 (2012): 155-164.

[60] Futoma, Joseph. *Gaussian process-based models for clinical time series in healthcare.* Ph.D. diss., Duke University, 2018.

[61] Janczura, Joanna, Sebastian Orzeł, and Agnieszka Wyłomańska. *Subordinated α-stable Ornstein–Uhlenbeck process as a tool for financial data description.* Physica A: Statistical Mechanics and its Applications 390, no. 23-24 (2011): 4379-4387.

[62] Abadie, Luis M. *Current expectations and actual values for the clean spark spread: The case of Spain in the Covid-19 crisis.* Journal of Cleaner Production 285 (2021): 124842.

[63] Valdivieso L., Schoutens W., Tuerlinckx F. *Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type.* Statistical Inference for Stochastic Processes. 2009,12(1):1-9.

[64] Tian, Meng-Yu, Can-Jun Wang, Ke-Li Yang, Peng Fu, Chun-Yan Xia, Xiao-Jing Zhuo, and Lei Wang. *Estimating the nonlinear effects of an ecological system driven by Ornstein-Uhlenbeck noise.* Chaos, Solitons & Fractals 136 (2020): 109788.

[65] Mandal, Manotosh, Soovoojeet Jana, Swapan Kumar Nandi, Anupam Khatua, Sayani Adak, and T. K. Kar. *A model-based study on the dynamics of COVID-19: Prediction and control.* Chaos, Solitons & Fractals 136 (2020): 109889.

[66] Calafiore, Giuseppe C., Carlo Novara, and Corrado Possieri. *A time-varying SIRD model for the COVID-19 contagion in Italy.* Annual reviews in control (2020).

[67] Covid, I. H. M. E. *Modeling COVID-19 scenarios for the United States.* Nature Medicine 27, no. 1 (2021): 94.

[68] Caprini, Lorenzo, Umberto Marini Bettolo Marconi, Andrea Puglisi, and Angelo Vulpiani. *The entropy production of Ornstein–Uhlenbeck active particles: a path integral method for correlations.* Journal of Statistical Mechanics: Theory and Experiment 2019, no. 5 (2019): 053203.

[69] Eliazar, Iddo, and Joseph Klafter. *Lévy, Ornstein–Uhlenbeck, and subordination: Spectral vs. jump description.* Journal of statistical physics 119, no. 1 (2005): 165-196.

[70] Donado, F., R. E. Moctezuma, L. López-Flores, M. Medina-Noyola, and J. L. Arauz-Lara. *Brownian motion in non-equilibrium systems and the Ornstein-Uhlenbeck stochastic process.* Scientific Reports 7, no. 1 (2017): 1-7.

[71] Tesnjak, Irena. *Superpositions of Ornstein-Uhlenbeck Type Processes: Intermittency and Applications to Finance.* Michigan State University, 2017.

[72] Robock, Alan. *Volcanic eruptions and climate.* Reviews of geophysics 38, no. 2 (2000): 191-219.

[73] Linde, Alan T., and I. Selwyn Sacks. *Triggering of volcanic eruptions.* Nature 395, no. 6705 (1998): 888-890.

[74] Sparks, R. Steve J. *Forecasting volcanic eruptions.* Earth and Planetary Science Letters 210, no. 1-2 (2003): 1-15.

[75] Voight, Barry. *A method for prediction of volcanic eruptions.* Nature 332, no. 6160 (1988): 125-130.

[76] Brenguier, Florent, Nikolai M. Shapiro, Michel Campillo, Valérie Ferrazzini, Zacharie Duputel, Olivier Coutant, and Alexandre Nercessian. *Towards forecasting volcanic eruptions using seismic noise.* Nature Geoscience 1, no. 2 (2008): 126-130.

[77] Shu, Yin, Qianmei Feng, Edward PC Kao, and Hao Liu. *Lévy-driven non-Gaussian Ornstein–Uhlenbeck processes for degradation-based reliability analysis.* IIE Transactions 48, no. 11 (2016): 993-1003.

[78] Endres, Sylvia, and Johannes Stübinger. *Optimal trading strategies for Lévy-driven Ornstein–Uhlenbeck processes.* Applied Economics 51, no. 29 (2019): 3153-3169.

[79] Mariani Maria, and Osei Kofi Tweneboah. *Modeling high-frequency stock market data by using stochastic models.* Stochastic Analysis and Applications (2021): 1-16.

[80] Jakub Obuchowski and Agnieszka Wylomanska *Ornstein-Uhlenbeck Process with Non-Gaussian Structure.*, Acta Physica Polonica Vol.44 (2013) 1123-1133.

[81] Barndorff-Nielsen O.E., Shephard N. *Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics.* Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63(2):167-241.

[82] Mariani, M.C., Tweneboah, O.K. *Stochastic Differential Equations Applied to the Study of Geophysical and Financial Time Series.*, Physica A 443 (2016) 170-178

[83] *https://www.adn.com/science/article/alaskas-biggest-volcanic-eruptions/2012/02/28/*

[84] Antonio Punzo *A new look at the inverse Gaussian distribution with applications to insurance and economic data.* (2019) Journal of Applied Statistics, 46:7, 1260-1287, DOI:10.1080/02664763.2018.1542668

[85] *https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html*

[86] Kevin B. Hendricks, Brian W. Jacobs, Vinod R. Singhal (2020) *Stock Market Reaction to Supply Chain Disruptions from the 2011 Great East Japan Earthquake.* Manufacturing and Service Operations Management 22(4):683-699. https://doi.org/10.1287/msom.2019.0777

[87] $https : //www.r-bloggers.com/2018/01/r-function-for-simulating-gaussian-processes/$

[88] Taqqu, Murad S., and Vadim Teverovsky. *On estimating the intensity of long-range dependence in finite and infinite variance time series.* A practical guide to heavy tails: statistical techniques and applications 177 (1998): 218.

[89] $https : //en.wikipedia.org/wiki/Rescaled_range$

[90] $https : //blogs.cfainstitute.org/investor/2013/01/30/rescaled-range-analysis-a-method-for-detecting-persistence-randomness-or-mean-reversion-in-financial-markets/$

[91] Liu, Shuhong. *Stochastic Calculus and Stochastic Differential Equations (2019).*

[92] Cox, David Roxbee, and Hilton David Miller. *The theory of stochastic processes. Routledge, 2017.*

[93] F. Baudoin,*Stochastic Processes, Editor(s): Penelope Peterson, Eva Baker, Barry McGaw, International Encyclopedia of Education (Third Edition), Elsevier,2010, Pages 451-452, ISBN 9780080448947,https : //doi.org/10.1016/B978 − 0 − 08 − 044894 − 7.01369 − 5. (https : //www.sciencedirect.com/science/article/pii/B9780080448947013695)*

[94] Maria Francesca Carfora, *Stochastic Processes, Editor(s): Shoba Ranganathan, Michael Gribskov, Kenta Nakai, Christian Schönbach, Encyclopedia of Bioinformatics and Computational Biology, Academic Press,2019, Pages 747-752, ISBN*

9780128114322, $https : //doi.org/10.1016/B978 - 0 - 12 - 809633 - 8.20362 - 2$. $(https : //www.sciencedirect.com/science/article/pii/B9780128096338203622)$

[95] Fabrizio Gabbiani, Steven James Cox, *Chapter 18 - Stochastic Processes, Editor(s): Fabrizio Gabbiani, Steven James Cox, Mathematics for Neuroscientists (Second Edition), Academic Press,2017, Pages 335-349, ISBN 9780128018958, https : //doi.org/10.1016/B978 - 0 - 12 - 801895 - 8.00018 - X. (https : //www.sciencedirect.com/science/article/pii/B978012801895800018X)*

[96] Clas Blomberg,*18 - STOCHASTIC PROCESSES, Editor(s): Clas Blomberg, Physics of Life, Elsevier Science B.V., 2007, Pages 178-182, ISBN 9780444527981, https : //doi.org/10.1016/B978 - 044452798 - 1/50018 - X. (https : //www.sciencedirect.com/science/article/pii/B978044452798150018X)*

[97] J. L. Doob (1971) What is a Martingale? The American Mathematical Monthly, 78:5, 451-463, DOI: 10.1080/00029890.1971.11992788

[98] Karatzas, Ioannis, and Steven E. Shreve. *Brownian motion.* In Brownian motion and stochastic calculus, pp. 47-127. Springer, New York, NY, 1998.

[99] Möters, Peter, and Yuval Peres. *Brownian motion. Vol. 30. Cambridge University Press, 2010.*

[100] Matsuda, Kazuhisa. *Introduction to Brownian Motion.* Department of Economics, The City University of New York (2005).

[101] Cyrille Rossant *IPython Cookbook, Second Edition, a guide to numerical computing and data science in the Jupyter Notebook. https : //ipython − books.github.io/133 − simulating − a − brownian − motion/*

[102] Maria C. Mariani, Peter K. Asante, William Kubin, Osei K. Tweneboah, Maria Beccar-Varela; *Determining the background driving process of the Ornstein-Uhlenbeck model*, Electron. J. Diff. Eqns., Special Issue 02 (2023), pp. 193-207.

[103] Box, G. E.; Jenkins, G. M.; and Reinsel, G. C. (1994). *Time series analysis, forecasting, and control.* Englewood Clifs.

[104] Hassani, H.; and Yeganegi, M. R. (2020). *Selecting optimal lag order in Ljung-Box test.* Physica A: Statistical Mechanics and its Applications, 541, 123700.

Curriculum Vita

Peter K. Asante was born on May 21, 1990, and is currently a Ph.D. candidate in the Computational Science Program at the University of Texas at El Paso (UTEP), supervised by Prof. Maria C. Mariani. He holds two master's degrees, one in mathematics from Kansas State University and the other in computational science from UTEP, and a bachelor's degree in mathematics from the Kwame Nkrumah University of Science and Technology in Kumasi, Ghana. Peter's interest in mathematical models and their ability to explain natural and artificial complex systems led him to pursue his doctorate in computational science.

During his Ph.D. studies, Peter worked as a teaching and research assistant at UTEP and as adjunct faculty at Dona Ana Community College. He has co-authored over ten papers in peer-reviewed journals and presented over twenty research papers and posters. He has received numerous research and travel grants from organizations such as the NIH, the DoE, the College of Science at UTEP, the UTEP Graduate School, and the Computational Science Program supporting his research. His current research focuses on applying data analytics to financial, health, and geophysical time series. He was the RSO and President of the Association of Computational Science Students at UTEP during the 2019–2020 and 2020–21 academic years, respectively. After graduation, Peter plans to pursue a career as a professor at a research university or as a research scientist.

Contact: asantepeterkwadwo@gmail.com

Website:*https://sites.google.com/view/peter-k-asante/home*