

# Analysis and Prediction of High Frequency Foreign Exchange Data

by

Adrian Patrick Kennedy

A thesis  
submitted to the Victoria University of Wellington  
in fulfilment of the  
requirements for the degree of  
Master of Science  
in Statistics.

Victoria University of Wellington  
2018



## **Abstract**

This thesis investigates the stochastic properties of high frequency foreign exchange data. We study the exchange rate as a process driven by Brownian motion, paying particular attention to its sampled total variation, along with the variance and distribution of its increments. The normality of its increments is tested using the Khmaladze transformation-2, which we show is straightforward to implement for the case of testing centred normality. We found that while the process exhibits properties characteristic of Brownian motion, increments are non-Gaussian and instead come from mixture distributions. We also introduce a technical analysis trading strategy for predicting price movements, and employ it using the exchange rate dataset. This strategy is shown to offer a statistically significant advantage, and provides evidence that exchange rates are predictable to a greater extent than current mathematical models suggest.



# Acknowledgments

I would like to thank my supervisor, Prof. Estate Khmaladze for introducing me to probability theory and the study of stochastic processes, which has since become a great interest of mine. The work presented in this thesis would not have been possible without his guidance and advice. I would also like to thank Dr. Yuichi Hirose, Dr. Budhi Surya, and Dr. John Haywood for the help and advice they have given me during my studies. To my family and friends, thank you for your continued support and encouragement. Finally, I would like to give a special thank you to Olsen Data, for providing me with the exchange rate data used in this thesis.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Real Analysis: An Introduction</b>	<b>3</b>
2.1	Topological Spaces . . . . .	3
2.1.1	Introduction . . . . .	3
2.1.2	Compact Spaces . . . . .	7
2.1.3	Compact Metric Spaces . . . . .	9
2.2	Linear Spaces . . . . .	11
2.2.1	Introduction . . . . .	11
2.2.2	Subspaces . . . . .	14
2.2.3	Linear Functionals . . . . .	14
2.2.4	Convex Sets and Functionals . . . . .	16
2.2.5	Normed Linear Spaces . . . . .	17
2.2.6	Euclidean Spaces . . . . .	18
2.3	Linear Functionals . . . . .	22
2.3.1	Introduction . . . . .	22
2.3.2	Norms of Linear Functionals . . . . .	24
2.3.3	The Conjugate Space . . . . .	24
2.3.4	The Weak Topology . . . . .	26
2.4	Linear Operators . . . . .	27
2.4.1	Introduction . . . . .	27
2.4.2	The Adjoint Operator . . . . .	30

<b>3</b>	<b>The Exchange Rate Data</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Quote Volume by Year . . . . .	34
3.3	Minimum Time Discretisation . . . . .	35
<b>4</b>	<b>Brownian Motion</b>	<b>39</b>
4.1	Introduction . . . . .	39
4.2	Total Variation of Brownian Motion . . . . .	41
4.3	The Exchange Rate Process . . . . .	44
4.3.1	Total Variation . . . . .	44
4.3.2	Are Increments Gaussian? . . . . .	46
4.3.3	Variance of Increments . . . . .	59
4.4	Mixtures of Normal Distributions . . . . .	64
<b>5</b>	<b>The Ornstein-Uhlenbeck Process</b>	<b>67</b>
5.1	Introduction . . . . .	67
5.2	Maximum likelihood estimation . . . . .	70
5.3	Ornstein-Uhlenbeck with Periodic Variance . . . . .	72
5.3.1	Maximum Likelihood Estimation . . . . .	72
5.4	The Exchange Rate Process . . . . .	74
<b>6</b>	<b>Fractal Sets</b>	<b>81</b>
6.1	What are Fractals? . . . . .	81
6.2	Dimension . . . . .	82
6.2.1	Box-counting Dimension . . . . .	82
6.2.2	Properties of Box-Counting Dimension . . . . .	85
6.2.3	Hausdorff Dimension . . . . .	87
6.2.4	Properties of Hausdorff Dimension . . . . .	90
6.3	Calculating and Estimating Dimension . . . . .	94
6.3.1	Box-counting . . . . .	94
6.3.2	Potential Theoretic Methods . . . . .	95
6.3.3	Rescaling of Self-similar Fractals . . . . .	96



6.4	Example: The Middle Third Cantor Set . . . . .	97
6.5	Fractal Graphs . . . . .	99
6.6	Brownian Motion as a Fractal . . . . .	103
6.6.1	Fractal Brownian motion . . . . .	106
6.6.2	Estimation of a Fractal Brownian Motion . . . . .	107
6.6.3	A Preliminary Analysis of the Estimator . . . . .	109
6.7	Estimating the Dimension of $X(t)$ . . . . .	112
<b>7</b>	<b>Predicting Future Price Movements</b>	<b>119</b>
7.1	Introduction . . . . .	119
7.2	Trend Reversion . . . . .	120
7.3	Method . . . . .	121
7.3.1	Introduction . . . . .	121
7.3.2	Identifying Periods of Trend Reversion . . . . .	123
7.3.3	Testing significance . . . . .	125
7.4	Trading Strategy . . . . .	127
7.5	Testing the Strategy . . . . .	128
7.5.1	Accuracy of Predictions . . . . .	129
7.5.2	Profitability of Trades . . . . .	129
7.6	Results . . . . .	130
7.6.1	Accuracy of Predictions . . . . .	130
7.6.2	Profitability of Trades . . . . .	133
<b>8</b>	<b>Conclusions</b>	<b>137</b>
8.1	Summary . . . . .	137
8.2	Limitations . . . . .	139
8.3	Further Research . . . . .	140
	<b>Appendix A Notation</b>	<b>143</b>



# List of Figures

3.1	Bid quotes recorded in early 2009 . . . . .	35
3.2	Total quote volume by year . . . . .	36
3.3	Minimum time discretisation . . . . .	37
4.1	Total variation estimates . . . . .	47
4.2	Empirical distributions of increments . . . . .	49
4.3	Density histograms of $\Delta X_i$ . . . . .	50
4.4	Distribution of test statistic $D(v_G^r)$ . . . . .	58
4.5	Example trajectory of $\tilde{v}_n(x; \hat{\theta})$ . . . . .	58
4.6	Trajectories $\tilde{v}_n(x; \hat{\theta})$ of $\Delta X_i$ . . . . .	60
4.7	Variance of increments . . . . .	62
4.8	Weekly periodicity of variance . . . . .	63
4.9	Daily periodicity of variance . . . . .	63
5.1	The exchange rate process on different time scales . . . . .	75
5.2	Variance of Ornstein-Uhlenbeck increments . . . . .	79
6.1	The middle third Cantor set . . . . .	98
6.2	Density histograms of minimal point estimator . . . . .	111
6.3	Variance vs sample size of minimal point estimator . . . . .	112
6.4	Histograms of minimal point and Shevchenko estimators . . . . .	113
6.5	Bid quote process throughout 2010 . . . . .	114
6.6	Counting $N_\delta(X)$ when estimating dimension . . . . .	116
6.7	Estimating the box-counting dimension of $X(t)$ . . . . .	116

7.1	Prediction using trend reversion . . . . .	121
7.2	Example plot of $V(t)$ . . . . .	122
7.3	Diagram for Algorithm 1 . . . . .	125
7.4	Accuracy of predictions vs trading time threshold . . . . .	132
7.5	Accuracy of predictions vs trading depth threshold . . . . .	132

# List of Tables

4.1	Total variation estimates . . . . .	45
4.2	Quantiles of test statistic $D(v_G^r)$ . . . . .	58
4.3	Test Statistics $D(\tilde{v}_n)$ for $\Delta X_i$ . . . . .	59
7.1	Predictions using trading time criteria . . . . .	131
7.2	Predictions using trading depth criteria . . . . .	131
7.3	Profitability of trading time criteria . . . . .	134
7.4	Profitability of trading depth criteria . . . . .	134
7.5	Profits for baseless strategy with trading time criteria . . . . .	134
7.6	Profits for baseless strategy with trading depth criteria . . . . .	135



# Chapter 1

## Introduction

In this thesis, we investigate the properties of high frequency foreign exchange data. We begin in Chapter 2 by reviewing some real analysis, as it plays a fundamental role in probability theory and the study of fractal sets, both of which are topics covered in this thesis. Then, after introducing the dataset and discretising the exchange rate process, we take three different approaches in our investigation.

Firstly, we study the exchange rate as a stochastic process, with the objective of understanding the underlying properties and structure of the exchange rate process. Areas of particular interest to us concern the stationarity, correlation, and distribution of the process. The first question we ask is whether the exchange rate behaves as a process that is driven by Brownian motion. To this end, we review some well known results for Brownian motion, and then compare these results with what we observe empirically. We test the normality of increments using the Khmaladze transformation-2 introduced in [11], deriving the relevant formulae so that this transformation can be employed easily. We also consider the Ornstein-Uhlenbeck process, a mean-reverting process that is driven by Brownian motion. We do so as one could reasonably expect some degree of mean-reversion in exchange rates, given that the relative value of two currencies changes very slowly over time.

The second approach we take is to view the exchange rate data as a fractal set, and ask what we can infer from its fractal properties. Many phenomena observed in financial data are characteristic of fractal sets, in particular statistical self-similarity which manifests in the form of scaling laws. We demonstrate this by reviewing the fractal properties of Brownian motion trajectories, and also deriving their dimension. We introduce a new estimator for the Hurst exponent of a fractal Brownian motion, and compare it empirically to an existing strongly consistent estimator presented in [18]. We also review two separate methods for the estimation of a graph's box-counting dimension, and apply these two methods to the exchange rate process. These estimates then also serve as evidence for whether or not the exchange rate process is driven by Brownian motion.

The final part of this thesis introduces a technical analysis method for predicting future price movements. We first present the underlying ideas of the method before formalising it in the form of an algorithm, and then test it empirically using the exchange rate dataset. We test both the accuracy of our predictions, and also the profitability after taking into account the primary fee associated with foreign exchange trading. Research into the development and efficacy of technical analysis methods goes beyond their immediate use to traders. It also contributes to the growing amount of empirical evidence surrounding the efficient market hypothesis.

The overall objective of this thesis is to better understand exchange rate dynamics within the context of the theory of stochastic processes. Both the fractal and technical analysis components of this thesis also contribute to this objective, as they are fundamentally related to ideas concerning correlation and mean-reversion. We hope to present a complete view of how the evolution of exchange rates relate to familiar Brownian motion-driven processes.



# Chapter 2

## Real Analysis: An Introduction

In this chapter, we review some real analysis, primarily following *Introductory Real Analysis* by Kolmogorov and Fomin. We review some important results concerning topological linear spaces and linear functionals.

### 2.1 Topological Spaces

#### 2.1.1 Introduction

When studying limits, we generally think of metric spaces with some familiar function  $\rho(x, y)$  providing the distance between two points  $x$  and  $y$ . Metric spaces are a special case of a topological space, in which the familiar notions of open sets, limits, and continuity are defined without the need for any metric. In this section, we introduce the idea of a topological space, and review some basic results concerning open sets, limits, and continuity.

**Definition 2.1.** Given any set  $X$ , we define a topology on  $X$  as a system  $\tau$  of subsets  $G \subset X$  with the following properties:

- The set  $X$  itself, along with the empty set  $\emptyset$ , belong to  $\tau$ .

- The system of sets  $\tau$  is closed under countably many unions and finitely many intersections.

Sets belonging to the system  $\tau$  are called *open sets*.

A topological space  $T$  consists of a set  $X$  along with a topology  $\tau$  defined on  $X$ .

The definition of a topological space allows us to define systems of open sets without the need of a metric. Just as in a metric space, we say a set is closed in the topological space if and only if its complement is open. This results in an important duality principle. While by definition open sets are closed under countably many unions and finitely many intersections, closed sets are closed under finitely many unions and countably many intersections.

**Theorem 2.1.** Let  $T = (X, \tau)$  be a topological space. Then any system of closed sets in the topological space  $T$  is closed under finitely many unions and countably many intersections.

*Proof.* The result follows by the Duality Principle for Sets

$$\begin{aligned} T - \cup_i X_i &= \cap_i (T - X_i) \\ T - \cap_i X_i &= \cup_i (T - X_i) \end{aligned}$$

along with the fact that the complement of any closed set is open.  $\square$

With the topology  $\tau$  of a topological space, we have natural ways of defining terms often used in the context of a metric space, without the need of a metric  $\rho(x, y)$ .

- A set  $G$  is a *neighbourhood* of the point  $x$  if  $G$  is an open set containing  $x$ .
- A point  $x$  is a *contact point* of the set  $M$  if every neighbourhood of the point  $x$  contains at least one point belonging to  $M$ .

- A point  $x$  is called a *limit point* of the set  $M$  if every neighbourhood of  $x$  contains infinitely many points belonging to  $M$ .
- The *closure* of a set  $M$ , denoted by  $[M]$ , is the union of all contact points of the set  $M$ .

With these terms well defined in a topological space, we can then generalise many of the important ideas of a metric space. Perhaps the most fundamental being the idea of a convergent sequence  $\{x_n\}$  that gets closer and closer to some limit point  $x$  as  $n \rightarrow \infty$ . This presents a difficulty in defining how we determine whether a sequence converges to some point. In a metric space, we say it converges if the distance from that point gets smaller and smaller, but clearly another approach must be taken in a topological space.

**Definition 2.2.** Let  $\{x_n\}$  be a sequence of points in a topological space  $T$ . Then the sequence is said to converge to a point  $x$  if every neighbourhood of  $x$  contains all points  $x_n$  starting from a certain index.

Hence, a sequence  $\{x_n\}$  is said to converge to a point  $x$  if for any neighbourhood  $G(x)$ , we can find some index  $n$  such that  $x_m \in G(x)$  for all  $m \geq n$ . While we can generalise convergence to a topological space fairly easily, the role it plays is slightly different. For example, in a metric space  $R$ , a point  $x$  is said to be a contact point of a set  $M$  if and only if  $M$  contains a sequence converging to  $x$ . In general, this is not true for a topological space. So while the general idea of convergence remains the same, not all of the properties of convergence in a metric space hold true in the more general topological spaces.

Similarly, the concept of continuity also generalises to topological spaces fairly easily.

**Definition 2.3.** Let  $f : X \rightarrow Y$  be a mapping from one topological space to another. Then we say  $f$  is continuous at a point  $x_0 \in X$  if for any

neighbourhood  $V_{y_0}$  of the point  $y_0 = f(x_0)$ , there exists a neighbourhood  $U_{x_0}$  of the point  $x_0$  such that

$$f(U_{x_0}) \subset V_{y_0}.$$

We say  $f$  is a continuous mapping if it is continuous at every point belonging to  $X$ .

Since in a topological space a neighbourhood  $U_{x_0}$  is simply any open set containing  $x_0$ , continuity can be defined in terms of open sets.

**Theorem 2.2.** Let  $f : X \rightarrow Y$  be a mapping from one topological space to another. Then  $f$  is continuous if and only if the preimage  $f^{-1}(G)$  of any open set  $G \subset Y$  is open.

*Proof.* Firstly, suppose  $f$  is continuous. Then for any open set  $G \subset Y$ , take some point  $x$  belonging to its preimage  $f^{-1}(G)$ , and let  $y = f(x)$ . Since  $G$  is an open set, it is by definition a neighbourhood of  $y$ . Hence by the definition of a continuous function we can find some neighbourhood  $U_x$  of  $x$  such that  $f(U_x) \subset G$ . Hence every point  $x \in f^{-1}(G)$  has an open set wholly contained in  $f^{-1}(G)$ , making  $f^{-1}(G)$  an open set.

For the other direction, suppose that the preimage of every open set is open. For any point  $x \in X$ , we can find some neighbourhood  $V_y$  of the point  $y = f(x)$ . By assumption, since  $V_y$  is open, its preimage  $f^{-1}(V_y)$  must be open. Moreover, by construction we have that  $U_x = f^{-1}(V_y)$  is a neighbourhood of  $x$  such that  $f(U_x) \subset V_y$ , making  $f$  continuous at the point  $x$ , and hence a continuous mapping.  $\square$

A consequence of Theorem 2.2 is that the composition of two continuous functions must also be continuous.

If  $f : X \rightarrow Y$  is a bijective mapping such that both  $f$  and  $f^{-1}$  are continuous, then we say  $f$  is a *homeomorphism* from  $X$  to  $Y$ . We say two topological spaces are homeomorphic if there exists a homeomorphism between them. Homeomorphisms serve as a kind of equivalence in topology since

two homeomorphic spaces have identical topological properties. In this way, two homeomorphic spaces are simply two different representations of the same space. This is demonstrated by the fact that homeomorphism is clearly an equivalence relation, and hence divides all topologies into a set of equivalence classes.

### 2.1.2 Compact Spaces

An important concept in general topology is the idea of a *compact space*. Compactness serves as a generalisation of the properties of closed and bounded subsets of  $\mathbb{R}^n$ .

**Definition 2.4.** We say a topological space  $T$  is compact if every open cover of  $T$  has a finite subcover.

By an open cover of a set, we mean any collection of open sets  $\{U_i\}$  such that  $T \subset \cup_i U_i$ . An alternative definition of compactness can be obtained by considering *centred* systems of sets.

**Definition 2.5.** We say that a system of subsets  $\{U_i\}$  is centred if every finite intersection  $\cap_{i=1}^n U_i$  is non-empty.

**Theorem 2.3.** Let  $T$  be a topological space. Then  $T$  is compact if and only if every centred system of closed subsets of  $T$  has a non-empty intersection.

*Proof.* First, suppose that  $T$  is compact, so that every open cover of  $T$  has a finite subcover. Take any centred system of closed subsets  $\{U_i\}$ . By definition this implies the compliments  $F_i = T - U_i$  are open. If the system of sets  $\{U_i\}$  has an empty intersection, then this implies that  $\{F_i\}$  is an open cover of  $T$ , and hence has a finite subcover. This is a contradiction as it implies that  $\{U_i\}$  is not a centred system of sets.

Conversely, suppose that any system of closed subsets of  $T$  has a non-empty intersection, and take any open cover  $\{F_i\}$  of  $T$ , and let  $U_i = T - F_i$ . Then the system of sets  $\{U_i\}$  must have an empty intersection, since  $\{F_i\}$  is

an open cover, and hence  $\{U_i\}$  cannot be centred. This implies there exists a finite collection of  $F_i$  with an empty intersection, and so by taking their complement we obtain a finite subcover, proving  $T$  is compact.  $\square$

The equivalence given to us by Theorem 2.3 tells us that any closed subset  $F$  of a compact topological space  $T$ , must itself be compact. This follows from the fact that any centred system of closed subsets in  $F$  is also a centred system of closed subsets in  $T$ , and so it must have a non-empty intersection. A similar result holds for continuous mappings.

**Theorem 2.4.** For any compact space  $X$  and continuous function  $f : X \rightarrow Y$ , the topological space  $Y = f(X)$  is compact.

*Proof.* Take any open cover  $\{V_i\}$  of the topological space  $Y$ , and let  $U_i = f^{-1}(V_i)$ . Since  $f$  is continuous, the preimage of any open set is open, and thus  $\{U_i\}$  is a collection of open sets. Furthermore, they cover the space  $X$ , making them an open cover. By the compactness of  $X$  we can then find a finite open subcover of  $\{U_i\}$  and hence a finite subcover of  $\{V_i\}$ .  $\square$

A weaker form of compactness is that of a *countably compact* space. The definition of a countably compact space is motivated by the following important result for any compact topological space  $T$ .

**Theorem 2.5.** Let  $T$  be a compact topological space. Then any infinite subset of  $T$  has at least one limit point.

*Proof.* Let  $X$  be an infinite subset of a topological compact space  $T$ . Assume towards contradiction that  $X$  contains no limit points. This implies that  $X$  contains a countable set  $\{x_1, x_2, \dots\}$  without a limit point. Letting  $X_n = \{x_n, x_{n+1}, \dots\}$ , it follows that  $\{X_n\}$  is a centred system of sets. Furthermore, they are closed since the set  $X$  has no limit points. This contradicts the compactness of  $T$ , as  $X_n$  forms a centred system of closed subsets with an empty intersection.  $\square$

**Definition 2.6.** We say a topological space  $T$  is countably compact if every infinite subset of  $T$  has at least one limit point in  $T$ .

Theorem 2.5 tells us that every compact space is countably compact, however the reverse is not generally true. Hence countable compactness is a weaker form of compactness. This can be seen by noting that in a countably compact space, it can be proven that any countable open cover of  $T$  has a finite subcover, while in a compact space this property holds true for any open cover of  $T$ .

### 2.1.3 Compact Metric Spaces

As mentioned, metric spaces are simply a special kind of a topological space. Hence all of the results thus far hold true in metric spaces. However, since the topology of a metric space is entirely determined by its metric  $\rho(x, y)$ , we can frame the concept of compactness in terms of this metric. We start by introducing the concept of *total boundedness*.

**Definition 2.7.** Let  $R$  be a metric space and  $\varepsilon > 0$  any positive number. Then an  $\varepsilon$ -net of a set  $M \subset R$  is any set  $A$  such that for every  $x \in M$  there exists an  $a \in A$  such that  $\rho(x, a) < \varepsilon$ .

So that an  $\varepsilon$ -net of a set  $M$  is simply another set  $A$  with the property that every element of  $M$  is within  $\varepsilon$  distance of some element of  $A$ . We then define a set as being totally bounded if for any  $\varepsilon > 0$  we can find at least one finite  $\varepsilon$ -net.

**Definition 2.8.** Let  $R$  be a metric space with some subset  $M \subset R$ . Then we say  $M$  is totally bounded if, given any  $\varepsilon > 0$ , we can find a finite  $\varepsilon$ -net of  $M$ .

It is tempting to assume that any totally bounded set is necessarily finite dimensional. However a novel example of a totally bounded and infinite dimensional set is the Hilbert cube. The Hilbert cube consists of all points  $(x_1, x_2, \dots, x_n, \dots)$  that satisfy

$$|x_n| \leq \frac{1}{2^{n-1}} \tag{2.1}$$

for all  $n$ . It is clear that the Hilbert cube is infinite dimensional, the fact it is also totally bounded is more subtle. Suppose that given some  $\varepsilon > 0$ , we wish to construct a finite  $\varepsilon$ -net of the Hilbert cube. We start by choosing  $n$  as the smallest integer satisfying

$$\frac{1}{2^{n-1}} < \frac{\varepsilon}{2}. \quad (2.2)$$

Then, for each point

$$x = (x_1, x_2, \dots, x_n, \dots)$$

belonging to the Hilbert cube, we associate with it a new point

$$x^* = (x_1, x_2, \dots, x_n, 0, 0, \dots).$$

It then follows that

$$\begin{aligned} \rho(x, x^*) &= \sqrt{\sum_{i=1}^{\infty} |x_i - x_i^*|^2} \\ &= \sqrt{\sum_{i=n+1}^{\infty} x_i^2} \\ &\leq \sqrt{\sum_{i=n}^{\infty} 4^{-k}} \\ &< \frac{1}{2^{n-1}} \\ &< \frac{\varepsilon}{2}. \end{aligned}$$

The set of all points  $x^*$  is totally bounded as it is a bounded set in  $\mathbb{R}^n$ , where total boundedness is equivalent to boundedness. Hence we can find an  $\varepsilon/2$ -net of the set of points  $x^*$ , which by the above inequality is an  $\varepsilon$ -net of the Hilbert cube.

We now conclude this section by showing how the concepts of total boundedness and compactness are related.



**Theorem 2.6.** A metric space  $R$  is compact if and only if it's totally bounded and complete.

*Proof.* For brevity we prove only one direction. Let  $R$  be a compact metric space, and let  $\{x_n\}$  be any Cauchy sequence belonging to  $R$ . If the sequence  $\{x_n\}$  does not converge to any element of  $R$ , then this implies the sequence has no limit point belonging to  $R$ . But this contradicts Theorem 2.5, and hence the space is complete.

To prove that  $R$  is totally bounded, assume towards contradiction that it is not. Hence there is some  $\varepsilon > 0$  such that no  $\varepsilon$ -net for  $R$  exists. Hence for any point  $x_1 \in R$ , we can find some  $x_2 \in R$  such that

$$\rho(x_1, x_2) > \varepsilon.$$

Furthermore, we can continue this procedure sequentially to obtain a countable set

$$\{x_1, x_2, \dots, x_n, \dots\}$$

such that  $\rho(x_i, x_j) > \varepsilon$  for all  $i \neq j$ . By construction this set clearly has no limit points in  $R$ , again contradicting Theorem 2.5. Hence  $R$  is totally bounded.  $\square$

## 2.2 Linear Spaces

### 2.2.1 Introduction

Linear spaces play an important role in many areas of mathematics. They are very general yet powerful spaces which include, among others, all Hilbert and Banach spaces. As many important spaces in mathematics are simply specialisations of linear spaces, the structure inherent to all linear spaces is of great interest.

**Definition 2.9.** A non-empty set  $L$  of elements is called a linear space if the following three axioms hold:

1. Any two elements  $x, y \in L$  uniquely determine a third element  $x + y \in L$  referred to as the sum of  $x$  and  $y$ . Furthermore, summation has the following properties:

- Associative:  $(x + y) + z = x + (y + z)$ .
- Commutative:  $x + y = y + x$
- There exists a unique zero element  $0 \in L$  such that  $x + 0 = x$  for all  $x \in L$ .
- For every  $x \in L$  there exists an inverse element  $-x \in L$  such that  $x + (-x) = 0$ .

2. Any number  $\alpha$  and any element  $x \in L$  uniquely determine an element  $\alpha x$ , called the product of  $\alpha$  and  $x$ , such that

- For all numbers  $\alpha, \beta$  and elements  $x \in L$ , we have  $\alpha(\beta x) = (\alpha\beta)x$ .
- $1x = x$ .

3. The above operations obey two distributive laws:

- $(\alpha + \beta)x = \alpha x + \beta x$ .
- $\alpha(x + y) = \alpha x + \alpha y$ .

The elements of a linear space  $L$  are often referred to as "points" or "vectors", while the numbers as "scalars." We call  $L$  a *real* linear space if  $\alpha, \beta \in \mathbb{R}$ . One example of a linear space is the set  $l_2$  of all sequences

$$x = (x_1, x_2, \dots, x_n, \dots)$$

that satisfy the condition

$$\sum_{i=1}^{\infty} |x_i|^2 < \infty, \quad (2.3)$$

with addition and scalar multiplication defined in the obvious way,

$$\alpha(x_1, x_2, \dots) + \beta(y_1, y_2, \dots) = (\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \dots).$$

The fact that  $x + y$  necessarily belongs to  $l_2$  follows from the inequality

$$(x_k + y_k)^2 \leq 2x_k^2 + 2y_k^2.$$

We say that a set of elements  $x_1, x_2, \dots, x_n \in L$  are *linearly dependent* if there exists a set of scalars  $\lambda_1, \lambda_2, \dots, \lambda_n$ , not all zero, such that

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n = 0.$$

If no such scalars exist, then we say the set of elements is *linearly independent*. The concept of linear independence is used to define the *dimension* of a linear space. We say a linear space  $L$  is  $n$ -dimensional if there exists  $n$  linearly independent elements in  $L$ , but all collections of  $n + 1$  elements are linearly dependent. This can be extended to include infinite dimensional linear spaces, in which we can find  $n$  linearly independent elements for all finite  $n$ . Any  $n$ -dimensional linear space can be completely described by a *basis* of  $n$  linearly independent elements.

**Definition 2.10.** A basis of an  $n$ -dimensional linear space is any collection of  $n$  linearly independent elements  $x_1, x_2, \dots, x_n$ .

A basis completely describes a linear space  $L$  in that any vector belonging to  $L$  can be represented as a linear combination of the elements in the basis. The coefficients in this linear combination serve as coordinates, so that every basis serves as a coordinate system for the linear space.

**Theorem 2.7.** Let  $L$  be an  $n$ -dimensional linear space with basis  $x_1, x_2, \dots, x_n$ . Then any element  $y \in L$  can be expressed as a linear combination of  $x_1, x_2, \dots, x_n$ .

*Proof.* If  $y$  cannot be expressed as a linear combination of  $x_1, x_2, \dots, x_n$ , then this implies that  $x_1, x_2, \dots, x_n, y$  is linearly independent. This is a contradiction as  $L$  is  $n$ -dimensional, and hence any collection of  $n + 1$  elements is linearly dependent.  $\square$

## 2.2.2 Subspaces

Suppose a subset  $M$  of a linear space  $L$  is a linear space itself. Then we say that  $M$  is a *subspace* of  $L$  with respect to the same operations of addition and scalar multiplication. A proper subspace is a subspace that is neither the empty set nor  $L$  itself.

As an example, given any linear space  $L$  and fixed scalar  $\lambda$ , the set

$$M = \{\lambda x : x \in L\}$$

is a subspace of  $L$ . This follows from the fact that if  $x, y \in M$ , then for some  $x', y' \in L$  we have

$$\begin{aligned} \alpha x + \beta y &= \alpha \lambda x' + \beta \lambda y' \\ &= \lambda(\alpha x' + \beta y'), \end{aligned}$$

where  $\alpha x' + \beta y' \in L$ .

It is straightforward to see that the intersection of any number of linear spaces must itself be a linear space. As a result, given any collection of elements  $\{x_i\}$  belonging to a linear space  $L$ , we can find a minimal subspace containing  $\{x_i\}$ . This is because there exists at least one subspace containing  $\{x_i\}$ , namely  $L$  itself, and the intersection of all such subspaces gives us a minimal subspace.

## 2.2.3 Linear Functionals

A *functional* on a linear space  $L$  is any numerical function defined on  $L$ . We say that a functional is linear if for all  $x, y \in L$  and scalars  $\alpha$  and  $\beta$ ,

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y). \quad (2.4)$$

For example, consider the space  $C_{[a,b]}$  of all continuous functions on the interval  $[a, b]$ . On this space, the functional

$$I(x) = \int_a^b x(t) dt$$

is linear due to the linearity of integration

For any linear functional  $f$  on a linear space  $L$ , we define the corresponding *null space* as the set

$$L_f = \{x \in L : f(x) = 0\}. \quad (2.5)$$

The null space of a linear functional is necessarily a subspace since, for any  $x, y \in L_f$ , we have

$$\begin{aligned} f(\alpha x + \beta y) &= \alpha f(x) + \beta f(y) \\ &= 0 \end{aligned}$$

so that  $\alpha x + \beta y \in L_f$ .

**Theorem 2.8.** Given a fixed element  $x_0 \in L - L_f$ , every  $x \in L$  has a unique representation of the form

$$x = \alpha x_0 + y \quad (2.6)$$

for some  $y \in L$ .

*Proof.* By definition  $f(x_0) \neq 0$  and hence  $x_0 \neq 0$  by linearity of  $f$ . We can assume  $f(x_0) = 1$  without loss of generality, since otherwise we need only rescale accordingly using  $\alpha$ .

For any  $x \in L$ , we set

$$y = x - f(x)x_0,$$

so that  $y \in L_f$ , since

$$\begin{aligned} f(y) &= f(x - f(x)x_0) \\ &= f(x) - f(x)f(x_0) \\ &= 0. \end{aligned}$$

Hence we have the representation (2.6). The representation not being unique leads to the contradiction that  $x_0 \in L_f$ , hence the result follows.  $\square$

**Corollary 2.8.1.** If two non-trivial linear functionals have the same null space, then they are proportional.

*Proof.* We again take  $f(x_0) = 1$  without loss of generality. Then, since  $f$  and  $g$  share the same null space, it follows that  $g(x_0) \neq 0$ . By Theorem 2.8 every  $x \in L$  can be expressed in the form

$$x = f(x)x_0 + y$$

for some  $y \in L_f$ . Hence,

$$g(x) = f(x)g(x_0)$$

proving the result. □

## 2.2.4 Convex Sets and Functionals

An important concept in the study of linear spaces is that of *convexity*. Loosely speaking, we think of a space as being convex if it contains the line joining any two of its points. To express the notion of convexity more formally, we say that the *line segment* joining two points  $x, y \in L$  is the set of all points  $z$  that can be expressed in the form

$$z = \alpha x + \beta y \tag{2.7}$$

such that  $\alpha + \beta = 1$ , and  $\alpha, \beta \geq 0$ .

**Definition 2.11.** Let  $L$  be a real line space, and  $M \subset L$  a subset. Then  $M$  is said to be convex if whenever  $M$  contains two points  $x$  and  $y$ , it contains the line segment joining  $x$  and  $y$ .

As with subspaces, it follows from the properties of intersection that the intersection of an arbitrary number of convex sets is also convex. The notion of convexity is by no means restricted to sets however, and is easily extended to functionals.

**Definition 2.12.** A functional  $p$  defined on a real linear space  $L$  is said to be a *convex functional* if the following hold:

1.  $p(x) \geq 0$  for all  $x \in L$ .
2.  $p(\alpha x) = \alpha p(x)$  for all  $x \in L$  and  $\alpha \geq 0$ .
3.  $p(x + y) \leq p(x) + p(y)$ .

An important result in linear analysis involving convex functionals is the Hahn-Banach theorem. For any linear space  $L$  with subspace  $M$ , we say that a functional  $f$  defined on  $L$  is an *extension* of  $f_0$  defined on  $M$  if

$$f(x) = f_0(x) \text{ for all } x \in M.$$

The Hahn-Banach theorem then says that if a functional  $f$  is bounded by some convex functional  $p$  on some subspace, then it has an extension that is bounded on the entire linear space. We omit the rather lengthy proof.

**Theorem 2.9** (Hahn-Banach). Let  $L$  be a real linear space with subspace  $M$ , and  $p$  a finite convex functional defined on  $L$ . Then if  $f_0$  is a functional defined on  $M$  such that

$$f_0(x) \leq p(x) \text{ for all } x \in M,$$

then  $f_0$  has an extension on  $L$  that is also bounded by  $p$ .

### 2.2.5 Normed Linear Spaces

Often times a linear space studied in mathematics has some notion of length inherent to its structure. This leads to the idea of a *normed linear space*, a linear space  $L$  along with some norm  $\rho(x)$  that gives the length of any vector  $x \in L$ .

**Definition 2.13.** A normed linear space is a linear space  $L$  equipped with a functional  $\rho$  such that for all  $x \in L$  the following properties hold:

- $\rho$  is a finite and convex functional.
- $\rho(x) = 0$  if and only if  $x = 0$ .

- $\rho(\alpha x) = |\alpha|\rho(x)$  for all scalars  $\alpha$ .

The *norm* of an element  $x$  is  $\|x\| = \rho(x)$ .

Every normed linear space is a metric space when equipped with the metric  $\rho(x, y) = \|x - y\|$ . As a result, properties of metric spaces immediately carry over to normed linear spaces.

An example of a normed linear space is  $C_{[a,b]}$  equipped with the norm

$$\|f\| = \max_{a \leq t \leq b} |f(t)|.$$

A particularly important consequence of the Hahn-Banach theorem given in the previous section concerns functionals in normed linear spaces.

**Theorem 2.10 (Hahn-Banach).** If a bounded linear functional  $f$  is defined on some subspace  $L$  of a normed linear space  $E$ , then  $f$  can be extended to the entire space  $E$  without increasing its norm.

*Proof.* An immediate consequence of applying the Hahn-Banach theorem with the convex functional

$$p(x) = \|f\|_{\text{on } L} \|x\|.$$

□

Any normed linear space that is also complete is called a *Banach space*. With the introduction of length to a linear space, it is natural to ask whether or not we can also generalise the idea of an angle between two vectors. As it turns out we can, in what is called a *Euclidean space*.

## 2.2.6 Euclidean Spaces

It is easy to see how one might generalise the angle between two vectors to  $\mathbb{R}^n$ , it is less clear how to define it in, for example,  $l_2$  or  $C_{[a,b]}$ . To do so, we introduce what is called the *scalar product*.



**Definition 2.14.** A Euclidean space is any real linear space  $R$  equipped with a scalar product, denoted by  $(x, y)$ , defined for all  $x, y \in R$  such that the following conditions hold:

1.  $(x, x) \geq 0$  and  $(x, x) = 0$  if and only if  $x = 0$ .
2.  $(x, y) = (y, x)$ .
3.  $(\lambda x, y) = \lambda(x, y)$ .
4.  $(x, y + z) = (x, y) + (x, z)$ .

for all  $x, y, z \in R$  and  $\lambda \in \mathbb{R}$ .

Some examples of Euclidean spaces include:

- The space  $\mathbb{R}^n$  equipped with  $(x, y) = \sum_{i=1}^n x_i y_i$ .
- The space  $l_2$  equipped with  $(x, y) = \sum_{i=1}^{\infty} x_i y_i$ .
- The space  $C_{[a,b]}$  equipped with  $(x, y) = \int_a^b x(t)y(t)dt$ .

Hence Euclidean spaces are far more general than the familiar Euclidean  $n$ -space  $\mathbb{R}^n$ .

Every Euclidean space is a normed linear space when equipped with the norm

$$\|x\| = \sqrt{(x, x)},$$

where the properties of norm can be proven using the Schwarz inequality,

$$|(x, y)| \leq \|x\| \|y\|.$$

The notion of a scalar product is fundamental as it allows us to define the angle between two vectors.

**Definition 2.15.** For any two vectors  $x$  and  $y$  belonging to a Euclidean space  $R$ , the angle  $\theta$  between  $x$  and  $y$  is defined by the equation

$$\cos \theta = \frac{(x, y)}{\|x\| \|y\|}. \quad (2.8)$$

A consequence of equation (2.8) is that

$$(x, y) = 0 \implies \theta = \frac{\pi}{2}.$$

Hence, we say that two vectors  $x$  and  $y$  are *orthogonal* if  $(x, y) = 0$ . Furthermore, we say a set of vectors  $x_1, x_2, \dots, x_n$  is an *orthogonal system* if  $(x_i, x_j) = 0$  whenever  $i \neq j$ . An orthogonal system is said to be *orthonormal* if  $(x_i, x_i) = 1$  for all  $i$ .

**Theorem 2.11.** Any orthogonal system of vectors  $x_1, x_2, \dots, x_n$  is linearly independent.

*Proof.* Suppose that for some scalars  $\alpha_i$ , we have

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0.$$

Then for any  $x_i$ , consider the scalar product

$$(x_i, \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n) = (x_i, 0) = 0.$$

But then the orthogonality of the vectors implies that

$$(x_i, \alpha_i x_i) = 0$$

so that  $\alpha_i = 0$ . Since this holds for all  $i$ , all the scalars  $\alpha_i$  must be 0. Hence the vectors are linearly independent.  $\square$

Orthonormal systems serve as a basis of the Euclidean space comprised of vectors of unit length. For example, if  $e_1, e_2, \dots, e_n$  is an orthonormal system in  $\mathbb{R}^n$ , then any vector  $x \in \mathbb{R}^n$  can be expressed as

$$x = \sum_{i=1}^n (x, e_i) e_i. \tag{2.9}$$

A similar result also holds for infinite dimensional Euclidean spaces, which we now prove.

**Theorem 2.12.** Let  $R$  be an infinite dimensional Euclidean space with orthonormal system

$$e_1, e_2, \dots, e_n, \dots$$

Then for any element  $x \in R$ , the expression

$$\left\| x - \sum_{i=1}^n \alpha_i e_i \right\|$$

achieves its minimum with  $\alpha_i = (x, e_i)$ . Furthermore, Bessel's inequality says that

$$\sum_{i=1}^{\infty} (x, e_i)^2 \leq \|x\|^2. \quad (2.10)$$

*Proof.* Let  $S_n$  be the partial sum process

$$S_n = \sum_{i=1}^n \alpha_i e_i. \quad (2.11)$$

Then it follows that

$$\begin{aligned} \|x - S_n\|^2 &= \left( x - \sum_{i=1}^n \alpha_i e_i, x - \sum_{i=1}^n \alpha_i e_i \right) \\ &= (x, x) - 2 \left( x, \sum_{i=1}^n \alpha_i e_i \right) + \left( \sum_{i=1}^n \alpha_i e_i, \sum_{i=1}^n \alpha_i e_i \right) \\ &= \|x\|^2 - 2 \sum_{i=1}^n \alpha_i (x, e_i) + \sum_{i=1}^n \alpha_i^2 \end{aligned}$$

which can then easily be rewritten as

$$\|x - S_n\|^2 = \|x\|^2 - \sum_{i=1}^n (x, e_i)^2 + \sum_{i=1}^n (\alpha_i - (x, e_i))^2. \quad (2.12)$$

So that the difference  $\|x - S_n\|^2$  achieves its minimum when the last term of the right hand side vanishes, when  $\alpha_i = (x, e_i)$  for  $i = 1, 2, \dots, n$ . Furthermore, since  $\|x - S_n\|^2 \geq 0$ , Bessel's inequality follows from (2.12).  $\square$

We have seen that any Euclidean space can be made into a normed linear space. An interesting question is then under what condition is a normed linear space Euclidean.

**Theorem 2.13.** A normed linear space  $R$  is Euclidean if and only if

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \quad (2.13)$$

for all  $x, y \in R$ .

*Proof.* We prove only one direction, as the other requires a very lengthy proof. Suppose that  $R$  is Euclidean. Then it follows that

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= (x + y, x + y) + (x - y, x - y) \\ &= (x, x) + 2(x, y) + (y, y) + (x, x) - 2(x, y) + (y, y) \\ &= 2(x, x) + 2(y, y) \\ &= 2(\|x\|^2 + \|y\|^2). \end{aligned}$$

□

## 2.3 Linear Functionals

### 2.3.1 Introduction

With a linear functional  $f$  defined as in the previous section, we now study the properties of real linear functionals defined on some topological linear space  $E$ .

**Theorem 2.14.** Let  $f$  be a linear functional defined on a topological linear space  $E$ . If  $f$  is continuous at any point  $x_0 \in E$ , then  $f$  is continuous at every point  $x \in E$ .

*Proof.* Suppose that  $f$  is continuous at some point  $x_0 \in E$ , and we wish to show this implies it's continuous at any point  $y \in E$ . For any  $\varepsilon > 0$ , let  $U$

be a neighbourhood of  $x_0$  such that  $|f(x) - f(x_0)| < \varepsilon$  for all  $x \in U$ . We then define the neighbourhood  $V$  of  $y$  as

$$V = \{x + y - x_0 : x \in U\}. \quad (2.14)$$

It follows that if  $x \in V$ , then  $x + x_0 - y \in U$ , and hence by the definition of  $U$  we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x + x_0 - y) - f(y + x_0 - y)| \\ &= |f(x + x_0 - y) - f(x_0)| \\ &\leq \varepsilon. \end{aligned}$$

Hence for any  $\varepsilon > 0$  we can find some neighbourhood  $V$  of  $y$  such that  $|f(x) - f(y)| < \varepsilon$  for all  $x \in V$ . That is,  $f$  is continuous at  $y$ .  $\square$

Hence to prove that a real linear functional is continuous on some topological linear space  $E$ , we need only verify it at a single point. An important consequence of Theorem 2.14 is that the continuity of a linear functional  $f$  is related to boundedness.

**Theorem 2.15.** Let  $f$  be a linear functional on a topological linear space  $E$ . Then  $f$  is continuous if and only if  $f$  is bounded on some neighbourhood of 0.

*Proof.* If  $f$  is continuous on  $E$ , then it is continuous at 0. Hence by definition, given any  $\varepsilon > 0$  we can find some neighbourhood  $U$  of 0 such that  $|f(x)| \leq \varepsilon$  for all  $x \in U$ . Hence  $f$  is clearly bounded on  $U$ .

Conversely, suppose  $f$  is bounded on some neighbourhood  $U$  of 0, so that for some finite constant  $c$ , we have  $|f(x)| < c$  for all  $x \in U$ . Then for any  $\varepsilon > 0$  we can construct a new neighbourhood around 0 defined as

$$V_\varepsilon = \left\{ \frac{\varepsilon}{c}x : x \in U \right\}.$$

By construction it follows that for all  $x \in V_\varepsilon$ , we have  $|f(x) - f(0)| < \varepsilon$  so that  $f$  is continuous at 0. By Theorem 2.14  $f$  is therefore continuous on  $E$ .  $\square$

### 2.3.2 Norms of Linear Functionals

As the set of all functionals on a linear space forms a space itself, it is natural to ask whether we should introduce some sort of norm. In the case of functionals, we define the norm as the supremum on the unit sphere.

**Definition 2.16.** Let  $f$  be a bounded linear functional  $f$  on a normed linear space  $E$ . Then the norm of  $f$  is defined as

$$\| f \| = \sup_{\|x\| \leq 1} |f(x)|. \quad (2.15)$$

Since by definition a linear functional  $f$  is linear, it follows that the norm of a functional  $f$  does not depend on the entire unit sphere  $\| x \| \leq 1$ , but only the outer shell  $\| x \| = 1$ . Hence,

$$\| f \| = \sup_{\|x\|=1} |f(x)|. \quad (2.16)$$

Additionally, we can also express the norm of a functional in terms of all  $x \in E$  with an appropriate rescaling by  $\| x \|^2$ ,

$$\| f \| = \sup_{x \neq 0} \frac{|f(x)|}{\| x \|^2}. \quad (2.17)$$

With an appropriate norm introduced, we next ask what can be said about the structure of this set of functionals, and whether or not it is also linear.

### 2.3.3 The Conjugate Space

Given any two functions  $f$  and  $g$ , along some scalar  $\alpha$ , we can define addition of functions and multiplication by scalars in the obvious way. So that

$$(f + g)(x) = f(x) + g(x) \quad (2.18)$$

and

$$(\alpha f)(x) = \alpha f(x). \quad (2.19)$$

The properties of linearity and boundedness clearly carry over to  $f + g$  and  $\alpha f$ . We can then study the space made up of these functionals as opposed to the original topological linear space they're defined on.

**Definition 2.17.** Let  $E$  be a topological linear space. Then the conjugate space  $E^*$  is the set of all continuous linear functionals  $f$  defined on  $E$ .

Clearly the conjugate space itself is a linear space when equipped with the operations defined above. A point of interest is therefore how to introduce a topology to the conjugate space. We first consider this question when the original space  $E$  is a normed linear space.

If  $E$  is a normed linear space, then we have already defined the norm of a functional  $f$ . Hence  $E^*$  can be made into a normed linear space by taking the norm of each functional  $f \in E^*$  as

$$\| f \| = \sup_{\|x\| \leq 1} |f(x)|. \quad (2.20)$$

We call the resulting topology the *strong topology* of the conjugate space.

**Theorem 2.16.** The conjugate space  $E^*$  equipped with the norm (2.20) is complete.

*Proof.* Let  $\{f_n\}$  be a Cauchy sequence of functionals in  $E^*$ . Then for any  $\varepsilon > 0$ , there exists some integer  $N$  such that for all  $n, m > N$  we have

$$\| f_n - f_m \| < \varepsilon.$$

Hence by (2.17) it follows that for all  $x \in E$ ,

$$|f_n(x) - f_m(x)| \leq \| f_n - f_m \| \| x \| < \varepsilon \| x \|.$$

Hence for any  $x \in E$  the sequence  $\{f_n(x)\}$  is Cauchy and hence convergent. We let the functional  $f$  be defined as

$$f(x) = \lim_{n \rightarrow \infty} f_n(x).$$

Properties of limits ensure that  $f(x)$  is a linear functional, so that we need only show  $f$  is continuous. Taking  $n$  large enough such that  $\|f_n - f_{n+p}\| < 1$  for all  $p \geq 0$ , then it follows that

$$|f_{n+p}(x)| \leq (\|f_n\| + 1) \|x\|,$$

and hence

$$\lim_{p \rightarrow \infty} |f_{n+p}(x)| = |f(x)| \leq (\|f_n\| + 1) \|x\|.$$

So that  $f$  is bounded and hence continuous, and thus  $f \in E^*$ . We omit the proof that  $f_n \rightarrow f$ .  $\square$

### 2.3.4 The Weak Topology

If  $E$  is a topological linear space with conjugate space  $E^*$ , then we define the weak topology as the topology generated by all open sets of the form

$$U = \{x : |f_1(x)| < \varepsilon, \dots, |f_r(x)| < \varepsilon\} \quad (2.21)$$

for any  $\varepsilon > 0$  and any finite collection of functionals  $f_1, f_2, \dots, f_r \in E^*$ . If a subset is open in the weak topology, then it is also open in the original topology of  $E$ . The weak topology defines what is called *weak convergence*, in contrast to strong convergence defined by the original topology of  $E$ .

**Theorem 2.17.** Let  $E$  be a topological linear space, and  $\{x_n\}$  a sequence of elements belonging to  $E$ . Then the sequence  $\{x_n\}$  is weakly convergent to  $x_0 \in E$  if and only if the sequence  $\{f(x_n)\}$  converges to  $f(x_0)$  for every continuous linear functional  $f$  defined on  $E$ .

*Proof.* By the linearity of the functionals  $f$  we can assume that  $x_0 = 0$  without loss of generality. Suppose first that  $\{x_n\}$  is weakly convergent to  $x_0$ . Then for every open neighbourhood  $U$  of  $x_0$  of the form (2.21), we can find some integer  $N$  such that  $x_n \in U$  for all  $n > N$ . This itself implies  $f(x_n) \rightarrow 0$  for all  $f \in E^*$  as we can take any  $f$  in the definition of an open set (2.21).



Conversely, suppose  $f(x_n) \rightarrow 0$  for all  $f \in E^*$ . Then for any open neighbourhood of the form (2.21), let  $N_i$  be the smallest integer such that  $|f_i(x_n)| < \varepsilon$  for all  $n > N_i$ . We then take  $N = \max(N_1, N_2, \dots, N_r)$ . Then  $x_n \in U$  for all  $n > N$  by construction, so that  $\{x_n\}$  is weakly convergent to  $x_0$ .  $\square$

The conjugate space  $E^*$ , being a topological linear space itself, also has a weak topology defined in a similar way. The weak topology of the conjugate space is the topology generated by all open sets of the form

$$U_{A,\varepsilon} = \{f : |f(x)| < \varepsilon \text{ for all } x \in A\} \quad (2.22)$$

for some  $\varepsilon > 0$  and  $A \subset E$ . As with the weak topology of  $E$ , this topology determines a kind of weak convergence for functionals  $f \in E^*$ . We can view the conjugate space not just as a space conjugate to  $E$ , but also as a linear space in its own right. In this way, one can also study the second conjugate space  $E^{**}$ .

## 2.4 Linear Operators

### 2.4.1 Introduction

An *operator*  $A$  between two topological linear spaces  $E_1$  and  $E_2$  is a mapping of the form

$$y = Ax$$

for  $x \in E_1$  and  $y \in E_2$ . As with functionals, we say that an operator  $A$  is linear if

$$A(\alpha x_1 + \beta x_2) = \alpha Ax_1 + \beta Ax_2.$$

The *domain* of  $A$ , denoted  $D_A$ , is the set of elements  $x \in E_1$  for which  $A$  is defined. We generally assume that the domain of  $A$  is  $E_1$  itself, however this need not be the case. The *range* of  $A$ , denoted  $R_A$ , is the set of all elements that are the image under  $A$  of at least one element of  $D_A$ .

We say that the operator  $A$  is continuous at the point  $x_0$  if, given any neighbourhood  $V$  of the point  $y_0 = Ax_0$ , we can find a neighbourhood  $U$  of the point  $x_0$  such that  $Ax \in V$  for all  $x \in U \cap D_A$ . Similarly, we say the operator  $A$  is continuous if it is continuous at all points  $x_0 \in D_A$ .

In the previous section we showed that the continuity of linear functionals is related to whether they are bounded on certain sets. A similar result holds for linear operators. We say a linear operator is bounded if every bounded subset of  $E_1$  is mapped into a bounded subset of  $E_2$ . This definition then gives us a useful necessary condition for establishing continuity.

**Theorem 2.18.** Let  $E_1$  be a topological linear space and  $A : E_1 \rightarrow E_2$  an operator. Then a necessary condition for  $A$  to be continuous on  $E_1$  is that  $A$  is bounded.

*Proof.* Suppose  $A$  is continuous, and let  $M$  be a bounded subset of  $E_1$ . If the image of  $M$  under  $A$  is unbounded, then we can find a neighbourhood  $V$  of 0 such that, for  $n = 1, 2, \dots$ , none of the sets

$$\frac{1}{n}AM$$

is a subset of  $V$ . Hence there exists a sequence of elements  $\{x_n\}$  belonging to  $M$  such that

$$\frac{1}{n}Ax_n \notin V$$

for all  $n$ . But then this contradicts the continuity of  $A$  since  $\{\frac{1}{n}x_n\}$  clearly converges to 0 while  $A\frac{1}{n}x_n$  does not.  $\square$

As with linear functionals, it is often useful to associate a norm with the set of linear mappings. We therefore define the norm of an operator similar to that of linear functionals.

**Definition 2.18.** Let  $A$  be a bounded linear operator mapping a normed linear space  $E_1$  into another normed linear space  $E_2$ . Then the norm of the operator  $A$  is defined as

$$\|A\| = \sup_{\|x\| \leq 1} \|Ax\|. \quad (2.23)$$

As with the norm of a linear functional, the linearity of  $A$  implies that

$$\| A \| = \sup_{x \neq 0} \frac{\| Ax \|}{\| x \|}.$$

The set of linear operators is also a linear space when we define addition and scalar multiplication in the obvious way, as we did with linear functionals. Some linear operators are of greater interest than others, particularly those that are *invertible*.

**Definition 2.19.** Let  $A$  be a linear mapping. Then the inverse of  $A$ , if it exists, is the linear mapping  $A^{-1}$  such that

$$A^{-1}Ax = x \text{ for all } x \in E_1. \quad (2.24)$$

A linear mapping  $A$  only has an inverse if there is a unique solution to the equation

$$Ax = y \text{ for all } y \in R_A.$$

We now prove that the inverse of any linear operator is itself a linear operator. This is particularly useful as it means any result proven for a linear operator  $A$  must also hold for its inverse.

**Theorem 2.19.** The inverse  $A^{-1}$  of a linear operator  $A$  is linear.

*Proof.* Suppose that for  $x_1, x_2 \in D_A$  and  $y_1, y_2 \in R_A$  we have

$$Ax_1 = y_1,$$

$$Ax_2 = y_2.$$

Then an immediate consequence is that

$$\alpha_1 A^{-1}y_1 + \alpha_2 A^{-1}y_2 = \alpha_1 x_1 + \alpha_2 x_2. \quad (2.25)$$

For the other direction, we see that since

$$A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 y_1 + \alpha_2 y_2,$$

it follows that

$$A^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 x_1 + \alpha_2 x_2. \quad (2.26)$$

Combining (2.25) and (2.26) proves the linearity of  $A^{-1}$ .  $\square$

## 2.4.2 The Adjoint Operator

Given a linear operator  $A : E_1 \rightarrow E_2$ , we now consider how  $A$  interacts with the conjugate spaces  $E_1^*$  and  $E_2^*$ . While  $A$  is defined on the original topological space  $E_1$ , it can also be used to establish a mapping  $A^* : E_2^* \rightarrow E_1^*$ . This mapping is called the *adjoint operator*. Given any  $g \in E_2^*$ , we define an  $f \in E_1^*$  as

$$f(x) = g(Ax) \text{ for all } x \in E_1, \quad (2.27)$$

so that  $f$  is the functional resulting by first mapping  $x \rightarrow Ax$  and then applying the functional  $g$ . So, using the example above, the adjoint operator  $A^* : E_2^* \rightarrow E_1^*$  is defined by

$$A^*g = f.$$

We illustrate this with an example. Suppose  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear mapping with matrix  $\| a_{ij} \|$ , then we ask what can be said about the matrix of the adjoint operator. By definition of matrix multiplication, the linear mapping is defined by a system of linear equations

$$y_i = \sum_{j=1}^n a_{ij}x_j \text{ for } i = 1, 2, \dots, m. \quad (2.28)$$

We know that given a basis  $e_1, e_2, \dots, e_n$ , any vector  $x \in \mathbb{R}^n$  has a unique representation of the form

$$x = \sum_{j=1}^n x_j e_j,$$

so that any functional  $f$  defined on  $\mathbb{R}^n$  can be expressed as

$$f(x) = \sum_{j=1}^n f(e_j)x_j. \quad (2.29)$$

Similarly, if  $f = A^*g$  and  $u_1, u_2, \dots, u_m$  is a basis in  $\mathbb{R}^m$ , we have

$$f(x) = g(Ax) = \sum_{i=1}^m g(u_i)y_i = \sum_{j=1}^n x_j \sum_{i=1}^m g_i(u_i)a_{ij}. \quad (2.30)$$

By combining (2.29) and (2.30) we then see that

$$f(e_i) = \sum_{j=1}^m a_{ji} g(u_j)$$

which clearly takes the form of (2.28). Hence the matrix of the adjoint operator  $A^*$  is  $\| a_{ji} \|$ , the transpose of the matrix of the operator  $A$ .

It is easy to see that, like  $A$ , the adjoint operator is linear. We conclude this chapter with a more subtle property concerning bounded operators.

**Theorem 2.20.** Let  $A : E_1 \rightarrow E_2$  be a bounded linear operator from one Banach space to another. Then  $A^*$  is also bounded, with  $\| A^* \| = \| A \|$ .

*Proof.* By the definition of the norm of a functional  $f$ , it follows that

$$\| f \| \geq \frac{|f(x)|}{\| x \|},$$

with a similar result holding for linear operators. It therefore follows that for all  $x \in E_1$ , we have

$$\begin{aligned} |A^*g(x)| &= |g(Ax)| \\ &\leq \| Ax \| \| g \| \\ &\leq \| g \| \| A \| \| x \| . \end{aligned}$$

This implies that

$$\| A^*g \| \leq \| A \| \| g \|$$

and hence

$$\| A^* \| \leq \| A \| .$$

To complete the proof we need only show the reverse inequality also holds. To this end, take any  $x \in E_1$  such that  $Ax \neq 0$ , and set

$$y_0 = \frac{Ax}{\| Ax \|} .$$

Let  $g$  be the functional such that

$$g(\lambda y_0) = \lambda$$

on the set

$$L = \{\lambda y_0 : \lambda \in \mathbb{R}\}.$$

It immediately follows that

$$\|g\|_{\text{on } L} = \sup_{\lambda \neq 0} \frac{|g(\lambda y_0)|}{\|\lambda y_0\|} = 1.$$

The Hahn-Banach theorem for a normed linear space then says we can extend the functional  $g$  onto the entire space  $E_2$  without increasing its norm. Furthermore, since  $g(y_0) = 1$ , it follows that  $\|Ax\| = g(Ax)$ . Consequently, we see that

$$\begin{aligned} \|Ax\| &= g(Ax) \\ &= |A^*g(x)| \\ &\leq \|A^*g\| \|x\| \\ &\leq \|A^*\| \|g\| \|x\| \\ &= \|A^*\| \|x\|. \end{aligned}$$

Which, by the definition of the norm of an operator, proves the reverse inequality.  $\square$

# Chapter 3

## The Exchange Rate Data

In this section, we introduce the dataset that is studied in this thesis. We give a brief overview of the dataset and the foreign exchange market in general. We then discretise the data in order to accommodate statistical analysis of the exchange rate as a stochastic process.

### 3.1 Introduction

In this thesis, historical EUR/USD exchange rate data is used to investigate the behaviour of the exchange rate as a stochastic process. The dataset was provided by Olsen Data (formally Olsen & Associates), an organisation based in Switzerland that maintains high quality tick data for a wide range of financial assets. Our dataset starts at the beginning of 2004 and ends mid 2011.

The decentralised nature of the foreign exchange market makes it difficult to obtain comprehensive data on transactions, which presents great difficulties to researchers. Thus, exchange rate quotes published by market makers are generally used instead. Market makers are financial institutions that take on the risk of holding foreign currencies in order to profit. They buy and sell foreign currency from other market participants, profiting from the difference in price they buy and sell at. In order to ensure

a fair price is given, an individual will not tell the market maker whether he intends to buy or sell. Thus, the market maker provides the individual with simultaneous buy and sell prices for the currency involved. The price a market maker is willing to sell a currency at is referred to as the *ask quote*, while the price they are willing to buy currency at is referred to as the *bid quote*. The *spread* is the price difference between the bid and ask quotes, and is the primary fee associated with trading currencies.

The dataset used in this thesis consists of EUR/USD synchronous bid and ask quotes, provided alongside time stamps of one second resolution. The EUR/USD exchange rate is the price of one euro in terms of US dollars. Since Olsen Data compiles quotes from various different sources, and published quotes are non-binding, some may be of poor quality and therefore do not accurately represent the market's true exchange rate. Fortunately, data filtering was used prior to delivery so that low quality or abnormal quotes have been removed. Although regular spikes in price are still present, this is a common characteristic of financial data, and can be interpreted as the market rapidly processing new information. We therefore assume that the quotes provided are accurate and further filtering is unnecessary.

Since trading is closed during weekends, no quotes are published and recorded during these times. This is illustrated in Figure 3.1, which shows the bid quote price during early 2009, where weekends are clearly visible as prolonged periods of no change. Leaving in these periods would lead to underestimating the variance of the process, and so we remove them. As a result, we proceed as if trading resumes immediately after market closure.

## 3.2 Quote Volume by Year

The dataset used in this thesis consists of bid and ask quotes that are irregularly spaced in time. The number of recorded quotes varies considerably



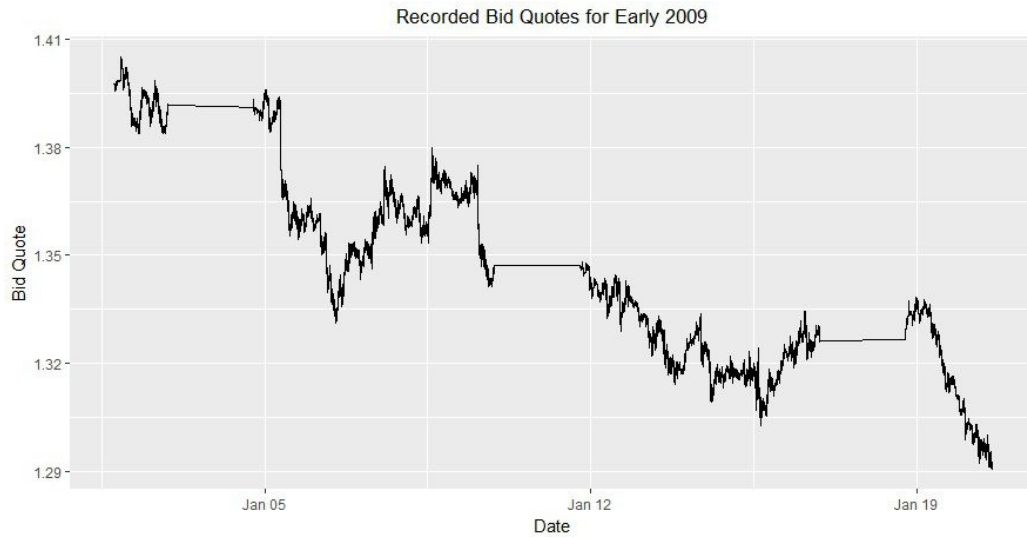


Figure 3.1: Bid quotes recorded early 2009.

by year. Figure 3.2 shows the number of quotes recorded in each year, with under 5 million during 2006, compared to over 17 million during 2009. In order to examine the exchange rate at very small intervals of time, we will generally restrict ourselves to later years, as the average time between quotes is much smaller. For this reason, unless otherwise stated, for the remainder of this thesis it is implied that we use data from 2010 alone. An exception to this is Chapter 7, when we use data from 2009 up to and including 2011.

### 3.3 Minimum Time Discretisation

Let  $X(t)$  and  $Y(t)$  denote the *bid* and *ask* prices at time  $t$  respectively. As the two prices behave in an almost identical manner, we will generally restrict ourselves to the bid price  $X(t)$ . An exception to this is Chapter 7 when we also make use of the ask quote process  $Y(t)$ . Let us denote the time of the  $i$ th quote as  $T_i$ , so that our dataset consists of the collection of quotes

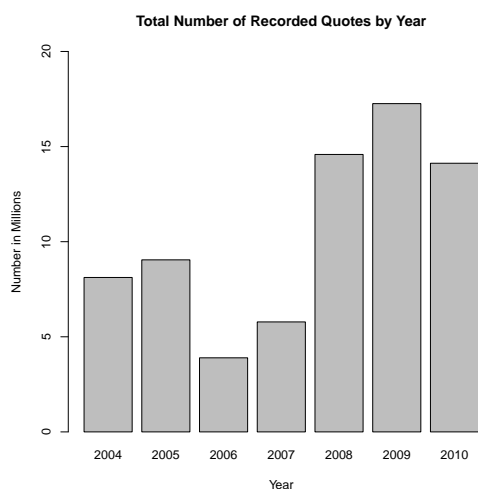


Figure 3.2: Total recorded quotes by year.

$$X(T_0), X(T_1), X(T_2), \dots, X(T_m)$$

where  $T_0 = 0$ . The fact that this process is irregularly spaced in time presents difficulties. For example, when studying the distribution of the increments  $X(T_{i+1}) - X(T_i)$ , it may be that the mean or variance depends on the time increment  $T_{i+1} - T_i$ . Hence it would be preferable to transform our data to the form

$$X(0), X(\Delta), X(2\Delta), \dots, X(n\Delta)$$

for various  $\Delta$ , where  $n = \lfloor T_m/\Delta \rfloor$ . Unfortunately we are unlikely to know the exchange rate at each of these exact times. Instead, we must use a discretisation method to approximate them.

In order to approximate the bid price at time  $i\Delta$ , for  $i = 0, 1, 2, \dots, n$ , we use the approximation

$$X(i\Delta) \approx X\left(\min_{0 \leq i \leq m} \{T_i : T_i \geq i\Delta\}\right). \quad (3.1)$$

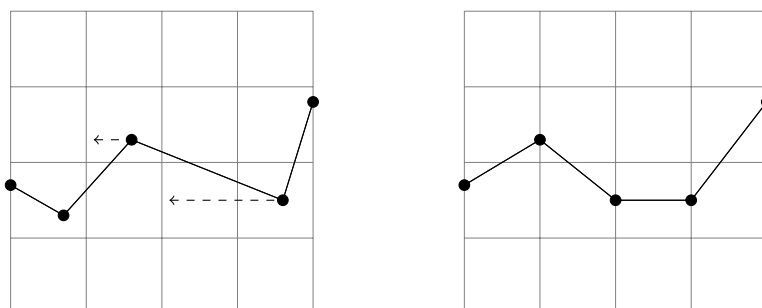


Figure 3.3: Minimum time discretisation of the process  $X(t)$ .

So that the price at time  $i\Delta$  is approximated by the next known price. For  $\Delta$  not smaller than 1 minute, there are sufficiently many quotes in later years so that this method is reasonably accurate. However, when we examine smaller  $\Delta$ , this approximation becomes increasingly inaccurate for two reasons. Firstly, there may be no known quotes on consecutive intervals, resulting in the same approximation being used several times. This is demonstrated in Figure 3.3. Secondly, as  $\Delta$  decreases in size, differences between  $i\Delta$  and our approximated times become larger, relatively speaking, and hence less accurate.

Following the discretisation method outlined above, we obtain a series of regularly spaced discrete time processes for varying  $\Delta$ . To simplify notation, we denote these discretised processes by

$$X_0^\Delta, X_1^\Delta, X_2^\Delta, \dots, X_n^\Delta$$

so that

$$X_i^\Delta = X(i\Delta). \quad (3.2)$$

We denote the increments of  $X_i^\Delta$  as

$$\Delta X_i = X_i^\Delta - X_{i-1}^\Delta. \quad (3.3)$$

For the remainder of this thesis, unless otherwise stated, we take  $X_i^\Delta$  as denoting the exchange rate process sampled at times  $i\Delta$  for the year 2010. We repeat this discretisation method for the ask quote process  $Y(t)$ , which we will make use of in Chapter 7.



# Chapter 4

## Brownian Motion

In this chapter, we introduce Brownian motion and the total variation of a stochastic process. We then compare the properties of Brownian motion to the behaviour exhibited by the exchange rate process. Additionally, we use the Khmaladze transformation-2 to test whether the increments  $\Delta X_i$  come from centred normal distributions.

### 4.1 Introduction

Brownian motion, also called the Wiener process in honour of Norbert Wiener, is a fundamental stochastic process and occurs frequently in many areas of applied mathematics.

**Definition 4.1.** Suppose the stochastic process  $B(t)$ , for  $0 \leq t \leq T$ , is such that

1.  $B(0) = 0$ ,
2. for any collection of points  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$  with finite  $k$ , the random variables  $B(t_1), B(t_2) - B(t_1), \dots, B(t_k) - B(t_{k-1})$  are independent Gaussian random variables with expected values 0 and variances

$$\text{Var}(B(t_j) - B(t_{j-1})) = t_j - t_{j-1} \quad (4.1)$$

for  $j = 1, 2, \dots, k$ .

Then the process  $B(t)$  is called *Standard Brownian Motion*.

The *Brownian bridge* process is then defined as a Brownian motion  $B(t)$  under the condition that  $B(T) = 0$ .

**Definition 4.2.** Let  $B(t)$  be standard Brownian motion on the interval  $[0, T]$ . Then the process

$$Z(t) = B(t) - \frac{t}{T}B(T), \quad (4.2)$$

for  $0 \leq t \leq T$ , is called *standard Brownian bridge*.

Thus Brownian motion is a process where future increments are unpredictable Gaussian random variables, with variance that is linear in time. The importance of Brownian motion in financial mathematics is due in part to the observation that financial prices behave as random walks. There is a clear connection between Brownian motion and random walks that can be seen through the following construction (Moral & Penev, 2016, pp.393-395).

Suppose we start by placing a point at 0 on the real line, and let  $B(t)$  denote the location of this point at time  $t$ . Then, after each interval of time  $\Delta$ , suppose this point moves in either direction by a step of size  $\sqrt{\Delta}$  with equal probability. We assume the steps taken are all independent. Under this construction, we can express  $B(t)$  as

$$B(t) = \sqrt{\Delta}(Y_1 + Y_2 + \dots + Y_k) \quad (4.3)$$

where  $k = \lfloor t/\Delta \rfloor$  and  $Y_i$  are random variables that take values 1 and  $-1$  with equal probability. It then follows from linearity of expectation and properties of variance that

$$\mathbf{E}[B(t)] = 0$$

and

$$\text{Var}(B(t)) = \Delta \lfloor t/\Delta \rfloor.$$

The central limit theorem then says that if  $t \gg \Delta$ , then  $B(t)$  is almost Gaussian with expected value 0 and variance  $t$ . One can then see that Brownian motion is the limit of this random walk as  $\Delta \rightarrow 0$ . This very general construction as the limit of a random walk is why Brownian motion arises often in the physical sciences. Indeed, seminal work on Brownian motion was motivated by a desire to explain the irregular paths that tiny particles make when suspended on liquid, these paths being the result of tiny "steps" on the molecular level.

## 4.2 Total Variation of Brownian Motion

When modelling a stochastic process  $X(t)$ , there are several quantities that can be used to discern whether the process is driven by a Brownian motion. Here, we say a process  $X(t)$  is driven by a Brownian motion if it obeys a stochastic differential equation of the form

$$dX(t) = a(X(t), t)dt + b(t)dB(t), \quad (4.4)$$

where  $B(t)$  is a standard Brownian motion. The form (4.4) is a specific type of Brownian motion-driven process, called a *Markov diffusion process*. We start with the *total variation* of a stochastic process. In what follows, we limit ourselves to equidistant partitions of the interval  $[0, T]$ , though these ideas also hold for more general partitions of the form  $0 = t_0 < t_1 < \dots < t_n = T$ .

**Definition 4.3.** Let  $X(t)$  be a stochastic process with  $0 \leq t \leq T$ . Then the total variation of  $X(t)$  is defined by the limit

$$V(X(t)) = \lim_{\Delta \rightarrow 0} \sum_{i=1}^n |X(i\Delta) - X((i-1)\Delta)|, \quad (4.5)$$

where  $n = T/\Delta$ .

If a process is driven by a Brownian motion, then we would expect its total variation to behave similar to that of a standard Brownian motion.

This is because the non-Brownian term in (4.4) is of order  $dt$ , while the Brownian term is of order  $\sqrt{dt}$ , being a Gaussian random variable with variance  $dt$ .

The total variation of a Brownian motion path is particularly interesting since it seems to contradict our natural instincts. It can be shown that Brownian motion paths are almost surely continuous (Klebaner, 2005, p.57; Moral & Penev, 2016, p.395), and so it would be natural to expect they almost surely have finite total variation. Instead, the total variation of a Brownian motion trajectory is almost surely infinite.

**Theorem 4.1.** Let  $B(t)$  be a Brownian motion on the interval  $[0, T]$ . Then the total variation of  $B(t)$  is almost surely infinite.

*Proof.* Let

$$V_n(B(t)) = \sum_{i=1}^n |B(i\Delta) - B((i-1)\Delta)| \quad (4.6)$$

where  $\Delta = T/n$ . Since each  $\Delta$ -increment of  $B(t)$  is Gaussian with mean 0 and variance  $\Delta$ , it follows that the absolute values of these increments have a half-normal distribution, and thus

$$\mathbf{E}[V_n(B(t))] = c_1\sqrt{n} \quad (4.7)$$

and

$$\mathbf{Var}(V_n(B(t))) = c_2 \quad (4.8)$$

for some constants  $c_1, c_2 > 0$ . Hence,

$$\begin{aligned} \mathbb{P}\left(V_n(B(t)) \geq \frac{1}{2}c_1\sqrt{n}\right) &= \mathbb{P}\left(V_n(B(t)) - c_1\sqrt{n} \geq -\frac{1}{2}c_1\sqrt{n}\right) \\ &= \mathbb{P}\left(-V_n(B(t)) + c_1\sqrt{n} \leq \frac{1}{2}c_1\sqrt{n}\right) \\ &\geq \mathbb{P}\left(|-V_n(B(t)) + c_1\sqrt{n}| \leq \frac{1}{2}c_1\sqrt{n}\right) \\ &\geq 1 - \frac{c_2}{\left(\frac{1}{2}c_1\sqrt{n}\right)^2} \end{aligned}$$



where the last inequality follows by applying Chebyshev's inequality. Hence,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( V_n(B) \geq \frac{1}{2} c_1 \sqrt{\Delta} \right) = 1, \quad (4.9)$$

which proves convergence in probability. Now, let us choose a subsequence  $\{n_k\}$  of  $n = 1, 2, 3, \dots$  such that

$$\sum_{k=1}^{\infty} \mathbb{P} \left( V_{n_k}(B(t)) \geq \frac{1}{2} c_1 \sqrt{n_k} \right) < \infty. \quad (4.10)$$

The result then follows by applying the Borel-Cantelli lemma.  $\square$

It is then worth considering the rate at which the total variation goes to infinity, as this gives us a way of comparing the irregularity of an observed stochastic process to that of Brownian motion.

**Proposition 4.1.** Let  $B(t)$  be a standard Brownian motion on  $[0, T]$ , with  $V_n(B(t))$  defined as in (4.6). Then  $\Delta^{\frac{1}{2}} V_n(B(t))$  converges in probability to some constant  $c > 0$ ,

$$\Delta^{\frac{1}{2}} V_n(B(t)) \xrightarrow{P} c. \quad (4.11)$$

*Proof.* Using the same notation as before, we have

$$\mathbf{E}[\Delta^{\frac{1}{2}} V_n(B(t))] = \sqrt{T} c_1$$

and

$$\text{Var}(\Delta^{\frac{1}{2}} V_n(B(t))) = \Delta c_2.$$

Hence, by applying Chebyshev's inequality we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \Delta^{\frac{1}{2}} V_n(B(t)) - \sqrt{T} c_1 \right| > \varepsilon \right) &\leq \lim_{n \rightarrow \infty} \frac{\Delta c_2}{\varepsilon^2} \\ &= 0 \end{aligned}$$

so that  $\Delta^{\frac{1}{2}} V_n(B(t)) \xrightarrow{P} \sqrt{T} c_1$ .  $\square$

A consequence of this is that  $V_n(B(t))$  should obey, at least approximately, a power law of the form

$$V_n(B(t)) \approx c\Delta^{-\frac{1}{2}}, \quad (4.12)$$

with the approximation becoming increasingly accurate as  $n \rightarrow \infty$ . This provides us with a means of checking whether a process, say some  $Z(t)$ , behaves as a Brownian motion. If  $V_n(Z(t))$  does not obey the power law (4.12), then this suggests that  $Z(t)$  is not a Brownian motion-driven process.

### 4.3 The Exchange Rate Process

In this section, we investigate whether the exchange rate  $X(t)$  behaves as a Brownian motion-driven process. We start by examining the sampled total variation of our discretised exchange rate process  $X_i^\Delta$ , and then investigate the distribution and stationarity of the increments  $\Delta X_i$ .

#### 4.3.1 Total Variation

We have seen that the total variation of a Brownian motion, sampled at equidistant times, obeys a power law of the form

$$\sum_{i=1}^n |B(i\Delta) - B((i-1)\Delta)| \approx c\Delta^{-\frac{1}{2}}. \quad (4.13)$$

So if the exchange rate process  $X(t)$  is driven by a Brownian motion, we would expect it to also. Let us denote the total variation of  $X(t)$  sampled at intervals of length  $\Delta$  as

$$V_\Delta(X(t)) = \sum_{i=1}^n |X_i^\Delta - X_{i-1}^\Delta|, \quad (4.14)$$

where we again take  $n = T/\Delta$ . We wish to estimate the parameter  $\alpha$  such that

$$V_\Delta(X(t)) \approx c\Delta^{-\alpha}, \quad (4.15)$$

Estimates of $\alpha$ for varying $\Delta_1, \Delta_2$				
$\Delta_1$	$V_{n_1}(X(t))$	$\Delta_2$	$V_{n_2}(X(t))$	$\hat{\alpha}$
1 Day	2.17	12 Hours	2.84	0.388
12 Hours	2.84	1 Hour	8.78	0.450
1 Hour	8.78	10 minutes	21.09	0.489
10 minutes	21.09	1 minute	64.03	0.482
1 minutes	64.03	10 seconds	130.17	0.396

Table 4.1: The total variation of  $X(t)$  sampled at intervals of length  $\Delta$  along with estimates of  $\alpha$ .

hoping to verify that indeed  $\alpha = 1/2$ . Clearly, given any two  $V_{\Delta_1}(X(t))$  and  $V_{\Delta_2}(X(t))$ , we can simultaneously solve (4.15) to obtain the estimator

$$\hat{\alpha} = \frac{\log(V_{\Delta_1}(X(t))) - \log(V_{\Delta_2}(X(t)))}{\log(\Delta_2) - \log(\Delta_1)}. \quad (4.16)$$

We hope that since  $\Delta^{\frac{1}{2}}V_n(B(t)) \xrightarrow{P} c$ , this estimate becomes increasingly accurate as  $\Delta_1, \Delta_2 \rightarrow 0$ . In Table 4.1, we give estimates  $\hat{\alpha}$  using a wide range of values for  $\Delta_1$  and  $\Delta_2$ . However, instead of seeing our estimates  $\hat{\alpha}$  converge to  $1/2$  as  $\Delta_1, \Delta_2 \rightarrow 0$ , we instead see that they begin to approach  $1/2$ , before diverging downwards. While this seems to contradict  $X(t)$  being driven by a Brownian motion, this behaviour can also be explained by the lack of exchange rate quotes for very small  $\Delta$ . Referring back to the minimum time discretisation used in Chapter 3, for  $\Delta$  not smaller than 1 minute, we have a price on at least 99% of intervals to take as  $X_i^\Delta$ . Once we go below 1 minute, the number of intervals of time without a known price increases substantially. As a result, we end up underestimating  $V_\Delta(X(t))$ , as the change in price on these unknown intervals cannot be measured.

With this in mind, suppose we restrict ourselves to  $\Delta$  no less than 1 minute, so that the number of such intervals is insignificant. The estimates given in Table 4.1 are few, and so we wish to present a more complete picture of whether or not  $\hat{\alpha} \rightarrow 1/2$ . To this end, let us measure  $\Delta$  in minutes,

and suppose we estimate  $\alpha$  using  $\Delta$  and  $\Delta + 1$ , so that we take

$$\hat{\alpha} = \frac{\log(V_{\Delta}(X(t))) - \log(V_{\Delta+1}(X(t)))}{\log(\Delta) - \log(\Delta + 1)}.$$

Our estimator then depends on a single  $\Delta$ , and we can examine the behaviour of  $\hat{\alpha}$  as  $\Delta$  decreases. In Figure 4.1 we give the complete plot of all our estimates, along with a zoomed in plot showing how  $\Delta$  indeed appears to approach  $1/2$ .

Since these estimates appear to converge to  $1/2$  so long as  $\Delta$  is kept large enough that we have sufficient data, we conclude that the total variation of the exchange rate process reflects that of a stochastic process driven by Brownian motion.

### 4.3.2 Are Increments Gaussian?

The widespread use of Brownian motion in financial mathematics relies on assumptions that returns, or at least some function of returns, are normally distributed. We now test this assumption in order to better understand the distribution of price changes for the exchange rate process. More specifically, we wish to test whether  $\Delta$ -increments have distributions of the form

$$\Delta X_i \sim N(0, \sigma_{\Delta}^2). \quad (4.17)$$

We give the empirical distribution functions of  $\Delta$ -increments in Figure 4.2, for various  $\Delta$ . These clearly resemble normal distributions of the form (4.17), however we wish to test this formally, as many distributions appear normal at face value.

When testing whether a random sample comes from some distribution  $F_{\theta}(x)$ , one usually considers the empirical process

$$v_n(x; \theta) = \sqrt{n} [F_n(x) - F_{\theta}(x)] \quad (4.18)$$

where  $F_n(x)$  is the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq x\}. \quad (4.19)$$

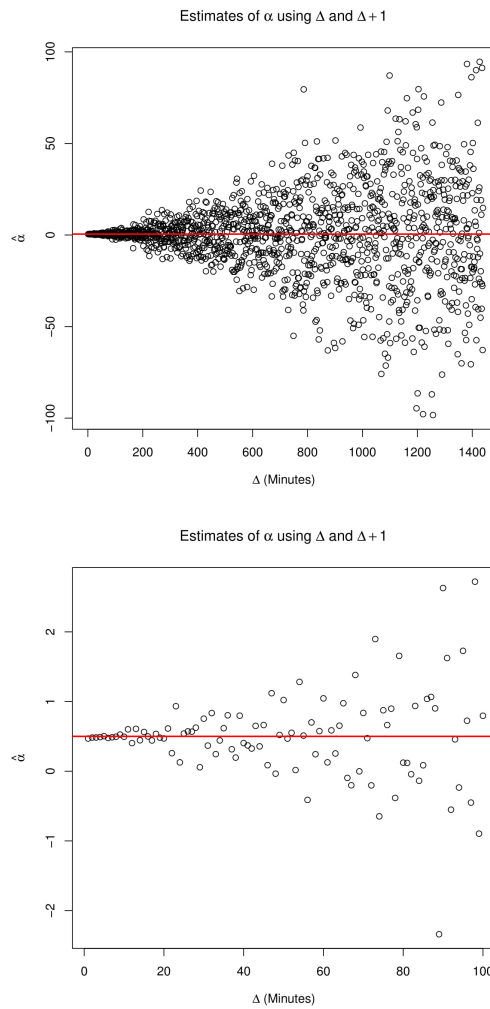


Figure 4.1: Estimates  $\hat{\alpha}$  plotted against  $\Delta$ , with  $f(\Delta) = \frac{1}{2}$  shown in red.

The empirical process has been extensively studied, and its asymptotic behaviour is used as the basis of many statistical tests, most notably the Kolmogorov-Smirnov test. However, in practice one does not know the parameter  $\theta$ , and thus an estimate  $\hat{\theta}$  must be used in its place. Hence, we instead need a statistical test based on the parametric empirical process

$$v_n(x; \hat{\theta}) = \sqrt{n} [F_n(x) - F_{\hat{\theta}}(x)]. \quad (4.20)$$

This presents a problem in that the asymptotic behaviour of the parametric empirical process differs from that of the empirical process. Namely, its asymptotic behaviour depends on the score function of  $F_{\theta}(x)$ , which differs between families of distributions and, in some cases, within families as well. This problem motivates the development of tests which are asymptotically distribution-free. To this end, Khmaladze (1981, 1988, and 1993) proposed a transformation of the parametric empirical process, called the Khmaladze transformation, to make testing asymptotically distribution-free. In this thesis, we apply the more recent Khmaladze transformation-2, introduced by Khmaladze in [11].

### The Khmaladze Transformation-2

Under the hypothesis that increments come from a centred normal distribution, there exists some  $\sigma \in \mathbb{R}_{>0}$  such that

$$\Delta X_i \sim N(0, \sigma^2)$$

for all  $i = 1, 2, \dots, n$ . The exact value of  $\sigma$  is unknown, and so we instead use the maximum likelihood estimate  $\hat{\sigma}$  to construct the parametric empirical process

$$v_n(x; \hat{\theta}) = \sqrt{n} [F_n(x) - F_{\hat{\theta}}(x)].$$

The idea behind the Khmaladze transformation-2 is to apply a unitary transformation to the parametric empirical process such that its asymptotic behaviour depends on neither the distribution  $F_{\theta}(x)$ , nor on the estimated distribution  $F_{\hat{\theta}}(x)$ . Before introducing this transformation, we first

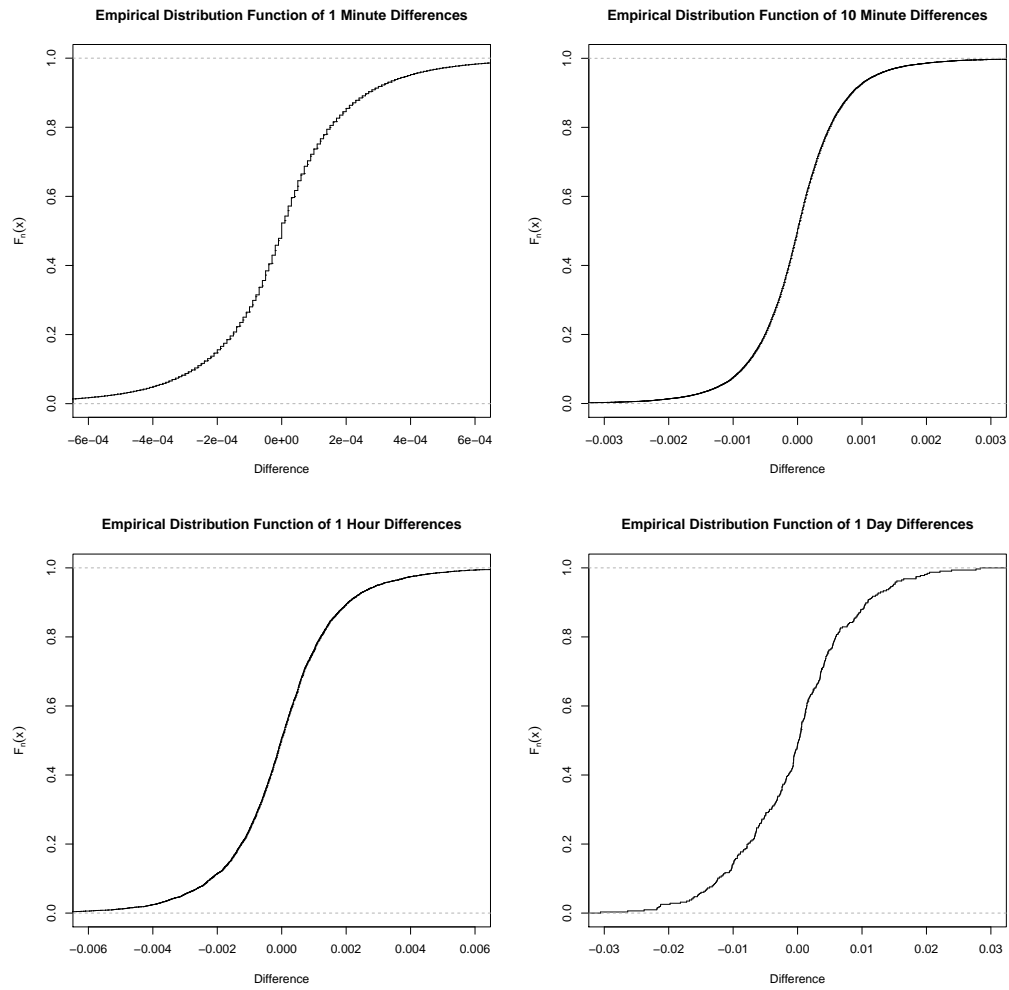


Figure 4.2: Empirical distribution functions of  $\Delta X_i$  for different  $\Delta$ .

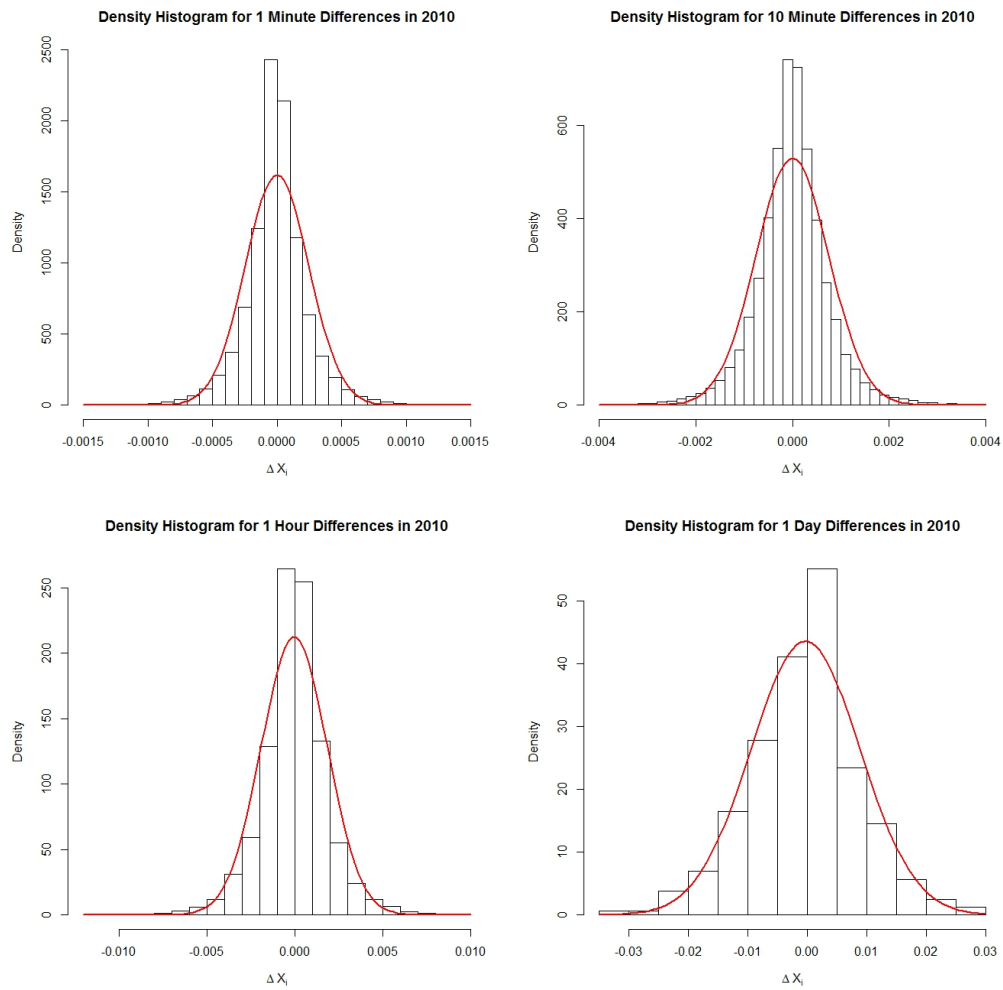


Figure 4.3: Density histograms of  $\Delta X_i$  compared to the maximum likelihood normal density functions.



need to introduce some important concepts concerning Brownian motions and Brownian bridges on the space  $L_2(F)$ , the set of functions that are square-integrable with respect to  $F$ .

**Definition 4.4.** Let  $w_F(\phi)$  be a linear functional in  $\phi$ , where  $\phi \in L_2(F)$ . Then  $w_F(\phi)$  is called *F-Brownian motion* if  $w_F(\phi)$  is a Gaussian random variable with mean 0 and variance

$$\int_{y \in \mathbb{R}} \phi^2(y) dF(y) \quad (4.21)$$

for every  $\phi \in L_2(F)$ .

This then leads to a natural generalisation of Brownian bridge.

**Definition 4.5.** Let  $w_F(\phi)$  be *F-Brownian motion*, and define  $v_F(\phi)$  as

$$v_F(\phi) = w_F(\phi) - \langle \phi, q_0 \rangle_F w_F(q_0), \quad (4.22)$$

where  $q_0$  is the function identically equal to 1. Then  $v_F(\phi)$  is called *F-Brownian bridge*.

It is straightforward to see that equation (4.22) represents  $v_F(\phi)$  as an orthogonal projection of  $w_F(\phi)$  parallel to the function  $q_0$ . For the purposes of statistical testing it is easier to consider the point parametric versions of  $w_F$  and  $v_F$ , which are obtained by setting  $\phi(y) = \mathbb{1}\{y \leq x\}$ , yielding

$$w_F(\mathbb{1}_x(y)) = \int_{y \in \mathbb{R}} \mathbb{1}_x(y) w_F(dy) = w_F(x)$$

and

$$v_F(\mathbb{1}_x(y)) = \int_{y \in \mathbb{R}} \mathbb{1}_x(y) v_F(dy) = v_F(x).$$

Now, let  $G(x)$  be a distribution that shares the same support as the distribution  $F_\theta(x)$ . In our example,  $F_\theta(x)$  is some centred normal distribution, and thus we can take  $G(x)$  as being any distribution that has positive density on the real line. In what follows, we take  $G(x)$  as the standard normal

distribution, though any suitable distribution may be used. Consider then the normalised score function  $\beta_F(x; \theta)$  of the distribution  $F_\theta(x)$ , defined as

$$\beta_F(x; \theta) = \frac{\dot{f}_\theta(x)}{f_\theta(x)} \left( \int_{y \in \mathbb{R}} \frac{\dot{f}_\theta(y)^2}{f_\theta(y)} dy \right)^{-\frac{1}{2}}, \quad (4.23)$$

where  $\dot{f}_\theta(x)$  is the derivative of the density function  $f_\theta(x)$  with respect to the parameter  $\theta$ . Khmaladze (2016) shows that the parametric empirical process has asymptotic expansion

$$v_n(x; \hat{\theta}) = v_n(x; \theta) - \int_{y \leq x} \beta_F(y; \theta) dF_\theta(y) \int_{y \in \mathbb{R}} \beta_F(y; \theta) v_n(dy, \theta) + o_p(1), \quad (4.24)$$

so that the main part of this expansion represents the parametric empirical process as an orthogonal projection of  $v_n(x; \theta)$  parallel to the normalised score function  $\beta_F(x; \theta)$ . Thus, the limit in distribution of  $v_n(x, \hat{\theta})$  can be written, in function parametric form, as

$$v_F(\phi) - \langle \beta_F, \phi \rangle v_F(\beta_F). \quad (4.25)$$

By then substituting in the definition of  $v_F(\phi)$  given by (4.22), and noting that  $\beta_F$  is necessarily orthogonal to  $q_0$ , we see that (4.25) can then be written as

$$w_F(\phi) - \langle q_0, \phi \rangle w_F(q_0) - \langle \beta_F, \phi \rangle w_F(\beta_F). \quad (4.26)$$

This is clearly not  $F$ -Brownian bridge, however the form is very similar, and can be regarded as a "different kind" of Brownian bridge. Let us now define the vector  $q$  as

$$q = [q_0, q_1]^T = [1, \beta_F]^T \quad (4.27)$$

and say that (4.26) is  $q$ -projected  $F$ -Brownian motion. Similarly, let us define the vector  $r$  as

$$r = [r_0, r_1]^T = [1, \beta_G]^T, \quad (4.28)$$

where  $\beta_G(x) \in L_2(G)$  is any function that is orthonormal to  $r_0$  in  $L_2(G)$ . This  $\beta_G(x)$  can be regarded as a kind of score function for the family of

distributions to which  $G(x)$  belongs. In this thesis, we take

$$\beta_F(x; \theta) = \frac{\sigma^{-2}x^2 - 1}{\sqrt{2}} \quad \text{and} \quad \beta_G(x) = \frac{x^2 - 1}{\sqrt{2}}.$$

We then define the function  $l(x; \theta)$  as

$$l(x; \theta) = \sqrt{\frac{dG(x)}{dF_\theta(x)}}. \quad (4.29)$$

From the definition of  $l(x; \theta)$  it follows that

$$\phi \in L_2(G) \implies l\phi \in L_2(F) \quad (4.30)$$

and

$$\|\phi\|_G = \|l\phi\|_F. \quad (4.31)$$

Hence the vector  $[lr_0, lr_1]^T$  is orthonormal in  $L_2(F)$ , since

$$\int l(x; \theta)r_i(x)l(x; \theta)r_j(x)dF_\theta(x) = \int r_i(x)r_j(x)dG(x).$$

Now, consider the unitary operator

$$K_{l, q_0} = I - \frac{2}{\|l - q_0\|_F^2} (l - q_0) \langle l - q_0, \cdot \rangle_F \quad (4.32)$$

on  $L_2(F)$ . We define  $\tilde{lr}_1$  as the image of  $lr_1$  under this operator,

$$\tilde{lr}_1 = K_{l, q_0}(lr_1), \quad (4.33)$$

and then define a second operator, identical in form to (4.32), as

$$K_{\tilde{lr}_1, q_1} = I - \frac{2}{\|\tilde{lr}_1 - q_1\|_F^2} (\tilde{lr}_1 - q_1) \langle \tilde{lr}_1 - q_1, \cdot \rangle_F. \quad (4.34)$$

We then take the product of these operators,  $\hat{K}$ , as the transformation that we will apply to the parametric empirical process,

$$\hat{K} = K_{\tilde{lr}_1, q_1} K_{l, q_0}. \quad (4.35)$$

Finally, the transformed process  $\tilde{v}_n$  is defined by

$$\tilde{v}_n(\phi; \hat{\theta}) = v_n(\hat{K}l\phi; \hat{\theta}). \quad (4.36)$$

Khmaladze proves that, provided  $G$  is absolutely continuous with respect to  $F$ , if  $v_F^q(\phi)$  is  $q$ -projected  $F$ -Brownian motion, then the process  $v_G^r(\phi) = v_F(\hat{K}l\phi)$  is  $r$ -projected  $G$ -Brownian motion. Thus, since the parametric empirical process is asymptotically  $q$ -projected  $F$ -Brownian motion, the process  $\tilde{v}_n$  defined by (4.36) is asymptotically  $r$ -projected  $G$ -Brownian motion. We can then investigate the distribution of statistics based on  $r$ -projected  $G$ -Brownian motion using computational methods, for use in distribution-free statistical testing. In this thesis we take the supremum as our test statistic,

$$D(\tilde{v}_n) = \sup_x |\tilde{v}_n(x; \hat{\theta})|, \quad (4.37)$$

where the point parametric version of (4.36) is defined by

$$\begin{aligned} \tilde{v}_n(x; \hat{\theta}) &= v_n(\hat{K}l\mathbb{1}_x; \hat{\theta}) \\ &= \int_{y \in \mathbb{R}} \hat{K}(l\mathbb{1}_x)(y; \hat{\theta}) v_n(dy; \hat{\theta}) \\ &= \sqrt{n} \left( \int_{y \in \mathbb{R}} \hat{K}(l\mathbb{1}_x)(y; \hat{\theta}) dF_n(y) - \int_{y \in \mathbb{R}} \hat{K}(l\mathbb{1}_x)(y; \hat{\theta}) dF_{\hat{\theta}}(y) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{K}(l\mathbb{1}_x)(X_i; \hat{\theta}) - \sqrt{n} \int_{y \in \mathbb{R}} \hat{K}(l\mathbb{1}_x)(y; \hat{\theta}) dF_{\hat{\theta}}(y) \end{aligned} \quad (4.38)$$

where  $X_1, X_2, \dots, X_n$  is the random sample being tested. For multi-dimensional distributions, this integral can be computed straightforwardly using computational methods. However, since our situation is rather simple, we derive the closed form expression of  $\tilde{v}_n$  solely in terms of  $G(x)$ , making computation of the process trivial.

### Deriving the Transformed Process

We begin by deriving the function  $\hat{K}l\mathbb{1}_x$ . Once this is derived, it is straightforward to obtain a closed form expression of  $\tilde{v}_n$ . First, we define the fol-

lowing constants to ease notation:

$$c_1 = \int_{y \in \mathbb{R}} l(y; \theta) dF_\theta(y) = \sqrt{\frac{2\sigma}{\sigma^2 + 1}}, \quad (4.39)$$

$$c_2 = \sqrt{\frac{2\sigma^2}{\sigma^2 + 1}}, \quad (4.40)$$

$$c_3 = \int_{y \in \mathbb{R}} l(y) r_1(y) dF(y) = \sqrt{\frac{\sigma}{\sigma^2 + 1}} \left( \frac{\sigma^2 - 1}{\sigma^2 + 1} \right), \quad (4.41)$$

$$c_4 = \int_{y \in \mathbb{R}} l(y) q_1(y) r_1(y) dF(y) \quad (4.42)$$

$$= \sqrt{\frac{\sigma}{2(\sigma^2 + 1)}} \left( \frac{10\sigma^2 - \sigma^4 - 1}{(\sigma^2 + 1)^2} \right), \quad (4.43)$$

$$c_5 = \frac{2}{\|l - q_0\|_F^2} = \frac{1}{1 - c_1}, \quad (4.44)$$

$$c_6 = \frac{2}{\|\tilde{l}r_1 - q_1\|_F^2} = \frac{1}{1 - c_4 + c_5^2 c_3^2 (1 - c_1)}. \quad (4.45)$$

For our chosen  $r_1$ , the function  $\tilde{l}r_1$  can then be written as

$$\tilde{l}r_1(x) = l(x)r_1(x) + c_3 c_5 (l(x) - 1), \quad (4.46)$$

while the operator  $\hat{K}$  is

$$\begin{aligned} \hat{K}\phi &= \phi - c_5 (l - q_0) \langle l - q_0, \phi \rangle_F \\ &\quad - c_6 (\tilde{l}r_1 - q_1) \langle \tilde{l}r_1 - q_1, \phi - c_5 (l - q_0) \rangle_F. \end{aligned}$$

Setting  $\phi = l\mathbb{1}_x$ , it is straightforward to derive the scalar products in this expression. We omit the working as it is long and would offer little benefit. We find that

$$\langle l - q_0, l\mathbb{1}_x \rangle_F = G(x) - c_1 G(x/c_2), \quad (4.47)$$

again noting that  $G(x) = \Phi(x)$  is the standard normal distribution function. The second scalar product is longer, but none of the integrals in-

volved are particularly difficult,

$$\begin{aligned}
& \langle \tilde{l}r_1 - q_1, l\mathbb{1}_x - c_5(l - q_0) \langle l - q_0, l\mathbb{1}_x \rangle_F \rangle_F \\
&= \frac{-x}{2\sqrt{\pi}} \exp\left(-\frac{x^2}{2}\right) - c_3c_5(G(x) - c_1G(x/c_2))(2c_5(1 - c_1) - 1) \\
&- \frac{c_2}{\sqrt{2\sigma}} \left( \left( \frac{c_2^2}{\sigma^2} - 1 \right) G(x/c_2) - \frac{c_2x}{\sigma^2\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c_2^2}\right) \right). \tag{4.48}
\end{aligned}$$

Again in the interest of easing notation, let us define the functions  $\alpha_1(x; \theta)$  and  $\alpha_2(x; \theta)$  as

$$\alpha_1(x; \theta) = c_5 \langle l - q_0, l\mathbb{1}_x \rangle_F \tag{4.49}$$

and

$$\alpha_2(x; \theta) = c_6 \langle \tilde{l}r_1 - q_1, l\mathbb{1}_x - c_5(l - q_0) \langle l - q_0, l\mathbb{1}_x \rangle_F \rangle_F \tag{4.50}$$

so that we can write the function  $\hat{K}l\mathbb{1}_x$  more simply as

$$\hat{K}(l\mathbb{1}_x)(y; \theta) = l(y; \theta)\mathbb{1}_x(y) - \alpha_1(x; \theta)(l(y; \theta) - 1) - \alpha_2(x; \theta)(\tilde{l}r_1(y; \theta) - q_1(y; \theta)). \tag{4.51}$$

Since we will be integrating over  $y$ , the functions  $\alpha_1(x; \theta)$  and  $\alpha_2(x; \theta)$  are of little concern to us, and so there is no point writing them out in full. Finally, we need only to derive the integral part of (4.38),

$$\int_{y \in \mathbb{R}} \hat{K}(l\mathbb{1}_x)(y; \theta) dF_\theta(y), \tag{4.52}$$

which is now straightforward to derive, and is equal to

$$c_1G(x/c_2) - \alpha_1(x; \theta)(c_1 - 1) - \alpha_2(x; \theta)c_3(1 + c_5(c_1 - 1)). \tag{4.53}$$

Thus, the closed form expression of  $\tilde{v}_n$ , in point parametric form, is

$$\begin{aligned}
\tilde{v}_n(x; \hat{\theta}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( l(X_i; \hat{\theta})\mathbb{1}_x(X_i) \right. \\
&\quad \left. - \alpha_1(x; \hat{\theta})(l(X_i; \hat{\theta}) - 1) - \alpha_2(x; \hat{\theta})(\tilde{l}r_1(X_i; \hat{\theta}) - q_1(X_i; \hat{\theta})) \right) \\
&\quad \left. - \sqrt{n} \left( c_1G(x/c_2) - \alpha_1(x; \hat{\theta})(c_1 - 1) - \alpha_2(x; \hat{\theta})c_3(1 + c_5(c_1 - 1)) \right) \right). \tag{4.54}
\end{aligned}$$

Noting that since the functions and constants depend on  $\theta$ , we must use the maximum likelihood estimate  $\hat{\theta}$  in its place. Nevertheless, Khmaladze (2016) notes that provided simple continuity assumptions hold for  $l(x; \theta)$  and  $\beta_F(x; \theta)$  in  $\theta$ , the difference resulting from using  $\hat{\theta}$  in place of  $\theta$  is asymptotically small, and therefore convergence to  $r$ -projected  $G$ -Brownian motion is still true.

### The Limiting Distribution of Test Statistic $D(\tilde{v}_n)$

While similar in form,  $r$ -Projected  $G$ -Brownian motion is not  $G$ -Brownian bridge. Hence, the limiting distribution of  $D(\tilde{v}_n)$  is not the usual Kolmogorov distribution used in the Kolmogorov-Smirnov test. The exact limiting distribution of  $D(\tilde{v}_n)$ , the supremum of  $\tilde{v}_n$ , is not known, but we can investigate it using computational methods. Namely, we generate a series of  $G$ -Brownian motions and, for each such  $w_G$ , construct an  $r$ -projected  $G$ -Brownian motion using the formula

$$v_G^r(\phi) = w_G(\phi) - \langle \phi, r_0 \rangle_G w_G(r_0) - \langle \phi, r_1 \rangle_G w_G(r_1), \quad (4.55)$$

or in point parametric form,

$$v_G^r(x) = w_G(x) - G(x)w_G(1) - \int_{y \leq x} r_1(y) dG(y) w_G(r_1). \quad (4.56)$$

We generate 10,000 such processes and give the empirical distribution function of test statistics  $D(v_G^r)$ , the supremums of these  $r$ -projected  $G$ -Brownian motions, in Figure 4.4. For comparison, we also give the Kolmogorov distribution function alongside it. We also give quantiles for use in statistical testing in Table 4.2. Finally, an example trajectory of  $\tilde{v}_n(x; \hat{\theta})$  is given in Figure 4.5, based off a random sample generated from a  $N(0, 4)$  distribution.

### Testing Normality of Increments

We test the normality of  $\Delta$ -increments  $\Delta X_i$  for the same  $\Delta$  considered in Figures 4.2 and 4.3. The processes  $\tilde{v}_n(x; \hat{\theta})$  for each  $\Delta$  are given in Figure

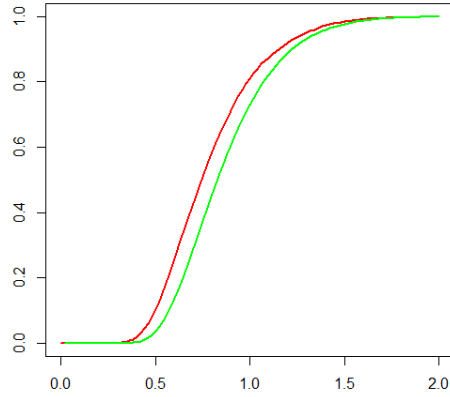


Figure 4.4: The empirical distribution function of test statistics  $D(v_G^r)$  (red) compared with Kolmogorov distribution function (green).

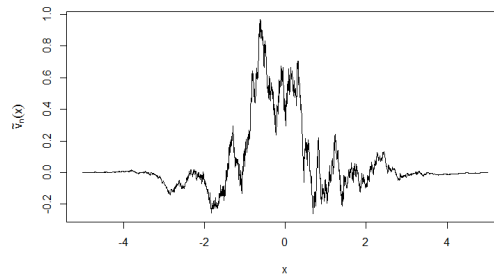


Figure 4.5: An example trajectory of  $\tilde{v}_n(x; \hat{\theta})$  using generated  $N(0, 4)$  data.

Quantiles of Test Statistic $D(v_G^r)$	
Quantile	$D(v_G^r)$
80%	0.986
85%	1.053
90%	1.153
95%	1.291
99%	1.573

Table 4.2: Quantiles of test statistic  $D(v_G^r)$ .



Test statistics $D(\tilde{v}_n)$ for increments $\Delta X_i$	
Increment Length	$D(\tilde{v}_n)$
1 Day	0.1462
1 Hour	1323
10 Minutes	$2.994 \times 10^9$
1 Minute	$6.470 \times 10^{68}$

Table 4.3: Test Statistics  $D(\tilde{v}_n)$  for increments  $\Delta X_i$ .

4.6, and the corresponding statistics  $D(\tilde{v}_n)$  are given in Table 4.3.

Using the quantiles for  $D(v_G^r)$  given in Table 4.2, we find that normality is strongly rejected for 1 minute, 10 minute, and 1 hour increments at the 5% level, but not rejected for daily increments. In fact, it is immediately clear by the trajectories of  $\tilde{v}_n$  for  $\Delta = 1$  minute, 10 minutes, and 1 hour that they are not  $r$ -projected  $G$ -Brownian motions. The smooth shape of these trajectories, with no apparent randomness, being caused by the increments' lack of normality. This leaves us with the question of daily increments. Although we failed to reject normality, we see in Figure 4.3 there is still a large discrepancy between the normal density function and the density histogram. Nevertheless, if we suppose that daily increments are indeed normally distributed, then we should hope to explain why intra-daily increments are not, as these statements seem to contradict one another. We address this question later in the chapter by considering mixture distributions. Next, we examine whether the variance of increments is linear in time.

### 4.3.3 Variance of Increments

We know that for any Brownian motion, the variance of its increments is proportional to increment length. Hence, if our exchange rate process  $X(t)$  behaves as a Brownian motion, the variance of its increments should

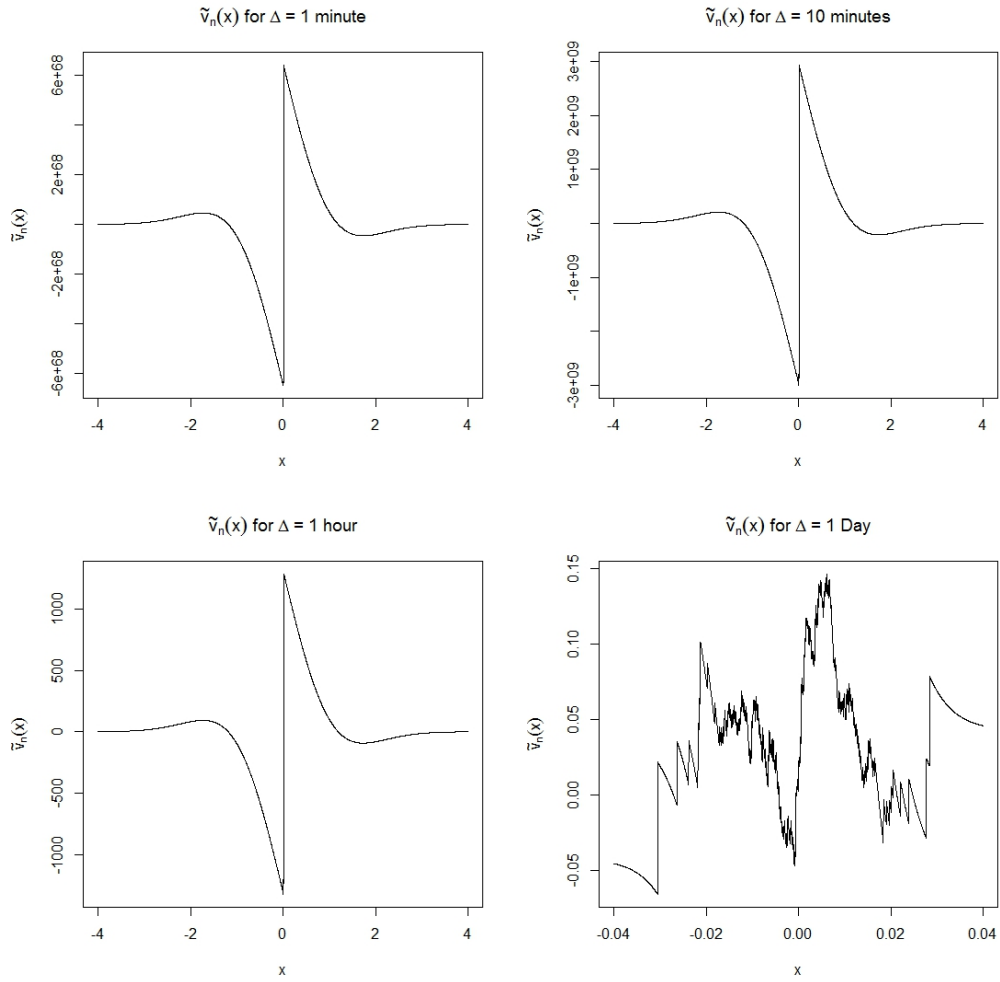


Figure 4.6: Trajectories  $\tilde{v}_n(x; \hat{\theta})$  of  $\Delta X_i$  for various  $\Delta$ .

be linear in time. We verify this using the empirical variance

$$\frac{1}{n} \sum_{i=1}^n \Delta X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n \Delta X_i \right)^2 .$$

In Figure 4.7 we plot the empirical variance of the increments  $\Delta X_i$  against increment length  $\Delta$ . We do so for two different time scales. The first looks at the increment length measured in minutes, while the second looks at a larger scale with increment length measured in days. On the scale of minutes, there is a clear linear relationship between variance and increment length. We also see a linear relationship on the scale of days, though it is unclear whether this linearity holds for larger  $\Delta$ . It appears as though the slope begins to decrease, though this might be due to the decreasing sample size giving inaccurate estimates.

If this linearity is indeed beginning to break down as  $\Delta$  increases, this suggests that  $X(t)$  might behave as Brownian motion on small scales, but is asymptotically stationary. We consider this in the next chapter with respect to the asymptotically stationary Ornstein-Uhlenbeck process.

While the variance of increments  $\Delta X_i$  appears to be linear in time, we did not take time of day into consideration. This assumes that the variance of an increment is independent of the time of day it occurs. Since volatility of financial prices is known to be correlated with market activity (Gelati, 2000), and market activity varies significantly by geographic region (BIS, 2016), it would be reasonable to expect variance displays some degree of periodic behaviour.

To investigate whether periodicity exists, we consider the variance of 10 minute price increments. We wish to see whether the variance of these increments displays daily or weekly periodic behaviour. To this end, we sort these increments depending on the hour of day and hour of week they occur. This allows us to see how the variance of 10 minute increments changes throughout the day and week. The results are shown in Figures 4.8 and 4.9.

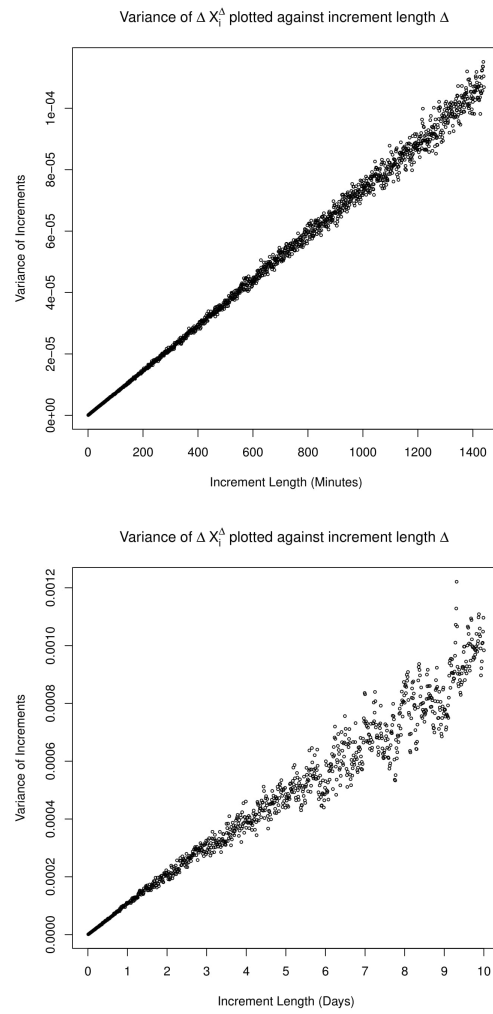


Figure 4.7: Variance of increments vs increment length.

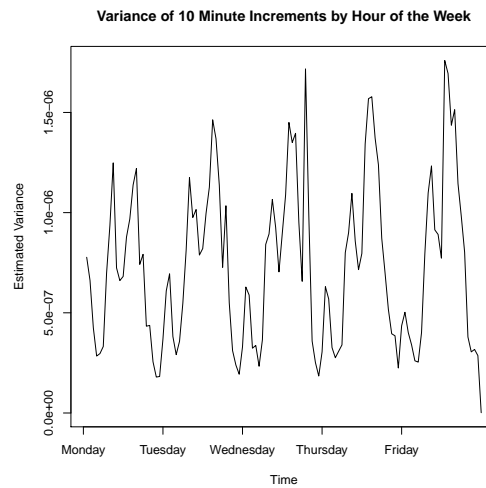


Figure 4.8: Empirical variance of 10 minute increments in each hour of the week in 2010.

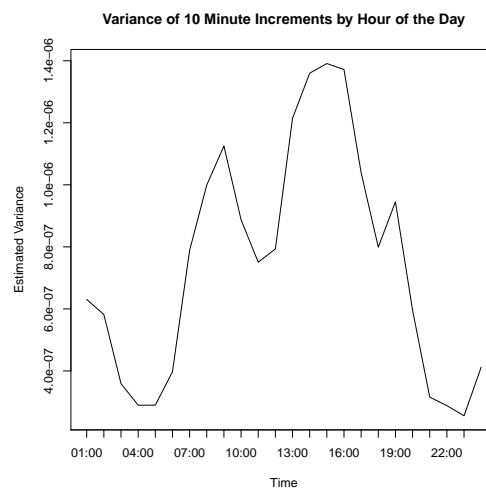


Figure 4.9: Empirical variance of 10 minute increments in each hour of the day in 2010.

A daily cycle is immediately clear, with several peaks in variance occurring at regular times throughout the day. Shiryaev explains this periodic behavior in terms of time zones and active regional markets (1999, p. 318-320). We are not concerned with the cause of this periodicity, but instead use it to explain why daily increments might be Gaussian, while intra-daily increments are clearly not. In the next section, we discuss the Gaussianity of increments in the context of mixture distributions.

## 4.4 Mixtures of Normal Distributions

Let  $f_i(x)$ , for  $i = 1, 2, \dots, m$ , be a finite collection of density functions, and  $w_i$ , for  $i = 1, 2, \dots, m$ , a collection of weights such that  $\sum_i w_i = 1$  and  $w_i \geq 0$  for all  $i$ . Then the function  $f(x)$  defined by

$$f(x) = \sum_i w_i f_i(x) \quad (4.57)$$

is the density function of a mixture distribution. Mixture distributions arise when you collect data from a population that is not homogeneously distributed, and offer one explanation for why daily increments might be normally distributed, while intra-daily increments are not. To see why this is, suppose we have a finite collection of independent and normally distributed random variables  $Z_i$ , each with a specific variance  $\sigma_i^2$ ,

$$Z_i \sim N(0, \sigma_i^2).$$

We then define a new random variable  $\varepsilon$  by randomly choosing one of these  $Z_i$  such that

$$P(\varepsilon = Z_i) = p_i,$$

and

$$\sum_i p_i = 1.$$

The distribution of  $\varepsilon$  is then a mixture of the distributions of  $Z_i$ , weighted according to the probabilities  $p_i$ . Although each  $Z_i$  is normally distributed, the random variable  $\varepsilon$  is not,

$$f_\varepsilon(x) = \sum_i p_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{x^2}{2\sigma_i^2}}.$$

On the other hand, it is well known that the random variable

$$X = \sum_i Z_i$$

is normally distributed, as it is simply the summation of finitely many independent normal random variables.

With this in mind, consider both the hour and daily increments discussed in the previous section. Suppose, hypothetically, that each hour increment is normally distributed with a specific variance according to the time of day it occurs, so that hour increments come from 24 normal distributions. Under this assumption, the distribution of all hour increments is a mixture of normal distributions, making it non-normal and consistent with what we observe. On the other hand, daily increments will be normally distributed as they are the sum of 24 independent and normally distributed hourly increments. In this way, periodic behaviour of variance can result in the normality of increments depending on increment length.

More generally, this leads to a natural model for the exchange rate process. Namely, suppose  $X(t)$  is of the form

$$X(t) = X(0) + \int_0^t \sigma(s) dB(s), \quad (4.58)$$

where  $\sigma(s)$  is a periodic function with period 1 day and  $B(t)$  is standard Brownian motion. This model seems consistent with what we have seen of the total variation, distribution, and variance of the process  $X(t)$ . Thus, we conclude this chapter by noting that while it is unlikely the exchange rate process is a Brownian motion of the form  $\sigma B(t)$ , the more general form

(4.58) seems consistent with what we observe. In the next chapter, we consider periodic variance in the context of the mean-reverting Ornstein-Uhlenbeck process.



# Chapter 5

## The Ornstein-Uhlenbeck Process

In this chapter, we introduce the mean-reverting Ornstein-Uhlenbeck process. We review some basic results, and also give maximum likelihood estimators. We finish by considering how periodic variance, as seen in the previous chapter, might be incorporated into the Ornstein-Uhlenbeck model of mean-reversion.

### 5.1 Introduction

Mean-reverting processes play an important role in modelling financial prices. This is particularly true when considering the exchange rate of two currencies. Since, under normal circumstances, the economies of two countries changes very slowly over time, it is reasonable to expect the "true" relative worth of their currencies to change slowly over time as well. One such process is the Ornstein-Uhlenbeck process, which is characterised by a constant rate of mean-reversion.

**Definition 5.1.** The Ornstein-Uhlenbeck process  $X_t$  is the process satisfying the stochastic differential equation

$$dX_t = a(\mu - X_t)dt + \sigma dW_t, \quad (5.1)$$

where  $a, \sigma > 0$ .

The behaviour of the Ornstein-Uhlenbeck process can be seen through the stochastic differential equation (5.1). In every interval of time  $dt$ , the process  $X_t$  reverts back to the mean  $\mu$  at a rate of  $a$ , while a random (possibly negative) amount  $dW_t$  is added. Depending on the rate of reversion  $a$ , the Ornstein-Uhlenbeck process behaves as Brownian motion at sufficiently small scales, but has an asymptotically stationary distribution with mean  $\mu$ . The solution to (5.1) can be found using suitable substitutions and applying Ito's lemma (Moral & Penev, 2016, p. 421).

**Theorem 5.1.** The solution to the differential equation (5.1) is

$$X_t = e^{-at}X_0 + \mu(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dW_s. \quad (5.2)$$

*Proof.* Using the change of variables  $Y_t = X_t - \mu$  and applying Ito's lemma to the function  $f(X, t) = X_t - \mu$  we obtain the stochastic differential equation

$$dY_t = -aY_t dt + \sigma dW_t. \quad (5.3)$$

A second change of variables  $Z_t = e^{at}Y_t$  and applying Ito's lemma to the function  $f(Y, t) = e^{at}Y_t$  gives us a third stochastic differential equation

$$dZ_t = ae^{at}Y_t dt + e^{at} dY_t. \quad (5.4)$$

Substituting (5.3) into (5.4) then gives

$$\begin{aligned} dZ_t &= ae^{at}Y_t dt + e^{at}(-aY_t dt + \sigma dW_t) \\ &= \sigma e^{at} dW_t. \end{aligned} \quad (5.5)$$

Integrating (5.5) from 0 to  $t$  we obtain

$$Z_t - Z_0 = \int_0^t \sigma e^{as} dW_s. \quad (5.6)$$

Substituting  $Y_t$  back in place of  $Z_t$ , keeping in mind that  $Y_0 = Z_0$ , gives

$$Y_t = e^{-at}Y_0 + \sigma e^{-at} \int_0^t e^{as} dW_s. \quad (5.7)$$

Finally, by substituting back  $Y_t = X_t - \mu$  we obtain the solution to (5.1)

$$X_t = e^{-at}X_0 + \mu(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dW_s.$$

□

Using the solution to (5.2) it is straightforward to derive the expected value and covariance of the Ornstein-Uhlenbeck process defined by (5.1).

**Theorem 5.2.** Let  $X_t$  be the Ornstein-Uhlenbeck process defined by equation (5.1). Then, assuming  $X_0$  is non-random, it has expected value

$$\mathbf{E}[X_t] = X_0 e^{-at} + \mu(1 - e^{-at}) \quad (5.8)$$

and covariance

$$\text{cov}(X_s, X_t) = \frac{\sigma^2}{2a} (e^{-a|t-s|} - e^{-a(t+s)}). \quad (5.9)$$

*Proof.* The expected value of  $X_t$  follows trivially since, using the solution obtained in Theorem (5.1),

$$\begin{aligned} \mathbf{E}[X_t] &= \mathbf{E}[e^{-at}X_0 + \mu(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dW_s] \\ &= e^{-at}X_0 + \mu(1 - e^{-at}) + \sigma e^{-at} \mathbf{E}[\int_0^t e^{as} dW_s] \\ &= e^{-at}X_0 + \mu(1 - e^{-at}) \end{aligned} \quad (5.10)$$

where we use the fact that for a continuous and bounded function  $g(t)$ , the integral  $\int_0^t g(s) dW_s$  has expected value 0. The covariance is obtained in a

similar manner.

$$\begin{aligned}
\text{cov}(X_s, X_t) &= \mathbf{E}[(X_s - \mathbf{E}[X_s])(X_t - \mathbf{E}[X_t])] \\
&= \mathbf{E}[(\sigma e^{-as} \int_0^s e^{au} dW_u)(\sigma e^{-at} \int_0^t e^{av} dW_v)] \\
&= \sigma^2 e^{-a(t+s)} \mathbf{E}[\int_0^s e^{au} dW_u \int_0^t e^{av} dW_v] \\
&= \sigma^2 e^{-a(t+s)} \mathbf{E}[(\int_0^{\min(s,t)} e^{au} dW_u)^2] \tag{5.11} \\
&= \sigma^2 e^{-a(t+s)} \int_0^{\min(s,t)} e^{2au} du \\
&= \sigma^2 e^{-a(t+s)} \left( \frac{1}{2a} e^{2a \min(s,t)} - \frac{1}{2a} \right) \\
&= \frac{\sigma^2}{2a} (e^{-a|t-s|} - e^{-a(t+s)}) \tag{5.12}
\end{aligned}$$

where (5.11) follows by the independence of disjoint intervals of  $W_t$ .  $\square$

**Corollary 5.2.1.** The variance of the Ornstein Uhlenbeck process defined by (5.1) is

$$\text{Var}(X_t) = \frac{\sigma^2}{2a} (1 - e^{-2at}). \tag{5.13}$$

The asymptotic stationarity of the Ornstein-Uhlenbeck process can then be seen by taking the limit of (5.8) and (5.13) as  $t \rightarrow \infty$ , giving

$$\lim_{t \rightarrow \infty} \mathbf{E}[X_t] = \mu \quad \text{and} \quad \lim_{t \rightarrow \infty} \text{Var}(X_t) = \frac{\sigma^2}{2a}.$$

## 5.2 Maximum likelihood estimation

Estimating the parameters of an Ornstein-Uhlenbeck process is straightforward. In general, we can derive maximum likelihood estimates for diffusion processes using likelihood ratios (Klebaner, 2005, pp.282-285). Alternatively, we can also use conditional distributions to construct a likelihood function. If  $X(t)$  is the Ornstein-Uhlenbeck process defined by the

stochastic differential equation (5.1), and is observed at times

$$X(0), X(\Delta), X(2\Delta), \dots, X(n\Delta)$$

which we again denote as

$$X_0^\Delta, X_1^\Delta, X_2^\Delta, \dots, X_n^\Delta$$

to simplify notation. Then we can construct a likelihood function using the conditional distribution of each  $X_i^\Delta$  given  $X_{i-1}^\Delta$ , which is

$$X_i^\Delta | X_{i-1}^\Delta \sim N \left( X_{i-1}^\Delta e^{-a\Delta} + \mu(1 - e^{-a\Delta}), \frac{\sigma^2}{2a}(1 - e^{-2a\Delta}) \right) \quad (5.14)$$

for  $i = 1, 2, \dots, n$ . Additionally, suppose the initial condition has distribution

$$X_0^\Delta \sim N \left( \mu, \frac{\sigma^2}{2a} \right). \quad (5.15)$$

Thus, using the fact that

$$f(x_0, x_1, x_2, \dots, x_n) = f_0(x_0) \prod_{i=1}^n f_i(x_i | x_{i-1})$$

we obtain the likelihood function

$$L(a, \mu, \sigma; X) = f_0(X_0) \prod_{i=1}^n f_i(X_i | X_{i-1}). \quad (5.16)$$

The resulting maximum likelihood estimates are given in [2]. They are

$$\hat{a} = -\Delta^{-1} \log(\hat{\beta}_1), \quad \hat{\mu} = \hat{\beta}_2, \quad \hat{\sigma} = 2\hat{a}\hat{\beta}_3(1 - \hat{\beta}_1^2)^{-1} \quad (5.17)$$

where we define

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i^\Delta X_{i-1}^\Delta - n^{-1} \sum_{i=1}^n X_i^\Delta \sum_{i=1}^n X_{i-1}^\Delta}{\sum_{i=1}^n (X_{i-1}^\Delta)^2 - n^{-1} (\sum_{i=1}^n X_{i-1}^\Delta)^2}, \\ \hat{\beta}_2 &= \frac{\sum_{i=1}^n (X_i^\Delta - \hat{\beta}_1 X_{i-1}^\Delta)}{n(1 - \hat{\beta}_1)}, \\ \hat{\beta}_3 &= n^{-1} \sum_{i=1}^n (X_i^\Delta - \hat{\beta}_1 X_{i-1}^\Delta - \hat{\beta}_2(1 - \hat{\beta}_1))^2. \end{aligned}$$

### 5.3 Ornstein-Uhlenbeck with Periodic Variance

So far we have considered the Ornstein-Uhlenbeck process driven by a Brownian motion with constant  $\sigma$ . However, as we have seen in Chapter 4, the variance of the exchange rate process displays periodic behaviour, correlated in some way with market activity. Therefore, if it behaves as a mean-reverting process, it would be better described by the stochastic differential equation

$$dX_t = a(\mu - X_t)dt + \sigma_t dW_t \quad (5.18)$$

where  $\sigma_t$  is a positive continuous and periodic function with period 1 day. To account for this, it is straightforward to generalise the results introduced thus far. We omit the proofs of the following results as the working is nearly identical. Firstly, the solution (5.2) generalises to

$$X_t = e^{-at}X_0 + \mu(1 - e^{-at}) + e^{-at} \int_0^t \sigma_s e^{as} dW_s. \quad (5.19)$$

From this, once again assuming  $X_0$  is non-random, it immediately follows that

$$\mathbf{E}[X_t] = X_0 e^{-at} + \mu(1 - e^{-at}) \quad (5.20)$$

and

$$\text{Var}(X_t) = e^{-2at} \int_0^t \sigma_s^2 e^{2as} ds. \quad (5.21)$$

This then leads to the conditional distribution

$$X_t | X_s \sim N \left( X_s e^{-a(t-s)} + \mu(1 - e^{-a(t-s)}), e^{-2a(t-s)} \int_s^t \sigma_v^2 e^{2a(v-s)} dv \right). \quad (5.22)$$

#### 5.3.1 Maximum Likelihood Estimation

Generalising the maximum likelihood estimators introduced earlier is less straightforward. In particular, the presence of the unknown function  $\sigma_t$

makes estimation difficult. At the same time however, periodic variance does not seem so strange, and one can imagine it arising in many natural processes. Therefore the question of how to estimate parameters of an Ornstein-Uhlenbeck process with periodic variance has some practical importance. Suppose we observe  $X_t$  at times  $i\Delta$ , for  $i = 1, 2, \dots, n$ , such that  $\Delta$  is the period of  $\sigma_t$ , so that here we take  $\Delta = 1$  day. This then gives us

$$\text{Var}(X_i^\Delta | X_{i-1}^\Delta) = e^{-2a\Delta} \int_0^\Delta \sigma_s^2 e^{2as} ds \quad (5.23)$$

for all  $i = 1, 2, \dots, n$ . Thus, provided we observe the process so that increment length is equal to the period of  $\sigma_t$ , the variance of increments are once again equal. The exact form of  $\sigma_i$  over the interval  $[0, \Delta]$  is of little interest to us, instead we simply wish to estimate the mean, rate of reversion, and variance of increments. To this end, we hope to reduce the problem by replacing the integral in (5.23) with a single parameter. That is, we define the parameter  $\kappa$  as

$$\kappa = \sqrt{\int_0^\Delta \sigma_s^2 e^{2as} ds}, \quad (5.24)$$

so that the variance (5.23) becomes

$$\text{Var}(X_i^\Delta | X_{i-1}^\Delta) = e^{-2a\Delta} \kappa^2. \quad (5.25)$$

This then yields the likelihood function

$$L(a, \mu, \kappa | X) = \prod_{i=1}^n \frac{e^{a\Delta}}{\sqrt{2\pi\kappa}} \exp\left(-\frac{(X_i^\Delta - e^{-a\Delta} X_{i-1}^\Delta - \mu(1 - e^{-a\Delta}))^2}{2e^{-2a\Delta} \kappa^2}\right), \quad (5.26)$$

with which estimates  $\hat{a}$ ,  $\hat{\mu}$ ,  $\hat{\kappa}$  can be derived. We conclude this section by making one approximation in order to significantly simplify the maximum likelihood estimates obtained from (5.26). It is straightforward to see that this likelihood function leads to

$$\begin{aligned} \hat{\mu} &= \frac{e^{a\Delta} n^{-1} \sum_{i=1}^n X_i^\Delta - n^{-1} \sum_{i=1}^n X_{i-1}^\Delta}{e^{a\Delta} - 1} \\ &= \bar{X}_n + \frac{X_n^\Delta - X_0^\Delta}{n(e^{a\Delta} - 1)}, \end{aligned}$$

so that  $\hat{\mu} \rightarrow \bar{X}_n$  as  $n \rightarrow \infty$ , noting again that  $\Delta = 1$  day is fixed as the period of  $\sigma_t$ . Let us therefore take the sample mean as our estimate of  $\mu$ ,

$$\hat{\mu} = \bar{X}_n, \quad (5.27)$$

as this simplifies estimates  $\hat{a}$  and  $\hat{\kappa}$  substantially. Using this approximation we then obtain

$$\hat{a} = \frac{1}{\Delta} \log \left( \frac{\sum_{i=1}^n (X_{i-1}^\Delta - \bar{X}_n)^2}{\sum_{i=1}^n (X_i^\Delta - \bar{X}_n)(X_{i-1}^\Delta - \bar{X}_n)} \right) \quad (5.28)$$

and

$$\hat{\kappa} = \sqrt{\frac{1}{n} \sum_{i=1}^n (e^{\hat{a}\Delta} (X_i^\Delta - \bar{X}_n) - (X_{i-1}^\Delta - \bar{X}_n))^2}. \quad (5.29)$$

Thus, by observing the process (5.18) at the period of  $\sigma_t$ , we can estimate its mean, rate of reversion, and variance fairly easily, without having to know or estimate the function  $\sigma_t$ . However, these estimators have a significant draw back in that they make use of only a fraction of the data at hand, as data between the period of  $\sigma_t$  is not incorporated into estimates  $\hat{a}$ ,  $\hat{\mu}$ , and  $\hat{\kappa}$ . Hence, these estimators work better for processes with a shorter period of  $\sigma_t$ , or when data is scarce, as the proportion of unused data is smaller.

## 5.4 The Exchange Rate Process

We now examine the exchange rate process  $X(t)$  as a mean-reverting Ornstein-Uhlenbeck process with periodic  $\sigma_t$ . It is worth noting that the rate of reversion might be very weak, so that mean-reversion is only evident on very large time scales. This is demonstrated in Figure 5.1, where we show the exchange rate process over two different time scales. The first shows  $X(t)$  over a single month, where there is no indication of mean-reversion. The second shows  $X(t)$  over two years, where we see behaviour that could indeed be described as mean-reversion.

We mention this in order to justify modelling  $X(t)$  as an Ornstein-Uhlenbeck process with periodic  $\sigma_t$  despite the linearity of variance seen



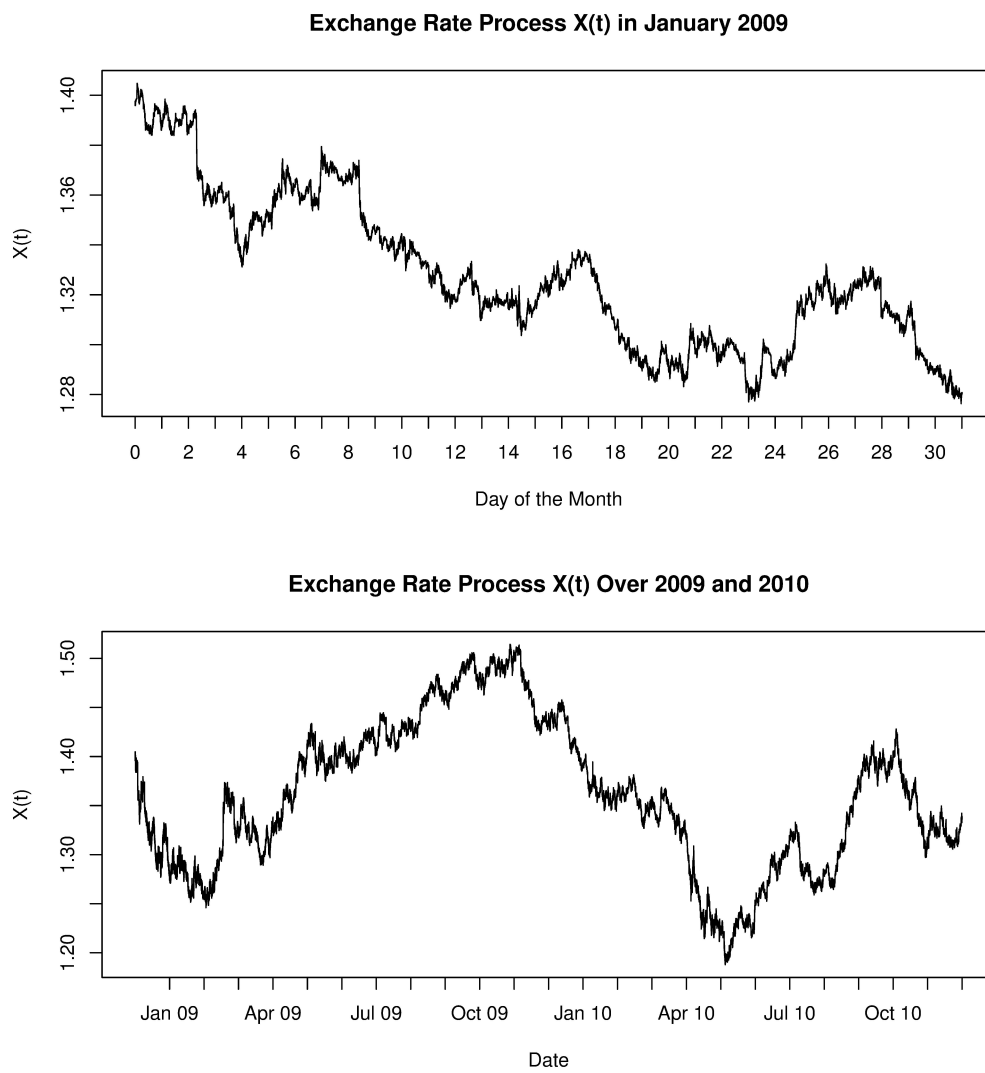


Figure 5.1: The exchange rate process  $X(t)$  over two different time scales.

in Chapter 4. Namely, in Figure 4.7 on page 62 we saw that increment variance was linear in time, while the Ornstein-Uhlenbeck process is asymptotically stationary. To see why this is, let us simplify the problem by considering the Ornstein-Uhlenbeck process with constant  $\sigma$ . Suppose that  $a\Delta$  is very small, and then consider the variance of the Ornstein-Uhlenbeck process given by (5.13). We see that

$$\begin{aligned}\text{Var}(\Delta X_i | X_{i-1}^\Delta) &= \frac{\sigma^2}{2a}(1 - e^{-2a\Delta}) \\ &\approx \frac{\sigma^2}{2a}2a\Delta \\ &= \sigma^2\Delta.\end{aligned}$$

Hence, depending on the size of  $a$ , the variance of increments will be approximately linear for sufficiently small  $\Delta$ .

Since the variance of the exchange rate process displays daily periodic behaviour, we use the estimators introduced in Section 5.3.1, where we take  $\Delta = 1$  day as the period of  $\sigma_t$ , and measure time in days. This yields estimates

$$\hat{\mu} = 1.33, \quad \hat{a} = 0.017, \quad \hat{\kappa} = 0.0092 \quad (5.30)$$

which implies that for  $\Delta = 1$  day, we have

$$\mathbf{E}(X_i^\Delta | X_{i-1}^\Delta) = X_{i-1}^\Delta e^{-0.017} + 1.33(1 - e^{-0.017}) \quad (5.31)$$

and

$$\text{Var}(X_i^\Delta | X_{i-1}^\Delta) = 8.18 \times 10^{-5}. \quad (5.32)$$

It is then worth considering what can be said about the increment variance for an Ornstein-Uhlenbeck process with periodic variance, so that we can compare this with what we observe empirically in the exchange rate process. Without loss of generality, assume that  $\mu = 0$ , since otherwise we can simply centre the process. Again assuming  $X(0)$  is non-random, the

variance of an increment  $\Delta X(t)$  is then

$$\begin{aligned}
\text{Var}(X(t + \Delta) - X(t)) &= \text{Var}\left(e^{-a(t+\Delta)} \int_0^{t+\Delta} \sigma_s e^{as} dW_s - e^{-at} \int_0^t \sigma_s e^{as} dW_s\right) \\
&= \text{Var}\left((e^{-a\Delta} - 1)e^{-at} \int_0^t \sigma_s e^{as} dW_s + e^{-a(t+\Delta)} \int_t^{t+\Delta} \sigma_s e^{as} dW_s\right) \\
&= (e^{-a\Delta} - 1)^2 \text{Var}(X_t) + e^{-2a(t+\Delta)} \int_t^{t+\Delta} \sigma_s^2 e^{2as} ds \\
&= (e^{-a\Delta} - 1)^2 \text{Var}(X_t) + e^{-2a\Delta} \int_0^\Delta e^{2az} \sigma_{z+t}^2 dz \\
&= (e^{-a\Delta} - 1)^2 \text{Var}(X_t) + \frac{1}{a} \int_0^{a\Delta} e^{-2y} \sigma_{\Delta - \frac{y}{a} + t}^2 dy
\end{aligned} \tag{5.33}$$

where we have used the substitutions

$$z = s - t \quad \text{and} \quad y = a(\Delta - z).$$

Since  $\sigma_t$  is a periodic and bounded function, the integral in (5.33) converges to some constant  $c$  as  $\Delta \rightarrow \infty$ . Hence the variance of increments  $\Delta X(t)$  is such that

$$\lim_{\Delta \rightarrow \infty} \text{Var}(\Delta X(t)) = \text{Var}(X(t)) + c \tag{5.34}$$

where

$$c = \frac{1}{a} \int_0^\infty e^{-2y} \sigma_{\Delta - \frac{y}{a} + t}^2 dy.$$

Thus if the exchange rate process indeed behaves as an Ornstein-Uhlenbeck process with periodic variance, then it should be asymptotically stationary. To see whether this is the case, we again estimate the variance of increments as in Chapter 4, but now use the expected value

$$\mathbf{E}[\Delta X_i | X_{i-1}^\Delta] = X_{i-1}^\Delta (e^{-a\Delta} - 1) + \mu(1 - e^{-a\Delta}) \tag{5.35}$$

to estimate the variance as

$$\frac{1}{n} \sum_{i=1}^n (\Delta X_i - X_{i-1}^\Delta (e^{-\hat{a}\Delta} - 1) - \hat{\mu}(1 - e^{-\hat{a}\Delta}))^2. \tag{5.36}$$

Note that we no longer need restrict ourselves to  $\Delta = 1$  day, having already obtained estimates  $\hat{a}$  and  $\hat{\mu}$  using daily prices. We again give two different time scales for comparison in Figure 5.2. The plots look very similar to those given in Chapter 4, with a strong linear relationship for small  $\Delta$  that gets weaker as  $\Delta$  increases. While the plot looks as though linearity begins to break down for larger  $\Delta$ , it is difficult to draw conclusions as the change is small.

We therefore conclude this chapter by noting that if the exchange rate does behave as an Ornstein-Uhlenbeck process with periodic variance, the rate of reversion  $a$  is simply too small to see increment variance converge on the time scale of days. This suggests that even if the exchange rate is a mean-reverting process, the rate of reversion is small enough that it behaves as Brownian motion unless considering much longer increments, perhaps on the scale of months or years.

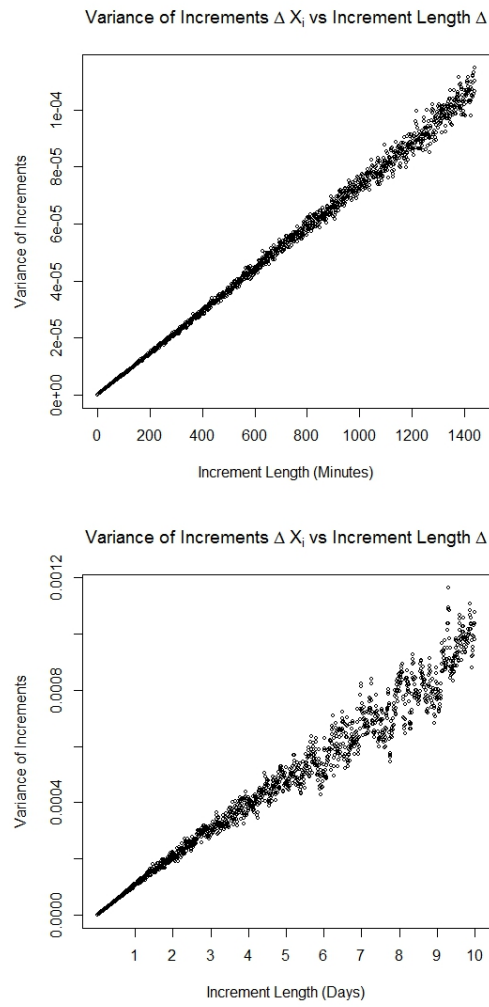


Figure 5.2: Estimated variance of Ornstein-Uhlenbeck increments vs increment length.



# Chapter 6

## Fractal Sets

In this chapter, fractal geometry is introduced along with the box-counting and Hausdorff dimensions, following Kenneth Falconer's *Fractal Geometry: Mathematical Foundations and Applications*. We give some methods of deriving or estimating the dimension of a fractal, and also introduce a new estimator for the Hurst exponent of a fractal Brownian motion. We conclude the chapter by estimating the box-counting dimension of the exchange rate process  $X(t)$ .

### 6.1 What are Fractals?

Fractal geometry is the study of sets which are highly irregular and complex at all scales, and has been applied in many fields where complex structures arise, including of course financial markets. Historically, the term fractal has generally referred to sets with a fractional dimension. This definition is problematic in that dimension describes only the complexity of a set, and not the structure or means by which it is constructed. We will see later that many fractals share a common set of properties which dimension alone does not describe. However this is the most widely used definition, and does give an impression of what to expect.

## 6.2 Dimension

Dimension is a central concept in fractal geometry. In traditional geometry, we think of a set  $F$  as having dimension  $n$  if  $\mathbb{R}^n$  is the smallest Euclidean space containing  $F$ . Clearly this intuitive definition does not generalise to the case of fractional dimensions. It is instead more useful to think of dimension as a measure of how much space a set takes up close to any of its points. In this sense, dimension is usually defined by measuring a set at some scale  $\delta$ , and then observing how these measurements change as  $\delta \rightarrow 0$ . In this thesis, we look at the *box-counting dimension* and the *Hausdorff dimension*. The former is more practical to use and allows us to easily estimate the dimension of a set, while the latter is a more mathematically sound definition and possesses properties we would expect of dimension.

### 6.2.1 Box-counting Dimension

The box-counting dimension is one of the most widely used dimensions in fractal geometry. Its formulation gives us a practical method of estimating a set's dimension, which we will use later in this chapter. As the name implies, the box-counting dimension is based on counting the minimum number of "boxes" needed to cover a set, and then examining how this number changes as we reduce their size. The notion of box-counting was studied by Kolmogorov in [13], where he calls the box-count the covering number, and refers to the logarithm of the covering number as metric entropy.

Suppose we have a set  $F \subset \mathbb{R}^2$ , and we wish use this idea of box-counting to define dimension. We start by dividing  $\mathbb{R}^2$  into a mesh of squares of side  $\delta$ , and define  $N_\delta(F)$  as the number of  $\delta$ -mesh squares that intersect the set  $F$ . The dimension of the set  $F$  should then reflect the manner in which  $N_\delta(F)$  grows as  $\delta \rightarrow 0$ . For example, suppose that  $N_\delta(F)$



approximately obeys a power law of the form

$$N_\delta(F) \approx c\delta^{-s}, \quad (6.1)$$

where  $c$  and  $s$  are both positive constants. Then it would be natural to say  $F$  has dimension  $s$ , as it reflects the rate at which  $N_\delta(F)$  grows as  $\delta \rightarrow 0$ . Following this line of thinking, we can then solve for  $s$  to obtain

$$s \approx \frac{\log N_\delta(F)}{-\log \delta} + \frac{\log c}{\log \delta}. \quad (6.2)$$

Based off (6.2), we might hope that as  $\delta$  decreases in size this approximation becomes increasingly accurate, so that  $s$  is given by

$$s = \lim_{\delta \rightarrow 0} \left( \frac{\log N_\delta(F)}{-\log \delta} + \frac{\log c}{\log \delta} \right) \quad (6.3)$$

$$= \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \quad (6.4)$$

provided the limit exists. In this way, we now give the formal definition for the box-counting dimension of a set  $F \subset \mathbb{R}^n$ .

**Definition 6.1.** Let  $F$  be a subset of  $\mathbb{R}^n$ , and suppose  $\mathbb{R}^n$  is divided into a mesh of  $n$ -dimensional cubes, each with side  $\delta$ . Then the box-counting dimension of the set  $F$  is defined as

$$\dim_{\text{B}} F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \quad (6.5)$$

provided that the limit exists, where  $N_\delta(F)$  is the number of  $\delta$ -mesh cubes that intersect the set  $F$ .

In Chapter 2 we introduced the concept of an open or closed cover. An important concept in fractal geometry is that of a  $\delta$ -cover.

**Definition 6.2.** Let  $F$  be a set in some metric space  $R$  with metric  $\rho$ , and  $\{U_i\}$  a collection of sets such that  $F \subset \cup_i U_i$ . We define the *diameter* of each set  $U_i$ , denoted  $|U_i|$ , as the greatest distance between any two points belonging to  $U_i$ . Furthermore, we say that  $\{U_i\}$  is a  $\delta$ -covering of  $F$  if

$$|U_i| \leq \delta \text{ for all } i.$$

Often when deriving the dimension of a set, it is easier to talk in terms of open and closed  $\delta$ -covers rather than  $\delta$ -meshes. This motivates the formulation of alternative but equivalent definitions of the box-count dimension.

**Theorem 6.1.** Equivalent definitions of the box-counting dimension can be obtained by defining  $N_\delta(F)$  as any of the following:

1. The smallest number of sets of diameter at most  $\delta$  needed to cover  $F$ .
2. The smallest number of closed balls of radius  $\delta$  needed to cover  $F$ .
3. The smallest number of cubes of side  $\delta$  needed to cover  $F$ .
4. The number of  $\delta$ -mesh cubes that intersect  $F$ .
5. The largest number of disjoint balls of radius  $\delta$  with centres in  $F$ .

*Proof.* We prove the equivalence of (1) and (4) since these are the definitions we will use most. Let  $N_\delta^1(F)$  be the smallest number of sets of diameter at most  $\delta$  needed to cover  $F$ , and  $N_\delta^2(F)$  be the number of  $\delta$ -mesh cubes that intersect  $F$ . We will show that  $N_\delta^1(F)$  and  $N_\delta^2(F)$  give equivalent definitions of the box-counting dimension. This allows us to use the most suitable definition for the task at hand.

Every  $\delta$ -cube has a diameter of  $\sqrt{n}\delta$ , which means that a  $\delta$ -mesh of cubes that covers  $F$  is a  $\sqrt{n}\delta$ -cover. Hence it follows that

$$N_{\sqrt{n}\delta}^1(F) \leq N_\delta^2(F).$$

Conversely, any set with a diameter at most  $\delta$  can be covered with  $3^n$   $\delta$ -mesh cubes, by placing it in an  $n$ -dimension cube with side  $3\delta$ , comprised of  $3^n$  cubes of side  $\delta$ . Hence, we also obtain the inequality

$$N_\delta^2(F) \leq 3^n N_\delta^1(F).$$

We then combine these two inequalities to obtain

$$N_{\sqrt{n}\delta}^1(F) \leq N_\delta^2(F) \leq 3^n N_\delta^1(F).$$

It is then trivial to see that

$$\frac{\log N_{\sqrt{n}\delta}^1(F)}{-\log(\sqrt{n}\delta) + \log \sqrt{n}} \leq \frac{\log N_\delta^2(F)}{-\log \delta} \leq \frac{\log 3^n + \log N_\delta^1(F)}{-\log \delta}.$$

Taking the limits as  $\delta \rightarrow 0$ , provided they exist, gives the desired result

$$\lim_{\delta \rightarrow 0} \frac{\log N_\delta^1(F)}{-\log \delta} \leq \lim_{\delta \rightarrow 0} \frac{\log N_\delta^2(F)}{-\log \delta} \leq \lim_{\delta \rightarrow 0} \frac{\log N_\delta^1(F)}{-\log \delta}.$$

□

### 6.2.2 Properties of Box-Counting Dimension

The practical nature of the box-counting dimension makes it an attractive definition; however, it should also obey some basic intuitive properties. It follows from the definition that:

- If  $E \subset F$ , then  $\dim_{\mathbb{B}} E \leq \dim_{\mathbb{B}} F$ .
- If  $F \subset \mathbb{R}^n$ , then  $0 \leq \dim_{\mathbb{B}} F \leq n$ .
- If  $F \subset \mathbb{R}^n$  is open, then  $\dim_{\mathbb{B}} F = n$ .
- If  $F$  has finitely many elements, then  $\dim_{\mathbb{B}} F = 0$ .
- If  $F$  is a smooth bounded  $m$ -dimensional surface of  $\mathbb{R}^n$ , then  $\dim_{\mathbb{B}} F = m$ .

These are all properties we would expect an appropriate definition of dimension to possess. However, one problem with the box-counting dimension arises when you consider a set's closure. We mentioned earlier that an equivalent definition can be obtained by taking  $N_\delta(F)$  as the smallest number of closed balls of radius  $\delta$  that cover  $F$ . As a result, the box-counting dimension of a set  $F$  must be equal to that of its closure.

**Proposition 6.1.** For any set  $F \subset \mathbb{R}^n$ ,  $\dim_{\mathbb{B}} F = \dim_{\mathbb{B}} \bar{F}$ .

*Proof.* Using the equivalent definition that  $N_\delta(F)$  is the smallest number of closed balls of radius  $\delta$  that cover  $F$ , it follows that  $N_\delta(F) = N_\delta(\bar{F})$  since any cover of  $F$  is necessarily a cover of  $\bar{F}$ . Hence the result follows. □

This is undesirable as it means that any dense subset of an open region in  $\mathbb{R}^n$  must have box-counting dimension  $n$ . For example, the set of rational numbers is dense in  $\mathbb{R}$  and hence must have box-counting dimension 1, even though it's a countable collection of points. Since a countable set is significantly smaller and less complex than an uncountable set, a good definition of dimension should distinguish between the two.

This also tells us that the box-counting dimension is not countably stable. Any single point  $P$  has box-counting dimension 0 since  $N_\delta(P) = 1$  for all  $\delta > 0$ , and so the dimension of any single point in the set of rational numbers must be 0. Hence it is not always true that

$$\dim_{\text{B}} \bigcup_i^\infty F_i = \sup_i \{\dim_{\text{B}} F_i\}.$$

This contradicts our traditional notions about how dimension should behave. For example, in Euclidean geometry, we see that  $\dim(\mathbb{R}^n \cup \mathbb{R}^m) = \max(n, m)$ . Box-counting dimension fails further in this regard, in that it is not even finitely stable.

In practice, it is often difficult to derive the exact dimension of a set. In such instances it's easier to find lower and upper bounds. To this end, we now give an important result concerning Lipschitz transformations which give a practical method of obtaining bounds for the box-counting dimension of a set.

**Theorem 6.2.** Let  $F \subset \mathbb{R}^n$  and  $f : F \rightarrow \mathbb{R}^m$  a Lipschitz transformation satisfying the Lipschitz condition

$$|f(x) - f(y)| \leq c|x - y| \tag{6.6}$$

for all  $x, y \in F$ , then  $\dim_{\text{B}} f(F) \leq \dim_{\text{B}} F$ .

*Proof.* Taking  $N_\delta(F)$  as the smallest number of sets of diameter at most  $\delta$  that cover  $F$ , let  $\{U_i\}$  be a  $\delta$ -cover of the set  $F$ . Clearly the collection of intersections  $\{U_i \cap F\}$  is also a  $\delta$ -cover, and so by (6.6) it follows that  $\{f(U_i \cap F)\}$  is a  $c\delta$ -cover of  $f(F)$ . This tells us that given any  $\delta$ -covering of

the set  $F$ , we can find a  $c\delta$ -covering of  $f(F)$ , and so  $N_{c\delta}(f(F)) \leq N_\delta(F)$  for all  $\delta > 0$ . Hence

$$\frac{\log N_{c\delta}(f(F))}{-\log(c\delta) + \log c} \leq \frac{\log N_\delta(F)}{-\log \delta}$$

and so the result follows by taking the limit as  $\delta \rightarrow 0$ .  $\square$

**Corollary 6.2.1.** Let  $F \subset \mathbb{R}^n$  be a set and  $f : F \rightarrow \mathbb{R}^m$  a bi-Lipschitz transformation satisfying the bi-Lipschitz condition

$$c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y| \quad (6.7)$$

for all  $x, y \in F$ , then  $\dim_{\mathbb{B}} F = \dim_{\mathbb{B}} f(F)$ .

By Corollary 6.2.1, it follows that for any affine, congruence, or similarity transformation  $f : F \rightarrow \mathbb{R}^m$ , we have  $\dim_{\mathbb{B}} F = \dim_{\mathbb{B}} f(F)$  since all such transformations are bi-Lipschitz. Furthermore, since projection mappings do not increase distances, they are Lipschitz, and hence for any projection we have

$$\dim_{\mathbb{B}} \text{proj } F \leq \dim_{\mathbb{B}} F. \quad (6.8)$$

These results can be used to quickly find lower and upper bounds. Consider for example any continuous curve in  $\mathbb{R}^2$ . Any such curve can be projected onto  $\mathbb{R}$  to obtain a union of open and closed intervals, each with dimension 1. Hence it follows by (6.8) that any continuous curve in  $\mathbb{R}^2$  has a box-counting dimension greater than or equal to 1.

### 6.2.3 Hausdorff Dimension

The Hausdorff dimension has the advantage of being defined for any set, while the box-counting dimension might not be. It is similar to the box-counting dimension in that it is based on  $\delta$ -coverings of a set  $F$ , but instead of defining dimension by a limit, it is defined using *Hausdorff measures*. Hausdorff measures are generalisations of the familiar Lebesgue measure, and play a fundamental role in the study of fractal sets. Therefore, before defining the Hausdorff dimension, we first introduce Hausdorff measures.

Suppose we have a set  $F \subset \mathbb{R}^n$  and some  $s \geq 0$ . Then for each  $\delta > 0$  we define

$$\mathcal{H}_\delta^s(F) = \inf \left\{ \sum_i |U_i|^s : \{U_i\} \text{ is a } \delta\text{-cover of } F \right\} \quad (6.9)$$

so that  $\mathcal{H}_\delta^s(F)$  looks at all  $\delta$ -coverings of  $F$ , and then minimises the sum of the diameters taken to the power of  $s$ . As we decrease  $\delta$ , fewer coverings are available over which to minimise this sum, and so  $\mathcal{H}_\delta^s(F)$  must be non-decreasing as  $\delta \rightarrow 0$ . The  $s$ -dimensional Hausdorff measure is then defined by the limit

$$\mathcal{H}^s(F) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^s(F). \quad (6.10)$$

We now give some useful results for Hausdorff measures in general.

**Proposition 6.2.** If  $E \subset F$ , then  $\mathcal{H}^s(E) \leq \mathcal{H}^s(F)$ .

*Proof.* If  $E \subset F$ , then any  $\delta$ -covering of  $F$  must also be a  $\delta$ -covering of  $E$ . Hence for any  $\delta > 0$  it must be that  $\mathcal{H}_\delta^s(E) \leq \mathcal{H}_\delta^s(F)$ , since the set of permissible  $\delta$ -coverings for  $F$  is contained within that of  $E$ . By taking the limit as  $\delta \rightarrow 0$  the result follows.  $\square$

**Proposition 6.3.** If  $F_1, F_2, \dots$  is a countable collection of sets, then

$$\mathcal{H}^s(\cup_i F_i) \leq \sum_i \mathcal{H}^s(F_i). \quad (6.11)$$

*Proof.* Let  $\{U_j^i\}$  be any  $\delta$ -covering of  $F_i$ , then clearly  $\cup_{i,j} U_j^i$  is a  $\delta$ -covering of  $\cup_i F_i$ . From this it follows that for all  $\delta > 0$ , we have  $\mathcal{H}_\delta^s(\cup_i F_i) \leq \sum_i \mathcal{H}_\delta^s(F_i)$ . Taking the limit as  $\delta \rightarrow 0$  gives the result.  $\square$

**Proposition 6.4.** Let  $F \subset \mathbb{R}^n$  and  $f : F \rightarrow \mathbb{R}^m$  be an  $\alpha$ -Lipschitz transformation. That is, a transformation satisfying

$$|f(x) - f(y)| \leq c|x - y|^\alpha \quad (6.12)$$

for all  $x, y \in F$  and positive constants  $\alpha$  and  $c$ . Then for all  $s \geq 0$  it follows that

$$\mathcal{H}^{s/\alpha}(f(F)) \leq c^{s/\alpha} \mathcal{H}^s(F). \quad (6.13)$$

*Proof.* For any  $\delta$ -covering  $\{U_i\}$  of  $F$ , it follows by (6.12) that  $|f(F \cap U_i)| \leq c|F \cap U_i|^\alpha \leq c|U_i|^\alpha$ . This implies that  $\{f(F \cap U_i)\}$  is a  $c\delta^\alpha$ -covering of  $f(F)$  and, since  $\sum_i |f(F \cap U_i)|^{s/\alpha} \leq c^{s/\alpha} \sum_i |U_i|^s$ , it must be that  $\mathcal{H}_{c\delta^\alpha}^{s/\alpha}(f(F)) \leq c^{s/\alpha} H_\delta^s(F)$ . Taking the limit as  $\delta \rightarrow 0$  gives the result.  $\square$

As a generalisation of Lebesgue measure, Hausdorff measures also possess scaling properties much like that of length, area, and volume. In general, if you double the size of an  $n$ -dimensional cube, its  $n$ -dimensional volume is multiplied by  $2^n$ . Hausdorff measures possess the same scaling property.

**Proposition 6.5.** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a similarity transformation of scale  $\lambda > 0$  so that for all  $x, y \in F$ , we have

$$|f(x) - f(y)| = \lambda|x - y|. \quad (6.14)$$

Then for any  $F \subset \mathbb{R}^n$ ,

$$\mathcal{H}^s(f(F)) = \lambda^s \mathcal{H}^s(F). \quad (6.15)$$

*Proof.* By applying Proposition 6.4 to  $f$  with  $c = \lambda$  and  $\alpha = 1$ , we get  $\mathcal{H}^s(f(F)) \leq \lambda^s \mathcal{H}^s(F)$ . Applying Proposition 6.4 to its inverse  $f^{-1}$  with  $c = \lambda^{-1}$  and  $\alpha = 1$  gives us  $\mathcal{H}^s(F) \leq \lambda^{-s} \mathcal{H}^s(f(F))$ . The result follows from combining these inequalities.  $\square$

Since  $\mathcal{H}_\delta^s(F)$  is non-increasing for all  $\delta < 1$ , it follows that  $\mathcal{H}^s(F)$  is also non-decreasing. Furthermore, we obtain a useful inequality by noting for any  $\delta$ -covering  $\{U_i\}$  of a set  $F$ ,

$$\sum_i |U_i|^t = \sum_i |U_i|^{t-s} |U_i|^s \leq \delta^{t-s} \sum_i |U_i|^s, \quad (6.16)$$

which implies that

$$\mathcal{H}_\delta^t(F) \leq \delta^{t-s} \mathcal{H}_\delta^s(F). \quad (6.17)$$

The inequality given by (6.17) is very important because it means that if  $\mathcal{H}^s(F)$  is finite, then  $\mathcal{H}^t(F) = 0$  for all  $t > s$ . Hence as a function of  $s$ ,  $\mathcal{H}^s(F)$

is infinite and then, at some critical value, jumps down to 0. The critical value at which  $\mathcal{H}^s(F)$  jumps from  $\infty$  to 0 is how we define the Hausdorff dimension.

**Definition 6.3.** Let  $F \subset \mathbb{R}^n$  be a set. Then the Hausdorff dimension of the set  $F$  is defined as

$$\dim_{\text{H}} F = \inf\{s \geq 0 : \mathcal{H}^s(F) = 0\} = \sup\{s \geq 0 : \mathcal{H}^s(F) = \infty\}. \quad (6.18)$$

Unlike with the box-counting dimension, estimating a set's Hausdorff dimension is not straightforward. Furthermore, it is often difficult to derive the Hausdorff dimension of even simple sets.

## 6.2.4 Properties of Hausdorff Dimension

As with the box-counting dimension, it is trivial to see that:

- If  $E \subset F$ , then  $\dim_{\text{H}} E \leq \dim_{\text{H}} F$ .
- If  $F \subset \mathbb{R}^n$ , then  $0 \leq \dim_{\text{H}} F \leq n$ .
- If  $F \subset \mathbb{R}^n$  is open, then  $\dim_{\text{H}} F = n$ .
- If  $F$  has finitely many elements, then  $\dim_{\text{H}} F = 0$ .
- If  $F$  is a smooth bounded  $m$ -dimensional surface of  $\mathbb{R}^n$ , then  $\dim_{\text{H}} F = m$ .

However, the Hausdorff dimension goes further by correcting the two undesirable properties of the box-counting dimension. Firstly, the Hausdorff dimension is countably stable.

**Proposition 6.6.** If  $F_1, F_2, \dots$  is a countable collection of sets, then

$$\dim_{\text{H}} \cup_i F_i = \sup_i \{\dim_{\text{H}} F_i\}. \quad (6.19)$$



*Proof.* Clearly  $\dim_{\mathbb{H}} \cup_i F_i \geq \sup_i \{\dim_{\mathbb{H}} F_i\}$  since  $F_i \subset \cup_i F_i$  for all  $i$ . On the other hand, if  $s > \dim_{\mathbb{H}} F_i$  for all  $i$ , then by definition  $\mathcal{H}^s(F_i) = 0$  for all  $i$ . This implies that  $\mathcal{H}^s(\cup_i F_i) \leq \sum_i \mathcal{H}^s(F_i) = 0$  so that  $\dim_{\mathbb{H}} \cup_i F_i \leq \sup_i \{\dim_{\mathbb{H}} F_i\}$ . The result then follows by combining the inequalities.  $\square$

Secondly, the Hausdorff dimension of any countable set must be 0. This means that the Hausdorff dimension can distinguish between countable dense subsets and the uncountable sets they're contained in.

**Proposition 6.7.** If  $F$  is a countable set, then  $\dim_{\mathbb{H}} F = 0$ .

*Proof.* For any single point  $P$ ,  $\dim_{\mathbb{H}} P = 0$ . Thus by applying Proposition 6.6 to a countable set, the result follows.  $\square$

Although Hausdorff dimension has properties more resembling what we would expect of dimension, it comes at the cost of practicality. Deriving or estimating a set's Hausdorff dimension is generally much more difficult than the box-counting dimension. Results for quickly finding lower and upper bounds are therefore of great interest.

**Theorem 6.3.** Let  $F \subset \mathbb{R}^n$  and  $f : F \rightarrow \mathbb{R}^m$  a transformation satisfying the Hölder condition

$$|f(x) - f(y)| \leq c|x - y|^\alpha \quad (6.20)$$

for all  $x, y \in F$ . Then it follows that

$$\dim_{\mathbb{H}} f(F) \leq \alpha^{-1} \dim_{\mathbb{H}} F. \quad (6.21)$$

*Proof.* Suppose that for some  $s \geq 0$ , we have  $s > \dim_{\mathbb{H}} F$ . Then by applying Proposition 6.4 it follows that  $\mathcal{H}^{s/\alpha}(f(F)) \leq c^{s/\alpha} \mathcal{H}^s(F) = 0$ . This means that the Hausdorff dimension of  $f(F)$  must be no greater than  $s/\alpha$  for all  $s > \dim_{\mathbb{H}} F$  by definition of Hausdorff dimension. Hence the result follows.  $\square$

**Corollary 6.3.1.** Let  $F \subset \mathbb{R}^n$  and  $f : F \rightarrow \mathbb{R}^m$  a bi-Lipschitz transformation satisfying

$$c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y| \quad (6.22)$$

for all  $x, y \in F$ . Then it follows that

$$\dim_{\mathbb{H}} f(F) = \dim_{\mathbb{H}} F. \quad (6.23)$$

As with the box-counting dimension, Theorem 6.3 implies that Hausdorff dimension is invariant under congruence, similarity, or affine transformations. Further, we also see that for any projection transformation,

$$\dim_{\mathbb{H}} \text{proj } F \leq \dim_{\mathbb{H}} F.$$

Given how similar the properties of box-counting and Hausdorff dimensions are, it is natural to ask how they themselves are related.

**Proposition 6.8.** For any non-empty and bounded set  $F \subset \mathbb{R}^n$ ,

$$\dim_{\mathbb{H}} F \leq \dim_{\mathbb{B}} F. \quad (6.24)$$

*Proof.* Take  $N_{\delta}(F)$  as the smallest number of sets of diameter at most  $\delta$  needed to cover  $F$ , and suppose that for some  $s \geq 0$  we have  $\mathcal{H}^s(F) > 1$ , which is true for all  $s < \dim_{\mathbb{H}} F$ . Then this implies that for all sufficiently small  $\delta$ ,  $\mathcal{H}_{\delta}^s(F) > 1$ . Since  $\mathcal{H}_{\delta}^s(F)$  is defined by minimising the sum of diameters taken to the power of  $s$ , it follows that

$$1 < \mathcal{H}_{\delta}^s(F) \leq N_{\delta}(F)\delta^s,$$

for sufficiently small  $\delta$ . By taking taking logarithms we obtain

$$0 < \log N_{\delta}(F) + s \log \delta$$

and hence

$$s < \frac{\log N_{\delta}(F)}{-\log \delta}.$$

By taking the limit as  $\delta \rightarrow 0$  we see that  $s \leq \dim_{\mathbb{B}} F$ . Since this holds for all  $s < \dim_{\mathbb{H}} F$ , it follows that  $\dim_{\mathbb{H}} F \leq \dim_{\mathbb{B}} F$ .  $\square$

Thus the box-counting dimension itself provides us with an upper bound for a set's Hausdorff dimension. We now give two final results for obtaining upper and lower bounds of dimension.

**Proposition 6.9.** Let  $F$  be a set which can be covered by  $n_k$  sets of diameter not greater than  $\delta_k$  such that  $\delta_k \rightarrow 0$  as  $k \rightarrow \infty$ . Then an upper bound for the Hausdorff and box-counting dimensions of  $F$  is

$$\lim_{k \rightarrow \infty} \frac{\log n_k}{-\log \delta_k},$$

provided the limit exists.

*Proof.* If  $N_\delta(F)$  is taken as the smallest number of sets of diameter at most  $\delta$  that cover  $F$ , then by definition it follows that  $N_\delta(F) \leq n_k$ . By taking the limit as  $\delta_k \rightarrow 0$  it immediately follows that

$$\dim_{\text{B}} F = \lim_{\delta_k \rightarrow 0} \frac{N_{\delta_k}(F)}{-\log \delta_k} \leq \lim_{k \rightarrow \infty} \frac{\log n_k}{-\log \delta_k}.$$

It then follows by Proposition 6.8 that it is also an upper bound for the Hausdorff dimension of  $F$ .  $\square$

The next result relies on the idea of a *mass distribution*, a measure defined on a bounded set that represents the "mass" concentrated at each point. In probabilistic terminology, a mass distribution is simply a finite measure.

**Definition 6.4.** A mass distribution  $\mu$  on a bounded set  $F$  is a  $\sigma$  additive measure on  $F$  such that  $0 < \mu(F) < \infty$ .

**Proposition 6.10.** Let  $\mu$  be a mass distribution,  $F \subset \mathbb{R}$  a Borel set, and  $0 < c < \infty$  a constant. If

$$\lim_{r \rightarrow 0} \mu(B_r(x))/r^s < c \text{ for all } x \in F, \quad (6.25)$$

then  $\mathcal{H}^s(F) \geq \mu(F)/c$ .

*Proof.* For all  $\delta > 0$  we define a set  $F_\delta$  as

$$F_\delta = \{x \in F : \mu(B_r(x)) < cr^s \text{ for all } 0 < r \leq \delta\}.$$

Take any  $\delta$ -cover of  $F$ , which is also a  $\delta$ -cover of  $F_\delta$ . Then for each  $U_i$  that contains some point  $x$  belonging to  $F_\delta$ , the ball  $B$  centred at  $x$  with radius  $|U_i|$  contains  $U_i$ . Then using the inequality in the definition of  $F_\delta$  it follows that

$$\mu(U_i) \leq \mu(B) < c|U_i|^s$$

since  $\{U_i\}$  is a  $\delta$ -cover and hence  $|U_i| \leq \delta$ . Hence

$$\mu(F_i) \leq \sum_i \mu(U_i) \leq c \sum_i |U_i|^s. \quad (6.26)$$

Since this holds for any  $\delta$ -cover of  $F$ , this implies  $\mu(F_\delta) \leq c\mathcal{H}_\delta^s(F)$ . Taking the limit as  $\delta \rightarrow 0$  then gives the final result.  $\square$

## 6.3 Calculating and Estimating Dimension

### 6.3.1 Box-counting

The definition of box-counting dimension allows us to estimate any set's dimension quite easily. Since it is defined by the limit

$$\dim_{\text{B}} F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta},$$

we can estimate a set's box-counting dimension by the slope of the graph of  $\log N_\delta(F)$  plotted against  $-\log \delta$ . Hence, if we take  $N_\delta(F)$  as the smallest number of  $\delta$ -mesh squares that intersect  $F$ , then we estimate the box-counting dimension of a set  $F$  in the following way:

1. Construct a  $\delta$ -mesh for some  $\delta > 0$ .
2. Count the number  $N_\delta(F)$  of  $\delta$ -mesh squares that intersect  $F$ .
3. Repeat steps 1 and 2 for a range of values of  $\delta$ .
4. Plot  $\log N_\delta(F)$  against  $-\log \delta$ .
5. Estimate the box-counting dimension by the slope of the graph obtained in step 4.

### 6.3.2 Potential Theoretic Methods

Later in this chapter we will derive the dimension of a Brownian motion trajectory. To do this, we need to introduce *s-potential* and *s-energy*, ideas from physics which have found very practical use in fractal geometry.

**Definition 6.5.** Let  $\mu$  be a mass distribution in  $\mathbb{R}^n$ . Then the *s-potential* at a point  $x \in \mathbb{R}^n$  resulting from the mass distribution  $\mu$  is defined as

$$\phi_s(x) = \int \frac{d\mu(y)}{|x - y|^s}. \quad (6.27)$$

**Definition 6.6.** Let  $\mu$  be a mass distribution in  $\mathbb{R}^n$ . Then the *s-energy* of  $\mu$  is defined as

$$I_s(\mu) = \int \phi_s(x) d\mu(x). \quad (6.28)$$

It is generally easy to find upper bounds for the Hausdorff dimension of a set by using Proposition 6.9. Finding lower bounds is much more difficult. Fortunately the ideas of potential and energy provide a method of doing so.

**Theorem 6.4.** Let  $F$  be any set in  $\mathbb{R}^n$ . If there exists a mass distribution  $\mu$  on  $F$  such that  $I_s(\mu) < \infty$ , then  $\mathcal{H}^s(F) = \infty$  and  $\dim_{\text{H}} F \geq s$ .

*Proof.* Suppose there exists some mass distribution  $\mu$  with support contained in  $F$  such that  $I_s(\mu) < \infty$ . Under this assumption, we note that for any  $x \in F$  it must be that  $\mu(\{x\}) = 0$ , since otherwise would imply infinite *s-energy*. Then define the set  $F_1$  as

$$F_1 = \{x \in F : \lim_{r \rightarrow 0} \mu(B_r(x))/r^s > 0\}. \quad (6.29)$$

This definition implies that for any  $x \in F_1$ , we can find some  $\epsilon > 0$  along with a strictly decreasing sequence of positive numbers  $r_1, r_2, \dots$  such that  $\mu(B_{r_i}(x)) \geq \epsilon r_i^s$  for all  $i$ . Then, for  $q_i \in (0, r_i)$ , we consider the annulus  $A_i = B_{r_i}(x) \setminus B_{q_i}(x)$ . The continuity of  $\mu$ , along with the fact  $\mu(\{x\}) = 0$ , implies we can take  $q_i$  small enough such that  $\mu(A_i) \geq \frac{1}{4}\epsilon r_i^s$  for all  $i$ . Without loss

of generality we assume that  $q_i > r_{i+1}$  so that each  $A_i$  is disjoint from one another. Then for all  $x$ , it follows that

$$\begin{aligned} \phi_s(x) &= \int \frac{d\mu(y)}{|x-y|^s} \\ &\geq \sum_i \int_{A_i} \frac{d\mu(y)}{|x-y|^s} \end{aligned} \quad (6.30)$$

$$\geq \sum_i \frac{1}{4} \epsilon r_i^s r_i^{-s} \quad (6.31)$$

$$= \infty \quad (6.32)$$

where (6.31) follows from the fact that  $|x-y|^s \geq r_i^s$  on  $A_i$ , since  $A_i$  is contained within a ball of radius  $r_i$ . Since  $I_s(\mu)$  is assumed finite, this implies that  $\mu(F_1) = 0$ . Since for all  $x \in F \setminus F_1$  we have  $\lim_{r \rightarrow 0} \mu(B_r(x))/r^s = 0$ , then Proposition 6.10 gives us

$$\mathcal{H}^s(F) \geq H^s(F \setminus F_1) \geq \mu(F \setminus F_1)/c \geq (\mu(F) - \mu(F_1))/c = \mu(F)/c$$

for all  $c > 0$ . Since  $c$  can be made arbitrarily small, this implies  $\mathcal{H}^s(F) = \infty$  and so  $\dim_{\text{H}} F \geq s$ .  $\square$

### 6.3.3 Rescaling of Self-similar Fractals

Self-similarity is a common property of many fractals, though it may manifest in varying forms. Many of the most famous fractal sets, such as the Middle Third Cantor set considered next, are comprised of disjoint rescaled versions of themselves. By this we mean that a fractal  $F$  can be expressed as a union of disjoint sets  $F_i$ ,

$$F = \cup_i^n F_i,$$

where each  $F_i$  is  $F$  rescaled by some common scale factor  $\lambda$ . Calculating the set's Hausdorff dimension then becomes straightforward under the assumption that

$$0 < \mathcal{H}^s(F) < \infty$$

for some  $s$ . We can use Proposition 6.5 to obtain

$$\begin{aligned}\mathcal{H}^s(F) &= \sum_{i=1}^n \mathcal{H}^s(F_i) \\ &= \sum_{i=1}^n \lambda^s \mathcal{H}^s(F).\end{aligned}\tag{6.33}$$

Which, after dividing both sides by  $\mathcal{H}^s(F)$ , can be solved to obtain

$$s = \frac{\log n}{-\log \lambda}.\tag{6.34}$$

Since the  $s$ -dimensional Hausdorff measure of a set can only be both non-zero and finite if it has Hausdorff dimension  $s$ , this implies its Hausdorff dimension is given by (6.34).

## 6.4 Example: The Middle Third Cantor Set

The middle third Cantor set is a fractal that arises from a very simple recursive construction. We start with the interval  $[0, 1]$  which we denote by  $F_0$ . We then obtain  $F_1$  by removing the middle third of  $F_0$ , giving us

$$F_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Similarly,  $F_2$  is obtained by removing the middle third of the remaining segments,

$$F_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

In general,  $F_k$  is obtained by removing the middle third of all remaining segments of  $F_{k-1}$ , as demonstrated in Figure 6.12. The middle third Cantor set is then defined by the limit

$$F = \lim_{k \rightarrow \infty} F_k.\tag{6.35}$$

Intuitively, it is tempting to think such a set would be empty. However not only is the set non-empty, it is in fact uncountably infinite. This can be seen by noting that at any stage  $F_k$ , the end points of any of the  $2^k$  intervals of  $F_k$  can never be removed, since they will always remain end points at every subsequent iteration.

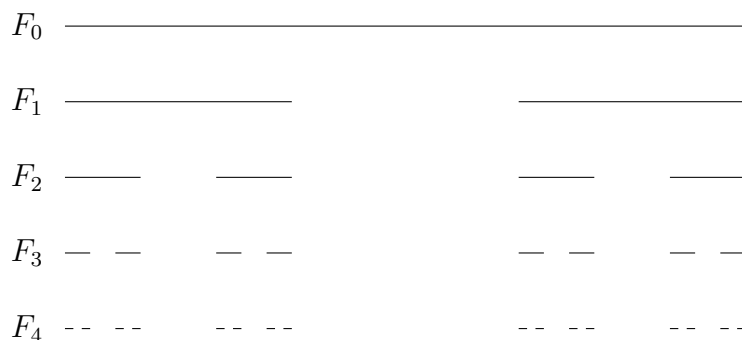


Figure 6.1: Construction of the middle third Cantor set.

We would expect the dimension of the middle third Cantor set to be greater than 0 since it uncountable, however it must also be less than or equal to 1 since it a subset of the real line. As it turns out, the box-counting and Hausdorff dimensions of the middle third Cantor set are equal, and can be derived quite easily.

To calculate its box-counting dimension, first consider the  $k$ th stage in its construction  $F_k$ . It contains a total of  $2^k$  intervals each of length  $3^{-k}$ . Hence for any  $\delta$  such that  $3^{-k} < \delta \leq 3^{-k+1}$ , the intervals of  $F_k$  can be used as a  $\delta$ -cover of the middle third Cantor set, and hence  $N_\delta(F) \leq 2^k$ , where we are taking  $N_\delta(F)$  as the smallest number of sets of diameter at most  $\delta$  needed to cover  $F$ . Hence it follows that

$$\dim_{\text{B}} F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \leq \lim_{k \rightarrow \infty} \frac{\log 2^k}{-\log 3^{-k+1}} = \frac{\log 2}{\log 3}.$$

Conversely, if we take  $\delta$  such that  $3^{-k-1} \leq \delta < 3^{-k}$  then any interval of length  $\delta$  cannot intersect more than 1 interval of  $F_k$ , since the gaps between them are at least of length  $3^{-k}$ . Since all of the  $2^k$  intervals of  $F_k$  contain



points in  $F$ , this means that  $N_\delta(F) \geq 2^k$  and hence

$$\dim_{\text{B}} F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \geq \lim_{k \rightarrow \infty} \frac{\log 2^k}{-\log 3^{-k-1}} = \frac{\log 2}{\log 3}.$$

By combining these inequalities, we see that the middle third Cantor set must have box-counting dimension  $\log 2 / \log 3$ .

To derive its Hausdorff dimension, we note that the middle Cantor set can be decomposed into two disjoint parts  $F \cap [0, \frac{1}{3}]$  and  $F \cap [\frac{2}{3}, 1]$ , each of which is simply  $F$  rescaled by a factor of  $\frac{1}{3}$ . Using equation (6.34) then gives  $\dim_{\text{H}} F = \log 2 / \log 3$ .

## 6.5 Fractal Graphs

As the subject of this thesis is the exchange rate process  $X(t)$ , it is worthwhile reviewing results specific to fractal graphs. Such fractals often arise when recording data such as financial prices, temperature, population levels, and so on. A graph  $f : [a, b] \rightarrow \mathbb{R}$  is a subset of  $\mathbb{R}^2$  which takes the form

$$\{(x, f(x)) : a \leq x \leq b\}.$$

We begin by defining the *maximum range* of a function  $f$  over the interval  $[t_1, t_2]$  as

$$R_f[t_1, t_2] = \sup_{t_1 \leq t, u \leq t_2} |f(t) - f(u)|. \quad (6.36)$$

We can then derive a useful inequality for any continuous function  $f$  in terms of its maximum range  $R_f$ .

**Proposition 6.11.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function and choose any  $\delta \in (a, b)$ . Let  $m$  be the smallest integer greater than or equal to  $(b-a)/\delta$  and  $N_\delta(f)$  the smallest number of  $\delta$ -mesh squares that intersect the graph  $f$ . Then the following inequality holds

$$\delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta] \leq N_\delta(f) \leq 2m + \delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta]. \quad (6.37)$$

*Proof.* On any interval  $[i\delta, (i+1)\delta]$ , no more than  $2 + R_f[i\delta, (i+1)\delta]/\delta$  mesh squares can intersect the graph  $f$ , since  $R_f$  is its maximum vertical change over that interval. Since this is true for all  $m$  intervals, summing them up gives us the upper bound. Conversely,  $f$  must intersect at least  $R_f[i\delta, (i+1)\delta]/\delta$  mesh squares on this interval, as doing otherwise implies having a smaller maximum range. By again summing this up across all intervals we obtain the lower bound, and so the result follows.  $\square$

This inequality allows us to prove a useful result that provides an upper bound for the box-counting dimension of any continuous function  $f$  satisfying a certain Hölder condition. Furthermore, by Proposition 6.8 it then also provides an upper bound for the Hausdorff dimension of  $f$ .

**Theorem 6.5.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function such that

$$|f(x) - f(y)| \leq c|x - y|^{2-s} \quad (6.38)$$

for all  $x, y \in [a, b]$ , where  $c$  is a positive constant and  $s \in [1, 2]$ . Then it follows that  $\dim_B f \leq s$ .

*Proof.* On any interval  $[t_1, t_2] \subset [a, b]$ , the Hölder condition (6.38) tells us that  $R_f[t_1, t_2] \leq c|t_1 - t_2|^{2-s}$ . It follows from Proposition 6.11 that

$$\begin{aligned} N_\delta(f) &\leq 2m + \delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta] \\ &\leq 2m + \delta^{-1} mc\delta^{2-s} \\ &\leq 2(1 + \delta^{-1}) + \delta^{-1}(1 + \delta^{-1})c\delta^{2-s} \\ &\leq c_1\delta^{-s} \end{aligned}$$

where  $c_1$  is some constant independent of  $\delta$ , and we have used the fact that  $m \leq (1 + \delta^{-1})$ . The result follows by then applying Proposition 6.9 since this gives

$$\dim_B f \leq \lim_{\delta \rightarrow 0} \frac{\log(c_1\delta^{-s})}{-\log \delta} = s.$$

$\square$

The usefulness of this will become clear later when we derive the box-counting dimension of a Brownian motion path. A similarly useful result provides a lower bound for the box-counting dimension of a graph  $f$ , although not for its Hausdorff dimension.

**Theorem 6.6.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function such that there exists positive constants  $c$  and  $\delta_0$  with the property that for all  $t \in [a, b]$  and  $\delta \in (0, \delta_0]$ , there exists some  $u$  such that  $|t - u| \leq \delta$  and

$$|f(t) - f(u)| \geq c\delta^{2-s}. \quad (6.39)$$

Then it follows that  $\dim_{\mathbb{B}} f \geq s$ .

*Proof.* On any interval  $[t_1, t_2] \subset [a, b]$ , the condition (6.39) implies  $R_f[t_1, t_2] \geq c|t_1 - t_2|^{2-s}$ . Proposition 6.11 then gives us

$$\begin{aligned} N_\delta(f) &\geq \delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta] \\ &\geq \delta^{-1} m c \delta^{2-s} \\ &\geq \delta^{-1} \delta^{-1} c \delta^{2-s} \\ &= c \delta^{-s} \end{aligned}$$

where we use the fact that  $m \geq \delta^{-s}$ . Hence

$$\dim_{\mathbb{B}} f = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(f)}{-\log \delta} \geq \lim_{\delta \rightarrow 0} \frac{\log(c\delta^{-s})}{-\log \delta} = s.$$

□

The final property of fractal graphs we discuss is that of autocorrelation. The autocorrelation function is often used to investigate the properties of a stochastic process, and so it would be beneficial if it could also be used in the study of fractal graphs. As it turns out, the autocorrelation functions of fractal graphs often display power law behaviour of some form.

To see why this is, suppose we have a continuous and bounded function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , and we examine this function on the interval  $[-T, T]$ . We then define the average of the function  $f$  by

$$\bar{f} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t) dt. \quad (6.40)$$

The autocorrelation function of  $f$  is then defined as

$$C(h) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (f(t+h) - \bar{f})(f(t) - \bar{f}) dt \quad (6.41)$$

$$= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t+h)f(t) dt - (\bar{f})^2. \quad (6.42)$$

Then, by substituting

$$f(t+h)f(t) = \frac{1}{2} (f(t+h)^2 + f(t)^2 - (f(t+h) - f(t))^2)$$

into (6.42), we can rewrite the autocorrelation function of  $f$  as

$$\begin{aligned} C(h) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \frac{1}{2} (f(t+h)^2 + f(t)^2 - (f(t+h) - f(t))^2) dt - (\bar{f})^2 \\ &= \bar{f}^2 - (\bar{f})^2 - \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (f(t+h) - f(t))^2 dt \\ &= C(0) - \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (f(t+h) - f(t))^2 dt, \end{aligned} \quad (6.43)$$

where we have used the fact that

$$\bar{f}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t)^2 dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t+h)^2 dt.$$

The autocorrelation function's significance is then made clear in the form (6.43) by considering Theorems 6.5 and 6.6. We know that if a function  $f$  satisfies both conditions (6.38) and (6.39), then it must have box-counting dimension  $s$ . Additionally, if both conditions are satisfied, then we can find constants  $c_1$  and  $c_2$  such that

$$c_1 h^{2-s} \leq |f(t+h) - f(t)| \leq c_2 h^{2-s},$$

which, when squared and integrated from  $-T$  to  $T$ , gives us

$$c_3 h^{4-2s} \leq \frac{1}{2T} \int_{-T}^T (f(t+h) - f(t))^2 dt \leq c_4 h^{4-2s}, \quad (6.44)$$

provided conditions (6.38) and (6.39) are satisfied in a reasonable way. By combining (6.44) and (6.43) we then see that if the autocorrelation function satisfies a powerlaw of the form

$$C(0) - C(h) \approx ch^{4-2s} \quad (6.45)$$

then it is likely to have box-counting dimension  $s$ . Based off this result, we have an alternative method for estimating the box-counting dimension of a graph  $f$  based on its autocorrelation function. If the power law (6.45) holds, then this implies

$$s \approx 2 - \frac{\log(C(0) - C(h)) - \log c}{2 \log h},$$

which, if not for the unknown constant  $c$ , could be used to estimate  $s$ . To overcome this, we might hope that as  $h \rightarrow 0$  we see that  $s \rightarrow \dim_{\text{B}} f$  so that we can then estimate the box-counting dimension by

$$\dim_{\text{B}} f = \lim_{h \rightarrow 0} \left( 2 - \frac{\log(C(0) - C(h)) - \log c}{2 \log h} \right) \quad (6.46)$$

$$= \lim_{h \rightarrow 0} \left( 2 - \frac{\log(C(0) - C(h))}{2 \log h} \right) \quad (6.47)$$

which then frees us of needing to estimate the constant  $c$ .

## 6.6 Brownian Motion as a Fractal

The importance of fractal geometry in financial mathematics is due to the fundamental role of Brownian motion. The fractal nature of Brownian motion can be seen by considering the length of its paths. Given a Brownian motion on  $[0, 1]$ , the length of its trajectory is almost surely infinite. At the

same time, the area of its trajectory is certainly 0. This is similar to the construction of Hausdorff dimension in Section 6.2.3, where

$$\mathcal{H}^s(F) = \begin{cases} \infty, & \text{if } s < \dim_{\text{H}} F \\ 0, & \text{if } s > \dim_{\text{H}} F \end{cases} \quad (6.48)$$

suggesting the Hausdorff dimension of a Brownian motion trajectory is somewhere between 0 and 1. In order to derive the Hausdorff and box-counting dimensions of a Brownian motion trajectory, we use the following result.

**Proposition 6.12.** Let  $B(t)$  be a Brownian motion on some interval of time  $[0, 1]$ . Then for any  $\lambda \in (0, \frac{1}{2})$ , there is a random constant  $B$  such that

$$|B(t+h) - B(t)| \leq B|h|^\lambda, \quad (6.49)$$

with probability 1.

*Proof.* By symmetry of the normal distribution we have that for any  $h > 0$ ,

$$\begin{aligned} P(|B(t+h) - B(t)| > h^\lambda) &= \frac{2}{\sqrt{2\pi h}} \int_{h^\lambda}^{\infty} e^{-\frac{x^2}{2h}} dx \\ &= c_1 \int_{h^{\lambda-1/2}}^{\infty} e^{-\frac{y^2}{2}} dy \\ &\leq c_2 \int_{h^{\lambda-1/2}}^{\infty} e^{-y} dy \\ &\leq c_3 h^2 \end{aligned} \quad (6.50)$$

where we use the substitution  $y = xh^{-1/2}$ , and the constants  $c_1, c_2, c_3$  do not depend on  $h$  or  $y$ . With this inequality in mind, we then consider intervals of the form  $[(m-1)2^{-j}, m2^{-j}]$ . The inequality (6.50) tells us that for any positive integer  $k$  we have

$$\begin{aligned} P(|B(m2^{-j}) - B((m-1)2^{-j})| > 2^{-j\lambda} \text{ for some } j \geq k) &\leq c_3 \sum_{j=k}^{\infty} 2^j 2^{-2j} \\ &= c_3 2^{-k+1} \end{aligned} \quad (6.51)$$

so that there is, with probability 1, some integer  $K$  such that for all  $j > K$  we have

$$|B(m2^{-j}) - B((m-1)2^{-j})| \leq 2^{-j\lambda}. \quad (6.52)$$

If  $h < 2^{-k}$  then the interval  $[t, t+h]$  can be expressed as a countable union of intervals of the form  $[(m-1)2^{-j}, m2^{-j}]$  such that no more than two intervals share the same length. Taking  $k$  as the least integer satisfying  $2^{-k} \leq h$ ,

$$|B(t+h) - B(t)| \leq 2 \sum_{j=k}^{\infty} 2^{-j\lambda} = \frac{2^{-k\lambda+1}}{1-2^{-\lambda}} \leq \frac{2h^\lambda}{1-2^{-\lambda}}. \quad (6.53)$$

Since Brownian motion paths are almost surely continuous, and hence bounded, inequality (6.52) holds for some value  $b \neq 2^{-k}$  when  $2^{-k} \leq h \leq 1$ , and hence the result follows.  $\square$

We can now derive the box-counting and Hausdorff dimensions of a Brownian motion trajectory.

**Theorem 6.7.** Let  $B(t)$  be a Brownian motion trajectory on the interval  $[0, 1]$ . Then, with probability 1,

$$\dim_{\text{B}} B(t) = \dim_{\text{H}} B(t) = \frac{3}{2}. \quad (6.54)$$

*Proof.* By Proposition 6.12, any Brownian motion path satisfies a Hölder condition of the form

$$|B(t+h) - B(t)| \leq B|h|^\lambda, \quad (6.55)$$

and so by applying Theorem 6.5 we see that the Hausdorff and box-counting dimension of any Brownian motion path cannot exceed  $\frac{3}{2}$ .

Finding a lower bound is more difficult, and requires the use of Theorem 6.4. For any two points on the Brownian motion path, say  $(t, B(t))$  and  $(t+h, B(t+h))$ , we consider the following probability

$$p(x) = P(|B(t+h) - B(t)| \leq x) = \frac{1}{\sqrt{2\pi h}} \int_0^x e^{-\frac{y^2}{2h}} dy.$$

For any  $s \in (1, 2)$ ,

$$\begin{aligned}
\mathbf{E}((|B(t+h) - B(t)|^2 + h^2)^{-s/2}) &= \frac{1}{\sqrt{2\pi h}} \int_0^\infty (x^2 + h^2)^{-s/2} e^{-\frac{x^2}{2h}} dx \\
&\leq ch^{-1/2} \int_0^\infty (x^2 + h^2)^{-s/2} dx \\
&\leq ch^{-1/2} \left( \int_0^h h^{-s} dx + \int_h^\infty x^{-s} dx \right)
\end{aligned} \tag{6.56}$$

where we use the fact that if  $a < b$ , then

$$(a^2 + b^2)^{-s/2} \leq (2b^2)^{-s/2} \leq b^{-s}. \tag{6.57}$$

We then define a mass distribution on the Brownian motion path  $B(t)$  as

$$\mu(A) = \mathcal{L}\{t : (t, B(t)) \in A\} \tag{6.58}$$

where  $\mathcal{L}$  is the usual Lebesgue measure. Using Pythagoras' Theorem for the distance between two points  $x$  and  $y$  on a graph, we then obtain

$$\begin{aligned}
\mathbf{E} \left[ \int \int \frac{d\mu(x)d\mu(y)}{|x-y|^s} \right] &= \int_0^1 \int_0^1 \mathbf{E} [(|B(t) - B(u)|^s + |t-u|^s)^{-s/2}] dt du \\
&\leq \int_0^1 \int_0^1 c_1 |t-u|^{1/2-s} dt du \\
&< \infty
\end{aligned}$$

provided that  $s < 1\frac{1}{2}$ . Hence the mass distribution  $\mu$  has finite  $s$ -energy with probability 1, and hence by Theorem 6.4 we see that  $\dim_{\text{H}} B(t) \geq 1\frac{1}{2}$ . Combining the upper and lower bounds gives the result.  $\square$

Thus if a stochastic process is driven by a Brownian motion, then it should have dimension close to  $\frac{3}{2}$ .

### 6.6.1 Fractal Brownian motion

Fractal Brownian motion is a generalisation of standard Brownian motion, where the assumption of independent increments is dropped.



**Definition 6.7.** A stochastic process  $B(t)$ , with  $0 \leq t \leq T$ , is an  $\alpha$ -fractal Brownian motion with  $0 < \alpha < 1$  if the following conditions hold:

1. With probability 1, the process is continuous with  $B(0) = 0$ .
2.  $B(t)$  is a Gaussian process with mean 0 and covariance function

$$\text{cov}(B(s), B(t)) = \frac{1}{2}(|t|^{2\alpha} + |s|^{2\alpha} + |t - s|^{2\alpha}). \quad (6.59)$$

From this it can immediately be seen that the correlation between disjoint increments is positive when  $\alpha > \frac{1}{2}$  and negative when  $\alpha < \frac{1}{2}$ . When  $\alpha = \frac{1}{2}$  the correlation is 0 and we get standard Brownian motion. Hence the process is smoother for larger  $\alpha$ , and more irregular for smaller  $\alpha$ . This is reflected in the box-counting and Hausdorff dimensions of an  $\alpha$ -fractal Brownian motion path, which is almost surely  $2 - \alpha$ .

### 6.6.2 Estimation of a Fractal Brownian Motion

If  $B(t)$  is an  $\alpha$ -fractal Brownian motion, then there are a number of methods for estimating the parameter  $\alpha$ . For example, it is straightforward to derive the maximum likelihood estimator of  $\alpha$ . With that said, we now introduce a new method of estimation and examine its efficacy on generated fractal Brownian motions. It is well known that for any  $a > 0$ , an  $\alpha$ -fractal Brownian motion obeys a scaling law of the form

$$B(t) \sim a^{-\alpha} B(at), \quad (6.60)$$

where  $\sim$  denotes similarity in distribution. We hope to use this property alone in order to estimate  $\alpha$ . If we assume that

$$\mathbf{E}[B(t)^2] < \infty \quad (6.61)$$

for all  $t$ , then a straightforward method of estimation exists. Let us fix some  $a > 0$ , noting the exact value we choose doesn't concern us. Then,

given an  $\alpha$ -fractal Brownian motion  $B(t)$ , we consider a new process of the form

$$B_\beta(t) = a^{-\beta} B(at) \quad (6.62)$$

for some  $\beta \in (0, 1)$ . From this, we construct the differences of the squared increments

$$\gamma_\beta(t) = (B(t + \Delta) - B(t))^2 - (B_\beta(t + \Delta) - B_\beta(t))^2. \quad (6.63)$$

We can then estimate  $\alpha$  using the following result concerning the expected value of  $\gamma_\beta(t)$ .

**Proposition 6.13.** Let  $\gamma_\beta(t)$  be defined by (6.63) for some stochastic process  $B(t)$  obeying (6.61) and (6.60). Then it follows that

$$\mathbf{E}[\gamma_\beta(t)] = \begin{cases} < 0, & \text{if } \beta < \alpha \\ 0, & \text{if } \beta = \alpha \\ > 0 & \text{if } \beta > \alpha \end{cases} \quad (6.64)$$

for all  $\beta \in (0, 1)$ .

*Proof.* By definition of  $\gamma_\beta(t)$ , it follows that

$$\begin{aligned} \mathbf{E}[\gamma_\beta(t)] &= \mathbf{E}[\Delta B(t)^2 - \Delta B_\beta(t)^2] \\ &= \mathbf{E}[\Delta B(t)^2] - a^{-2\beta} \mathbf{E}[(B(a(t + \Delta)) - B(at))^2] \\ &= \mathbf{E}[\Delta B(t)^2] - \frac{a^{-2\beta}}{a^{-2\alpha}} \mathbf{E}[a^{-2\alpha}(B(a(t + \Delta)) - B(at))^2] \\ &= \mathbf{E}[\Delta B(t)^2] - (a^2)^{\alpha-\beta} \mathbf{E}[\Delta B_\alpha(t)^2] \\ &= \mathbf{E}[\Delta B(t)^2] - (a^2)^{\alpha-\beta} \mathbf{E}[\Delta B(t)^2] \\ &= \mathbf{E}[\Delta B(t)^2](1 - (a^2)^{\alpha-\beta}) \end{aligned}$$

proving the result, since  $\mathbf{E}[\Delta B(t)^2] > 0$ .  $\square$

Thus, the parameter  $\alpha$  is the turning point at which the expected value of  $\gamma_\beta(t)$  goes from being negative to positive. To use this as a basis of estimation, we consider partitions of the form

$$0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < 1$$

for some fixed  $m > 0$ . For each  $\alpha_i$ , we then calculate the average of  $\gamma_{\alpha_i}(t)$ ,

$$\bar{\gamma}_{\alpha_i} = \frac{1}{n} \sum_{j=1}^n \gamma_{\alpha_i}(j\Delta),$$

where  $\Delta = T/n$ . Proposition 6.13 tells us that for  $\alpha_i < \alpha$  these will, on average, be negative. While for  $\alpha_i > \alpha$  they will, on average, be positive. Our estimate of  $\alpha$  then need only be a statistic which reflects this turning point. To this end, we consider the partial sum process

$$S_n(\alpha_i) = \sum_{\alpha_j \leq \alpha_i} \bar{\gamma}_{\alpha_j}. \quad (6.65)$$

For  $\alpha_i < \alpha$ , these summands will generally be negative and hence the process will be decreasing. Once  $\alpha_i > \alpha$  however, these summands will generally be positive, resulting in a now increasing process. Hence our estimate  $\alpha$  is the point at which this partial sum process achieves a minimum:

$$\hat{\alpha} = \arg \min S_n(\alpha_i). \quad (6.66)$$

### 6.6.3 A Preliminary Analysis of the Estimator

We now briefly examine the efficacy of our estimator using generated fractal Brownian motions. The following is not meant to be an in depth analysis, but rather a short demonstration that the underlying idea is correct.

We start by studying the distributions of our estimator  $\hat{\alpha}$  for various  $\alpha$ . We do this by generating a number of fractal Brownian motions using the *somebm* package in R. We obtain an estimate for each generated fractal Brownian motion, providing us with an empirical distribution to examine. For each value of  $\alpha$ , we generate 1,000 discrete time fractal Brownian motions of length 1,000. When applying our method of estimation, we take  $a = 2$  and use the partition

$$\alpha_1 = 0.001 < 0.002 < \dots < 0.999 = \alpha_m.$$

The resulting density histograms of our estimates  $\hat{\alpha}$  are given in Figure 6.2. We see that for all considered values, the resulting distributions are centred around the true value. This suggests our estimator has small or negligible bias, which is obviously a desirable property. We also see the shape of the distributions appear to change with  $\alpha$ . This is to be expected since  $\alpha$  is bounded above and below, and so we would certainly not expect a bell curve shape to hold near the upper and lower bounds. The other point of interest is whether the estimator is consistent. That is, does it converge in probability,

$$\hat{\alpha} \xrightarrow{P} \alpha.$$

We can investigate this empirically as follows. Let  $n$ , the sample size, be the length of each generated fractal Brownian motion. If  $\hat{\alpha}$  is consistent, we would expect it to become increasingly accurate as  $n$  increases. Hence, for varying  $n$ , we generate 1,000 fractal Brownian motions of length  $n$  in order to obtain a collection of estimates  $\hat{\alpha}_{i,n}$ . We then calculate the empirical variance

$$\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\alpha}_{i,n} - \alpha)^2$$

and verify it approaches 0 as  $n$  increases. In this example we fix  $\alpha = 1/2$ . The plot of the empirical variance against sample size  $n$  is given in Figure 6.3. It is immediately clear that the variance appears to approach 0 as  $n$  increases.

Finally, we briefly compare our estimator to an existing alternative. A strongly consistent estimator for the Hurst exponent of a fractal Brownian motion is

$$\hat{\alpha} = \frac{\frac{1}{n-2} \sum_{i=2}^n (B(i\Delta) - B((i-2)\Delta))^2 - \frac{1}{n-1} \sum_{i=1}^n (B(i\Delta) - B((i-2)\Delta))^2}{2 \log 2}, \quad (6.67)$$

where  $B(t)$  is an  $\alpha$ -fractal Brownian motion. The derivation and proof of strong consistency can be found in [18]. For clarity, let us refer to (6.66) and (6.67) as the minimal point and Shevchenko estimators respectively. For

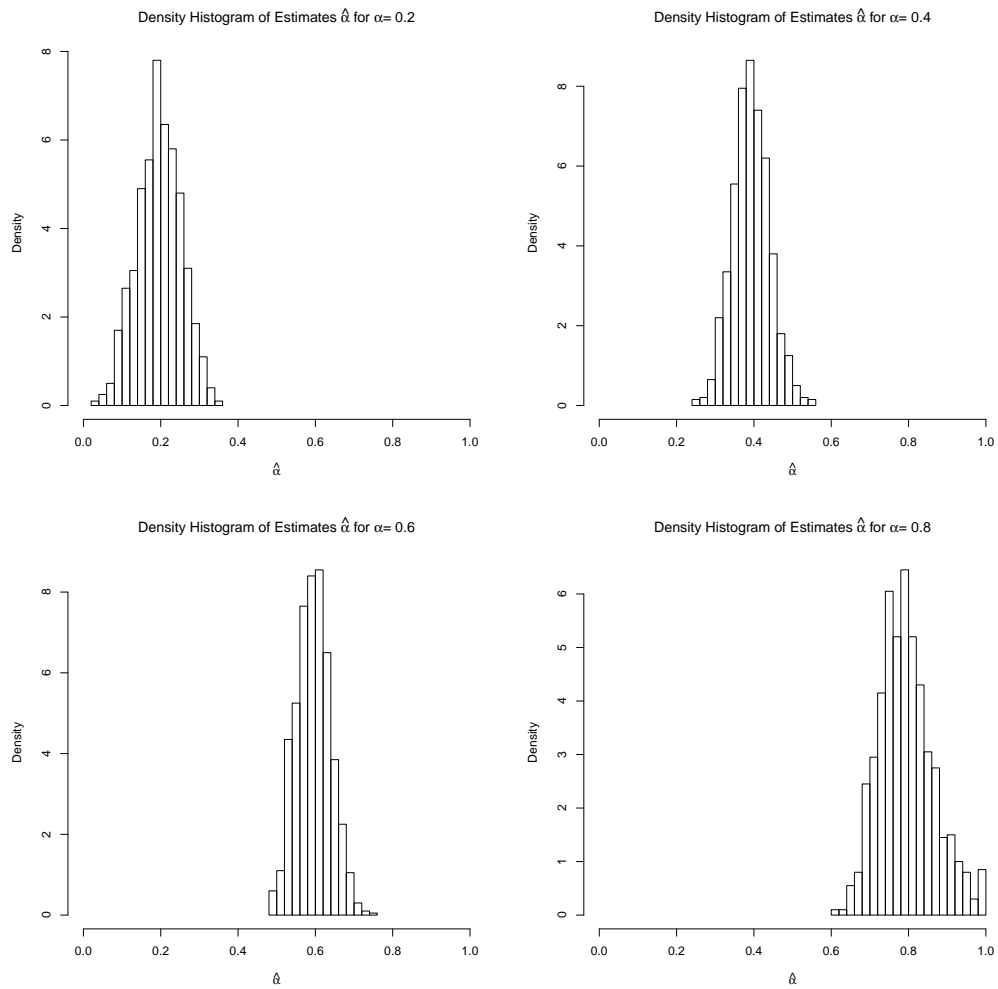


Figure 6.2: Density histograms of the estimates  $\hat{\alpha}$  for various  $\alpha$ .

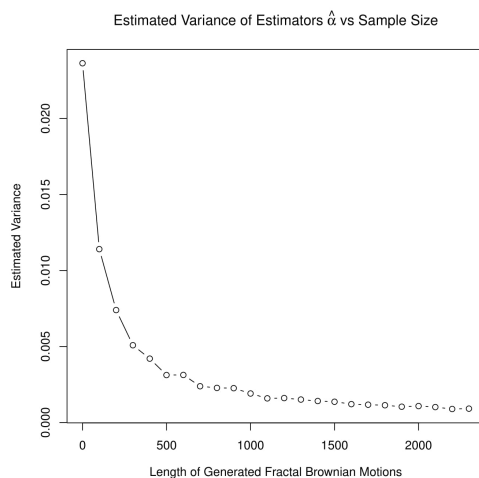


Figure 6.3: Variance of the estimator  $\hat{\alpha}$  vs sample size.

two different sample sizes  $n$ , we generate 1,000 fractal Brownian motions of length  $n$  in order to obtain an empirical distribution of our estimators  $\hat{\alpha}$ . The resulting density histograms are given in Figure 6.4. From these we can clearly see that Shevchenko's estimator (6.67) is superior to our minimal point estimator for both small and large sample sizes. Hence, while the idea underlying our estimator has shown to be correct, it is clearly inferior to an existing alternative.

## 6.7 Estimating the Dimension of $X(t)$

We now return our attention to the discretised EUR/USD exchange rate process  $X_i^\Delta$ . In previous chapters, for reasons already discussed, we have generally restricted ourselves to increments not smaller than 1 minute. However, if we wish to study the fractal properties of the exchange rate process, we will need to examine it on as fine a scale as possible. For this reason, we now consider 10 second data taken from over the year 2010, as shown in Figure 6.5.

It is trivial to see that, since we only have the process at a finite collec-

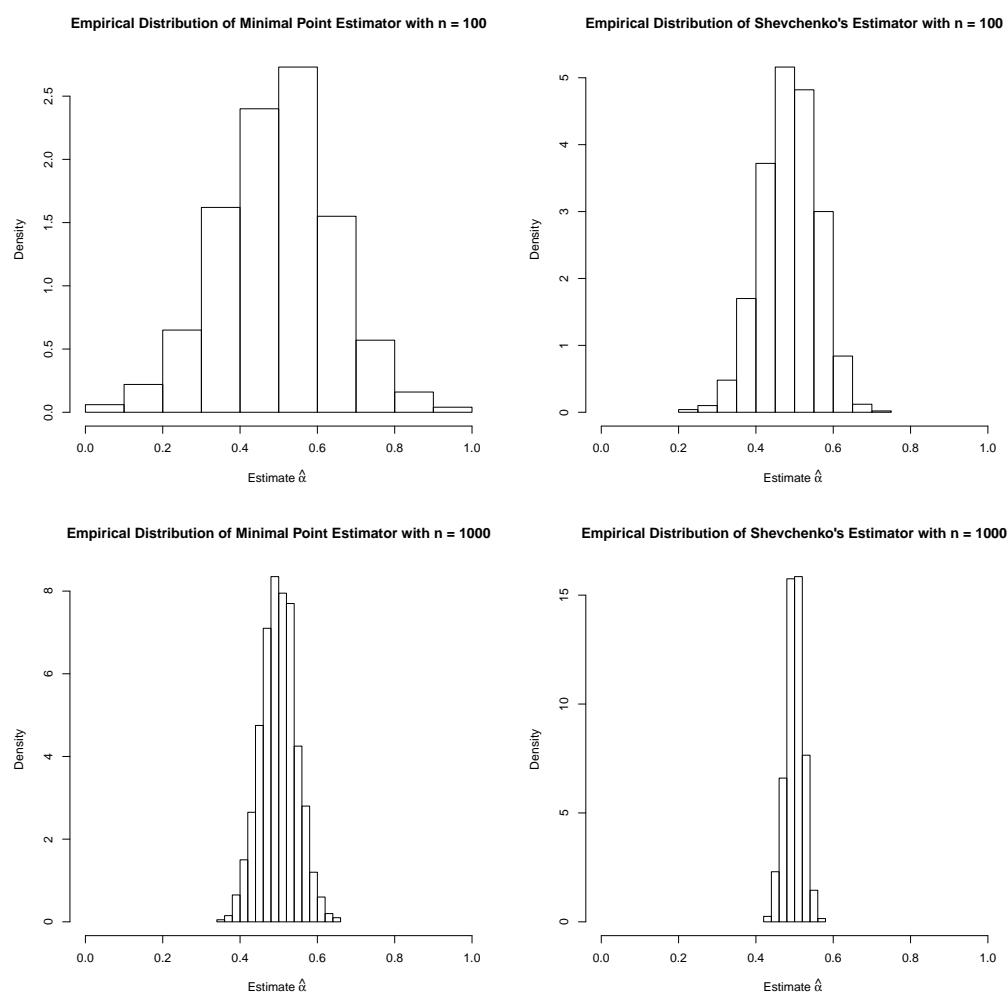


Figure 6.4: Density histograms of the minimal point and Shevchenko estimators for two sample sizes.

tion of times, its box-counting and Hausdorff dimensions are both 0. However, when we examine the discrete time process at scales much larger than  $\Delta$ , we hope that it will appear similar to its continuous time version. That is, provided we do not look "too closely" at the discrete time process, it might well appear fractal.

In Section 6.3.1 we outlined a simple method for estimating the box-

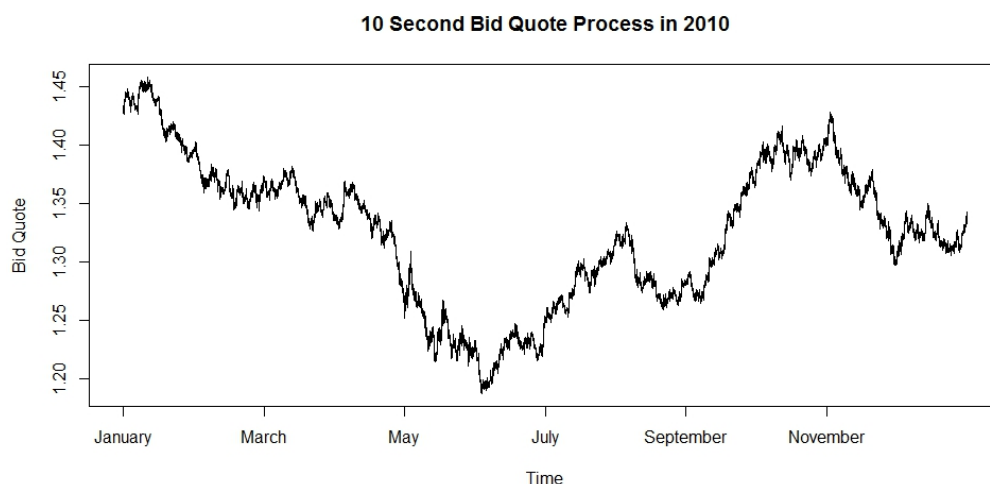


Figure 6.5: The 10 second bid quotes recorded in 2010.

counting dimension of any set  $F$ . To use this approach, we consider  $X_i^\Delta$  on the interval of time  $[0, 1]$ , so that time is measured in years. We then take the maximum and minimum of the process on this interval,

$$U = \max_{0 \leq i \leq n} X_i^\Delta \quad \text{and} \quad L = \min_{0 \leq i \leq n} X_i^\Delta, \quad (6.68)$$

and construct a grid between the points  $(0, L), (1, U) \in \mathbb{R}^2$ , consisting of squares of side  $\delta$ . The quantity  $N_\delta(X)$  is then taken as the number of such squares which contain at least one point belonging to the discretised process  $X_i^\Delta$ . This is demonstrated in Figure 6.6. Repeating this process for varying  $\delta$ , we can then employ the estimation method introduced in Section 6.3.1. We outline the procedure below for clarity:

1. Choose some  $\delta > 0$ , such that  $\delta \gg \Delta$ .
2. Construct a  $\delta$ -mesh of squares that covers the process  $X_i^\Delta$ .
3. Count the number  $N_\delta(X)$  of  $\delta$ -mesh squares that intersect  $X_i^\Delta$ .
4. Repeat steps 1 – 4 for a range of values of  $\delta$ .



5. Plot  $\log N_\delta(X)$  against  $-\log \delta$ .
6. Estimate the box-counting dimension by the slope of the resulting graph.

The resulting graph is given in Figure 6.7, where the line of best fit is shown in red. The slope of the graph, obtained using least squares, is  $\hat{\beta} = 1.271$ . Hence our estimate for the box-counting dimension of the process is

$$\hat{\dim}_B X_i^\Delta = 1.271 \quad (6.69)$$

which is much lower than what we would expect of a process driven by a Brownian motion. This estimate suggests that while our exchange rate process is highly irregular, it is not as irregular as a Brownian motion path. This is somewhat unexpected since, for example, if  $X(t)$  is an Ornstein-Uhlenbeck process then it is driven by a Brownian motion, and it is known behave as Brownian motion at sufficiently small scales.

We now use an alternative approach hoping to verify this result. In Section 6.5 we proved some important results concerning the autocorrelation functions of fractal graphs which can applied to the discretised process  $X_i^\Delta$ . Namely, we can estimate the box-counting dimension of a graph  $f$  using its autocorrelation function  $C(h)$  as

$$\dim_B f \approx 2 - \lim_{h \rightarrow 0} \frac{\log(C(0) - C(h))}{2 \log h}, \quad (6.70)$$

where we approximate the limit by taking the smallest  $h$  available to us. Using this approach we obtain a new estimate

$$\hat{\dim}_B X_i^\Delta = 1.554 \quad (6.71)$$

which is much closer to the dimension of a Brownian motion. There is clearly a large discrepancy between the two estimates which warrants discussion.

One possible explanation is that using the discretised process  $X_i^\Delta$ , even while restricting ourselves to  $\delta \gg \Delta$ , results in an underestimation of the

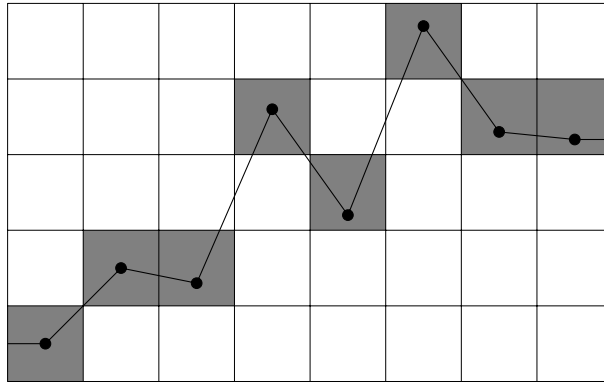


Figure 6.6: Counting the squares for which  $X_i^\Delta$  intersects, but not those that intersect the interpolated lines.

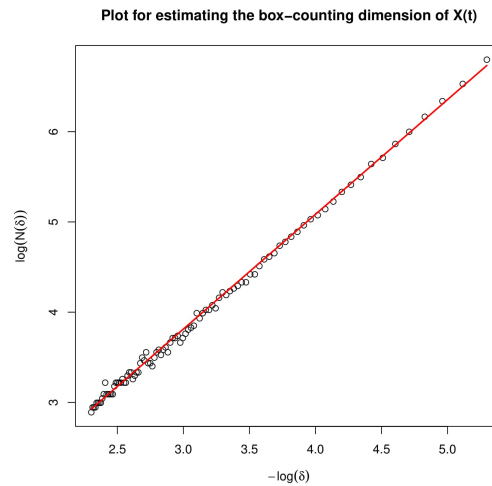


Figure 6.7: Plot for estimating the box-counting dimension of  $X(t)$ .

dimension using the first graphical method. Since we have only sampled the process  $X(t)$  at finitely many points in time, the count  $N_\delta(X)$  is underestimated. This is demonstrated in Figure 6.6 where we count the squares intersecting the points  $X_i^\Delta$ , but not all squares which intersect the interpolated lines between them. This would explain the discrepancy, since

$$N_\delta(X_i^\Delta) < N_\delta(X(t)) \implies \frac{\log N_\delta(X_i^\Delta)}{-\log(\delta)} < \frac{\log N_\delta(X(t))}{-\log(\delta)}. \quad (6.72)$$

Regardless of the cause of this discrepancy, it is clear that at least one estimate is unreliable. Therefore we cannot make any reliable conclusions, except that the process  $X_i^\Delta$  is sufficiently irregular to have a box-counting dimension greater than 1.



# Chapter 7

## Predicting Future Price Movements

In this chapter, we introduce a technical analysis method for predicting future price movements of foreign exchange rates. The method is first introduced, and then formalised in the form of an algorithm. Its efficacy and profitability is then tested using the EUR/USD exchange rate dataset.

### 7.1 Introduction

Technical analysis generally refers to methods which make use of recent financial data in order to predict price reversals or continuations. There is a great deal of variation in the techniques employed, and the efficacy of technical analysis is still debated. Some technical traders employ simple rules involving moving averages where objective decisions can be made; others make subjective trading decisions based on patterns they believe exist in the process.

In this chapter, we take the approach that  $X(t)$  is a process with behaviour that changes slowly over time. If the rate of change is slow, we hope that recent historical data can be used to detect anomalous behaviour while it still persists, in order to exploit it for profit. To take this approach,

we simply need to develop statistical methods for identifying types of anomalous behaviour. The method we introduce is designed to identify periods of the day when the process tends towards past values.

## 7.2 Trend Reversion

The method we introduce attempts to exploit what we call *trend reversion*. Loosely speaking, we say a process is trend reverting if future price movements tend toward recent historical prices. We formalise this in the following definition.

**Definition 7.1.** Let  $\{X(t) : t > 0\}$  be a stochastic process. The process  $X(t)$  is said to be trend reverting on the interval  $[t_1, t_2]$  with reference time  $t_0$  if the following holds:

$$X(t_1) - X(t_0) > 0 \iff E[X(t_2) - X(t_1) | X(t_1) - X(t_0)] < 0. \quad (7.1)$$

Note that the property of trend reversion is specific to times  $t_0, t_1, t_2$ . We do not say that for a stochastic process  $X(t)$ , property (7.1) holds for any collection of times. This implies that the process, on average, tends towards  $X(t_0)$  on the interval of time  $[t_1, t_2]$ . It is then trivial to use this information in order to predict whether the process will rise or fall, as demonstrated in Figure 7.1. As such, in a financial context the interval  $[t_1, t_2]$  is the time period over which we will trade.

Formally, if  $X(t)$  is trend reverting on the interval  $[t_1, t_2]$  with reference time  $t_0$ , then we make predictions as follows:

1. If  $X(t_0) < X(t_1)$ , then we predict that  $X(t_1) > X(t_2)$ .
2. If  $X(t_0) > X(t_1)$ , then we predict that  $X(t_1) < X(t_2)$ .
3. If  $X(t_0) = X(t_1)$ , then we make no prediction.

Thus, the indicator we use for our predictions is the sign of the quantity  $X(t_1) - X(t_0)$ . Consequently, if there has been no change on the interval

$[t_0, t_1]$ , then we cannot make a prediction based on the notion of trend reversion.

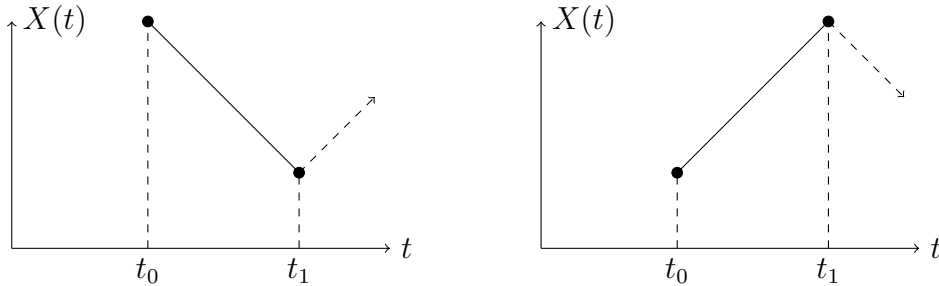


Figure 7.1: Predicting price movements using trend reversion.

## 7.3 Method

### 7.3.1 Introduction

Suppose we assume that the process  $X(t)$  takes the form

$$X(t) = X(0) + \int_0^t \sigma(s) dW(s). \quad (7.2)$$

In this form, little can be done to accurately estimate the function  $\sigma(s)$ , however if we assume it is a periodic function this becomes straightforward. Such an assumption is certainly justified as we have seen in Chapter 4 the process's empirical variance exhibits strong daily periodic behaviour. Hence, assuming that  $\sigma(s)$  is a periodic function, with period 1 day, we can use the expectation

$$\mathbf{E}[(X(t) - X(0))^2] = \int_0^t \sigma^2(s) ds \quad (7.3)$$

to estimate  $\int_0^t \sigma(s)^2 ds$  for  $0 < t < 1$ , where time  $t$  is measured as a fraction of a day.

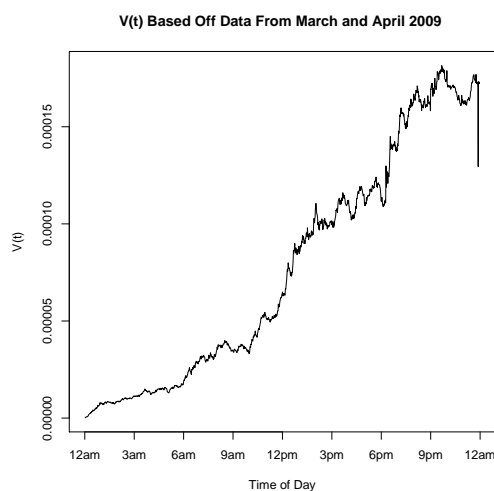


Figure 7.2: The process  $V(t)$  constructed using data from March and April of 2009.

Let  $X_i(t)$  denote the price at time  $t$  on day  $i$ , for  $i = 1, 2, \dots, n$ . We can then estimate the right hand side of equation (7.3) by

$$V(t) = \frac{1}{n} \sum_{i=1}^n (X_i(t) - X_i(0))^2. \quad (7.4)$$

If equation (7.2) holds, then the process  $V(t)$  should be monotonically non-decreasing, since  $\sigma(s)^2$  is clearly non-negative. However, when we construct the process  $V(t)$  using data from March and April of 2009, as shown in Figure 7.2, we see several prolonged periods where  $V(t)$  decreases. Some of these decreases are small and are likely the result of error in our estimation; however, for large decreases which persist over a long period of time, it seems unlikely that this is the sole cause. This suggests that  $X(t)$  does not obey (7.2) with periodic  $\sigma(s)$ .

To interpret the decreasing periods of  $V(t)$ , suppose that  $V(t)$  starts decreasing at time  $t_1$  until time  $t_2$ , when it then resumes increasing. Then



by (7.4), this implies

$$\frac{1}{n} \sum_{i=1}^n (X_i(t_1) - X_i(0))^2 > \frac{1}{n} \sum_{i=1}^n (X_i(t_2) - X_i(0))^2, \quad (7.5)$$

meaning that  $X(t_2)$  is closer, on average, to  $X(0)$  than  $X(t_1)$  is. In fact, this behaviour is exactly what we would expect to observe if  $X(t)$  was trend reverting on the interval  $[t_1, t_2]$  with reference time  $t_0 = 0$ . Hence, we can develop a technical analysis method based on the idea that if  $V(t)$  significantly decreases over some interval of time  $[t_1, t_2]$ , then this is evidence that the process undergoes trend reversion over this period with reference time  $t_0 = 0$ . The process  $V(t)$  can therefore be used to identify periods of trend reversion, allowing us to make predictions in the immediate future.

This leads to an interesting trade off with regards to how far back we should go to construct the process  $V(t)$ . On the one hand, it is clear that only the most recent data will have all or most of the present reversion periods in common, and as we go further back in time, we increase our risk of identifying reversion periods which no longer exist. On the other hand, we need a sufficient amount of data in order to construct  $V(t)$  with a reasonable degree of accuracy. In this thesis, we choose to use two months worth of data, so that when we refer to  $V(t)$  for a particular month, it is implied the process was constructed using data from the previous two months.

### 7.3.2 Identifying Periods of Trend Reversion

We now present an algorithm for identifying periods of time over which  $V(t)$  decreases. Due to the highly irregular plots of  $V(t)$ , as shown in Figure 7.2, we do not find intervals  $[t_1, t_2]$  over which  $V(t)$  is strictly decreasing. Instead, we seek intervals when  $V(t)$  has a general downward trend, taking  $t_1$  and  $t_2$  as local maxima and minima respectively.

The algorithm given below works by going along the process  $V(t)$  and, at each time  $t_1$ , checking whether the process next rises or falls. If it rises,

the algorithm does nothing and goes forward a minute in time. If it falls however, it then attempts to find the time  $t_2$  when  $V(t)$  achieves a minimum on the interval  $[t_1, t_3]$ , where  $t_3$  is the smallest time at which  $V(t)$  is greater than or equal to its current value  $V(t_1)$ . This is demonstrated in Figure 7.3 with times  $t_1, t_2, t_3$  labelled, with the algorithm described formally as Algorithm 1.

After applying this algorithm to  $V(t)$ , we identify a collection of time periods  $\{[t_{i,1}, t_{i,2}]\}$ . From these, we need to determine which are significant. After all, many intervals are very short and unlikely to be caused by trend reversion. In order for our trading strategy to be successful, we should only take intervals which provide the best evidence of trend reversion, based on some predetermined criteria. In the next section, we introduce two different criteria based on the length of time and the amount that  $V(t)$  decreases.

**Data:** The process  $V(t)$ .

**Result:** A collection of time periods  $\{[t_{i,1}, t_{i,2}]\}$ .

Initialise  $t_1 = 1$ ;

**while**  $t_1 < 1440$  **do**

**if**  $V(t_1) > V(t_1 + 1)$  **then**

        Set  $t_3 = \min\{\min\{s : V(s) \geq V(t_1)\}, 1440\}$ ;

        Set  $t_2 = \operatorname{argmin}\{V(s) : t_1 < s < t_3\}$ ;

        Record  $[t_1, t_2]$  as interval where  $V(t)$  decreases;

        Set  $t_1 = t_2 + 1$ ;

**else**

        Set  $t_1 = t_1 + 1$ ;

**end**

**end**

**Algorithm 1:** Finding periods of time where  $V(t)$  decreases.

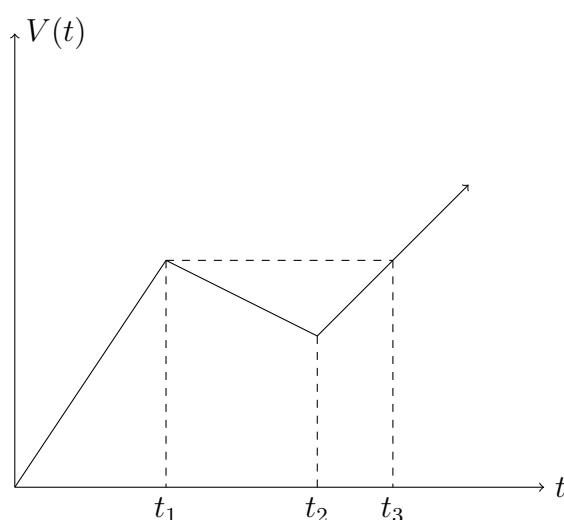


Figure 7.3: Times  $t_1, t_2, t_3$  as identified by Algorithm 1.

### 7.3.3 Testing significance

Consider the time intervals  $\{[t_{i,1}, t_{i,2}]\}$  identified using Algorithm 1. Among these, some may be the result of trend reversion, but many will not be. Therefore we need some predetermined criteria for determining which of these time intervals should be considered significant enough to include in our trading strategy. The first criteria we consider is based on the length of the interval.

In finance it is generally understood that, all else being equal, longer trades are more likely to be profitable. This is because the fee associated with each trade becomes less significant as time goes on. The length of time we trade is therefore one basis by which we can determine whether an identified interval is worth trading over.

**Definition 7.2.** Let  $[t_{i,1}, t_{i,2}]$  be a period of time identified by Algorithm 1. Then the *trading time* of this interval is defined as the difference

$$\Delta t_i = t_{i,2} - t_{i,1}. \quad (7.6)$$

We call them trading times since, if these time periods are determined to be significant, then we will "buy in" at time  $t_1$  and "cash out" at time  $t_2$ . The first criteria we consider is then of the form

$$\Delta t_i \geq t_{\min}, \quad (7.7)$$

so that if the trading time is above some threshold  $t_{\min}$ , then we determine it to be significant and include it in our trading strategy. It is unclear what an appropriate threshold is, and so we will use a range of values when testing the efficacy of our strategy.

The second criteria we consider is based on how much  $V(t)$  changes on the interval  $[t_{i,1}, t_{i,2}]$ . We consider such a threshold since, intuitively, the change in  $V(t)$  should reflect the strength of trend reversion over this interval. On the one hand, it seems unlikely that the difference

$$V(t_{i,1}) - V(t_{i,2}) \quad (7.8)$$

will be useful in determining which are significant, since Figure 7.2 shows  $V(t)$  tends to vary more as  $t$  increases. So a criteria of the form

$$V(t_{i,1}) - V(t_{i,2}) \geq v_{\min} \quad (7.9)$$

is likely to favour trades later in the day, while neglecting earlier ones. Instead, suppose we consider the difference in price at these times

$$|X(t_{i,2}) - X(t_{i,1})|. \quad (7.10)$$

Obviously the larger this difference is, the more we will profit from successfully predicting whether  $X(t)$  will rise or fall on the interval  $[t_{i,1}, t_{i,2}]$ . So a criteria related to this difference seems more relevant in determining which trading times are significant. Since we can't know the size of this difference prior to time  $t_{i,2}$ , we might instead hope to estimate it in some way using  $V(t)$ . To this end, consider the definition of  $V(t)$  given by (7.4), which shows that  $V(t)$  is an estimate of

$$(X(t) - X(0))^2$$

for days in the immediate future. Hence, by using the estimate

$$\sqrt{V(t)} \approx |X(t) - X(0)|, \quad (7.11)$$

we can estimate a lower bound of (7.10) by

$$\begin{aligned} \sqrt{V(t_{i,1})} - \sqrt{V(t_{i,2})} &= |\sqrt{V(t_{i,1})} - \sqrt{V(t_{i,2})}| \\ &\approx ||X(t_{i,1}) - X(0)| - |X(t_{i,2}) - X(0)|| \\ &\leq |X(t_{i,2}) - X(t_{i,1})|. \end{aligned} \quad (7.12)$$

So if  $\sqrt{V(t_{i,1})} - \sqrt{V(t_{i,2})}$  is greater than some threshold  $v_{\min}$ , then inequality (7.12) suggests an estimate of  $|X(t_{i,2}) - X(t_{i,1})|$  is also likely to be above that threshold. Hence, the difference  $\sqrt{V(t_{i,1})} - \sqrt{V(t_{i,2})}$  might serve as a good criteria.

**Definition 7.3.** Let  $[t_{i,1}, t_{i,2}]$  be a period of time identified by Algorithm 1. Then the *trading depth* of this interval is defined as

$$\Delta V_i = \sqrt{V(t_{i,1})} - \sqrt{V(t_{i,2})}. \quad (7.13)$$

The trading depth is, in some sense, a measure of how significant the decrease in  $V(t)$  is over the interval  $[t_{i,1}, t_{i,2}]$ . The second criteria we use to test whether a time period is significant is then of the form

$$\Delta V_i \geq v_{\min}, \quad (7.14)$$

so that if an interval's trading depth is above some minimal threshold, we determine it to be significant enough to trade over. Once again, it's unclear what an appropriate threshold might be. Hence we will also use a range of values for  $v_{\min}$  when testing our trading strategy.

## 7.4 Trading Strategy

The trading strategy we employ combines the ideas presented in the previous sections. Given two months worth of data, we attempt to make predictions on whether the exchange rate will rise or fall at particular times

of the day throughout the next month. We first construct the process  $V(t)$  using the previous two months worth of data, and then use Algorithm 1 to identify periods of time during which  $V(t)$  decreases. For each of these time periods, we use a predetermined criteria of the form (7.7) or (7.14) to identify which are significant. Each significant time period is then traded over the next month using the trend reversion strategy outlined in Section 7.2.

For example, suppose the time period  $[t_{i,1}, t_{i,2}]$  is one of those identified and determined to be significant. Then on each day throughout the next month, we look at the price at time  $t_1$  and compare it with the price at midnight earlier that day. Based on this comparison, we predict whether the price at time  $t_2$  will be higher or lower than at time  $t_1$ . If we predict the exchange rate will rise over this period, then we buy euros at time  $t_1$  and then exchange them back for US dollars at time  $t_2$ . Conversely, if we predict the exchange rate will fall over this period, then we buy US dollars at time  $t_1$  and then exchange them back for euros at time  $t_2$ .

## 7.5 Testing the Strategy

In this thesis, we test the efficacy of our technical analysis trading strategy using the EUR/USD exchange rate dataset. We examine both the accuracy of our predictions and also the profitability of our trades. Quotes are recorded too infrequently to use data prior to 2008, and the financial crisis of that year may cause abnormal behaviour in the process well into 2009. For these reasons, we start making predictions in June of 2009, and continue up until the end of our dataset, with the last set of predictions taking place in June of 2011. That is, we begin by using data from April and May of 2009 to make predictions throughout June of 2009, and end by using data from April and May of 2011 to make predictions throughout June of 2011.

### 7.5.1 Accuracy of Predictions

To test the accuracy of our predictions, we record the total number of correct and incorrect predictions across all identified time periods. If this method is baseless, then we would expect to see roughly 50% of our predictions being correct, as the process  $X(t)$  appears to rise and fall with equal probability. Hence, under the usual binomial assumptions, it is straightforward to test the significance of our method.

We denote the proportion of correct predictions by  $\hat{p}$  so that our test statistic will be

$$Z = \frac{\hat{p} - 0.5}{\sqrt{0.5^2 n}}, \quad (7.15)$$

where  $n$  is the total number of predictions made. If the method offers no advantage (or disadvantage), then  $Z$  should have a standard normal distribution. Therefore, the test statistic  $Z$  can be used to test

$H_0$  : The strategy offers no advantage

$H_1$  : The strategy offers an advantage

where we reject the null hypothesis at the 2.5% level of significance if  $Z > 1.96$ .

### 7.5.2 Profitability of Trades

Even if our strategy offers an advantage, there is no guarantee that it will be profitable. There are a number of fees associated with trading, the most significant of which being the spread between the bid and ask quotes. In order to test whether our trading strategy is profitable, we simply simulate trading by buying and selling at the quoted prices, as dictated by our strategy.

We imagine starting on June 2009, before making any predictions, with \$100 and €100. For each prediction made over the 25 month period, we trade according to the quoted prices in our dataset. If we predict the exchange rate will rise between times  $t_1$  and  $t_2$ , then we exchange our dollars

for euros at time  $t_1$ , and then back into euros at time  $t_2$ . And conversely if we predict the exchange rate will fall. At the end of the 25 month period, we then look at whether our dollars and euros have grown or shrunk as a result of these trades. Due to the fee associated with every trade, we would expect a baseless strategy to lose money.

We then repeat this procedure using a "baseless strategy" that offers no advantage in order to compare the amount our strategy profits to some baseline value. That is, we trade over the same intervals as before, except we now make predictions randomly instead of using our trend reversion criteria. On each of the identified and significant intervals  $[t_{i,1}, t_{i,2}]$ , we predict a rise or fall in the exchange with probability  $1/2$ . We would certainly expect such a strategy to lose money, but the exact amount we would lose is less clear, but nonetheless important as a point of comparison.

## 7.6 Results

### 7.6.1 Accuracy of Predictions

We give the correct and total number of predictions made in Tables 7.1 and 7.2, broken down by both criteria and threshold used. We see that for both criteria, along with all considered thresholds  $t_{\min}$  and  $v_{\min}$ , the majority of our predictions were correct. Additionally, the test statistics  $Z$  are all statistically significant, well above the critical value. Since the test statistics are significant for all criteria and thresholds considered, we conclude that our trading strategy offers some advantage.

Our attention then turns to how the proportion of correct predictions  $\hat{p}$  changes depending on the criteria and threshold used. This question is an important one since, if this strategy is to be employed, we should hope to optimise the criteria used in order to maximise profit. If the criteria we have chosen do indeed measure significance well, we would expect that increasing the thresholds  $t_{\min}$  and  $v_{\min}$  should improve the accuracy of our



Results using trading time criteria				
$t_{\min}$	Correct Predictions	Total Predictions	$\hat{p}$	$Z$
30 minutes	1020	1846	0.553	4.52
45 minutes	543	944	0.575	4.62
60 minutes	300	491	0.611	4.92
75 minutes	186	304	0.612	3.90
90 minutes	126	198	0.636	3.84

Table 7.1: Number of correct predictions from June 2009 to June 2011 using trading time as the criteria for significance.

Results using trading depth criteria				
$v_{\min}$	Correct Predictions	Total Predictions	$\hat{p}$	$Z$
0.0004	1668	3114	0.536	3.98
0.0005	998	1805	0.553	4.50
0.0006	644	1127	0.571	4.80
0.0007	446	748	0.596	5.27
0.0008	311	492	0.632	5.86
0.0009	205	328	0.625	4.53

Table 7.2: Number of correct predictions from June 2009 to June 2011 using trading depth as the criteria for significance.

predictions. This is shown in Figures 7.4 and 7.5, where we see a positive linear relationship between  $\hat{p}$  and the threshold used for each criteria. That is, as we increase  $t_{\min}$  and  $v_{\min}$ , the proportion of correct predictions  $\hat{p}$  tends to increase.

Intuitively, it makes sense why  $\hat{p}$  would increase with  $v_{\min}$ , since trading depth measures how significantly  $V(t)$  decreases over some interval. A significant decrease in  $V(t)$  suggests a strong rate of trend reversion, and hence our strategy is likely to perform better. It is less clear why  $\hat{p}$  would increase with  $t_{\min}$ . The reason we introduced a trading time criteria was

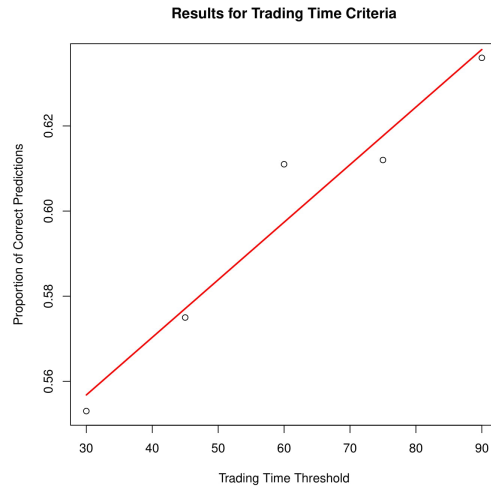


Figure 7.4: Proportion of correct predictions plotted against trading time threshold  $t_{\min}$ .

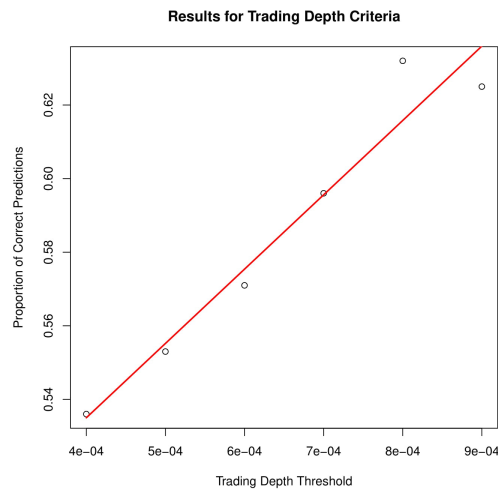


Figure 7.5: Proportion of correct predictions plotted against trading depth threshold  $v_{\min}$ .

based on profitability, not because we thought it would improve the accuracy of our predictions. One possible explanation for this is that longer trading times, on average, have a larger trading depth. However, regardless of the cause, it is worth noting how beneficial this is from a trading stand point. This allows us to exclude short trades without reducing the accuracy of our predictions. Since we take only long trades, where fees become less significant, we are more likely to profit in the long term.

### 7.6.2 Profitability of Trades

The final amounts in dollars and euros after the 25 month trading period are given in Tables 7.3 and 7.4. These results show that the thresholds with the highest proportion of successful predictions are not necessarily the most profitable. For example, we saw in the previous section that when using trading time as the criteria, we had the highest  $\hat{p}$  when using  $t_{\min} = 90$  minutes. However, Table 7.3 shows the most profitable threshold was  $t_{\min} = 45$  minutes, with about a 2.25% and 2.57% profit on our dollars and euros respectively. This discrepancy is likely caused by missed opportunities when using a threshold that is too high. Although using a higher threshold increases the accuracy of our predictions, it also results in ignoring trades which might, on average, still be profitable.

Secondly, we see that a high success rate  $\hat{p}$  is not necessarily enough to ensure profitability. For example, despite correctly predicting price movements over 60% of the time, we end up losing US dollars when using the trading time criteria with a threshold  $t_{\min}$  of 60 or 75 minutes. It is worth noting that overall we still profit, as in both cases our euro profits are more than enough to offset these losses; however the results are surprising nonetheless. It also suggests there is some degree of asymmetry in our method where, depending on the criteria, we more accurately predict price movements in one direction. The reverse can be seen with  $t_{\min} = 30$  minutes, where we profit in US dollars, but lose euros.

Profitability of trading using trading time criteria		
$t_{\min}$	Final US dollars	Final Euros
30 minutes	102.46	97.97
45 minutes	102.25	102.57
60 minutes	99.97	103.99
75 minutes	99.13	102.04
90 minutes	100.50	100.11

Table 7.3: Profitability of trades from June 2009 to June 2011 using trading time as the criteria for significance.

Profitability of trading using trading depth criteria		
$v_{\min}$	Final US dollars	Final Euros
0.0004	94.60	96.52
0.0005	100.14	101.24
0.0006	102.17	100.55
0.0007	102.23	102.44
0.0008	101.76	100.32
0.0009	100.35	100.13

Table 7.4: Profitability of trades from June 2009 to June 2011 using trading depth as the criteria for significance.

Profits using a baseless strategy with trading time criteria		
$t_{\min}$	Final US dollars	Final Euros
30 minutes	93.65	89.97
45 minutes	101.72	101.82
60 minutes	98.12	101.91
75 minutes	98.26	101.03
90 minutes	99.8	99.42

Table 7.5: Final capital when trading with a baseless strategy using trading time criteria.

Profits using a baseless strategy with trading depth criteria		
$v_{\min}$	Final US dollars	Final Euros
0.0004	84.72	85.68
0.0005	89.30	90.82
0.0006	95.06	93.32
0.0007	96.63	96.78
0.0008	99.50	98.25
0.0009	99.34	99.33

Table 7.6: Final capital when trading with a baseless strategy using trading depth criteria.

On this note, we now investigate the amount we would lose trading with no advantage. On each of the intervals traded according to our strategy, we now trade randomly, predicting rises and falls with equal probability. We again consider the same criteria and threshold values. The final amounts for each criteria and threshold using our baseless strategy are given in Tables 7.5 and 7.6.

These results serve as a better baseline for us to gauge how significant our profit is, as opposed to simply noting whether we profit or not. This is particularly apparent with the results for trading depth criteria, using  $v_{\min} = 0.0004$ , where our results showed a loss of 5.4% in US dollars. Such a significant loss at first appears to show our strategy fails to offer any advantage; however when we compare it to the results of a baseless strategy, it is clear some advantage still exists. Trading with a baseless strategy on the same intervals, we end up with \$84.72, corresponding to a loss of over 15%. Hence even when our strategy loses money, it still significantly outperformed a baseless strategy. This is also reflected in the trading time criteria results. Using a threshold of 30 minutes, we incurred a loss of around 2% in euros. However we lost over 10% when trading over those same intervals using a baseless strategy.

In summary, our strategy appears to be profitable for most criteria and thresholds considered. In all instances when our strategy lost money, it still out performed a baseless strategy in terms of the amount lost. The profit made is small considering trades take place over a 25 month period, however could be magnified in several ways. For example, by keeping our capital in US dollars, we could simultaneously apply our strategy to several currencies priced against the US dollar. Combined with the large amount of leverage available to foreign exchange traders, these seemingly small profits might result in significantly larger gains. Regardless, the important note is that our strategy offers enough of an advantage to consistently "beat the market" despite incurring regular transaction fees. The results of this section provide strong evidence that the exchange rate is more predictable than current mathematical models suggest.

# Chapter 8

## Conclusions

The objective of this thesis was to investigate the properties of the exchange rate process from several different stand points. In this chapter, we discuss the results of our investigation and acknowledge some limitations. We then conclude by suggesting several areas for further research.

### 8.1 Summary

In this thesis, we have studied the stochastic and fractal properties of real exchange rate data. By discretising the data, we were able to study it as a stochastic process observed at equidistant points in time.

We first investigated whether the exchange rate behaves as a process driven by Brownian motion. The process's sampled total variation  $V_{\Delta}(X(t))$  increased at a rate close to that of Brownian motion, which would be expected for a Brownian motion-driven process. Additionally, we saw that the variance of increments is linear in time, though might be asymptotically stationary. This is consistent with the Ornstein-Uhlenbeck process, which behaves very much like Brownian motion on sufficiently small time scales. However, despite possessing properties that are characteristic of a Brownian motion-driven process, we found that increments are unlikely to be normally distributed. Instead, we saw that increments are likely to

come from mixture distributions, these mixtures being caused by periodic behaviour in the variance of the process. Thus, if we wish to model the exchange rate as a Brownian motion-driven process, we must account for the non-constant behaviour of variance. To this end, we derived estimators for the parameters of an Ornstein-Uhlenbeck process with periodic variance, but note the significant draw back in that they make use of only a fraction of the data available to us.

As part of our investigation of the exchange rate as a Brownian motion-driven process, we used the Khmaladze transformation-2 to test whether the increments  $\Delta X_i$  come from a centred normal distribution. To this end, we derived a closed form expression, in terms of the standard normal distribution  $\Phi(x)$ , for the transformed process  $\tilde{v}_n$ . This formula can be applied to any random sample in order to test whether it comes from a centred normal distribution, and is trivial to implement using the given expression. This work therefore provides an easy to implement and distribution-free formula for testing centred normality.

In Chapter 6, we introduced some fractal geometry, and reviewed some results concerning trajectories of Brownian motion as fractal sets. Following this, we derived a new estimator for the Hurst exponent of a fractal Brownian motion, and compared it to an existing alternative. We found that while our estimator performs as expected, the existing alternative given in [18] is far superior, and easier to implement. We also estimated the box-counting dimension of the exchange rate process using two different methods. These estimates significantly differed from one another, suggesting at least one of our methodologies was flawed, and so no definitive conclusions can be drawn. However, both estimates were well above 1, suggesting the exchange rate process is, to some extent, fractal.

In the final part of this thesis, we introduced a technical analysis method for predicting future price movements. We tested both the accuracy and profitability of this method using the EUR/USD exchange rate dataset. There is significant reason to believe this method provides an advantage



when predicting price movements, as all of our test statistics were significant at the 1% level. The method also provided a small profit after taking the exchange rate spread into account, and significantly out performed the baseless strategy used as a point of comparison. These results show that exchange rates are predictable to a greater extent than current mathematical models suggest, and provides strong evidence against the efficient market hypothesis.

## 8.2 Limitations

One caveat of this research concerns the dataset used. As already stated, the foreign exchange market is decentralised, and so comprehensive transaction data is unavailable to researchers. Instead, over-the-counter quotes are used in lieu of actual trades. We should therefore note that the results of this thesis concern the behaviour of these published quotes specifically, and not necessarily the true price that participants are trading currency at.

Further limitations concern the testing of our technical analysis method introduced in Chapter 7. Firstly, as just mentioned, our dataset consists of non-binding quotes. However, when we simulated trading, we purchased and sold currency according to any quote in our dataset. This necessarily assumes that these quotes are binding, meaning we can purchase any amount of currency at these prices regardless of current market conditions, which is not true in general. Realistically, we would experience some degree of slippage, where the price we expect to trade currency at differs from the actual price at which the trade is executed.

Furthermore, when testing profitability we only took the spread into consideration, neglecting other fees such as trading commissions. The spread is by far the largest fee associated with foreign exchange trading, so that our simulation did account for the bulk of transactions fees we would incur, however we should note that the profitability would likely be lower if other fees were taken into consideration.

### 8.3 Further Research

Firstly, when testing normality we found that it was straightforward to derive a closed form expression of  $\tilde{v}_n$ . One area of further research is the more general, and far more practical, question of testing whether a sample comes from a  $N(\mu, \sigma^2)$  distribution, so that we are no longer restricted to centred normal distributions. A closed form expression of  $\tilde{v}_n$  for this more general problem would provide quick and computationally easy distribution-free testing of normality.

There are several areas of further research with regards to the technical analysis method discussed in Chapter 7. The fundamental idea underlying the method is the process  $V(t)$ , however the criteria we used to test significance of intervals were by no means optimal. Alternative statistics for testing how significantly  $V(t)$  decreases over some interval might improve the accuracy and profitability of the method substantially. Hence further research could focus on developing alternative tests of significance, or perhaps optimising the thresholds considered in this thesis.

An additional way that the methodology might be improved involves the time at which we exit trades. Due to the nature of the method, there is no obvious way to improve the time we enter trades. However, having entered a trade, we might hope to increase the profitability of our method by introducing conditions for exiting the trade earlier or later than as dictated by the process  $V(t)$ . This is motivated by an idea commonly held by traders that one should "cut their losses early" and "let their profits run." A significant body of work exists on optimal exit strategies, and any number of these strategies can be combined with the methodology presented in Chapter 7.

Lastly, the logic underlying the method is by no means restricted to the EUR/USD currency pair. An interesting question is therefore whether the technical analysis method is successful in predicting not just other currency pairs, but also other financial prices such as stocks or commodities.

The accurate predictions made for the EUR/USD dataset certainly suggests it might work for other currency pairs, since the EUR/USD currency pair is among the most liquid, and liquid markets are generally more efficient. Comprehensive datasets exist for exchange traded assets such as stocks, so research in this direction would also address the concern of using published quotes in lieu of traded prices.



# Appendix A

## Notation

$X(t)$  Bid quote process

$Y(t)$  Ask quote process

$X_t^\Delta$  Discretised bid quote process

$W(t)$  Standard Brownian motion

$V_n(X(t))$  Sampled total variation

$V_\Delta(X(t))$  Sampled total variation of  $X_t^\Delta$

$v_n(x; \theta)$  Empirical process

$v_n(x; \hat{\theta})$  Parametric empirical process

$\tilde{v}_n(x; \hat{\theta})$  Transformed parametric empirical process

$\dim_{\mathbb{B}}$  Box-counting dimension

$\dim_{\text{H}}$  Hausdorff dimension

$N_{\delta}(X)$  Number of  $\delta$ -mesh squares intersecting  $X$

$C(h)$  Autocorrelation function

$V(t)$  Estimated variance of  $X(t) - X(0)$

$\Delta t_i$  Trading time

$t_{\min}$  Trading time threshold

$\Delta V_i$  Trading depth

$v_{\min}$  Trading depth threshold

$\hat{p}$  Proportion of correct predictions

# Bibliography

- [1] Bank for International Settlements. (2016). *Triennial Central Bank Survey Foreign Exchange Turnover in April 2016*. Retrieved from: <http://www.bis.org/publ/rpfx16fx.pdf>.
- [2] Chen, S.X. & Tang, C.Y. (2006). *Parameters Estimation and Bias Correction for Diffusion Processes*. *Statistics Preprints*. 45.
- [3] Eldredge, N. (2016). *Analysis and Probability on Infinite-Dimensional Spaces [Lecture Notes]*. Retrieved from: <https://arxiv.org/pdf/1607.03591.pdf>.
- [4] Falconer, K. (2014). *Fractal Geometry: Mathematical Foundations and Applications*. Third Edition. West Sussex, United Kingdom: John Wiley & Sons, Ltd.
- [5] Galati, G. (2000). *Trading Volumes, Volatility and Spreads in Foreign Exchange Markets: Evidence from Emerging Market Countries*. Basle: BIS, Monetary and Economic Dep.
- [6] Glazman, I.M., & Ljubič, J.I. (2006). *Finite-Dimensional Linear Analysis: a Systematic Presentation in Problem Form*. Mineola, NY: Dover Publications.
- [7] Huang, J. (2013). *somebm: Some Brownian Motions Simulation Functions*. R package version 0.1. <https://CRAN.R-project.org/package=somebm>

- [8] Khmaladze, E. (1981). Martingale Approach in the Theory of Goodness-of-Fit tests. *Theory Probab. Appl.* **26**(2), 240-257. <https://doi.org/10.1137/1126027>
- [9] Khmaladze, E. (1988). An Innovation Approach to Goodness-of-Fit Tests in  $\mathbf{R}^m$ . *Ann. Statist.* **16**, 1503-1516. <https://doi.org/10.1214/aos/1176351051>
- [10] Khmaladze, E. (1993). Goodness of Fit Problem and Scanning Innovation Martingales. *Ann. Statist.* **21**, 798-829. <https://doi.org/10.1214/aos/1176349152>
- [11] Khmaladze, E. (2016). Unitary Transformations, Empirical Processes and Distribution Free Testing. *Bernoulli.* **22**(1), 563-588. <https://doi.org/10.3150/14-BEJ668>
- [12] Klebaner, F.C. (2005). *Introduction To Stochastic Calculus with Applications*. Second Edition. London, United Kingdom: Imperial College Press.
- [13] Kolmogorov, A.N., & Tikhomirov, V.M. (1954).  $\varepsilon$ -Entropy and  $\varepsilon$ -Capacity of Sets in Functional Spaces. *American Mathematical Society Translations*, 2(17), 277-364.
- [14] Kolmogorov, A.N., & Fomin, S.V. (1975). *Introductory Real Analysis*. New York, NY: Dover Publications, Inc.
- [15] Mandelbrot, B. (1967). *How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension*. *Science*, 156(3775), 636-638.
- [16] Mandelbrot, B. (1982). *The Fractal Geometry of Nature*. New York, NY: W.H. Freeman & Company.
- [17] Moral, P.D. & Penev, S. (2016). *Stochastic Processes: From Applications to Theory*. Boca Raton, FL: CRC Press.



- [18] Shevchenko, G. (2014). *Fractal Brownian Motion in a Nutshell* [Lecture Notes]. Retrieved from: <https://arxiv.org/pdf/1406.1956.pdf>.
- [19] Shiryaev, A.N. (1999). *Essentials of Stochastic Finance: Facts, Models, Theory*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- [20] Smith, R. (2010). Is High-Frequency Trading Inducing Changes in Market Microstructure and Dynamics?. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.1632077>