2012

Stochastic Modeling with Temporally Dependent Gaussian Processes: Applications to Financial Engineering, Pricing and Risk Management

Daniel Jonathan Scansaroli
Lehigh University

Stochastic Modeling with Temporally Dependent Gaussian Processes

Applications to Financial Engineering, Pricing and Risk Management

by

Daniel Jonathan Scansaroli

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Industrial Engineering

January 2012

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Date

Dissertation Co-Advisors:

_____

Vladimir Dobrić, Ph.D.

Department of Mathematics

_____

Robert H. Storer, Ph.D.

Department of Industrial & Systems Engineering

_____

Accepted Date

Committee Members:

_____

Robert H. Storer, Ph.D.

Committee Chair

_____

Vladimir Dobrić, Ph.D.

_____

Wei-Min Huang, Ph.D.

_____

Eugene Perevalov, Ph.D.

_____

Frank E. Curtis, Ph.D.

Acknowledgments

This dissertation is the result of the support and commitment of many individuals. To start, I would like to thank my co-advisors, Dr. Dobrić and Dr. Storer, for years of mentoring, education and research collaboration. Dr. Dobrić's rigorous understanding of mathematical theory and its application along with long late hours discussing new ideas and methods was critical in the formulation of research that encompasses the fields of mathematics, engineering and finance. Dr. Storer's ability to interconnect complicated ideas while maintaining "the big picture" was invaluable in forming the direction and language of this research. Without my co-advisors willingness to spend the time to share their mathematical and engineering expertise this dissertation would not have been possible. On this note, I would also like to recognize the Lehigh University Department of Mathematics and Industrial Engineering & Systems Department for facilitating the opportunity to collaborate my doctoral studies between the two departments.

I would like to thank Drs. Wei-Min Huang, Eugene Perevalov, and Frank E. Curtis for serving on my committee. My committee members, along with Dr. Pietro Belotti, provided invaluable feedback, comments and suggestions.

I would like to dedicate this dissertation to my family, especially my fiancé, parents and brothers. My parents, John and Cathy, for igniting and encouraging my love for learning and pursuit of knowledge from a young age. My brothers, Steve and Rob, for their never ending love and support. And Jennifer for her commitment to, and support and understanding of, the sacrifices required to pursue a doctorate. All of you have my most genuine appreciation and love for always encouraging me to go beyond my best.

Contents

List of Figures

Abstract

This dissertation studies and develops theory and techniques for the application of two popular classes of temporally dependent Gaussian processes. While many of our findings have application to physics, hydrology, logistics, biology and economics (to name a few), we focus on modeling, parameter estimation, risk management, pricing and optimal decision making for financial instruments.

The first part of this thesis focuses on the widely used fractional Brownian motion (fBm) process. We explore the advantages and disadvantages of modeling with the process and present new consistent estimation techniques for a fBm process that is influenced by drift and volatility. These techniques are grounded in ergodic theory for stationary processes. An empirical study is performed comparing the new estimators to leading methods. This study indicates that our estimators are highly competitive in terms of Root Mean Square Error (RMSE) and computational speed, making them ideal for fast paced financial markets. Additionally, we demonstrate how the long-range dependence structure of fBm can result in an underestimation of risk and volatility when the standard Brownian motion model is assumed. Similarly, we propose and demonstrate that the dependence structure (or memory) may explain some anomalies that are seen in derivative prices, such as a term structure to volatility. We conclude our study of fBm by implementing our new estimation method on three broad indexes to show how the memory parameter (known as the Hurst index/exponent) has behaved over time. Our study shows that factors like capitalization and liquidity may be influencing the Hurst index and therefore the propensity of serial runs up and down in market prices.

The second part of this thesis develops techniques to further understand and accurately model continuous sample path Gaussian Markov (GM) processes. Gaussian Markov processes are widely used to model dynamic systems. The majority of stochastic models in finance, particularly diffusion processes, are Markovian, and many are Gaussian. GM processes are commonly used

in the valuation of derivatives on equities, commodities, foreign exchange, and interest rate products. The GM process allows for a wide range of properties that can adapt to common properties observed in real market data, including long-memory, "fat tails," non-stationarity and the term structure of volatility observed in derivative prices.

We explore the properties of Gaussian Markov processes and derive new methods for modeling, pricing, risk management and parameter estimation. All of our results use Hida's [19] representation which establishes that GM processes can be decomposed into a time dependent scaling function and a Gaussian martingale.

To understand the range behavior of the GM process, we explicitly derive a closed form representation of quadratic variation. Our derivation reveals that the quadratic variation can be represented as a Riemann Stieltjes integral. Our result leads to a consistent estimation method for the diffusion parameter using the quadratic variation. We establish consistency of the estimator and use the Borel-Cantelli lemma to obtain almost sure convergence and we derive the confidence interval bounds on a fixed interval. We apply the new method to estimate the volatility in the Ornstein-Uhlenbeck (O-U) process and show that the estimator is significantly more accurate than the MLE on small to moderate size samples on a fixed interval. Additionally, we derive closed form Maximum Likelihood Estimators of all model parameters in the O-U process, eliminating the standard method that uses three dimensional numerical optimization methods.

The quadratic variation result is also important for understanding the pricing, hedging strategies and risk management with GM processes. We demonstrate how the quadratic variation can be adapted to Itô calculus for modeling with Gaussian Markov processes. This allows us to extend the Black-Scholes model to include many observed properties like long-range dependence, non-stationarity and heteroscedasticity. With the GM process we adapt modern pricing theory, to determine arbitrage free criteria, replicating portfolios, and Kolmogorov equations. This leads to a general closed form formula for pricing call options as well as hedging strategies for risk management with GM processes.

Motivated by the fact that closed form solutions do not exist for many American style derivative contracts, we introduce both an exact simulation technique and a recombining n-period binomial tree model for Gaussian Markov processes. The central limit theorem for stochastic processes established by Andersen and Dobrić [2] is used to prove that the tree process converges weakly to the continuous sample path GM process. This result is important in making decisions

and finding optimal timing for stochastically dynamic systems. The tree representation provides insight into the behavior of the stochastic process, particularly the tree allows for understanding the effect of temporal dependence separately from the time transformation effects of the underlying martingale.

Given a recombining tree, dynamic programming (DP) techniques are applied to obtain optimal policies and prices for maximizing an objective function. Many DP methods, such as backward inductions, are well understood, fast, and exact. (See Luenberger [30]) The accuracy of the solutions is determined by the number of periods, n, in the tree. We found that as $n \to \infty$, the tree process converges to the continuous stochastic process at a rate of $1/\sqrt{n}$. We demonstrate how this tree is constructed for the Vasicek interest rate model and price an American style put option.

Chapter 1

Introduction

> "Dam design was an important task in the nineteenth century, but one in which - like finance today - the mathematically easy path was preferred. Engineers assumed flood variations from one year to the next were statistically independent, as with Bachelier's coin-tossing [which is the basis of modern option pricing theory]...With coins you can get a run of heads or tails...and there is a simple formula for it: The range [of the height over] time varies by the square root of the number of tosses. [Harold Hurst] found the range from the highest Nile flood to the lowest widened faster than the coin-tossing rule predicted. The highs were higher and the lows, lower. But the problem was not the individual floods; looked at singly, the bell curve fit the data on each year's flooding reasonably well. Apparently, it was the runs of weather - the back-to-back floods or droughts - that were changing the game. It seems obvious, now: Not just the size of the floods, but also their precise sequence, matters."

(Mandelbrot "The (Mis)Behavior of Markets" [32] , p.177)

The excerpt above is not an isolated case, but an example of how modeling techniques tend to be oversimplified for the sake of tractability and convenience; however the result of such assumptions can be catastrophic. Statistical techniques on independent random variables are well understood and therefore have a tendency to be commonly used in financial modeling. The assumption of independence, even in the Gaussian case, can result in exposures to unrealized risks, failures, and misvaluations. The Nobel prize winning Black-Scholes-Merton option pricing

model assumes that equity log returns evolve independently. Modeling with the assumption of independence has played critical roles in the formation of portfolio insurance (which resulted in the crash of 1987), the loss of $4.6 billion in less than four months by the hedge fund Long-Term Capital Management (which threatened a global collapse after the 1998 Russian debt default), and the mortgage misvaluations that contributed to the 2008 financial crisis. Many argue that the modeling assumption of the Gaussian distribution is the problem (an argument of "heavy tails"), however, as with the flood variations of the Nile river, temporal dependence can incorporate the Gaussian assumption and still result in extreme fluctuations beyond the model's prediction of risk.

Recently, electronic trading technology and algorithms have given investors the ability to execute trades within 128 microseconds. Speed and accuracy of models and their estimates are crucial to risk management, quantitative valuation and security trading. Therefore, techniques which are both fast and accurate on small data sets are needed in today's marketplace. The main goal of this thesis is to develop and explore the tools for the financial application of two classes of continuous temporally dependent Gaussian processes; fractional Brownian motion and Gaussian Markov (GM) process. The tools presented in this thesis allow for quick and accurate estimation of model parameters, generation of efficient discrete representations and the understanding of the influence of dependency on pricing, optimal strategies and risk management.

In regards to risk management, the standard models engineers used for determining dam height worked around the world, but even with large safety factors, the Nile overcame the dam walls. In 1951, British hydrologist Harold Hurst documented the long-range dependence of water levels in reservoirs along the Nile river. He modeled the reservoir storage capacity using a fractional Brownian motion process. Fractional Brownian motion (fBm) was first introduced by Russian probabilist Andrey Kolmogorov in 1940. The auto-correlation function, and strength of temporal dependence, for fBm is defined by a single parameter, $H \in (0, 1)$, known as the Hurst index or exponent.

Applications of fBm are diverse, for example, it is commonly used to predict the vortex behavior in the turbulent flow of non-Newtonian (high Reynolds number) fluids. The process has also shown to be a reasonable model for solar activity, log returns of stock prices, geographical temperature, commodity prices in liberated markets, heart rate variability, and network trafficking. Mandelbrot popularized the use of fBm as a more suitable process for stock price modeling than the classical Brownian motion process, which has independent increments.

In 1997, Robert Merton and Myron Scholes received the Nobel prize in Economics for a model of stock pricing based on geometric Brownian motion. Under the Brownian motion model, relative log returns are independent identically distributed as Gaussian. Practitioners typically find the empirical distribution of relative log returns in the stock market assuming returns are independent or stationary. A comparison between the empirical and Gaussian distributions demonstrate that Brownian motion does not exhibit the "fat tails" seen in the empirical distribution (Cont [12]). Cont [12] also demonstrates that the relative log returns in the equity market exhibit signs of temporal dependence and volatility clustering, which are the predominant advantages of a fBm process over a Brownian motion process. These properties were also observed by Mandelbrot [32, 33] and are fundamental in his arguments for the use of of fBm in finance.

Given the financial crisis of 2008, there has been large debates over the use of Gaussian processes in finance. However, studies like those from Cont [12] on "heavy tails," and distributions draw conclusions from which they assume log returns are stationary and/or independent and identically distributed (i.i.d.) over a long time frame. Analysis of the markets in the presence of temporal dependence is difficult because of a need for knowledge about the auto-correlation structure. Without the assumption of stationarity, analysis becomes even more difficult because of constant distribution changes over time. Market behavior is effected by factors like news, events or policy which impact the market and cause a shift in valuation (Cont [12]). Jumps, non-stationarity and temporal dependence impact the perceived distribution. Given these observations, many models have both a continuous process coupled with a jump process. In this thesis we focus on developing tools for modeling the continuous path process that model "typical" movements.

This dissertation is structured in the following way. Chapter 2 summarizes the definitions and properties of Gaussian processes that are used throughout the dissertation. In Part I we review the application of fBm in finance and develop estimation techniques for the memory parameter (known as the Hurst index) and discuss the effect of dependence on pricing methods and properties. Currently there are many techniques to estimate the Hurst index. (See §3.1) However, the most accurate technique, Whittle's approximate MLE, takes considerable time to compute, while faster algorithms lack comparable precision. (Taqqu et al. [51]) The new methods in this thesis address the need for fast and accurate methods on the small sample sizes of an ever changing market. In Chapter 3, we summarize leading estimation techniques and introduce new estimators of the Hurst index based on ergodic theory for stationary processes. A numerical

comparison of our new estimators to Whittle's approximate MLE concludes that our techniques are highly competitive (in terms of RMSE) and explicitly faster because of reduced complexity (see Appendices I, II, and III for analysis results). We also develop techniques to estimate all parameters in a fractional Wiener process (fractional Brownian motion with drift and volatility). Results indicate that our techniques allow for faster and more accurate computation of the Hurst index on smaller data sets. In §4.3, we apply our new estimation techniques to three financial indexes to provide further empirical evidence that the Hurst index has varied significantly over time.

Chapter 4 begins with a literature review of the fBm call option pricing equations, arbitrage issues and binomial tree representations. To further the understanding of the influence of dependence, §4.2 investigates the consequences of using the standard Brownian motion model in the presence of a fBm model. Through analysis, we argue that processes with long-range dependence, particularly fBm, can appear to have larger tails when analyzed as if the process is a series of independent random variables, even though the model's distribution is still Gaussian. This result indicates that the assumption of independence in standard methods can underestimate the riskiness of a financial instrument when long-range temporal dependence is present. If the Hurst index is known, we show that it is trivial to formulate the exact affect the dependence has on the misvaluations of risk. This, in turn, influences the valuation of financial instruments. An analysis of the affect of dependence on option pricing is also presented in this section. To demonstrate how the dependence structure changes the Black-Scholes-Merton call option prices, we use the Hu and Oksendal [21] pricing equations for fBm. Other anomalies, like the term structure of implied volatility in the derivatives market, may also be explained by long-range dependence.

We conclude Part I with an application of our methods to real index data to show how the Hurst index has behaved over time. (§4.3) Bayraktar et Poor [3] estimated the Hurst index for S&P 500 from 1989 to 2000 and discovered that the Hurst index appears to take values as high as $H \approx 0.7$ and as low as $H \approx 0.48$. Willinger et Taqqu [57] perform a similar analysis in which they conclude that the stock market (on average) displayed significant signs of long range dependence ($H > 1/2$). By comparing three differently weighted portfolios (CRSP indexes), we find that the Hurst index may be influenced by the trade volume, capitalization or liquidity (measured by the bid-ask spread).

Despite the many applications of fBm, the popularity of fBm has been limited by a few barriers in theory and implementation of the model in finance. These barriers include, but are

not limited to:

- fBm admits arbitrage and therefore prevent proper risk management and hedging strategies that are necessary for implementation. (Rogers [44])
- Binomial trees are difficult to construct and do not recombine, making pricing of many American style and exotic options computationally expensive and inaccurate. (Konstantopoulos and Skhanenko [26])
- Pricing theory is derived using Riemann-Stieltjes integrals (or Wick Calculus) from which replicating portfolios and self-financing strategies have no economic meaning. (Sottinen [47])
- Market prices show times of non-stationarity. (Cont [12])

In an attempt to expand beyond the limitations of fBm, Part II of this thesis focuses on continuous Gaussian processes of the Markov type. General diffusion models in physics, thermodynamics, fluid mechanics, biology and finance are all Markovian and many are Gaussian. These processes are used in modeling and decision making under uncertainty. In finance, almost all interest rate models are Markovian. Gaussian Markov (GM) processes allow for a wide range of properties from stationary to non-stationary, and long-range dependence, short-range dependence or memoryless.

In Chapter 5 we review critical properties and theorems needed to understand Gaussian Markov properties. All the results in Part II use Hida's [19] representation of GM processes (Theorem 5.3). Hida's theorems establish that continuous Gaussian Markov processes can be constructed by the product of a scaling function and a Gaussian martingale. The scaling function can be shown to completely control the dependence structure of the process (Proposition 5.17), while the martingale controls the speed in which a Brownian motion process evolves over time (Theorem 5.7).

Chapter 6 develops a consistent estimator of the diffusion parameter using the quadratic variation of a GM process. The first section gives a review of the subject with well known examples of the quadratic variation for fBm and Brownian motion. In §6.2, we explicitly derive a closed form representation of quadratic variation for GM processes and prove that the quadratic variation can be represented as a Riemann-Stieltjes integral. We establish consistency of estimators and use Borel-Cantelli lemma to obtain almost sure convergence and confidence interval bounds on a fixed interval. This result allows for the use of quadratic variation to estimate model parameters.

To conclude the chapter we apply the quadratic variation estimator to the Ornstein-Uhlenbeck process, where we derive a closed form representation of the MLE for all model parameters and demonstrate that the quadratic variation estimator is more accurate on small to moderate size sample sizes.

Chapter 7 focuses on pricing and risk management with GM processes. A review of the influence of quadratic variation in stochastic calculus is given. The quadratic variation representation derived in Chapter 6 is used to modify option pricing theory and stochastic calculus to general representations of a geometric GM stock model. Hedging strategies, call option prices, and Kolmogorov backwards equations are derived for risk management and valuation practices.

Optimal policy and valuation of American style derivatives (contracts that can be exercised before expiration) typically cannot be explicitly derived. In fact, if a derivative payoff structure is non-convex or zero when the price of the underlying security is zero, then there is no guarantee that a closed form solution exists. (Shreve [46]) Additionally, it is possible (and in most cases likely) that the optimal strategy is to execute the contract early based on the path the security takes over time. Chapter 8 introduces techniques to aid in valuation of such dynamic systems through discrete representations. In cases of path dependent options, it is common practice to use simulation methods to evaluate optimal policy and pricing. In this motivation, we introduce an exact and fast method of generating simulations of GM processes. This work is presented in §8.2.

In §8.3 we use the central limit theorem for stochastic processes established by Andersen and Dobrić [2] to derive a binomial tree representation for general Gaussian Markov processes and prove that the tree process weakly converges to a continuous Gaussian Markov processes. Unlike the fBm tree, the tree is recombining at each node; the size of the tree increases linearly with the number of periods. Additionally, the algorithm to generate the GM process tree is simple and fast. An example of a binomial tree for the Vasicek interest rate model is provided. The Vasicek interest rate model utilizes the Ornstein-Uhlenbeck process which is mean reverting, path dependent, and non-stationary. The influence of the dependence, drift, and diffusion of the processes can be analyzed seen explicitly in our construction, allowing for further insights into a GM process. Most importantly, the tree structure allows for analysis with standard Dynamic Programming (DP) methods. DP methods, like backwards induction, are well understood, fast and exact. We apply the backwards induction method to the Vasicek tree to price an interest rate put option and find the optimal execution times and hedging strategies.

Chapter 2

Background

This chapter provides a review of various concepts, theory, terminology and notation that will be used throughout this thesis. The chapter is divided into five sections. Section 2.1 presents definitions and properties of Gaussian processes, §2.2 provides formal definitions and properties of fractional Brownian motion, §2.3 reviews simulation techniques for Gaussian processes and provides sample paths of fBm to aid in understanding the influence of fBm's Hurst index. A brief review of Martingale Theory is provided in §2.4. If the reader is familiar with these topics, they may skip this chapter and return to it as needed for reference.

2.1 Introduction to Gaussian Self-Similar Processes

Definition 2.1. (Gaussian Process)

A real valued stochastic process $(X_t)_{t\in T}$ is called *Gaussian*, if for any $t_1, t_2, \ldots, t_n$ in T the random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ are jointly normal.

Remark 2.2. A Gaussian process $(X_t)_{t\in T}$ is called centered if $\mathbb{E}[X_t] = 0$ for all $t \in T$.

Remark 2.3. A Gaussian process $(X_t)_{t\in T}$ is completely characterized by it covariance $\mathbb{E}[X_t X_s], \forall t, s \in T$.

Definition 2.4. (Self-Similar Process)

Let $X = (X(t))_{t\in T}$ be a stochastic process on probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X is said to be a *self-similar* process if for any $a > 0$, there exists a $b > 0$ such that for any $t_1 < t_2 < \cdots < t_n$ in T, the following holds

$$(X(at_1), X(at_2), \ldots, X(at_n)) \overset{d}{=} (bX(t_1), bX(t_2), \ldots, bX(t_n)),$$

where $\overset{d}{=}$ means equality of the distributions of the left hand side and the right hand side vectors.

Corollary 2.5. *(H Self-Similar)*

If $(X_t)_{t\in T}$ is a non-trivial, stochastically continuous at $t = 0$, and is a self-similar process, it can be shown that there exists a unique $H \geq 0$ such that $b = a^H$. Such a process is called H *self-similar*. It can be shown that if $H \geq 0$ and $(X_t)_{t\in T}$ is a H self-similar process, then $X_0 = 0$ almost surely.

Definition 2.6. (Brownian Motion)

Brownian motion (Bm) is a process $(B(t))_{t\geq 0}$ with the following properties:

1. $B(0) = 0$, almost surely
2. $B(t)$ has independent increments: $0 \leq t_1 < t_2 < \cdots < t_n$ then
   $(B(t_n) - B(t_{n-1})), \ldots, (B(t_2) - B(t_1))$ are independent
3. $B(t) - B(s)$ distributed as $N(0, |t - s|)$
4. $t \to B(t)$ is continuous, almost surely.

Remark 2.7. From the properties of Brownian motion, it follows that $B(at) \overset{d}{=} a^{\frac{1}{2}} B(t)$, $\forall a > 0$. Therefore, Brownian motion is a *centered Gaussian* $1/2$ *self-similar process*. Moreover, since $\mathbb{E}[B(t)]^2 = t$ then $\mathbb{E}[B(t) B(s)] = \min\{t, s\}$.

Remark 2.8. If we let $\Delta t \equiv t_j - t_{j-1}$ for all $j = 1, 2, \ldots, N$, then the increments of Brownian motion

$$(Z(t_i))_{i=1}^N \equiv (B(t_j) - B(t_{j-1}))_{j=1}^N \overset{i.i.d.}{=} N(0, \Delta t)$$

are known as *white noise* (Wn), where *i.i.d.* means independent identically distributed.

Definition 2.9. (Long/Short-Range Dependence)

Let $(X_t)_{t\in T}$ be a centered Gaussian process and let the auto-covariance between the n^{th} increment and the first increment of the process X be denoted as $\gamma_n = \mathbb{E}[(X_1 - X_0), (X_{1+n} - X_n)]$, $n \geq 1$. Then,

1. If $\gamma_n > 0$ for all $n \geq 1$, the process has positively correlated increments. If $\sum_{n=1}^{\infty} |\gamma_n| = \infty$, we say that X has *long-range dependence*.

2. If $\gamma_n < 0$ for all $n \geq 1$, the process has positively correlated increments. If $\sum_{n=1}^{\infty} |\gamma_n| = c < \infty, c \neq 0$, we say that X has *short-range dependence*.

3. If $\gamma_n = 0$ for all $n \geq 1$, $\sum_{n=1}^{\infty} |\gamma_n| = 0$ and we say that X is an independent process.

2.2 Definition and Properties of Fractional Brownian Motion

Definition 2.10. (Fractional Brownian Motion)

Fractional Brownian motion (fBm) is a centered Gaussian process $\left(B^H(t)\right)_{t\in\mathbb{R}}$ where $H \in (0,1)$ with the following properties:

1. $B^H(0) = 0$, almost surely

2. $B^H(t) - B^H(s)$ is distributed as $N\left(0, |t-s|^{2H}\right)$

3. $t \to B^H(t)$ is continuous, almost surely

Corollary 2.11. Since $\mathbb{E}\left[B^H(0)\right] = 0$ and $\mathbb{E}\left[B^H(t)\right]^2 = t^{2H}$,

$$\begin{aligned}\mathbb{E}\left[B^H(t)\,B^H(s)\right] &= \frac{\mathbb{E}\left[B^H(1)\right]^2}{2}\left\{\mathbb{E}\left[B^H(t)\right]^2 + \mathbb{E}\left[B^H(s)\right]^2 - \mathbb{E}\left[B^H(|t-s|)\right]^2\right\} \\ &= \frac{\mathbb{E}\left[B^H(1)\right]^2}{2}\left\{t^{2H} + s^{2H} - |t-s|^{2H}\right\}\end{aligned}$$

Remark 2.12. Throughout this thesis without loss of generality we assume a *standard* fractional Brownian motion, that is $\mathbb{E}\left[B^H(1)\right]^2 = 1$.

Remark 2.13. If we let $\Delta t \equiv t_j - t_{j-1}$ for all $j = 1, 2, \ldots, N$, then the increments of fractional Brownian motion

$$\left(Z^H(t_i)\right)_{i=1}^N \equiv \left(B^H(t_i) - B^H(t_{i-1})\right)_{i=1}^N \overset{d}{=} \left\{N_i\left(0, \Delta t^{2H}\right)\right\}_{i=1}^N,$$

where

$$\begin{aligned}\mathbb{E}\left[Z_{t_i}^H Z_{t_j}^H\right] &= \mathbb{E}\left[B_{t_i}^H B_{t_j}^H\right] - \mathbb{E}\left[B_{t_i}^H B_{t_{j-1}}^H\right] - \mathbb{E}\left[B_{t_{i-1}}^H B_{t_j}^H\right] + \mathbb{E}\left[B_{t_{i-1}}^H B_{t_{j-1}}^H\right] \\ &= \frac{\Delta t^{2H}}{2}\left\{|i-j+1|^{2H} + |i-j-1|^{2H} - 2|i-j|^{2H}\right\}\end{aligned}$$

$\left(Z_{t_i}^H\right)_{i=1}^N$ are known as *fractional Gaussian noise* (fGn).

Corollary 2.14. Fractional Brownian motion is a *Gaussian H self-similar* process, that is

$$\mathbb{E}\left[B^H(at)\, B^H(as)\right] \overset{d}{=} a^{2H}\mathbb{E}\left[B^H(t)\, B^H(s)\right], \forall t, s \in \mathbb{R}$$

Remark 2.15. Another property of fractional Brownian motion is that when $H = 1/2$,

$$\mathbb{E}\left[B^{\frac{1}{2}}(t)\left(B^{\frac{1}{2}}(t) - B^{\frac{1}{2}}(s)\right)\right] = 0,\ \forall t > s.$$

Fractional Brownian motion with $H = 1/2$ satisfies the definition of Brownian motion (definition 2.6). Additionally, when $H = 1$ the process is degenerate since,

$$\mathbb{E}\left[\left(B^H(t) - tB^H(1)\right)\right] = t^2 - 2t^2 + t^2 = 0 \implies B^H(t) \overset{d}{=} tB^H(1), \text{ almost surely.}$$

Definition 2.16. (First Order Stationary Process)

A stochastic process is defined as a first-order stationary process if its probability density function remains the same regardless of any shift in time, that is, given a time shift τ, the process $\{X_t\}_{t\in T}$ is a *first-order stationary process* if it satisfies the equation

$$P(X_t \leq x) = P(X_{t+\tau} \leq x),\ \forall t \in T.$$

Corollary 2.17. *Fractional Brownian motion has stationary increments.*

Proof. Since fBm is a centered Gaussian processes, $\forall t > s, \tau > 0$ we only need to consider the covariance function to prove the stationarity of increments,

$$\begin{aligned}\mathbb{E}\left[\left(B_{t+\tau}^H - B_\tau^H\right)\left(B_{s+\tau}^H - B_\tau^H\right)\right] &= \mathbb{E}\left[B_{t+\tau}^H B_{s+\tau}^H\right] - \mathbb{E}\left[B_{t+\tau}^H B_\tau^H\right] - \mathbb{E}\left[B_\tau^H B_{s+\tau}^H\right] + \mathbb{E}\left[B_\tau^H\right]^2 \\ &= \frac{1}{2}\left\{t^{2H} + s^{2H} - (t-s)^{2H}\right\} \\ &= \mathbb{E}\left[B_t^H B_s^H\right].\end{aligned}$$

□

This proves that $\left(B^H_{t+\tau} - B^H_\tau,\, t \in T\right) \overset{d}{=} \left(B^H_t,\, t \in T\right)$. (Embrechts[14], pp.7)

Remark 2.18. Using stationarity it can be shown that the auto-covariance function for fBm is given by

$$\gamma_n = \frac{1}{2}\left[(n+1)^{2H} - 2n^{2H} + (n-1)^{2H}\right]$$

therefore

$$\gamma_n \approx H\,(2H-1)\,n^{2H-2},\ \text{as } n \to \infty, H \neq \frac{1}{2}.$$

Notice that when

1. $H = \frac{1}{2}, \gamma_n = 0, \forall n$ therefore fractional Brownian motion has independent increments.

2. $H > \frac{1}{2}, \gamma_n > 0, \forall n$ and $\gamma_n \approx H\,(2H-1)\,n^{2H-2}$, as $n \to \infty$ therefore the increments of the fBm process are positively correlated and by p-series $\sum_{n=1}^{\infty} |\gamma_n| = \infty$, therefore has *long-range dependence.*

3. $H < \frac{1}{2}, \gamma_n < 0, \forall n$ and $\gamma_n \approx H\,(1-2H)\,n^{2H-2}$, as $n \to \infty$ therefore the increments of the fBm process are negatively correlated and by p-series $\sum_{n=1}^{\infty} |\gamma_n| = c < \infty$, therefore has *short-range dependence.*

Remark 2.19. For $\frac{1}{2} < H < 1$, H measures the intensity of long-range dependence. The closer H is to 1 the stronger long-memory the process exhibits.

2.2.1 Fractional ARIMA Processes and fBm

Definition 2.20. (ARMA)
A discrete time series $X = \{X_t : t = 1, 2, \ldots, N\}$ is an *Autoregressive Moving Average process,* $ARMA\,(p, q)$, if it is stationary and satisfies

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where Z_t, $t = 1, 2, \ldots, N$ are i.i.d. $N\left(0, \sigma^2\right)$ random variables and where ϕ and θ are polynomials of degrees p and q, respectively, satisfying $\phi + \theta \neq 0$ and the coefficients of ϕ_i and θ_i must lie inside the unit circle.

An alternative representation of an $ARMA$ process is given by

$$\phi(B) X_t = \theta(B) Z_t,$$

where B is known as the backward shift operator and is defined as,

$$B^j X_t \equiv X_{t-j},$$

and $\phi(B)$ and $\theta(B)$ are the p^{th} and q^{th} degree polynomials of the backward shift operator

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q. \end{aligned}$$

(Brockwell [6], pp.55)

Definition 2.21. (ARIMA)

An *Autoregressive Integrated Moving Average process,* $ARIMA(p, d, q)$, is defined as a process which has the form

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t,\ d \in \mathbb{N}.$$

$ARIMA$ processes can be used to represent non-stationary series and are used to incorporate trends. (Brockwell [6], pp.180)

The $ARIMA$ process can be generalized to non-integer d values, as in the next definition.

Definition 2.22. (FARIMA)

A *Fractional Integrated* $ARIMA(0, d, 0)$, $FARIMA(0, d, 0)$, process is defined by

$$X_t = (1 - B)^{-d} Z_t = \sum_{j=0}^{\infty} b_j Z_{t-j}.$$

According to Doukhan [13] if z is a complex number in the region $|z| < 1$ and $d \in \mathbb{R}$ then

$$(1 - z)^{-d} = \sum_{j=0}^{\infty} b_j z^j,$$

where $b_0 = 1$ and

$$b_j = \prod_{k=1}^{j} \frac{k-1+d}{k} = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \, j \in \mathbb{N}.$$

The Gamma function is defined as $\Gamma(p) \equiv \int_0^\infty t^{p-1}e^{-t}dt$, $p > 0$. Using the property $\Gamma(p) = p\Gamma(p)$ and Stirling's formula

$$\Gamma(p) \approx \sqrt{2\pi}e^{-p+1}(p+1)^{p-\frac{1}{2}}, \text{ as } p \to \infty,$$

it can be shown

$$b_j \approx (\Gamma(d))^{-1} j^{d-1}, \text{ as } j \to \infty.$$

Remark 2.23. Doukhan [13] shows that there is a decomposition of d for every FARIMA process has $-\frac{1}{2} < d - \lfloor d + \frac{1}{2} \rfloor < \frac{1}{2}$, where $\lfloor . \rfloor$ denotes the smallest integer. This means that d is defined on the basic range $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and therefore it can be shown that when $0 < d < \frac{1}{2}$, the process has long-range dependence and when $-\frac{1}{2} < d < 0$ the process has short-range dependence. (Doukhan [13], pp.19)

Corollary 2.24. *The Gaussian $FARIMA(0,d,0)$ process $\{X_t : t = 1, 2, \ldots, N\}$ with $d = H - \frac{1}{2}$ converges in distribution to fractional Brownian motion $\left\{B_t^H\right\}_{t\in\mathbb{R}}$. As $n \to \infty$,*

$$\left(\frac{1}{n^H}\sum_{t=1}^{[ns]} X_t\right)_{s\in[0,T]} \xrightarrow{w} \left(B_s^H\right)_{s\in[0,T]},$$

where $\xrightarrow{w}$ means the discrete process weakly converges to the continuous time process. (Doukhan [13], pp.21)

In the next section an algorithm to recursively simulate fractional Brownian motion using the $FARIMA$ representation to compute a partial auto-correlation function.

2.3 Simulation of Gaussian Processes

Simulation and prediction algorithms play a crucial role in modeling complex systems. Appendix IV reviews the three main algorithms that can be used to simulate Gaussian processes. Here we outline the main idea behind the algorithms for simulation of Gaussian processes.

The first algorithm is Choleski decomposition. Choleski decomposition can be used to exactly simulate (or predict) any Gaussian process. The major drawback of Choleski's procedure is its requirement to store and invert the entire auto-covariance matrix. The Lower Upper decomposition takes $N^3/3$ floating point operations, thus limiting the size of the simulations that can be performed. (Doukhan [13], pp.581)

The next two algorithms for simulating Gaussian processes are known as Linear Prediction algorithms. They include the Durbin-Levinson algorithm and the Innovations algorithm. The power of these algorithms is in their ability to recursively compute a one-step predictor and corresponding error function (errors are normally distributed). Each algorithm uses the auto-covariance function to create a partial auto-correlation function (a linear orthogonal decomposition of the auto-covariance function) and all the past observations in the time series to predict the next step. These algorithms are much more computationally efficient and faster than Choleski decomposition, allowing for large scale simulations.

The Durbin-Levinson algorithm is formulated for stationary time series and uses the Toeplitz structure of all stationary processes to form discrete convolution operations as matrix multiplication. This also means that only the first row (or column) of the auto-covariance matrix needs to be stored. The speed of the Durbin-Levinson algorithm is $\mathcal{O}\left(N^2\right)$,much faster and computationally efficient than Choleski decomposition methods which have complexity $\left(\mathcal{O}\left(N^3\right)\right)$. (Doukhan [13], pp.581)

The Innovations algorithm is formulated for any Gaussian time series. This algorithm is a generalization of the Durbin-Levinson algorithm, however has one main difference; linear predictors are formed on the forecast errors instead of the time series itself. These predictors are known as "innovations." Innovations are necessary because without stationarity, the auto-covariance matrix is no longer Toeplitz. The innovations allow for discrete convolution operations as matrix multiplication on the entire auto-covariance. Recursion of the algorithm is still maintained. The algorithm is slightly slower than the Durbin-Levinson algorithm, but still has the same complexity. (See Appendix IV for further details)

2.3.0.1 Simulations of Fractional Brownian Motion with the Durbin-Levinson Algorithm

By definition 2.10, the increments of fBm $\left(Z_{t_i}^H\right)_{i=1}^{N-1}$ (known as fractional Gaussian noise) is a centered stationary Gaussian process. Therefore, the Durbin-Levinson algorithm can be used to

generate simulations of fractional Gaussian noise. The cumulative sum of the fGn up to time t_i gives a simulation of the fractional Brownian motion process at time t_i:

$$B^H_{t_{n+1}} = \sum_{i=1}^{n} Z^H_{t_i},\ n = 0, \ldots, N-1,\ t_{n+1} = n\Delta t.$$

Below we use the Durbin-Levinson algorithm to generate a fractional Brownian motion time series of length $N = 1000$, where $\Delta t = 1/1000$ for nine different values of the Hurst index. The left graphic in Figure 2.1 depicts the simulation results for the short-range dependent process and Brownian motion, $H = 0.1, 0.2, \ldots, 0.5$, while the right graphic shows the results for the long-range dependent process and Brownian motion, $H = 0.5, 0.6, \ldots, 0.9$. The nine paths were generated in Matlab® using the method of common random numbers (where each path utilizes an identical set of standard normal r.v.s). Note that the Figures are on different scales.

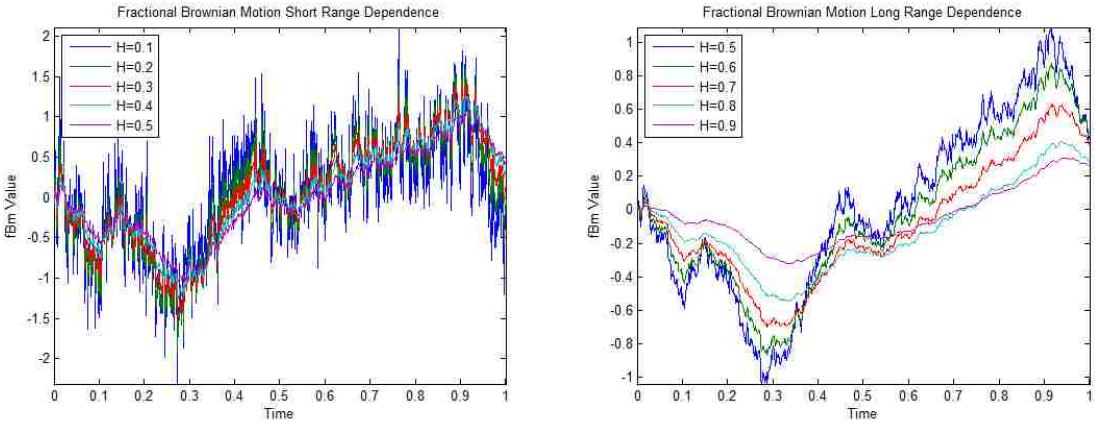


Figure 2.1: Simulations of fBm via Durbin Levinson

Figure 2.2 exhibits the fractional Gaussian noise processes and the resulting paths of the fractional Brownian motion (on different scales) for detail and comparison.

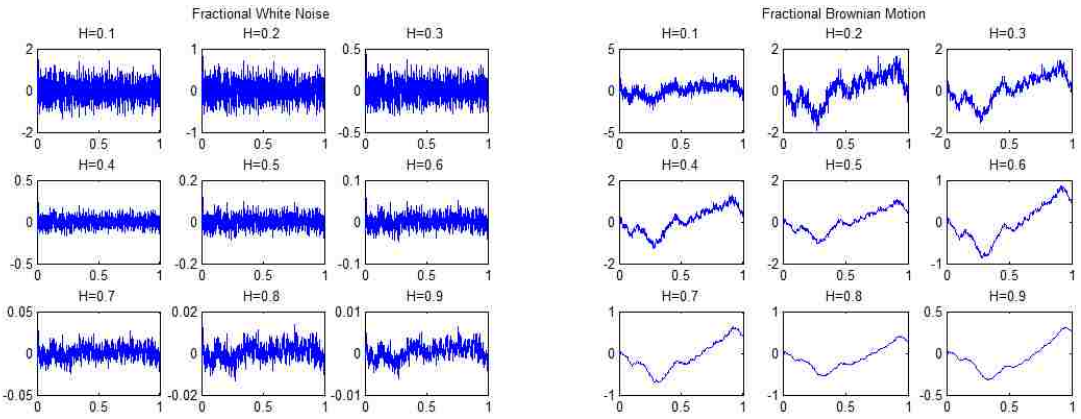


Figure 2.2: Simulations of fractional Gaussian noise via Durbin Levinson

Comparing these paths we notice that low values of the Hurst index result in a process whose sample paths oscillate wildly, while higher values of the Hurst index result in a smoother process. Additionally, the long-range dependence causes clusters of trends (or momentum effects) where the paths have a tendency to continue in a particular direction. Finally, notice that the terminal values and extreme values of the paths are significantly influenced by the Hurst index.

2.4 A Brief Review of Martingale Theory

Definition 2.25. (Martingale)

A collection of σ-algebras $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathbb{R}}$ such that if $s < t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$ is known as a filtration. The process $X_t = \{X\}_{t\in\mathbb{R}}$ is adapted to the filtration $\mathcal{F}$ if X_s is $\mathcal{F}_t$-measurable for every $s \leq t$. The stochastic process X is a martingale with respect to the σ-algebra $\mathcal{F}$ if

1. $\mathbb{E}\left[|X_t|\right] < \infty, \forall t < \infty$.
2. $\mathbb{E}\left[X_t \,|\mathcal{F}_s\right] = X_s, \forall t \geq s$.

Lemma 2.26. *Brownian motion is a martingale. (This falls directly from definition 2.6.)*

Remark 2.27. Fractional Brownian motion is not a martingale unless $H = \frac{1}{2}$. When $H < \frac{1}{2}$, Embrechts [14] shows that the fBm process, $\left(B_t^H\right)_{t\in\mathbb{R}}$ has infinite p variation on all compact intervals. Additionally, when $H > \frac{1}{2}$ the quadratic variation of B^H is a zero process (we provide our own proof in §6.1.2). Embrechts furthermore shows that fBm with $H > \frac{1}{2}$ does not have a Doob-Meyer decomposition $B_t^H = M_t + V_t$, where $(M_t)_{t\in\mathbb{R}}$ is a local martingale and $(V_t)_{t\in\mathbb{R}}$ is

a finite variation process. Therefore, fBm for $H > \frac{1}{2}$ cannot be a semi-martingale. Embrechts also demonstrates that when $H < \frac{1}{2}$ the quadratic variation is infinite (see §6.1.2 for our proof). Without finite quadratic variation, processes cannot be a semi-martingale. The semi-martingale property is important to the development of the option pricing theory. Since fractional Brownian motion is not a semi-martingale, Itô calculus does not hold for fBm.

Part I

Fractional Brownian Motion

Chapter 3

New Estimators of the Hurst Index in Fractional Brownian Motion

3.1 Literature Review: Estimators of the Hurst Index in fBm

Estimation of the Hurst index (typically denoted as H) is a well studied problem. A plethora of estimation techniques have been developed. In today's marketplace, electronic trading has dramatically increased the volume and speed of trades. The ability to execute trades in as little as 126 microseconds has created a highly competitive market. To gain a competitive edge, estimation techniques need to be both fast and accurate. Currently, there is a disproportionate trade-off between the accuracy of an estimate of the Hurst index and the time it takes to compute. Many techniques to estimate the Hurst index are computationally fast and simple, but typically have slow convergence rates and wide confidence intervals. (Taqqu et al. [51]) Additionally, some of these techniques are unable to detect short range dependence ($H < 1/2$) effectively. Whittle's approximate MLE, though quite accurate, is a computationally demanding optimization algorithm and therefore requires significant time to compute. We give an introduction to these techniques to illustrate the different methods for estimating H. For a description of these techniques refer to Taqqu et al. [51] in which he simulated 50 sample paths with $N = 10,000$ realizations of fBm for $H = 0.5, 0.6, \ldots, 0.9$. In this paper, Taqqu estimates the Hurst index using various estimation techniques and compares the accuracy of the estimates. There are three main types of estimators

of the Hurst index for fractional Brownian motion.

1. Aggregated Processes Analysis; Variance and Absolute convergence methods
2. Time Domain based Analysis; Rescaled range (R/S) method
3. Frequency based Analysis; Periodogram/Spectral methods, Wavelet methods and Whittle's MLE

Each of the methods for estimating the Hurst index is derived using the properties of the fractional Gaussian noise time series. The aggregated methods typically transform the time series through the aggregate sum of a function of the time series. The size of the summations, m, is increased such that a linear regression of the log of the aggregation and m yield the Hurst index as the slope. In §3.1.2 we show the Variance of Residuals method as an example of an aggregate statistic for estimating the Hurst index.

Many of the time domain methods take advantage of stationarity and the fact that the auto-covariance function of fractional Brownian motion satisfies

$$\gamma_n \approx H\left(2H-1\right) n^{2H-1}, \text{ as } n \to \infty.$$

Frequency domain methods often take advantage of the fact the Periodogram is an unbiased estimator of the spectral density function. The Periodogram of a time series $\{Y_t\}_{t=1}^{N}$is defined by

$$I\left(\lambda_k\right) = \frac{1}{N}\sum_{t=1}^{N}\left|Y_t e^{-it\lambda_k}\right|^2,$$

where λ is defined as the Fourier frequencies such that $\lambda_k = \frac{2\pi k}{N}$. The spectral density f, derived as a Fourier transform of γ, is defined as

$$f\left(\lambda\right) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_n e^{-i\lambda k},\ \lambda \in \left[-\pi, \pi\right].$$

The estimators in the frequency domain typically take advantage of the property

$$f\left(\lambda\right) \approx c_H \lambda^{1-2H}, \text{ as } \lambda \to 0,$$

where c_H is a constant dependent on H.

The empirical study in Taqqu et al. [51] compares the Root Mean Square Error (RMSE) of various estimators of the Hurst index (except for Wavelet methods). In their study they identify Whittle's approximate MLE as the most accurate and the Peng's Variance of Residuals method as the second most accurate. Bayrakatar et al. [3] evince a Wavelet based estimator that is equivalent to or better than Whittle's approximate MLE. Below we introduce these three estimators of the Hurst index.

3.1.1 Whittle's Approximate MLE

The Whittle estimator of the Hurst index is an approximation of the Maximum Likelihood Estimator (MLE), $L(H)$. In 1961, Whittle [56] derived the MLE using an orthonormal basis for a stationary Gaussian time series. Whittle's MLE is a function of the Periodogram $I(\lambda)$ with respect to the time series and the spectral density function $f(\lambda, H)$, where λ are Fourier frequencies in $[-\pi, \pi]$. Whittle shows that the Likelihood equation for stationary Gaussian processes is defined by

$$L(H) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, H)} d\lambda + \int_{-\pi}^{\pi} \log f(\lambda, H)\, d\lambda,\ \lambda \in [-\pi, \pi].$$

(Brockwell [6], pp.471-474)

In order to find the best estimator of the Hurst index, $L(H)$ must be minimized with respect to H numerically using an optimization algorithm. Ledesma and Liu [27] showed that the spectral density of the fGn process can be written as,

$$f(\lambda, H) = 2\Delta t^{2H} \sin(\pi H)\, \Gamma(2H+1)\, (1 - \cos(\lambda)) \left\{ |\lambda|^{-2H-1} + B(\lambda, H) \right\},$$

where

$$B(\lambda, H) = \sum_{j=1}^{\infty} \left\{ (2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1} \right\}.$$

Ledesma and Liu split $B(\lambda, H)$ into two pieces

$$B(\lambda, H) = \sum_{j=1}^{2} \left\{ (2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1} \right\} + B_{3:\infty},$$

where

$$B_{3:\infty} = \sum_{j=3}^{\infty} \left\{ (2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1} \right\}.$$

Consistent with Paxson's [37] approximation method, Ledesma and Liu noticed that $B_{3:\infty}$ can be approximated by a linear function $D(\lambda, H)$ where,

$$D(\lambda, H) = p\lambda + q,$$

where p and q are constants that depend on H only.
Minimizing the Mean Square Error (MSE)

$$MSE = \int_0^{\infty} \left[B_{3:\infty}(\lambda, H) - D(\lambda, H) \right]^2 d\lambda,$$

p and q are found as a linear combination of equations for $F(H)$ and $G(H)$. The minimization of the MSE yields:

$$F(H) = \frac{\pi^2}{2} p + \pi q,$$

where

$$F(H) = \sum_{k=3}^{\infty} \frac{(2\pi k - \pi)^{-2H} - (2\pi k + \pi)^{-2H}}{2H}$$

and

$$G(H) = \frac{\pi^3}{3} p + \frac{\pi^2}{2} q.$$

In these equations, when $H \neq \frac{1}{2}$ then

$$G(H) = \sum_{k=3}^{\infty} \left[\frac{(2\pi k)(2\pi k + \pi)^{-2H} + (2\pi k)(2\pi k - \pi)^{-2H} - 2(2\pi k)^{-2H+1}}{2H} \right],$$

and when $H = \frac{1}{2}$, then

$$G(H) = \sum_{k=3}^{\infty} \left[\frac{2\pi k}{2\pi k + \pi} + \frac{2}{2\pi k - \pi} + \ln(2\pi k + \pi) + \ln(2\pi k - \pi) - 2\ln(2\pi k) - 2 \right].$$

Therefore,

$$\begin{aligned} p &= -\frac{6}{\pi^2}F(H) + \frac{12}{\pi^3}G(H) \\ q &= \frac{4}{\pi}F(H) - \frac{6}{\pi^2}G(H). \end{aligned}$$

Ledesma and Liu plot p and q as a function of n. They illustrated that p converges very fast, a value of $n = 20$ is sufficient for estimating p. However, q converges slowly. Ledesma and Liu deduce that a value of $n = 200$ is needed for sufficient estimation of q when $H > \frac{1}{2}$. Ledesma and Liu statistically tested the changes in the estimates for q and concluded that there is no significant improvement in the q value when n is increased from $n = 200$ to $n = 400$, for $H > \frac{1}{2}$. Ledesma and Liu recommended using a minimum of $n = 200$ for the spectral density calculation.

For quick calculations of the spectral density, Paxson [37] showed that this infinite sum can be approximated by

$$B(\lambda, H) \approx (1.0002 - 0.000134\lambda)\left\{B_3(\lambda, H) - 2^{-7.65H-7.4}\right\},$$

where

$$\begin{aligned} B_3(\lambda, H) &= \sum_{j=1}^{3}\left[(2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1}\right] + \\ &\quad + \frac{(2\pi 3 + \lambda)^{-2H} + (2\pi 3 - \lambda)^{-2H} + (2\pi 4 + \lambda)^{-2H} + (2\pi 3 - \lambda)^{-2H}}{8\pi H}. \end{aligned}$$

Using Ledesma and Liu's approximation with $n = 400$ we can see how the spectral density function varies with frequency and the Hurst parameter. The spectral density function is symmetric about the y-axis and therefore only $\lambda > 0$ is shown in Figure 3.1.

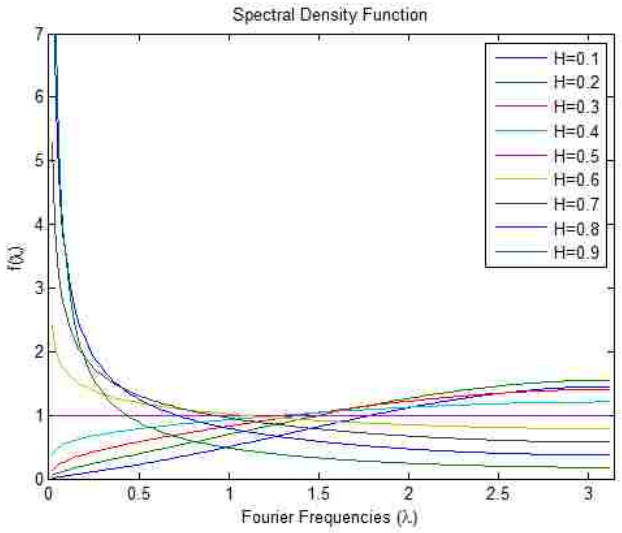


Figure 3.1: fGn Spectral Density Function

The next step in computing Whittle's approximate MLE is understanding how to compute the Periodogram for fractional Gaussian noise. Let $\{Y_j\}_{j=1}^{N}$ be a fractional Gaussian noise time series, and the Fourier frequencies be defined by,

$$\lambda_k = \frac{2\pi k}{N},\ k = 1, 2, \ldots, \frac{N-1}{2}.$$

Recall, the Periodogram $\{I\}_{\lambda\in(0,\pi)}$ is defined by

$$I(\lambda_k) = \frac{1}{N}\sum_{j=1}^{N}\left|Y_j e^{-ij\lambda_k}\right|^2.$$

Since the Periodogram is a Fourier transform of the data set, it can be calculated by using the fast Fourier transform algorithm. The Periodogram is a L^2 decomposition of the process X, that is

$$\|X\|^2 = \sum_{k=1}^{\frac{N-1}{2}} I(\lambda_k).$$

Also, note that the Periodogram is an unbiased estimate of the spectral density function where,

$$2\pi f(\lambda_k) = I(\lambda_k).$$

(Brockwell [6], pp.120,322)

Given the preceding methods of calculating the spectral density function and Periodogram, the MLE for H can now be calculated. Recall,

$$L(H) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, H)} d\lambda + \int_{-\pi}^{\pi} \log f(\lambda, H)\, d\lambda,\ \lambda \in [-\pi, \pi].$$

This function is symmetric about $\lambda = 0$. Additionally, since there is no closed form for the spectral density function, integration of $L(H)$ must be approximated numerically using Riemann sums. Therefore, the Likelihood approximation is given by

$$L(H) = \frac{1}{\pi} \left[\frac{2\pi}{N} \sum_{j=1}^{\frac{N-1}{2}} \frac{I(\lambda_j)}{f(\lambda_j, H)} + \frac{2\pi}{N} \sum_{j=1}^{\frac{N-1}{2}} \log f(\lambda, H) \right].$$

The term $\int_{-\pi}^{\pi} \log f(\lambda, H)\, d\lambda$ is a normalization of $f(\lambda, H)$. According to Taqqu et al. [52], if we normalize $f(\lambda, H)$ and call the normalized function $f^*(\lambda, H)$, then $\int_{-\pi}^{\pi} \log f(\lambda, H)\, d\lambda = 0$. This is the equivalent to $f^*(\lambda, H) = \beta f(\lambda, H)$ where β is a constant which does not depend on the other parameters. If $f(\lambda, H)$ is normalized then,

$$L^*(H) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, H)} d\lambda \approx \frac{2}{N} \sum_{j=1}^{\frac{N-1}{2}} \frac{I(\lambda_i)}{f^*(\lambda, H)}.$$

Minimization of $L(H)$ with respect to H yields the MLE estimate for the Hurst index:

$$\min_H L^*(H) = \hat{H}.$$

Additionally, Whittle showed that the first term in the objective function is an estimator for the variance of the stationary Gaussian process. Specifically he shows that the MLE of the fractional Gaussian noise process is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda, H)} d\lambda = \hat{\sigma} \Delta t^{2H}.$$

(Brockwell [6], pp.472-473)

It can be shown that $L^*(H)$ is a convex function. Therefore, to find the global minimum numerical, gradient methods like the golden section method can be efficiently applied to optimize over H. Additionally, since this is an approximation of the MLE, it can be shown that this estimate is unbiased (Brockwell [6], pp.472-473). In essence, Whittle's MLE finds the minimum

orthogonal distance between the decomposition of the estimated variance (using the Periodogram) and the decomposition of the known variance of fGn (using the spectral density representation of fGn) at each Fourier frequency. Whittle's method is an approximation of the MLE because both the spectral density function for fGn and the integrals of the Likelihood equation are approximated using Riemann sums. The computation of $L(H)$ is $\mathcal{O}\left(\frac{N-1}{2}\log N\right)$ where N is the length of the times series, however the minimization complexity is unknown and depends on the optimization tolerance. (Brockwell [6], pp.249, 471-474)

Taqqu et al. [52]stress that a major advantage of the Whittle's estimator is the ability to derive its convergence in distribution to the true Hurst parameter. It can be shown that

$$\sqrt{N}\left(\hat{H}-H\right)\rightarrow Z\sqrt{2\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{d}{dH}\log f^{*}(\lambda,H)\right)^{2}d\lambda\right]^{-1}},\text{ as } N\rightarrow\infty.$$

Taqqu et al. [51] empirically show that the Whittle Approximate MLE is far superior (in terms of mean square error and average deviation) to the existing techniques for estimating the Hurst index. Additionally, the variance or scale of the process can be computed at the same time. Whittle's estimator is asymptotically unbiased and normally distributed, however it comes at a price. The algorithm can take considerable time because the spectral density function must be recomputed at each iteration of the optimization algorithm. In §3.2.5 and Appendices I,II, and III we compare the computational time and accuracy of Whittle's estimates to techniques developed in this chapter using ergodic theory for stationary processes.

3.1.2 Peng's Variance of Residuals

Peng et al. [39] derive a Hurst index estimator based on aggregate blocks of size m, where within each block a partial sum of fGn $\{X_i\}_{i=1}^{N}$ is defined as $Y(t)=\sum_{i=1}^{t}X_i$. Peng proves that if a least-squares line, $\alpha+\beta t$ is fit to the partial sums within each block then the sample variance of the residuals,

$$\frac{1}{m}\sum_{t=1}^{m}\left(Y(t)-\alpha-\beta t\right)^{2}\overset{d}{\sim}m^{2H},\text{ as } m\rightarrow\infty.$$

Therefore the sample variance of residuals is computed for each block, and its mean (or median) is obtained over the N/m blocks. A linear least-squares line is then fit to the log-log plot versus m, where the slope theoretically converges to $2H$. The median is typically used with

cutoffs on m such that $m \in \left[10^{0.7}, 10^{2.5}\right]$. (Taqqu [52], pp.190)

3.2 New Estimators of the Hurst Index Using Ergodic Theory

In this section we introduce three new consistent estimators of the Hurst index for fractional Brownian motion (fBm) using ergodic theory for stochastic processes. We derive closed form solutions for the estimators that are computationally fast and accurate. These new estimators allow for the estimation of the parameters of a fractional Wiener process with unknown and constant drift, scale and Hurst index. Robustness of these estimators is also explored. Using Monte Carlo simulation, we perform an empirical study of the ergodic estimators, Peng's Variance of Residuals Method [39] and Whittle's approximate MLE [56, 4]. Our study demonstrates that the ergodic estimators outperform Peng's method and are very competitive to Whittle's estimates in terms of RMSE. We demonstrate the versatility of the ergodic estimation techniques to accommodate different data structures; i.e. standard fractional Brownian motion or a fractional Wiener process with unknown drift and scale.

3.2.1 Introduction

Modeling with fractional Brownian motion (fBm) requires reliable estimation of the Hurst index. Applications in finance, biology or network flows often require both speed and accuracy in parameter estimation for small samples in order to facilitate dynamic decision making and risk management. Fractional Brownian motion's weak derivative (or increments) with respect to time is known as fractional Gaussian noise (fGn). The self-similar and stationary properties of fractional Gaussian noise make the process a perfect candidate for the use of ergodic theory to estimate parameters influencing the behavior of these models.

Taqqu et al. [51] gives a summary of several previously proposed estimators of the Hurst index and estimates their relative accuracy for large sample sizes via Monte-Carlo simulations. These estimators typically are derived using the properties of the behavior of the spectral density of fBm, estimated through a Periodogram. Other simpler methods take advantage of the asymptotic behavior of the process in the time domain. The ergodic estimators of the Hurst index for fBm introduced in this paper are shown to be competitive to the top performers in Taqqu's paper in

terms of both RMSE and computational time.

3.2.2 Ergodic Estimators for the Hurst Index

We start by estimating the Hurst index for fractional Brownian motion using an L^2 norm calculation. We expand on this method by considering a more realistic model where the fractional Brownian motion process (fBm) is subject to unknown scale and drift. Throughout this section we will use the notation $\left(W_i^H\right)_{i=0}^N$ to represent a discrete realization of $N+1$ observations of a fractional Brownian motion process with Hurst index H.

The new estimators introduced in this chapter are all a result of the ergodic theory for stationary processes. According to ergodic theory, a stationary process is a collection $(\xi_n)_{n\in\mathbb{Z}}$ of random variables with values in some measure space $(X,\mathbb{B})$ such that the joint distribution of $(\xi_{n_1},\xi_{n_2},\ldots,\xi_{n_k})$ is the same as $(\xi_{n_1+n},\xi_{n_2+n},\ldots,\xi_{n_k+n})$ for every choice of $k\geq 1$ and $n,n_1,n_2,\ldots,n_k\in\mathbb{Z}$. Assuming that the space (X,B) is reasonable and Kolmogorov's consistency theorem applies, there exists a measure P on the countable product space Ω of sequences $\{x_n\}_{n\in\mathbb{Z}}$ with values in X, defined for sets in the product σ-field $\mathcal{F}$. On the space Ω there is the natural shift defined by $(B\omega)(n)=x_{n+1}$ for ω with $\omega(n)=x_n$. If we consider the space Ω, a σ-field, a one-to-one invertible measurable map $B:\Omega\to\Omega$ with measurable inverse B^{-1} and a probability measure P on $(\Omega,\mathcal{F})$ that is B-invariant where $P\left(B^{-1}A\right)=P(A)$ for every $A\in\mathcal{F}$, then P is an invariant measure for B and B is a measure-preserving transformation for P. It there exists a measurable map from $\xi:(\Omega,\mathcal{F})\to(X,\mathbb{B})$, then it is easily seen that $\xi_n(\omega)=\xi(B^n\omega)$ defines the stationary stochastic process.

Ergodic theorem states that for any Borel function $f\in L_p(P), 1<p<\infty$ the limit

$$\lim_{n\to\infty}\frac{f(\omega)+f(B\omega)+\ldots+f\left(B^{n-1}\omega\right)}{n}=g(\omega)$$

exists almost surely (B is a linear transformation commonly known as the backward shift operator); additionally, the function $g\in L_p$. Moreover $g(\omega)$ is given by the conditional expectation

$$g(\omega)=\mathbb{E}_P[f|\mathcal{E}](\omega), a.s.$$

where the invariant σ-field$\mathcal{E}$ is defined as

$$\mathcal{E}\equiv\{A:BA=A\}.$$

(Varadhan [55], pp.131-132)

By definition 2.10, $\left(W_{i+1}^H - W_i^H\right) \stackrel{d}{=} N_i\left(0, \Delta t^{2H}\right)$, where Δt is a constant unit of time between W_{i+1}^H and W_i^H. Let $X =< W_1^H - W_o^H, W_2^H - W_1^H, W_3^H - W_2^H, \ldots, W_N^H - W_{N-1}^H, \cdots >$ and f be any Borel function with $\mathbb{E}\left[f\left(W_1^H\right)\right] < \infty$, then since X is stationary and ergodic sequence

$$\frac{1}{N}\sum_{i=0}^{N} f\left(W_{i+1}^H - W_i^H\right) \rightarrow \mathbb{E}\left[f\left(W_1^H - W_o^H\right)\right]$$

almost surely (a.s.), since the fGn process.

3.2.2.1 Ergodic Theory and Hurst Index Estimation

Let us set $f(x) = |x|^k$, $k \in \mathbb{R}^+$. By ergodic theory and properties of fGn, we have

$$\frac{1}{N}\sum_{i=0}^{N-1} \left|W_{i+1}^H - W_i^H\right|^k \quad \rightarrow \quad \mathbb{E}[W_1^H]^k, \; a.s. \tag{3.1}$$

and since the increments of fGn are Gaussian

$$\mathbb{E}[W_1^H]^k = \Delta t^{2H}\left[\frac{2^{k/2}\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\right].$$

Note that the use of the k^{th} moment for estimating the Hurst index is not the result of the maximum likelihood estimation (MLE) formulations. Ergodic theory gives us no information about the bias of the estimate. If we are given any realization of a fractional Brownian motion time series $(W_i)_{i=0}^N$, we can apply ergodic theory to estimate the Hurst index by using the second moment of a normal distribution. Solving for H, we obtain:

$$\hat{H} = \frac{\log\left\{\frac{1}{N}\sum_{i=0}^{N-1}\left(W_{i+1}^H - W_i^H\right)^2\right\}}{2\log(\Delta t)}. \tag{3.2}$$

Peltier [**?**] shows (through the use of box dimension analysis) that absolute moment estimators of the Hurst index all perform well. However the second moment yields the most accurate estimators in terms of RMSE. In §3.2.5.1, we give numerical results in which we compare the "Second Moment" method to Whittle's approximate MLE and Peng's Variance of Residuals method. We empirically demonstrate that the ergodic estimator using the second moment is superior to Whittle's method in terms of RMSE and far better in terms of computational time, however this method can only be used when the scale and location of the fBm process are known.

3.2.3 Parameter Estimation in a Fractional Wiener Process

Real world data does not follow a standard fBm model. In this section we derive methods to estimate the Hurst index when the fBm is not standard, but is influenced by unknown scale and drift. Let $\{X_i\}_{i=1}^N$ be a fractional Wiener process that is $X_i \equiv \mu\Delta t + \sigma\left(W_{i+1}^H - W_i^H\right)$. Since $X_i \stackrel{d}{=} N_i\left(\mu\Delta t, \sigma^2(\Delta t)^{2H}\right)$, an estimate of the drift μ can be found using ergodic theory as

$$\hat{\mu} = \frac{1}{N\Delta t}\sum_{i=1}^{N} X_i \to \frac{\mathbb{E}\left[N\left(\mu\Delta t, \sigma^2\left(\Delta t\right)^{2H}\right)\right]}{\Delta t} \tag{3.3}$$

We can use the location estimate to obtain a scaled fractional Gaussian noise process, $X_i - \hat{\mu}\Delta t = \sigma\left(W_{i+1}^H - W_i^H\right)$. In the next sub-sections, we introduce new ergodic estimators of the the Hurst index when fGn is influenced by an unknown scale σ.

3.2.3.1 Ratio of Second Moments Method

If fBm is only affected by a scale factor, the second moment converges by ergodic theory to $\sigma^2(\Delta t)^{2H}$:

$$SS_1 \equiv \frac{1}{N}\sum_{i=0}^{N-1}\sigma^2\left(W_{i+1}^H - W_i^H\right)^2 \to \sigma^2(\Delta t)^{2H}. \tag{3.4}$$

If we form stationary processes on disjoint sets of length $2\Delta t$, then we can once again use the ergodic second moment to define two estimates; one formed from the even increments and the other from the odd increments:

$$SS_{even} \equiv \frac{1}{\lfloor N/2\rfloor}\sum_{i=0}^{\lfloor N/2\rfloor-1}\sigma^2\left(W_{2i+2}^H - W_{2i}^H\right)^2 \to \sigma^2(2\Delta t)^{2H},$$

$$SS_{odd} \equiv \frac{1}{\lfloor N/2\rfloor}\sum_{i=0}^{\lfloor N/2\rfloor-1}\sigma^2\left(W_{2i+3}^H - W_{2i+1}^H\right)^2 \to \sigma^2(2\Delta t)^{2H}.$$

To reduce the error of the $\sigma^2(2\Delta t)^{2H}$ estimate, and utilize all information available in the time series, the even and odd estimates are averaged. Both the even and the odd estimators use the data set and thus these two estimators have the same variance. Therefore, the average of theses two estimators reduces the variance and bias:

$$SS_2 \equiv \frac{SS_{even} + SS_{odd}}{2} \to \sigma^2(2\Delta t)^{2H}. \tag{3.5}$$

Notice that for a fractional Wiener process, the second moment estimator converges to

$$\begin{aligned}\mathbb{E}\left[X_i^2\right] &= \mu^2(\Delta t)^2 + 2\mu\Delta t\sigma\mathbb{E}\left[W_{i+1}^H - W_i^H\right] \\ &\quad +\sigma^2\mathbb{E}\left[\left(W_{i+1}^H - W_i^H\right)^2\right] \\ &= \mu^2\left(\Delta t\right)^2 + \sigma^2\left(\Delta t\right)^{2H}.\end{aligned}$$

Additionally, when Δt is small $\mu^2(\Delta t)^2 \ll \sigma^2(\Delta t)^{2H}$, if μ is comparable to σ in magnitude. Therefore, when estimating $\mathbb{E}\left[X_i^2\right]$ with small Δt, an estimate of μ may not be needed. In this situation the term $\mu^2(\Delta t)^2$ would contribute to the error ϵ of the estimate and we can proceed using equation 3.5 directly, where

$$\mathbb{E}\left[X_i^2\right] = \sigma^2(\Delta t)^{2H} + \epsilon.$$

Note that even if $\Delta t \ll 1$, as H increases the magnitude of $\sigma^2(\Delta t)^{2H}$ relative to the error ϵ becomes closer. Taking a ratio of the two moments SS_1 (equation 3.4) and SS_2 (equation 3.5) the scaling and time factors cancel and we obtain:

$$\frac{SS_2}{SS_1} = 2^{2H} \implies \hat{H} = \frac{\log\left(\frac{SS_2}{SS_1}\right)}{2\log(2)}.$$

This estimator of H is based on the ratio of two second moments, therefore we refer to this method as the "Ratio method". The Ratio method's estimate of H can be applied in equation 3.4 to estimate the scale influence the fractional Wiener process, $\hat{\sigma}$:

$$\hat{\sigma} = \sqrt{\frac{SS_1}{(\Delta t)^{2\hat{H}}}}$$

In §3.2.5.2 we show the results of Monte Carlo simulations of a fractional Wiener Process to evaluate the performance of the Ratio method estimator. It should be noted that application of this method on real data requires filtering of any identifiable outliers or jumps, since a large jump will skew SS_1 and SS_2 and therefore bias the estimation of the Hurst index. The Ratio method is sensitive to these types of anomalies in data, as discussed in §3.2.4. The error in the Ratio method's estimates of H and σ are highly correlated, which is evident from the method used. The same kind of estimators can be derived using different combinations of the higher moments in equation 3.1 to estimate the Hurst index for a fractional Wiener process. These estimators

can be shown to be equivalent to or worse than the Ratio method.

3.2.3.2 Quadrant Method

In this section we introduce an estimator which is more robust to outliers and jumps and which (unlike the Ratio method) does not depend on σ. Let us consider a fractional Wiener process with no drift (μ=0),

$$X_i = \sigma \left(W_{i+1}^H - W_i^H \right).$$

Note that the process $\{X_i\}_{i=1}^N$ is mean zero. If the data set being analyzed has drift, an estimate of drift will need to be made using equation 3.3.

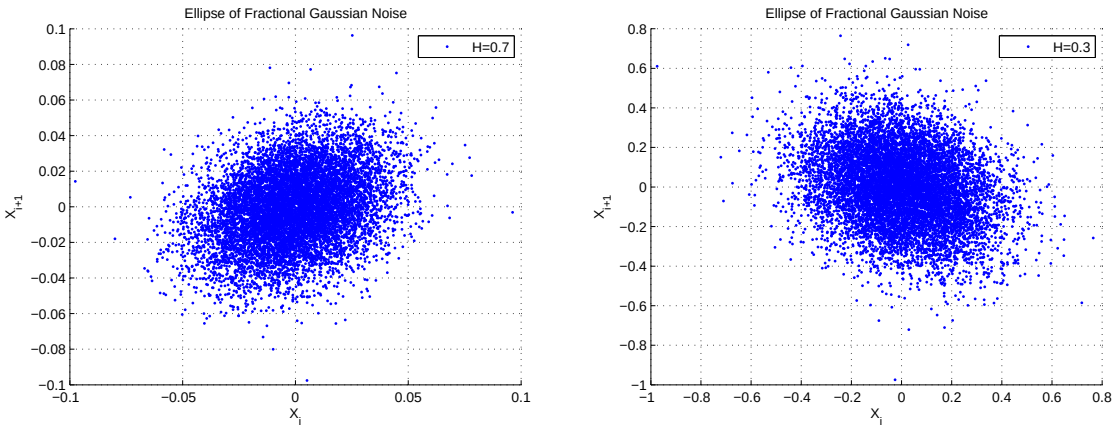


Figure 3.2: PDF X_{i+1} vs. X_i

Two consecutive observations of fGn are normally distributed with Pearson's correlation coefficient $\rho = 2^{2H-1} - 1$. A 2-D plot of consecutive random observations of fGn is shaped like an ellipse (or a circle when $H = 1/2$) at a constant probability level. The Hurst index of the process (and the probability level) directly dictates the length of the axes of the bi-variate normal distribution (see Figure 3.2). The shape of the ellipse (or in this case the relative density in any particular quadrant of the 2-D plot) can be used to estimate the Hurst index. The major axis of the ellipse is always at $\pm\frac{\pi}{4}$ with respect to the positive or negative auto-correlation of the process, respectively.

Let us define a new process $(Z_i)_{i=1}^N$ by

$$Z_i \equiv sgn\left(X_i\right) sgn\left(X_{i+1}\right),$$

where $sgn(x) = 1$ if $x > 0$, and $sgn(x) = -1$ if $x < 0$, and $sgn(x) = 0$ if $x = 0$.

The signum function only sees sign and not magnitude of X_i, therefore σ does not affect the estimation of H. To estimate the Hurst index we need to compute the expected value of the process Z_i. This can be accomplished using ergodic theory. Notice,

$$\mathbb{E}[Z_i] = \mathbb{E}\left[\frac{X_i}{\sqrt{X_i^2}}\frac{X_{i+1}}{\sqrt{X_{i+1}^2}}\right].$$

Since $\{X_i\}$ is a scaled fractional Gaussian noise, it is normally distributed with mean zero and variance $\sigma^2\Delta t^{2H}$, with correlation between X_i and X_{i+1} given by $\rho = 2^{2H-1} - 1$, therefore, the $\frac{1}{N}\sum_{i=1}^{N} X_i$ converges to

$$\mathbb{E}[Z_i] = \frac{1}{2\pi D^{1/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{xy}{|x||y|}e^{-\frac{1}{2D}\left(x^2-2xy\rho+y^2\right)}dx\,dy, \tag{3.6}$$

where $D = (1-\rho^2)$. Analytically, the expected value (equation 3.6) is the same as the probability that two consecutive observations of fractional Gaussian noise (X_i and X_{i+1}) are in the same quadrants of a two dimensional graph of X_i verses X_{i+1}. Each Z results in four outcomes. We refer to this technique as the "Quadrant method." Equation 3.6 becomes,

$$\begin{aligned}\mathbb{E}[Z_i] &= P(X \geq 0, Y \geq 0) + P(X < 0, Y < 0)\\ &\quad -P(X \geq 0, Y < 0) - P(X < 0, Y \geq 0).\end{aligned}$$

Utilizing the symmetry of the two dimensional Gaussian distribution,

$$\mathbb{E}[Z_i] = 2 * P(X \geq 0, Y \geq 0) - 2 * P(X \geq 0, Y < 0). \tag{3.7}$$

Let $u = \frac{x}{\sqrt{D}}$ and$v = \frac{y}{\sqrt{D}}$ then,

$$\mathbb{E}[Z_i] = \frac{\sqrt{D}}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{uv}{|u||v|}e^{-\frac{1}{2}\left(u^2-2uv\rho+v^2\right)}du\,dv,$$

and equation 3.7 becomes,

$$\mathbb{E}[Z_i] = 2\sqrt{D}\left[P(U \geq 0, V \geq 0) - P(U \geq 0, V < 0)\right].$$

The first term in equation 3.7 yields

$$\begin{aligned} P\left(U \geq 0, V \geq 0\right) &= \frac{1}{2\pi}\int_0^\infty \int_0^\infty e^{-\frac{1}{2}\left(u^2-2uv\rho+v^2\right)} du\, dv \\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(u-v\rho)^2} du\, e^{-\frac{1}{2}\left(v^2-\rho^2v^2\right)} dv. \end{aligned}$$

Let $x = u - v\rho$, then

$$\begin{aligned} P\left(U \geq 0, V \geq 0\right) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{1}{\sqrt{2\pi}}\int_{0-v\rho}^\infty e^{-\frac{x^2}{2}} dx\, e^{-\frac{1}{2}\left(v^2-\rho^2v^2\right)} dv, \\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \Phi\left(v\rho\right) e^{-\frac{1}{2}\left(v^2-\rho^2v^2\right)} dv. \end{aligned}$$

where $\Phi\left(v\rho\right) = P\left(N\left(0,1\right) < v\rho\right)$. Substituting $y = v\rho$,

$$P\left(U \geq 0, V \geq 0\right) = \frac{1}{\sqrt{2\pi}\rho}\int_0^\infty \Phi\left(y\right)\, e^{-\frac{y^2}{2}\left(\frac{1-\rho^2}{\rho^2}\right)} dy. \tag{3.8}$$

If

$$I(\alpha) \equiv \int_0^\infty \Phi\left(\alpha y\right)\, e^{-\frac{1}{2}y^2\left(\frac{1-\rho^2}{\rho^2}\right)} dy, \tag{3.9}$$

then,

$$\frac{\partial I(\alpha)}{\partial \alpha} = \frac{1}{\sqrt{2\pi}}\int_0^\infty y e^{-\frac{1}{2}\alpha^2y^2} e^{-\frac{1}{2}y^2\left(\frac{1-\rho^2}{\rho^2}\right)} dy.$$

Substituting $x = \frac{y^2}{2}$,

$$\begin{aligned} I^{'}\left(\alpha\right) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-x\left(\alpha^2+\left(\frac{1-\rho^2}{\rho^2}\right)\right)} dx \\ &= \frac{1}{\sqrt{2\pi}\left(\alpha^2+\left(\frac{1-\rho^2}{\rho^2}\right)\right)}. \end{aligned}$$

Therefore,

$$I(\alpha) = \frac{\rho}{\sqrt{2\pi\left(1-\rho^2\right)}} \arctan\left(\frac{\alpha\rho}{\sqrt{1-\rho^2}}\right) + C. \tag{3.10}$$

To solve for C we utilize equation 3.9 with $\alpha = 0$,

$$I(0) = \int_0^\infty \frac{1}{2} e^{-\frac{1}{2}y^2\left(\frac{1-\rho^2}{\rho^2}\right)} dy.$$

Substituting $x = y\sqrt{\left(\frac{1-\rho^2}{\rho^2}\right)}$,

$$I(0) = \frac{1}{2\sqrt{\frac{1-\rho^2}{\rho^2}}} \int_0^\infty e^{-\frac{1}{2}x^2} dx \Rightarrow I(0) = \frac{\sqrt{2\pi}}{4\sqrt{\frac{1-\rho^2}{\rho^2}}}.$$

Equating equation 3.9 and equation 3.10 with $\alpha = 0$,

$$C = \frac{\sqrt{2\pi}}{4\sqrt{\frac{1-\rho^2}{\rho^2}}}.$$

The function $I(\alpha)$ becomes,

$$I(\alpha) = \frac{\rho}{\sqrt{2\pi\left(1-\rho^2\right)}} \arctan\left(\frac{\alpha\rho}{\sqrt{\left(1-\rho^2\right)}}\right) + \frac{\sqrt{2\pi}\rho}{4\sqrt{\left(1-\rho^2\right)}}. \tag{3.11}$$

Substituting equation 3.11 with $\alpha = 1$into equation 3.8 yields

$$P\left(X \geq 0, Y \geq 0\right) = \frac{1}{2\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right) + \frac{1}{4}. \tag{3.12}$$

A similar procedure can be used with minor changes to find the second term in equation 3.7 which can be shown to be,

$$P\left(X \geq 0, Y < 0\right) = -\frac{1}{2\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right) + \frac{1}{4}. \tag{3.13}$$

Substituting equation 3.12 and equation 3.13 and D into equation 3.7, the expected value of Z_i is obtained,

$$\begin{aligned} \mathbb{E}\left[Z_i\right] &= \frac{2}{\pi} \arctan\left(\frac{\rho}{\sqrt{\left(1-\rho^2\right)}}\right) \\ &= \frac{2}{\pi} \arcsin\left(\rho\right). \end{aligned} \tag{3.14}$$

This expected value can be used to estimate both the correlation and the Hurst index. Solving equation 3.14 for ϱ we obtain estimates,

$$\hat{\rho} = \sin\left(\frac{\pi}{2} E\left[Z_i\right]\right).$$

Since $\varrho = 2^{2H-1} - 1$ then we obtain the ergodic "Quadrant method" estimator for H,

$$\hat{H} = \frac{\frac{\log(\hat{\rho}+1)}{\log(2)} + 1}{2}.$$

Computationally, this algorithm is very fast and fairly accurate (see §3.2.5 for numerical results). The major advantage of this method is that estimates are not largely affected by outliers, since the magnitude of the observed values does not disproportionately influence the estimator. This means the Quadrant method is robust to data that may not perfectly follow a fractional Wiener process, see §3.2.4.3. An ergodic estimator of H can also be derived using constant volume ellipsoids for the function $E[X_i X_{i+1}]$. The derivation of this statistic is very similar to the Quadrant method derivation, however it requires the use of a non-linear mixed integer optimization method.

Remark 3.1. The Quadrant and Ratio methods were designed such that an estimate of the scale is not required. Recall, that the Ratio method actually takes the ratio of two estimates of the variance and then divides to eliminate the scale affecting the process. However, taking the ratio of two estimates results in an increase in the error, since both estimates are correlated. This is why the ergodic Second moment method (which is related to the Ratio method) gives much better estimates when the scale is known. Both the Quadrant and the Ratio methods allow for the estimation of the scale of the process by computing the ergodic Second moment, which converges to $\sigma^2 \Delta t^{2H}$. The Ratio method requires no extra computations to perform this step, since the ergodic Second moment is already incorporated in the calculation. However, the Quadrant method requires a separate calculation to find the estimates of the scale parameter. This is not particularly a disadvantage for the Quadrant method since it utilizes two very different estimates to find $\hat{H}$ and $\hat{\sigma}$. Empirical studies indicated that the errors of $\hat{H}$ and $\hat{\sigma}$ are much less correlated than if the Hurst index is estimated using the Ratio method. This is because the Ratio method's estimation of σ uses information that it already incorporated in the estimation of H. Whittle's MLE allows for the simultaneous estimation of σ and H. The estimation of H is unaffected by any drift in the process since the optimization is unaffected by any scalar addition. Just like in the ergodic Ratio method, Whittle uses the estimate of H to find the scale of the process, and therefore is also affected if Δt is not small.

3.2.4 Robustness of Hurst Index Estimators

Robust statistics is the stability theory of statistical procedures. Statistical inferences are in part, even in the simplest cases, based on observations. Estimation of a statistic (or statistical tests for properties) lend themselves to explicit or implicit assumptions about the distribution of data, the relationships between observations (i.e. independence), and or the behavior of the real world. Assumptions are necessary to gain insight into the behavior or properties of observations. However, these assumptions are never supposed to be exactly true, but a small deviation from a mathematical model should only result in a small error in the final conclusions of any statistics. The purpose of "Robust Statistics" is to investigate the sensitivity of an inference to deviations from model assumptions. An inference that is insensitive to small deviations from the assumptions is called "Robust". I this section we investigate the robustness of the new ergodic estimators and compare them to the robustness of Whittle's approximate MLE and Peng's variance of residuals estimators.

The "Influence Curve" is a way to evaluate the sensitivity of an estimator to one contaminating point and therefore understand the "local robustness" of the estimators when the rest of the observations are assumed to come from the true distribution (Huber [22], pp.14); fGn is a Gaussian process with mean zero, variance $(\Delta t)^{2H}$ and covariance

$$\mathbb{E}\left[X_i X_j\right] = \frac{(\Delta t)^{2H}}{2}\left(\left|i-j+1\right|^{2H} + \left|i-j-1\right|^{2H} - 2\left|i-j\right|^{2H}\right)$$

where Δt is a known constant and $X_i = W_{i+1}^H - W_i^H$. In this section we create and compare influence curves for various estimators of the Hurst index. The influence curves $IC\left(x, H\right)$ are generated with contaminating values of $x = k\left(\Delta t\right)^H$, $k \in [-3, 3]$ for $H = 0.1, \ldots, 0.9$. Since the true distribution of the data is assumed to be normally distributed, this is equivalent to the contaminating observation falling within an interval of three sigma. The graphs are all generated with $\Delta t = 1/252$ and a sample size $n = 156$, therefore we see the influence of the 157^{th} observation. A summary of the sensitivity of the Hurst estimators to a single contaminator $x = \pm 3\left(\Delta t\right)^H$ appears in Figure 3.3.

Contaminator	H=0.1	H=0.2	H=0.3	H=0.4	H=0.5	H=0.6	H=0.7	H=0.8	H=0.9
2nd Moment Method									
- $3\Delta t^H$	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
+ $3\Delta t^H$	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
Ratio Method									
- $3\Delta t^H$	0.036	0.027	0.018	0.011	0.005	-0.001	-0.006	-0.010	-0.014
+ $3\Delta t^H$	0.036	0.027	0.018	0.011	0.005	-0.001	-0.006	-0.010	-0.014
Quadrant Method									
- $3\Delta t^H$	0.003	0.002	0.001	0.001	0.000	-0.001	-0.001	-0.001	-0.002
+ $3\Delta t^H$	0.003	0.002	0.001	0.001	0.000	-0.001	-0.001	-0.001	-0.002
Whittle's Method									
- $3\Delta t^H$	0.055	0.029	0.014	0.005	0.000	-0.005	-0.009	-0.014	-0.027
+ $3\Delta t^H$	0.060	0.032	0.017	0.007	0.001	-0.004	-0.009	-0.014	-0.027
Peng's Variance of Residuals Method									
- $3\Delta t^H$	0.073	0.046	0.026	0.012	0.003	-0.003	-0.007	-0.010	-0.012
+ $3\Delta t^H$	0.077	0.049	0.029	0.015	0.006	-0.001	-0.005	-0.008	-0.010

Figure 3.3: Robustness Summary

3.2.4.1 Influence Curve for the Second Moment Method

In the Second Moment method estimator, the addition of one extra term, x, in the series causes a change to the estimation of the Hurst index of:

$$H_{n+1} = \frac{\log\left[\frac{n(\Delta t)^{2H_n}+x^2}{n+1}\right]}{2\log(\Delta t)}$$

This gives the empirical influence function

$$IC(x,H) = H_{n+1} - H_n = \frac{\log\left[\left(\frac{1}{n+1}\right)\left\{n+\frac{x^2}{(\Delta t)^{2H_n}}\right\}\right]}{2\log(\Delta t)}.$$

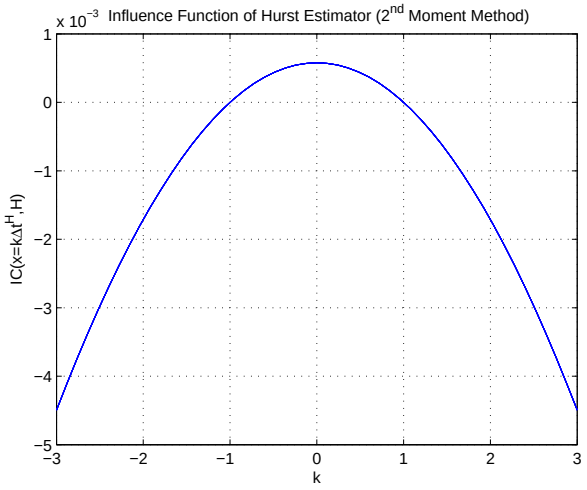


Figure 3.4: Influence Curve: 2^{nd} Moment Method

In Figure 3.4 we see that a contaminating point that is within three sigma of the true distribution has a maximum influence of $\pm 4x10^{-3}$ for all Hurst values. Given the scale, the influence curve for the Second Moment method is relatively flat over a $\pm 3\sigma$ range of values of x. Additionally, the maximum of the influence curve occurs at a height of $\log\left(\frac{n}{n+1}\right)/2\log(\Delta t) \geq 0$, $\Delta t < 1$.

3.2.4.2 Influence Curve for the Ergodic Ratio of Second Moments Method

The influence curve for the Ratio method estimator is similar to the Second Moment method in that it is a function of two Second Moments:

$$H_n = \frac{\log\left[\frac{SS_{2,n}}{SS_{1,n}}\right]}{2\log(2)},$$

where

$$SS_{2,n} = \frac{\sum_{i=1}^{n-1}(X_{i+1}+X_i)^2}{n-1}$$

$$SS_{1,n} = \frac{\sum_{i=1}^{n}(X_i)^2}{n}.$$

Since the influence curve is derived assuming that none of the $\{X_i\}_{i=1}^n$ deviate from the true

distribution, we know that $SS_{1,n} = \sigma^2 (\Delta t)^{2H_n}$ and $SS_{2,n} = \sigma^2 (2\Delta t)^{2H_n}$. Notice that $SS_{1,n}$ is the same as the ergodic Second Moment method, and therefore if we add one more term in the sequence, x, then we have

$$SS_{1,n+1} = \frac{nSS_{1,n} + x^2}{n+1}.$$

The term $SS_{2,n+1}$ is the same as the ergodic Second Moment method with half the sample rate, however to compute the influence of the contaminator, x, we need to consider the location of this extra observation. If the observation is at the beginning or the end of the sequence, it only affects the estimate in one term (notice in the formula for $S_{2,n}$ that the terms X_1 and X_n are only counted once, while all other X_i, $2 \leq i \leq n-1$ appear in two terms of $(X_{i+1} + X_i)^2$. Therefore, to see the maximum influence of an additional observation, we need to place the contaminating observation somewhere in between the first and last. Without loss of generality, we can place it right before the last observation, giving the sequence $\{X_1, X_{2,\ldots}, X_{n-1}, x, X_n\}$. Therefore,

$$SS_{2,n+1} = \frac{(n-2)SS_{2,n} + (x + X_{n-1})^2 + (X_n + x)^2}{n}.$$

Since X_i and X_{i+1} come from the true distribution,

$$\mathbb{E}\left[(x + X_{n-1})^2 + (X_n + x)^2\right] = 2\sigma^2 (\Delta t)^{2H_n} + 2x^2.$$

In this framework x is treated as a constant. Therefore, since the estimator

$$\mathbb{E}\left[SS_{2,n}\right] = \sigma^2 (2\Delta t)^{2H_n},$$

the expected influence of x has the form:

$$IC(x, H) = \log\left[\left(\frac{n+1}{n}\right)\left(\frac{(n-2)\sigma^2(2\Delta t)^{2H_n} + 2\sigma^2(\Delta t)^{2H_n} + 2x^2}{n\sigma^2(\Delta t)^{2H_n} + x^2}\right)\right] / 2\log(2) - H_n.$$

When $\sigma^2 = 1$, then we obtain the influence curves in Figure 3.5.

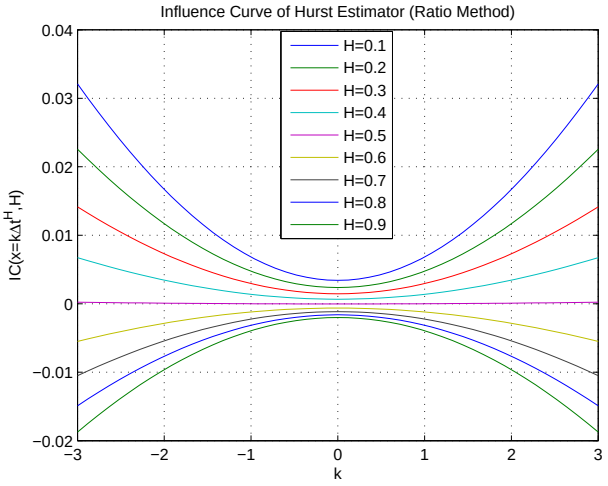


Figure 3.5: Influence Curve: Ratio Method

Figure 3.5 shows that the Ratio method's influence curve changes concavity when the process changes from long to short range dependence. The Ratio method has more sensitivity than the ergodic Second Moment for all Hurst values. The contaminating point's influence on the estimator increases as the Hurst index get further away from $H = 0.5$; more sensitivity occurs when the process has negative auto-correlation.

3.2.4.3 Influence Curve for the Ergodic Quadrant Method

Given the fractional Wiener process, $\{X_i\}_{i=1}^{n}$, the Hurst index estimator using the Quadrant method is a function of the statistic

$$T_n = \frac{\sum_{i=1}^{n-1} sgn(X_i)sgn(X_{i+1})}{n-1}.$$

The correlation of normals is then estimated by

$$\rho_n = \sin(\frac{\pi}{2}T_n).$$

Lastly, the Hurst index is computed

$$H_n = \frac{\left(\frac{\log(2\rho_n+2)}{\log(2)}\right)}{2}.$$

In order to compute the influence curve, we need to understand the estimator T. Once again to get the maximum contribution of an additional observation, we need to place the observation between the first and last X_i. If we place the contaminating data point, x, in the sequence as before $\{X_1, X_{2,\ldots}, X_{n-1}, x, X_n\}$:

$$T_{n+1} = \frac{(n-1)T_n + sgn(X_{n-1})sgn(x) + sgn(x)sgn(X_n)}{n}$$

The property of the signum function yields only three results, none of which are dependent on the magnitude of the contaminant, but only on the sign of the new observation and the sign of the immediately adjacent observations. This is because the Quadrant method attempts to find momentum in the time series. The function, T, looks for long term tendencies of the time series in a particular direction. The different outcomes are given in the following matrix.

$sgn(X_{n-1})sgn(x) + sgn(x)sgn(X_n)$	$x \geq 0$	$x < 0$
$X_i \geq 0, X_{i+1} \geq 0$	$1+1=2$	$-1-1=-2$
$X_i \geq 0, X_{i+1} < 0$	$1-1=0$	$-1+1=0$
$X_i < 0, X_{i+1} \geq 0$	$-1+1=0$	$1-1=0$
$X_i < 0, X_{i+1} < 0$	$-1-1=-2$	$1+1=2$

Therefore,

$$T_{n+1} = \begin{cases} \frac{(n-1)T_n - 2}{n} & \text{with probability } \frac{1}{4} \\ \frac{(n-1)T_n}{n} & \text{with probability } \frac{1}{2} \\ \frac{(n-1)T_n + 2}{n} & \text{with probability } \frac{1}{4} \end{cases}.$$

If all observations came from the true distribution, then

$$\mathbb{E}[T] = \frac{2}{\pi} Arctan\left(\frac{2^{2H-1}-1}{\sqrt{1-\left(2^{2H-1}-1\right)^2}}\right).$$

Therefore we can substitute the true statistic $\mathbb{E}[T]$ for T_n to show the expected influence of the contaminating term, x, on H_{n+1}. Performing this substitution, the influence curve can either be constant (when the contaminating point adds zero to the estimate of T_{n+1}) or the curve is $\pm$ a constant, with jumps left and right of the center (when the contaminating point adds $\pm 2/(n+1)$ to the estimate of T_{n+1}).

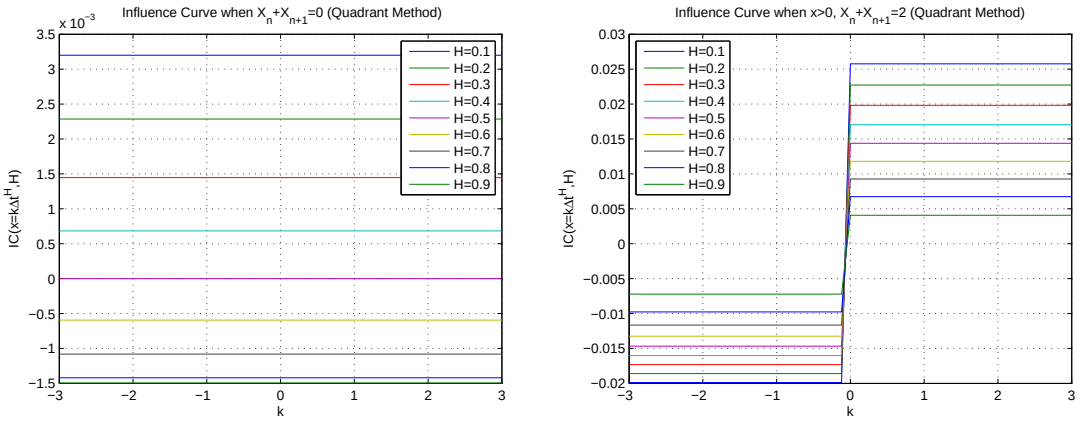


Figure 3.6: Influence Curve Decomposition: Quadrant Method

Note that the right graph in Figure 3.6 will always have an influence curve that jumps in the same pattern, (down on one side and up on the other or vise-versa). The jump pattern depends on the sign of the observations immediately adjacent to the contaminating point x. The analysis above shows that the short range dependent process $\left(H < \frac{1}{2}\right)$ has much more sensitivity to the contaminating observation than the long-memory process$\left(H > \frac{1}{2}\right)$.

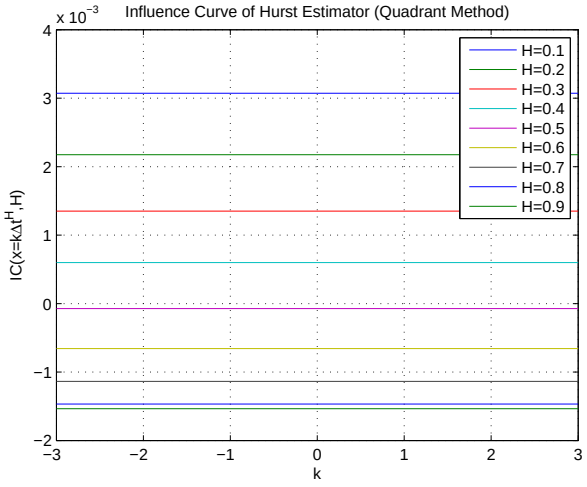


Figure 3.7: Influence Curve: Quadrant Method

In figure 3.7 we can see the expected influence curve for the Quadrant method's Hurst index estimator shows the extreme robustness to the size of the contaminating point. The Quadrant method is the most robust method discussed in this chapter.

3.2.4.4 Influence Curve for the Whittle's Approximate MLE and Peng's Variance of Residuals

Whittle's approximate MLE is calculated by minimizing the log ratio of the Periodogram (calculated from data) and the theoretical Spectral density function for fGn. The computation of the Spectral density function for fGn requires a truncated infinite sum (or linear approximation). Additionally, to calculate the estimates of the Hurst index, we need to numerically optimize a convex objective function. Whittle's objective function gives an estimator of the variance affecting the process, $\sigma^2(\Delta t)^{2H}$, at the optimal solution. This is accomplished using the Golden Section method.

To compute the influence curve we need to understand the influence curve of the Periodogram, which coupled with the optimization over the spectral density, complicates this calculation to an intractable degree since it is necessary to compute the contribution of the contaminant, x, for all $n/2$ Fourier frequencies. We have to resort to another way to evaluate the influence of x. One way to generate the influence curve is to use Monte-Carlo simulation.

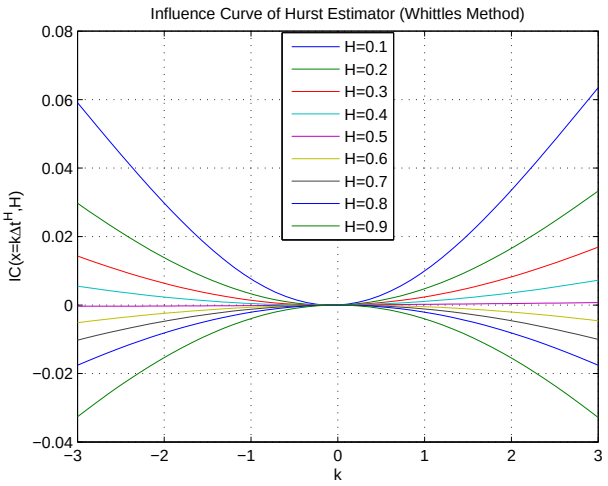


Figure 3.8: Influence Curve: Whittle's Approx. MLE

In Figure 3.8 we can see the average influence curve for Whittle's method. These curves were generated by simulating 500 replications of fractional Gaussian generated noise using the Durbin-Levinson algorithm with $N = 156$ observations. The Hurst index was then estimated via Whittle's algorithm, then the observation x was placed at position $n/4$. This created another

sample with $n = 157$ observations, which was used to estimate the Hurst index for the same x values used in the ergodic estimator influence curves. The estimated value of H_n for each replication was then subtracted from the estimate H_{n+1}, giving the influence curve. These 500 replications for each $H = 0.1, \ldots, 0.9$ were then averaged for each value of x to produce the average influence curves above. Whittle's method does not appear to be locally robust for $H < \frac{1}{2}$, while it is more robust when $H \geq \frac{1}{2}$. While there does not seem to be any literature on the influence curve of Whittle's method, Taqqu [53] on page 724 recognizes that

> "it is a parametric model in that it assumes the spectral density of the series is know with the exception of a few parameters, which are to be estimated. This assumption allows for very precise estimation when the series being examined fits the assumed model exactly. If, on the other hand, the actual series is not of the exact form specified in the model, the parametric estimators may give incorrect results."

In his paper, Taqqu discusses different techniques that have been developed to robustify Whittle's Approximate MLE. One such technique smooths out the higher frequencies in the data. The noise typically present in real data occurs at higher frequencies. This noise can skew the values of the spectral density function, resulting in a biased Hurst index estimate. The fact that there are at least four different methods that have been developed to robustify Whittle's MLE, indicates that this estimator may not be robust enough for certain real data sets. Taqqu [53] shows how each one of these robustified Whittle estimators changes for a given set of Ethernet data. Our simulations indicated that on average, a given contaminating point results in slightly worse deviations in Whittle's Approximate MLE than the ergodic Ratio method for all Hurst index values except $H = 0.3, 0.4$ and 0.5.

Peng's Variance of Residuals method estimates the Hurst index from the errors of a linear regression on a log of aggregated variance calculations. We perform a Monte-Carlo simulation in the same fashion as in the influence curve for Whittle's estimator.

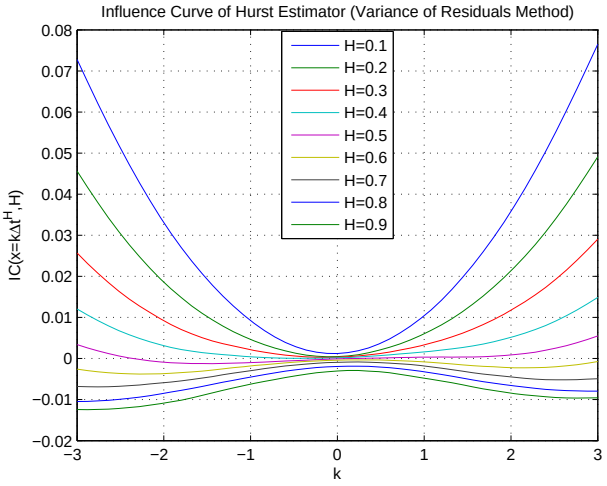


Figure 3.9: Influence Curve: Variance of Residuals

In Figure 3.9 we can see the average influence curve for Peng's Variance of Residuals method. Notice that this method is more robust (on average) to a single contaminating value when $H \geq \frac{1}{2}$. Its influence curve for these Hurst values is comparable to the Ratio method's influence curve. When $H < \frac{1}{2}$, this method is the least robust (on average) out of the new estimators presented in this chapter.

3.2.5 Simulation Study of the Performance of Estimators

In this section, we compare via Monte Carlo simulation, the performance of the newly introduced ergodic estimators of the Hurst index to Whittle's approximate MLE and Peng's Variance of Residuals estimators. Taqqu et al. [51] presents an empirical study of many estimators of the Hurst index in the same fashion. They show empirically that Whittle's approximate MLE estimator is the best estimator (of those tested) in terms of RMSE for a fBm time series. Their study indicates that Peng's Variance of Residuals method is the second best of the methods tested.

Taqqu generated 50 sample paths of fGn each with a sample size of $N = 10,000$ for $H = 0.5, 0.6, 0.7, 0.8, 0.9$ using Monte Carlo simulation (Durbin-Levinson algorithm). He computed the sample mean, sample variance and RMSE of the Hurst index estimators for each technique. Using the Durbin Levinson algorithm, we simulate 500 sample paths of fGn with length $N = 10,000$ and $\Delta t = 1/252$. We extend the analysis for processes with both short range ($H < 1/2$) and long range dependence ($H > 1/2$) by simulating $H = 0.1, 0.2, \ldots, 0.9$ using Matlab®. For each

$H = 0.1, 0.2, \ldots, 0.9$ we used the method of common random numbers with the same seed to generate 500 x 10,000 i.i.d. standard normal random variates for each set of paths. We increased the number of sample paths (compared to Taqqu et al. [51]) in order to increase the accuracy of our estimates of Root Mean Square Error (RMSE) and allow for the identification of significant differences in the estimators (see §3.2.5.2).

We implemented Whittle's algorithm using the spectral density approximation described by Ledesma and Liu [27], and use $n = 500$ terms in the linear approximation of the spectral density at each Fourier frequency. We found that even though Ledesma and Liu recommend $n = 200$ terms, at least $n = 500$ terms are needed in the linear approximation due to the slow convergence rate of the spectral density when $H \in (0, 0.3)$. Ledesma's recommendation was for $H \geq 1/2$. A Golden Section search algorithm is used to find the global maximum of Whittle's approximate MLE with a termination tolerance of 10^{-6} for the accuracy of the Hurst index estimate. The Golden Section method is initialized to search for the optimum on $H \in [0, 1]$. The ergodic algorithms do not require optimization and therefore are not constrained numerically on $H \in [0, 1]$. Peng's Variance of Residuals method is implemented for a minimum of 50 block sizes. Regression is performed on block sizes between $\left[10^{0.5}, 10^{0.7}\right]$. The median of residuals at each block size is used in the Hurst index estimator.

All computations were done using a Dell Optiplex 755 running Windows 7 with a 2.66 GHz Intel Core 2 Duo and 3326MB of RAM . The total time to compute all 4500 (500 paths by 9 Hurst index values) estimates of the Hurst index for different sample sizes are shown in Figure 3.10.

	Run Time(s)					
	Whittle		Variance	Ergodic	Ergodic	Ergodic
N	tol=10^{-6}, n=500	tol=10^{-4}, n=200	Residuals	2nd Moment	Ratio	Quadrant
39	152.0	51.6	12.4	0.000	0.000	0.000
78	158.8	57.4	38.7	0.000	0.016	0.000
156	176.1	72.5	88.2	0.016	0.016	0.015
312	222.2	114.3	93.0	0.016	0.124	0.093
625	367.0	249.5	110.8	0.047	0.234	0.156
1250	855.7	719.2	253.6	0.109	0.468	0.344
2500	2658.7	2477.8	464.4	0.218	0.952	0.671
5000	9307.5	9100.1	1055.9	0.421	1.841	1.342
10000	35771.0	35559.6	1879.6	0.889	3.837	2.683

Figure 3.10: Computational Time of Hurst Estimators

The power of closed-form representation of the ergodic estimators of the Hurst index can

be seen by the magnitude of difference in computational time for all nine simulations; ergodic estimators take seconds or less, while for large data sets Whittle's approximate MLE can take tens of hours. Whittle's lengthy computational time is primarily due to the re-computation of the spectral density function for each iteration of H in the optimization algorithm. Simple algorithms like the Variance of Residuals method can also be seen to take significantly more computational time than the ergodic methods.

3.2.5.1 Empirical Performance of Estimators

In this section we analyze the behavior of the estimators as the length of the fBm time-series is reduced, giving insight into the convergence rate. Difference analysis is used to demonstrate which estimators are more accurate. We provide a comparison of the various estimators for the 500 sample paths of fractional Brownian motion. Appendix I shows comparisons of the Hurst index estimators from the 500 x 9 simulated fBm paths via box-plots. The boxes represent the inter-quartile range (75 percent of the estimates fall in this range). The lines inside the boxes indicate the mean of the non-outlier points. The plus signs show outlier points, which are defined by values greater than the 'whiskers' length which is 1.5 times the distance outside the inter-quartile range.

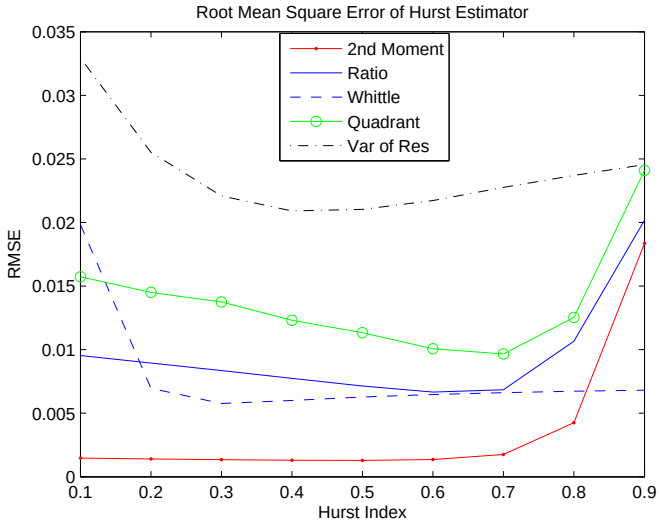


Figure 3.11: RMSE Summary

Figure 3.11 shows a comparison of each method for $N = 10,000$ data points in each time-series. In Appendix III, we provide a breakdown of the sample bias $\left(Bias(H, \hat{H}) = mean\left(\hat{H} - H\right)\right)$

and sample variance of the estimators since

$$RMSE(\hat{H}) = \sqrt{Var(\hat{H}) + (Bias(H, \hat{H}))^2}.$$

The ergodic estimators have similar performance to each other in that they have little bias for $H \in [0.1, 0.8]$ and increased bias for H values closer to 0.9. The ergodic estimates have the least standard deviation for $H \in [0.5, 0.7]$ and higher deviation as $H \to 1$. On the other hand, the Whittle estimates seem to underestimate the Hurst index on average, with more error as $H \to 0$. This is due to the slow convergence rate of the spectral density function. If the linear approximation of the spectral density is changed to include more terms, the accuracy of the estimators at $H = 0.1$ will improve slightly because of the slow convergence rate of the spectral density function, however this comes at the cost of computational time. We found that as $H \to 0$, the number of terms needed in the approximation of the spectral density explodes. However, setting $n = 500$ or more seems to have little affect on the convergence of the Hurst estimators when $H \geq 0.2$. Whittle's standard deviation increases as H becomes larger.

The simulation results show that the ergodic estimates are less biased for all values of H when compared to Whittle's estimates. It should be noted that while the Second moment method shows superior performance to all other methods, it assumes that the drift and scale affecting the fBm process are known. The other methods do not require this information to estimate the Hurst index. In the other methods the drift $\mu = 0$ and a scale $\sigma = 1$. These parameters are assumed to be unknown in the estimation of the Hurst index. The Quadrant method gives accurate estimates, however they are not as accurate as the Ratio method.

In the next sub-section we will see that the Quadrant method outperforms the Variance of Residuals method for almost all sample sizes and almost all Hurst index values, however it is not as accurate as Whittle's approximate MLE. We also demonstrate that the Ratio method is better than Whittle's approximate MLE on smaller sample sizes for central values of the Hurst index.

3.2.5.2 Difference Analysis and Numerical Convergence of Hurst Index Estimators

We use the various estimators discussed in this chapter to estimate the Hurst index on the simulated paths of fGn and then compare the estimator's absolute deviation using the paired t-test. If we let,

$$D = average\left[\left|\hat{H}_1 - H_{Actual}\right| - \left|\hat{H}_2 - H_{Actual}\right|\right]$$

$$\sigma_D^2 = Var\left[\left|\hat{H}_1 - H_{Actual}\right| - \left|\hat{H}_2 - H_{Actual}\right|\right]$$

The confidence interval on the statistic D can be shown to be approximately,

$$D \pm \sigma_D t_{\frac{\alpha}{2},N-1}. \tag{3.15}$$

We use equation 3.15 to construct 99% confidence interval for testing the null hypothesis

$$H_0 : \left|\hat{H}_1 - H\right| = \left|\hat{H}_2 - H\right|$$
$$H_1 : \left|\hat{H}_1 - H\right| \neq \left|\hat{H}_2 - H\right|$$

The results of the analysis can be seen below in Figure 3.12. The inclusion of zero in the confidence interval indicates that there is no significant difference in the estimators.

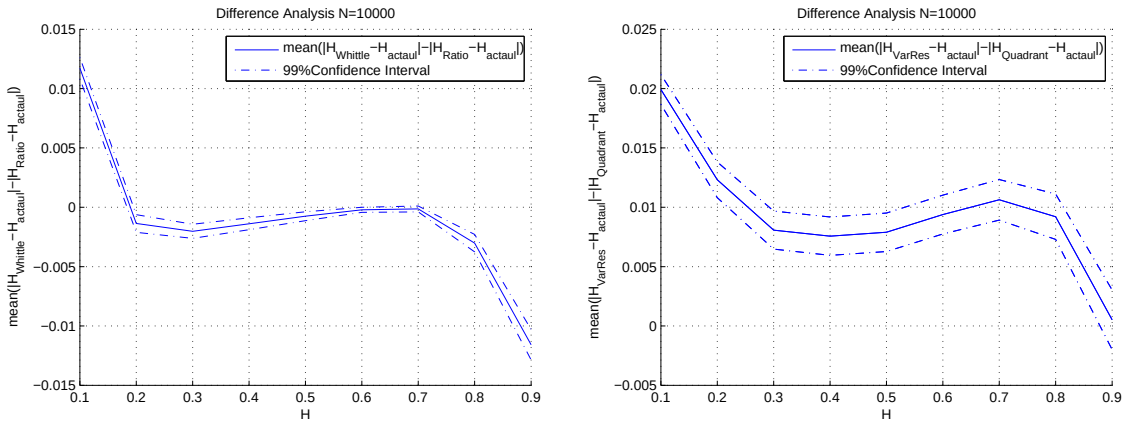


Figure 3.12: Difference Analysis: Large Sample Size

The analysis in Figure 3.12 (left) indicates for $N = 10,000$ that the Ratio method's estimates of H are significantly better than Whittle's estimates on average when $H = 0.1$, and that there is no significant difference between the estimators for $H = 0.6, 0.7$. Whittle's approximate MLE's estimates are slightly better than the Ratio's estimates for the H values between 0.2 and 0.6 and significantly better for 0.8 and 0.9. Furthermore in Figure 3.12 (right), the Quadrant method is shown to be statistically significantly more accurate than the Variance of Residuals method for all Hurst index values except for $H = 0.9$, where there is no statistical difference. The superiority

of the Quadrant method when compared to the Variance of Residuals method is fairly consistent as sample size is decreased (see Appendix II). In Figure 3.13, the difference analysis is expanded to estimates when the sample size (N) is reduced for the Ratio and Whittle estimators.

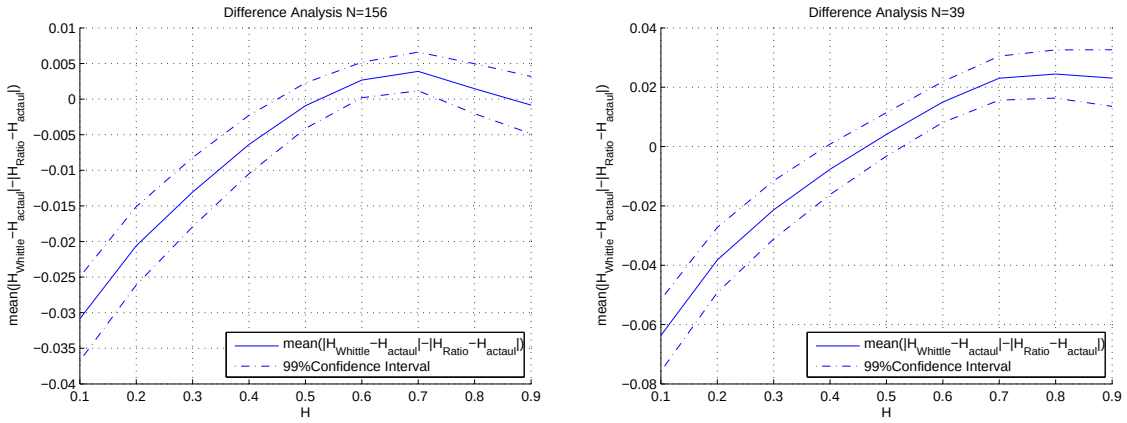


Figure 3.13: Difference Analysis: Small Sample Size

Figure 3.13 shows the difference analysis for Whittle and Ratio method sample paths with $N = 156$ (left) and 39 (right). The analysis indicates that the Ratio method produces more accurate estimates on average for H (in terms of RMSE) than Whittle's method for $H > 0.6$ and equivalent for $H = 0.5$ when $N = 39$ and 156. When N is increased to $N = 625$, the Ratio method still yields estimates that are not significantly different than Whittle's estimates for values of $H = 0.6, 0.7$ and 0.8. It is not until $N = 1,250$ that Ratio performs similarly to Figure 3.12 and falls behind Whittle's approximate MLE. Full details of the difference analysis can be found in Appendix II. The results in Appendix II have also be confirmed via the Wilcoxon signed-rank test.

The superior performance of the ergodic estimators for small sample size is a result of the convergence rate of the estimators. Whittle's approximate MLE converges at a rate of $\sqrt{N}$ (Taqqu et al. [53]) . Figure 3.14 provides a simulation based comparison of the numerical convergence rates of the RMSE.

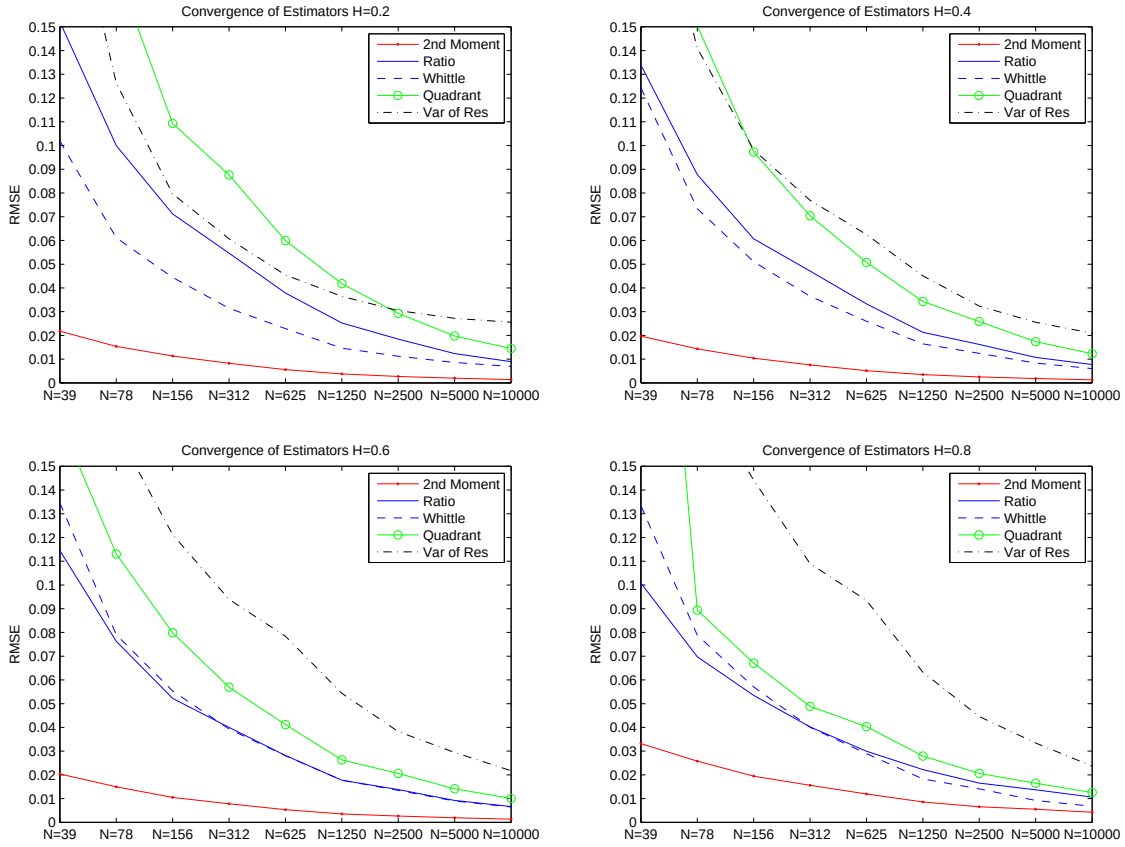


Figure 3.14: Convergence of Hurst Estimators

Figure 3.14 (above) compares the convergence rates for selected H values. Notice that the Ratio method performs similar to Whittle, while the Quadrant method requires $N > 78$ when $H \geq 0.8$. Full details can be found in Appendix III. Notice that the convergence rate for highly auto-correlated processes ($H = 0.2$ and$H = 0.8$) are significantly slower for the ergodic methods. The Ratio method performs similarly to Whittle's method for $H = 0.6$ and 0.7.

3.3 Discussion of Results

In this chapter we have introduced three new methods of estimating the Hurst index using ergodic theory. These methods have been shown to be comparable in performance to leading estimators in terms of RMSE. Our empirical analysis shows the robustness and computational speed of the ergodic estimators. The Second Moment method can be used for estimating the Hurst index when there is known location and scale. This method has been shown to be equivalent

to or more accurate than Whittle's approximate MLE with a computational speed 10^4 faster due to its simplicity. We have shown that the Ratio and Quadrant methods are consistent and competitive estimators of the Hurst index for fractional Wiener processes. The Ratio method becomes comparable to Whittle's estimator for $H \geq 1/2$ index for small sample sets ($N \leq 156$ data points), while the Quadrant method is robust and still outperforms most methods available. All methods introduced are statistically equivalent to or better than Peng's Variance of Residuals method (for most values of the Hurst index), the second best method reported in Taqqu et al. [51].

The primary advantage of the ergodic estimators introduced in this chapter is the availability of a closed-form solution for estimating the Hurst index. Methods like Whittle's approximate MLE require optimization algorithms which can take significant time to calculate. Simpler methods sacrifice accuracy for speed. The ergodic Ratio and Second Moment estimators have speed and simplicity with little sacrifice of accuracy. Additionally, the ergodic estimators show superior relative performance on small sample sizes. These properties are important in such fields as finance (Willinger et al. [57]) and network flow, where fractional Brownian motion models are being used, and reliable and fast estimates of the Hurst index are needed for decision making using small sample sizes.

Chapter 4

Option Pricing and Fractional Brownian Motion

4.1 Literature Review: Fractional Brownian Motion in Finance

Mandelbrot and Van Ness [33] observed that economics is "overwhelmingly devoted to sequences of independent random variables, to Markov processes, and to other random functions having the property that sufficiently distant samples of these functions are independent, or nearly so. [However] empirical studies of random chance phenomena often suggest, on the contrary, a strong interdependence to the total sample size." Mandelbrot and Van Ness proposed the incorporation of a Gaussian model with a "span of interdependence" between increments that is infinite. Mandelbrot claimed that the fractional Brownian motion process (first introduced by Kolmogorov) incorporates attributes that real world time series seem to portray. Consequently, fractional Brownian motion has been used to model financial securities.

4.1.1 Pricing Models with Fractional Brownian Motion

The most popular topic in finance regarding fractional Brownian motion is its applicability to stock returns and option pricing. Option pricing is motivated by the popular Black-Scholes-Merton option pricing model, which assumes that stocks $(S_t)_{t\in\mathbb{R}^+}$ follow the stochastic differential equation

$$dS_t = S_t\left(\mu dt + \sigma dB_t\right), \tag{4.1}$$

where B_t is a standard Brownian motion and drift μ and volatility σ are constants. Using Itô calculus, the solution for the stochastic differential equation in 4.1 gives a geometric Brownian motion which is of the form

$$S_t = S_0 e^{\mu t - \frac{1}{2}\sigma^2 t + \sigma B_t}, \tag{4.2}$$

where S_0 is the initial price of the security at time zero.

According to the Black-Scholes Option Pricing Theory, a European call option contract with T time units until expiration and a strike price K has the price

$$C = S_0\Phi\left(d_1\right) - Ke^{-rT}\Phi\left(d_2\right), \tag{4.3}$$

where Φ represents the standard cumulative normal distribution function

$$\Phi\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx,$$

$$d_1 = \frac{\ln\left(S_0/K\right) + rT + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}},$$

and

$$d_2 = d_1 - \sigma\sqrt{T} = \frac{\ln\left(S_0/K\right) + rT - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

Sottinen [48] and Hu and Oksendal [21] replace the standard Brownian motion B_t with fractional Brownian motion B_t^H in Equation 4.1

$$dS_t = S_t\left(\mu dt + \sigma dB_t^H\right). \tag{4.4}$$

The stochastic differential equation given by Equation 4.4 cannot be solved using Itô calculus since fractional Brownian motion is not a martingale (see 2.27). Instead, Hu and Oksendal [21] use of Wick calculus yields the solution

$$S_t = S_0 e^{\mu t - \frac{1}{2}\sigma^2 t^{2H} + \sigma B_t^H}, \tag{4.5}$$

The call option pricing formula using fBm is computed to be

$$C = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2), \tag{4.6}$$

where

$$d_1 = \frac{\ln(S_0/K) + rT + \frac{1}{2}\sigma^2 T^{2H}}{\sigma T^H},$$

and

$$d_2 = d_1 - \sigma T^H = \frac{\ln(S_0/K) + rT - \frac{1}{2}\sigma^2 T^{2H}}{\sigma T^H}.$$

Replicating portfolios and self-financing representations have been derived for fractional Brownian motion, however, unlike in the Brownian motion model, there is no economical or analytical interpretation the self-financing strategy in Wick calculus. Without these representations, modern pricing theory yields no direct method to hedge risks (or properly capture arbitrage opportunities), and create replicating portfolios of financial derivatives over time. (Hu and Oksendal [21])

Fractional Brownian motion has also been suggested to model interest rate fluctuations in short rate bond models. One commonly used bond pricing model is based on the Ornstein-Uhlenbeck (O-U) process, which takes the form

$$dr_t = \theta(\mu - r_t)\, dt + \sigma dB_t, \tag{4.7}$$

where $\sigma, \theta \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$.

The interest rate, r_t, follows a mean reversion to μ at speed θ. The instantaneous volatility of the process is dictated by σ. Unlike geometric Brownian motion, the process has a bounded variance and achieves a steady state with variance $\frac{\sigma^2}{2\theta}$. The process in discrete time can be modeled as AR(1) process. The solution to the O-U stochastic differential equation 4.7 can be derived using Itô calculus:

$$r_r = r_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \int_0^t \sigma e^{\theta(s-t)} dB_s.$$

Cheridito [11] uses Riemann-Stieltjes integrals to derive tractable solutions to a fractional Ornstein-Uhlenbeck process (with a fractional Brownian motion). We are unaware of any empirical studies of the bond market that test the efficacy of a fractional O-U process for short rates.

Other applications of fractional Brownian motion in finance include the commodities and currency markets (where momentum effects are strong and common). Fractional Brownian motion has shown promise for risk management and the hedging of price fluctuations in real assets.

4.1.2 Arbitrage and Fractional Brownian Motion

Definition 4.1. Arbitrage

1. Type A arbitrage is defined as any strategy that results in an immediate positive reward without entailing any future payoff or cost. It is also called self-financing arbitrage.

2. Type B arbitrage is defined as any strategy which requires a non-positive cost but has positive probability of yielding a positive payoff and has zero probability of yielding a negative payoff.

Mathematically arbitrage is defined in the following way:

Let $\theta(t)$ be a self financing trading strategy. Then there exists an arbitrage strategy (Type A and Type B respectively) if the following conditions hold for some fixed time t:

1. $\theta(0)^T S(0) < 0$ and $P\left(\theta(t)^T S(t) \geq 0\right) = 1$
2. $\theta(0)^T S(0) = 0$ and $P\left(\theta(t)^T S(t) \geq 0\right) = 1$ and $P\left(\theta(t)^T S(t) > 0\right) > 0$

(Glasserman [25], pp.26)

Definition 4.2. Statistical arbitrage is defined as a modification of Type B arbitrage, in which a strategy requires a non-positive cost but has positive probability of yielding a negative payoff. Mathematically, $\theta(0)^T S(0) < 0$ and $P\left(\theta(t)^T S(t) \geq 0\right) > 0$.

Cheridito [10] , Rogers [44], Sottinen and Valkeila [49], and Shiryaev [45] all develop arbitrage strategies for a fractional Brownian motion model. Rogers [44] shows that since fractional Brownian motion is not a semi-martingale (except when $H = 1/2$), "generally this means that there

must be arbitrage." Rogers [44] shows that there exists a trading strategy which can be defined as a continuous adapted process in which the expected value is positive. He then shows that there will be infinitely many promising periods, in any positive time interval, in which this property holds. He demonstrates that a unit amount invested in each promising time period then when the investor's gains has risen by one, the investor would sell and reap the benefit without risk. However, if the investment falls to a positive threshold α, investing one-half of the prior period's investment in each promising period will always result in a net gain of one. In the meantime (while playing the strategy), the investor's wealth can only fall to 4α. Rogers also shows that this strategy (while it may take a long time) will never take an infinite time to realize. Furthermore, he proves that the source of the arbitrage comes from the behavior of the fGn kernel function near zero (namely on small time scales $t^{H-1/2}$ is the reason for the arbitrage opportunities). He suggests a modification of the kernel function which created a new process that still has long range dependence. He concludes that since a model with fractional Brownian motion admits arbitrage, "fractional Brownian motion is an absurd candidate for a log-price process." (Rogers [44])

Cheridito [10] criticizes Rogers' [44] buy and hold strategy and points out that his arbitrage strategy only works for a zero drift process with positive volatility in which one knows the entire history of the stock process from the beginning of time to the present time. Currently, all arbitrage strategies utilize all information about the process back to the beginning of time. Cheridito [10] suggests that arbitrage should only be defined on information available, and therefore current arbitrage strategies are unrealizable. Additionally, Cheridito [10] proves that if B_t is defined as standard Brownian motion and B_t^H is defined as fractional Brownian motion with $H \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$ then the model in which

$$S_t = S_0 e^{\mu t + \sigma\left((1-\xi)B_t + \xi B_t^H\right)},$$

where

$$P\left(\xi = 0\right) = P\left(\xi = 1\right) = \frac{1}{2}$$

admits only statistical arbitrage.

Other arbitrage strategies have been developed based on Rogers [44], all of which derive self-financing arbitrage strategies that are based on stochastic integration and continuous re-

balancing. In the stock market, continuous re-balancing is impeded by the limit on the speed a trade can be executed. Cheridito [10] proves that (even with perfect information) all current arbitrage strategies can only be realized if it is possible to buy and sell within arbitrarily small time intervals. He shows that the introduction of an infinitesimal amount of time $\epsilon > 0$ between any two consecutive transactions eliminates all realizable arbitrage opportunities.

Even with trades taking place in microseconds the stock market is (and always will be) a discrete time process. There will always be a small amount of time between trades and therefore Cheridito's results prove that there is no realizable arbitrage in equity markets. However, this fact does not reconcile the hedging strategies and portfolio replication problems with fBm. While Cheridito proved there is only statistical arbitrage for a process which randomly changes from Brownian motion to fBm, no studies have investigated arbitrage strategies in a fractional Brownian process with changing Hurst index.

A major drawback of arbitrage strategies relates the replication and hedging of risk using derivatives. A model that admits arbitrage opportunities (like fBm) has an infinite number solutions for hedging and replications of a derivative security with a portfolio of stocks and bonds. In short, a model that admits arbitrage cannot allow for perfect hedging and risk management (or calculation of the "Greeks").

4.1.3 Fractional Random Walks

Binomial trees and their relationship to Brownian motion are well studied. Tree representations of processes are important to pricing instruments which can have which have paths that maximize their value (and require execution) before expiration. It can be mathematically proven that derivatives like the American call option maximize their value by executing at expiration, and therefore their is no need for a tree representation. However, the formulation of a binomial tree representation of processes play an important role in pricing complex financial instruments; such as pricing American put options. Construction of a binomial tree that converges to fractional Brownian motion is not as straight forward as the common tree that converges to Brownian motion. For the benefit of the reader, we provide some background on the standard binomial tree setup. Let $(\xi_i)_{i=1}^N$ be i.i.d. random variables with

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

Let us assume Δt is the step size between two consecutive times and let us define $(X_t)_{t\in\mathbb{R}^+}$ as the position at time t with

$$X_t \equiv \sqrt{\Delta t} \sum_{i=1}^{\left[\frac{t}{\Delta t}\right]} \xi_i$$

Suppose time evolves incrementally by size Δt. Then,

$$\mathbb{E}\left[X_t\right] = 0 \text{ and } \mathrm{Var}\left[X_t\right] = \left(\sqrt{\Delta t}\right)^2 \left[\frac{t}{\Delta t}\right] = t.$$

The process $(X_t)_{t\in[0,\infty)}$ is called a random walk. By the Functional Central Limit Theorem the process X_t converges to Brownian motion.

Below in Figure 4.1 we have implemented the binomial tree for Brownian motion.

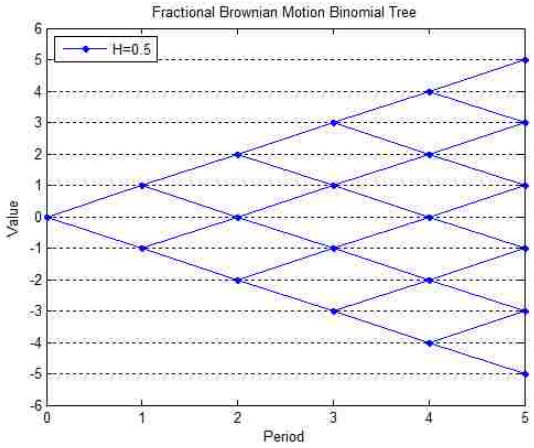


Figure 4.1: Brownian Motion Binomial Tree

Constructing a random walk for Brownian motion is simple since the $(\xi_i)_{i=1}^N$ of the random walk are independent. However, to construct a random walk for fractional Brownian motion, $(\xi_i)_{i=1}^N$ must be correlated. All increments, going back to negative infinity, influence the next increment in the sequence. The auto-correlation of increments decay according to a power function. Sottinen and Valkeila [49] present a binomial tree model that converges to fractional Brownian motion using the kernel representation of fBm, while Lindstrom [28] and Konstantopoulos and Skhanenko [26] both utilize the more tractable kernel function of fGn.

Konstantopoulos and Skhanenko [26] derives a moving average model for the fractional random walk. Given that the fractional noise correlation function decays as a power series when $H > \frac{1}{2}$. The fractional noise process $(Y_i)_{i=1}^N$ with $H \in \left(\frac{1}{2}, 1\right)$ is formed using weights a_j and the

i.i.d. random variables $(\xi_i)_{i=1}^{N}$

$$Y_j = \sum_{k=-\infty}^{\infty} a_{j-k}\xi_k,$$

where

$$a_n = \begin{cases} (n+1)^{H-\frac{1}{2}} - n^{H-\frac{1}{2}} & \text{for } n \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Take note that a convolution of the i.i.d. random variables ξ_i are needed from negative infinity to time j for the correct approximation of fGn. The fractional Brownian motion process $\left(X_t^{(n)}\right)_{t\in\mathbb{R}^+}$ is then given by the scaled sums of $(Y_i)_{i=1}^{N}$ such that

$$X_t^{(n)} = \frac{1}{n^H}\sum_{j=1}^{[nt]} Y_j.$$

We implemented the Konstantopoulos and Skhanenkos [26] weighted sum discrete approximation of fractional Brownian motion with a truncation for negative infinity of $k = -1000$ and the fractal scaling term $\frac{1}{n^H}$ is set to 1. Figure 4.2 illustrates our results for different H values.

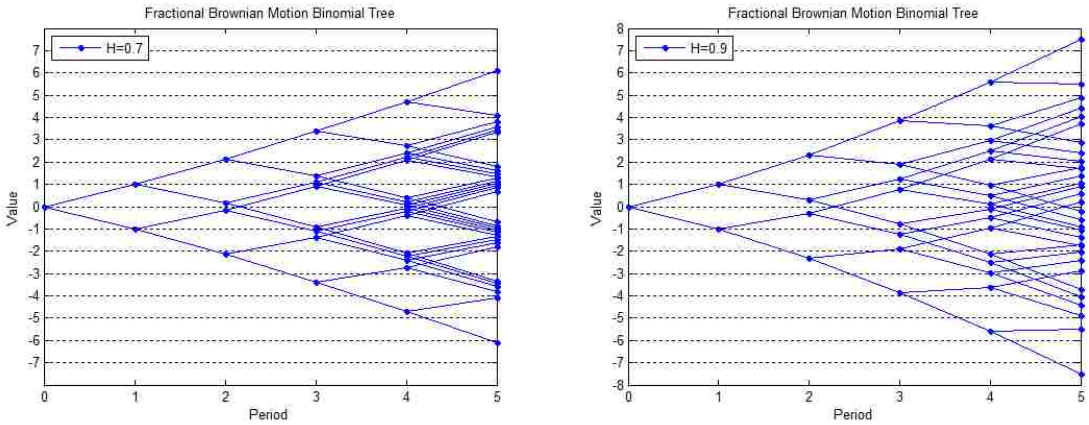


Figure 4.2: fBm Binomial Trees

The values of the binomial tree for Brownian motion have independent increments and therefore changes from one time step to the next are not influenced by past values in the tree and therefore, the tree recombines. However, in the case of long-range dependence the tree is not recombining. To generate a tree with N periods, the recombining tree requires $(N+1)^2/2+(N+1)/2$ values (notice that each period n adds $n+1$ values) and therefore grows linearly. However, for the non-combining tree N periods requires $2^{N+1}-1$ values (each period $n \leq N$ adds 2^n values) and

therefore grows exponentially. Therefore the ability to accurately price put options is limited to smaller time frames in which the values can be stored and backward recursion can be performed.

Contrary to the simulations given in Figure 4.2, the tree will not always be symmetrical since it is dependent on the path prior to time period zero (in fact the tree pivots around time zero). The auto-correlation also results in another interesting property. Notice that the upper path of the binomial tree for $H = 0.9$ at period four branches into two nodes; one where the change from period four is higher and the other where the change from period four is almost zero (and vice versa for the lower path). If these values represented the return on a stock, situations could arise where the return is non-negative with probability one; meaning there would be no risk but the opportunity for return above the risk free rate (or type B arbitrage) for a short time period. Sottinen and Valkeila [49] mathematically identifies two situations where arbitrage is possible in the tree representations; one where $\{\xi_i\}_{i=1}^N = (1, \ldots, 1)$ and the other where $\{\xi_i\}_{i=1}^N = (-1, \ldots, -1)$. He proves that these arbitrage opportunities do not occur on measure zero for long range dependent processes as $N \to \infty$. He also shows that given a constant n_H, the probability of these two arbitrage opportunities occurring is:

$$\lim_{N\to\infty} \frac{\#\text{arbitrage paths}}{2^N} \geq \frac{2}{2^{n_H - 1}} > 0,$$

where if $H = 0.6$ then $n_H \approx 350$ and $\frac{2}{2^{n_H - 1}} = 1.74x10^{-105}$. Note that as $H \to \frac{1}{2}, n_H \to \infty$.

The probability of observing Sottinen and Valkeila [49] arbitrage opportunities in the marketplace with $H = 0.6$ is immensely small and requires all information about the history of the process. To capitalize on Sottinen's arbitrage strategy one would need to observe $(\xi_i)_{i=1}^N = (1, \ldots, 1)$ for an infinitely large N and buy by the next tick. Additionally, by passing on a limit, Sottinen and Valkeila [49] have transformed the process into an analog of the continuous time fractional Brownian motion process, which is already known to exhibit arbitrage opportunities. As mentioned in this section, Cheridito [10] proves that there is no obtainable arbitrage in a binomial tree setting.

4.1.4 Estimation of the Hurst Index in Financial Data

The use of fractional Brownian motion in pricing has generated lengthy debates among economists and practitioners about market efficiency, and whether historical prices can yield information about future results. Some economists claim that a security modeled by fractional Brownian

motion with $H > 1/2$ violates the "Efficient Market Hypothesis" being that information from the past can give insights about the future behavior of the stock process. There has been a plethora of papers which explore the estimation of the Hurst index and more importantly, if the Hurst index indicates significant long-range dependence in real market data. Below we summarize the results of a select few.

Lo [29] utilized a modified R/S statistic to estimate the Hurst index for the Center for Research in Security Prices (CRSP) daily equal-weighted index from July 3rd, 1962 to December 31st, 1987. The sample size of this data was 6,409 observations. Lo's [29] empirical analysis gave an estimate of $H \approx 0.55$ for the 25 year time period. He concluded that there is little historical evidence of long-term memory in the U.S. stock market. He further suggests that the Hurst estimate is within the null hypothesis of $H \leq 0.5$. He established his confidence intervals using the asymptotic distribution of the statistic as N$\rightarrow\infty$. He justifies his claim by the hypothesizing that persistent informational asymmetries over long times must not exists given the frequency the financial markets clear.

Willinger et al. [57] revisited the same data as Lo [29] and performed an analysis using the R/S statistic, the modified R/S statistic and Whittle's MLE to estimate the Hurst index. Willinger et al. [57] demonstrated that the modified R/S statistic (while more robust) has a tendency to have high Type II error (wrongly accepting the null hypothesis of no long-range dependence). They attributed this to the finiteness of the data. He also performed Monte Carlo simulations to determine confidence intervals on the estimates. They calculated a value of $H = 0.58$ (95% confidence $(0.51, 0.65)$) with the R/S statistic with blocks of size 20 data points and $H = 0.63$ (95% confidence $(0.58, 0.68)$) with the R/S statistic with blocks of size 10 data points. Willinger et al. showed that the null hypothesis of no long range dependence was rejected for both of these statistics at a 5% significance level. Additionally, Willinger et al. performed an analysis using Whittle's approximate MLE and obtained an estimate of $H = 0.63$. They believe that there is high probability that there is long range dependence with a historical Hurst parameter for the market is around $H = 0.6$. They cautiously pointed out that the evidence is not absolutely conclusive since the Hurst estimates are so close to one-half. He also analyzed the CRSP daily value-weighted index and concluded that statistically there was no long range dependence for the same time period.

Bayraktar et al. [3] performed an analysis on one-minute intervals of the S&P 500 Index data from January 1989 to May 2000 (1,128,360 observations). Unlike Lo and Willinger they perform

the analysis over disjoint segments, showing how the Hurst index has behaved over time. They used a wavelet based method (log scale spectral analysis) and split the data into 275 segments of length 2^{12} (two weeks) and concluded that the Hurst estimates changed from segment to segment with a mean of $H = 0.6156$ and a standard deviation of 0.0531. When the data was divided into larger segment sizes they obtained similar Hurst estimates, however when they estimated for time intervals of 2^{15} data points (two months) or greater the Hurst estimators became exhibited extreme variance which were not consistent with other estimates. They believe the breakdown at 2^{15} data points is due to dominant non-stationarity (and significant changes in the underlying parameters). They concluded that on average $H \approx 0.6$ and historically stayed constant for approximately two months at a time. They show that "the Hurst parameter of this data set is significantly above the efficient markets value of $H = \frac{1}{2}$, it began to approach that level in 1997. Bayraktar et al. conjectured that the increased use of the Internet in disseminating stock quotes and in trading may explain why the Hurst parameter was found to be lower after 1997. In §4.3, we estimate the Hurst index for three stock market indexes using the ergodic Ratio method. Similar to Bayraktar, we give empirical support that the Hurst index changes over time.

4.2 The Tail Behavior of fBm from a Bm Lens

The use of a Brownian motion as a model for log stock returns has revealed many inadequacies of the model. Two of the most common problems relate to the volatility of the process. After the crash of 1987 a pricing phenomenon in the derivatives world started to emerge. Investors, realizing that the stock market had larger tails than standard Bm models, started valuing derivatives that were out of the money at higher values than theoretical prices. Therefore, when the Black-Scholes equation is used to price call options, the resulting prices deviate in greater magnitude as the strike price gets further away from the spot price (the stock's current price). This was the birth of the "volatility smile" and the "term structure" to volatility.

Many traders calculate the "implied volatility" or volatility that would be needed (in the Black-Scholes model) to match the market price for an option. The implied volatility is typically higher than the estimated volatility from the stock data and increases as the strike prices gets further from the spot price, a phenomenon known as a "volatility smile." Additionally, traders have noticed a volatility term structure when pricing options, where the implied volatility depends on the maturity of the option. Implied volatility tends to be an increasing function of maturity

when short dated volatilities are low and vice versa when short date volatilities are high. The current logic for this phenomenon is that Brownian motion and the Gaussian distribution do not have large enough tails to encompass long range events and/or extreme events. (Hull [23], pp.276, 377) We challenge this statement by proposing that temporal dependence could explain such a phenomenon.

Let us assume that a security follows the geometric fractional Brownian motion model equation 4.5

$$S_t = S_0 e^{\mu t - \frac{1}{2}\sigma^2 t^{2H} + \sigma B_t^H}.$$

Notice that the security prices are log-normally distributed, where

$$\ln \frac{S_t}{S_0} \stackrel{d}{=} N\left(\mu t - \frac{1}{2}\sigma^2 t^{2H}, \sigma^2 t^{2H}\right).$$

Let us examine the geometric Brownian motion model that is used in the derivation of the Black-Scholes-Merton option pricing theory. Recall that the underlying stock price model is given by equation 4.2

$$S_t = S_0 e^{\mu t - \frac{1}{2}\sigma^2 t + \sigma B_t}.$$

In this model security prices are log-normally distributed such that

$$\ln\left(\frac{S_t}{S_0}\right) \sim N\left(\mu t - \frac{1}{2}\sigma^2 t, \sigma^2 t\right).$$

Let us look into the volatility estimates if we assume the security follows this Brownian motion when it is actually (by assumption) fractional Brownian motion. Under this assumption, given a time series $\{S_t\}_{t=0}^{N}, N \in \mathbb{N}$, the estimate of the variance of the process, $\ln\left(\frac{S_t}{S_{t-1}}\right)$, is found by the Gaussian MLE estimate for the second moment:

$$\hat{\sigma}^2 \Delta t = \frac{1}{N-1}\sum_{t=1}^{N}\left[\ln\left(\frac{S_t}{S_{t-1}}\right) - \hat{\nu}\right]^2,$$

where

$$\hat{\nu} = \frac{1}{N}\sum_{t=1}^{N}\ln\left(\frac{S_t}{S_{t-1}}\right).$$

If we model the security prices using geometric Brownian motion we would estimate the variance as $\hat{\sigma}^2\Delta t$ when in actuality the variance of the fGn process is $\sigma^2\Delta t^{2H}$ (as demonstrated using ergodic theory in the last section). To find the call option prices we need the volatility explicitly. Therefore the estimated volatility (assuming a white noise model when the actual model is fGn) would be

$$\hat{\sigma} = \sigma\Delta t^{H-1/2}.$$

Notice that the Brownian motion estimate depends on the sampling resolution Δt, where Δt is assumed to be constant. Therefore the Brownian motion volatility estimate *underestimates* the actual volatility when the fBm process has long range dependence and overestimates the volatility when the fBm process has short range dependence

$$\hat{\sigma} < \sigma \text{ for } H \in \left(\frac{1}{2}, 1\right)$$

$$\hat{\sigma} > \sigma \text{ for } H \in \left(0, \frac{1}{2}\right).$$

Figure 4.3 depicts the difference between the actual volatility and the estimated volatility when Brownian motion is used to model log returns and the log returns time series follows a fractional Brownian motion with a specified H. This illustration assumes that data is sampled once a day, hence $\Delta t = 1/252$. This holds the convention that σ is scaled as an annual number (which is typically needed for the option pricing formula). The security in this example has an actual volatility of $\sigma = 20\%$.

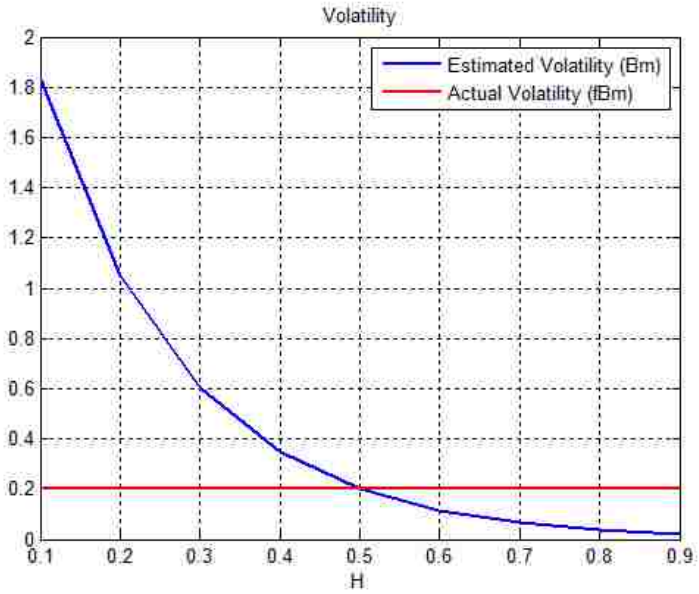


Figure 4.3: Estimated vs. Actual Volatility

Notice that the short range dependence of fractional Brownian motion has a much larger effect on the volatility estimate if Brownian motion is assumed. The bias in the volatility estimates are a result of the i.i.d. assumption made by the Brownian motion model (which results in a linear time scaling of volatility), when the actual model has dependence (which is formed by the non-linear structure of the variance of fBm). Given the Brownian motion assumption in the presence of dependence, an empiricist would forecast prices with less (when there is positive auto-correlation) risk than is actually present. Figure 4.4 shows the empirical (Bm) distribution when the "true" distribution is a fBm with $H = 0.6$, $\sigma = 1$ and $\Delta t = 1/252$. Even though both distributions are Gaussian, this underestimation of volatility can give the illusion of "large tails."

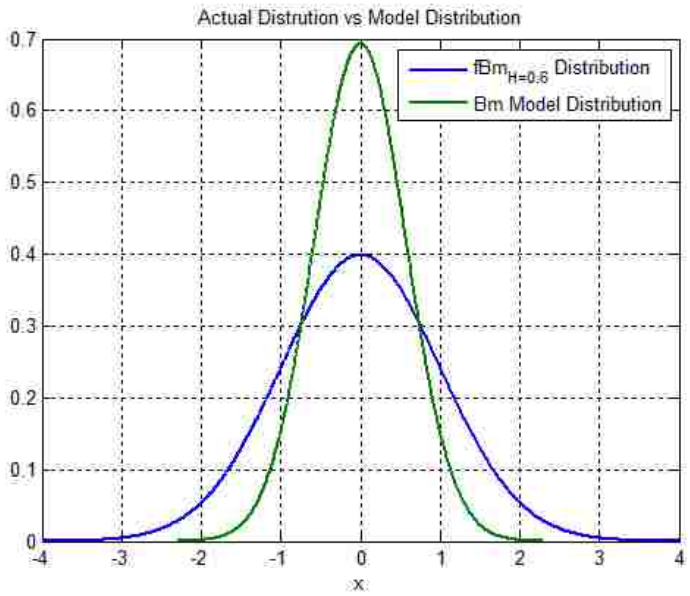


Figure 4.4: Illusion of "Heavy Tails"

The empirical studies in §4.1.4 suggests that the stock market on a whole exhibits long range dependence with an average Hurst index of approximately $H = 0.6$. In this situation, a Brownian motion model would price options lower than their actual value (high volatility translates to higher expected payoff and therefore a higher option price). In the Black-Scholes-Merton call option pricing formula (Equ. 4.3), the time to expiration T scales the volatility, while the volatility in the fBm option pricing formula (Equ. 4.6) is scaled by T^{2H}. The scalar T in the Brownian motion based model pushes the skewed volatility estimate toward the actual value since $\sqrt{T} > T^H$ when $H \in \left(\frac{1}{2}, 1\right)$ and $T < 1$. However since it is typical that $\Delta t < T$, the scaling term $\hat{\sigma}\sqrt{T}$ would still underestimate the actual term σT^H. Under the assumption that $\Delta t < T < 1$,

$$\hat{\sigma}\sqrt{T} < \sigma T^H \text{ for } H \in \left(\frac{1}{2}, 1\right)$$

$$\hat{\sigma}\sqrt{T} > \sigma T^H \text{ for } H \in \left(0, \frac{1}{2}\right).$$

In figure 4.5, we assume $r = 5\%$, $\sigma = 20\%$, $K = 100$, $\Delta t = 1/252$, $H = 0.6$ (for the upper figure) and $T = 1/12$ (for the lower figure).

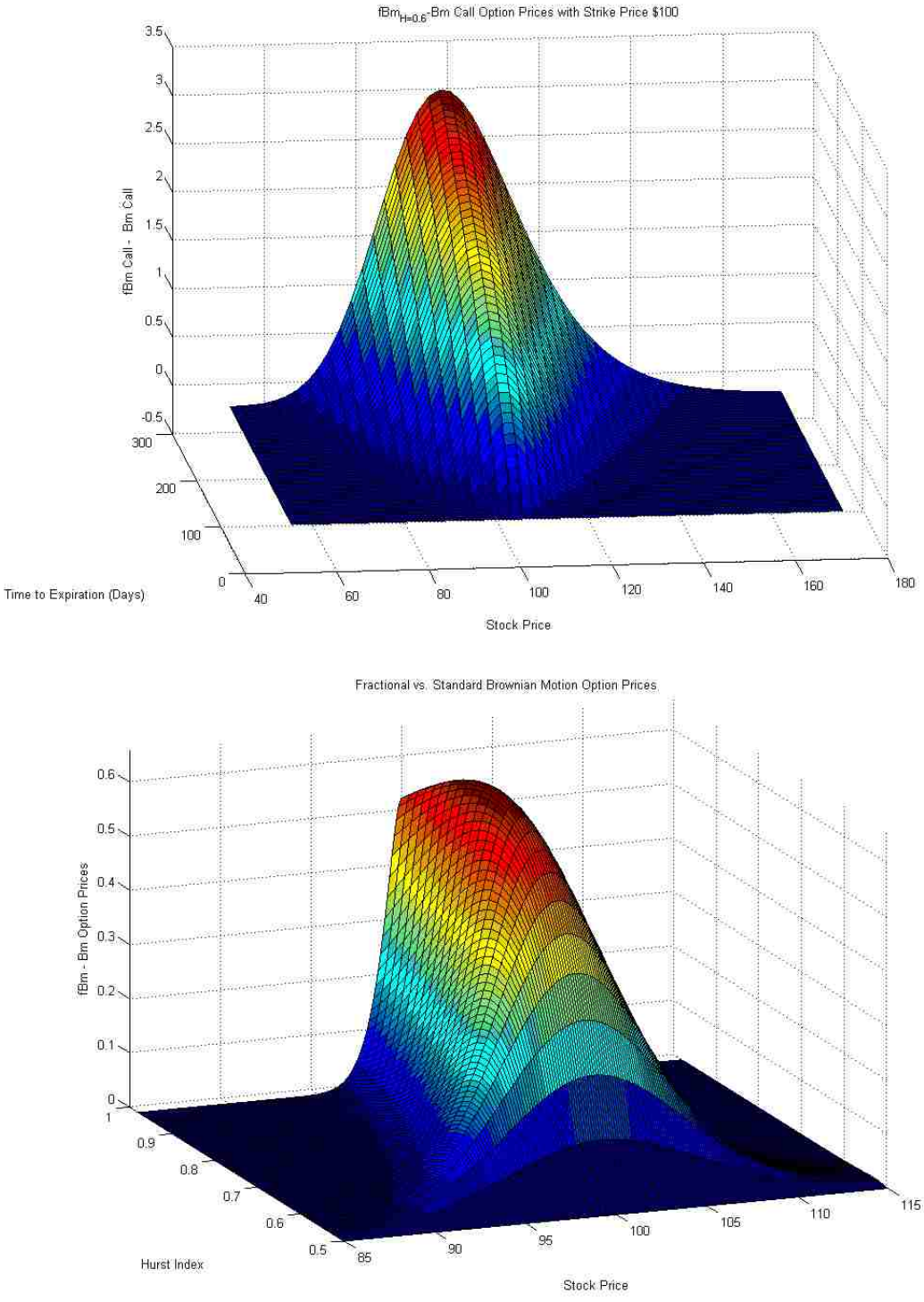


Figure 4.5: Bm vs. fBm Call Option Prices

If the market followed a fractional geometric Brownian motion process with $H \neq 0.5$, and the time series was analyzed using geometric Brownian motion, the call prices would *not* exhibit a

volatility smile (assuming the market accurately reflected the call option fractional Black-Scholes prices). However, Black-Scholes-Merton option pricing formula would *underestimate* the true call price for all strike prices, with the biggest difference being near the spot price. Therefore, the implied volatility would be higher when $H > 1/2$, but would be consistently skewed across all strike prices K. This is because the two models only differ by time scaling factors (T^H and $\sqrt{T}$), while all other inputs are constant. Notice that the implied volatility (assuming Black-Scholes-Merton) for all strike prices satisfies the equation,

$$\sigma_{Implied} = \sigma T^{H-1/2} \forall K. \tag{4.8}$$

In this situation, the implied volatility would be greater than the estimated volatility

$$\sigma_{Implied} = \sigma T^{H-1/2} > \hat{\sigma} \Delta t^{H-\frac{1}{2}},$$

when $H \in \left(\frac{1}{2}, 1\right), \Delta t < 1$ and $T > \Delta t$. In the figure below, we see that the implied volatility scaling factor that affects the volatility estimates over time to demonstrate the term structure of volatility when the Black-Scholes-Merton model is used in a fractional Brownian motion market.

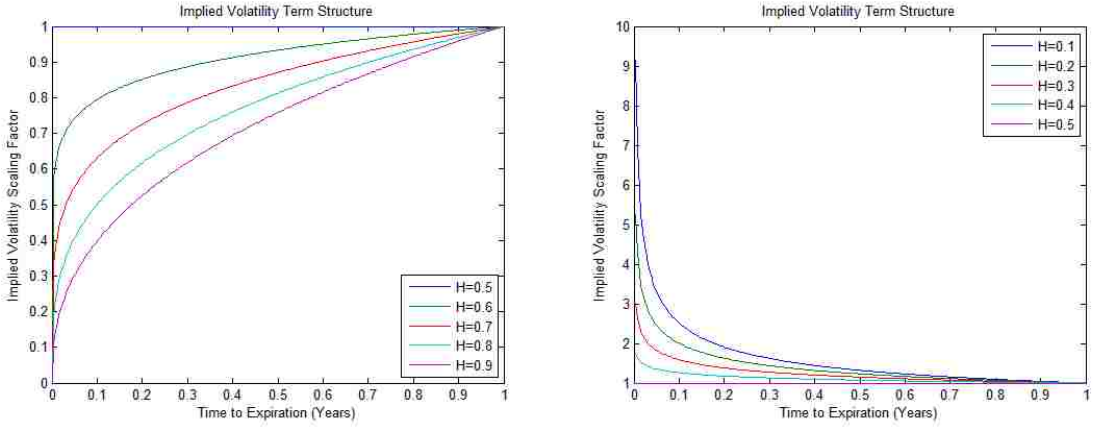


Figure 4.6: Implied Volatility Term Structure

This analysis shows that the time scaling effects of fBm cause when $H > 1/2$ could explain the term structure to implied volatility that is observed in the marketplace (when the fractional Wiener process is analyzed with the Black-Scholes-Merton assumptions). The fBm model does not scale time linearly as the Brownian motion model does, and this means that risk does not scale linearly with time. Equation 4.8 shows that as we approach expiration of the option, $T \to 0$,

the scaling factor $T^{H-1/2}$ has less of an effect on the true volatility when $H > 1/2$. This means that the implied volatility would be smaller as we approach expiration. Conversely, if $H < 1/2$ it has the opposite effect. Both situations have been observed in the marketplace. (Hull [23], pp.377)

4.3 An Empirical Study of the Hurst Index in Financial Indexes

In §4.1.4, we outline Lo [29] and Willinger's [57] analysis of the CRSP equal and value weighted index daily returns from July 3rd, 1962 to December 31st, 1987. They assumed that returns are log-normally distributed and found a Hurst estimate of $H = 0.63$ with a 95% confidence interval of $(0.58, 0.68)$ for the equal weighted index. This is one estimate of H using the entire 25.5 year time period. Using their estimates of H, they concluded that there was no (significant) long range dependence for the value weighted index. For the same data, we estimate the Hurst index using the ergodic Ratio of Second Moments method described in §3.2. We obtain

$$H = \begin{cases} 0.76 \ (0.72, 0.76) & \text{for the Equal Weighted Index} \\ 0.65 \ (0.61, 0.68) & \text{for the Value Weighted Index} \\ 0.61 \ (0.57, 0.69) & \text{for the S\&P 500 Index} \end{cases}$$

Utilizing the ergodic Ratio of Second moments method with a moving window of $N = 630$ data points (2.5 years), we extend the CRSP data set by 22 years and obtain daily returns from July 3rd, 1962 to December 31st, 2009 for three types of indexes from the Center for Research in Security Prices (CRSP). Figure 4.7 shows the time based estimation of the Hurst parameter on a daily rebalanced equal weighted return index of all US based stocks, a value (market capitalization) weighted index of all US based stocks, and the Standard and Poor's 500 index (500 largest-capital US stocks).

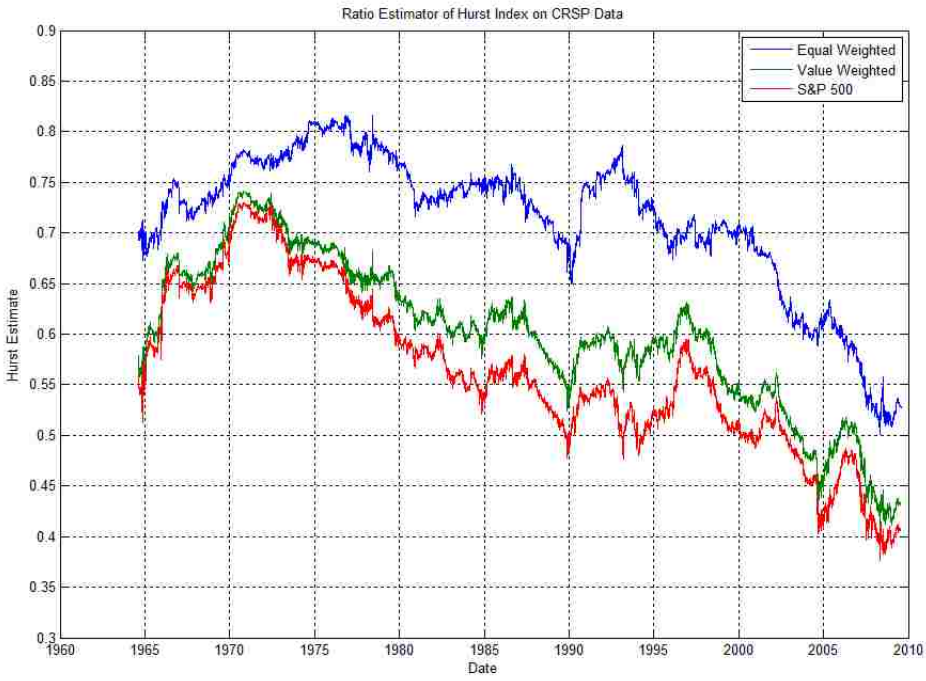


Figure 4.7: CRSP Data Analysis of the Hurst Index

The moving windows of $N = 630$ data points in Figure 4.7 have an approximate 95% confidence interval of ± 0.031 (approximated numerically using simulations). Our analysis shows that the Hurst index has been generally decreasing since the 1970's. This agrees with Bayraktar [3] who performed a similar analysis on one minute intervals of S&P 500 data from January 1989 to May 2000 using wavelets on segments of 2^{15} data points. He concludes the Hurst parameter is changing over time and seems to be tending toward $H = 0.5$. Figure 4.7 indicates that the Hurst index is changing significantly over time. When we change the number of data points in the moving window, we obtained similar trends and results. The moving average of 630 data points was chosen because it allows for some structure of the variations in H to be seen. As expected, an increase the number of data points results in a smoother path for the Hurst index, but the same trend remains.

A closer look at the analysis shows that there may be a form of dominance between the Hurst index values and the index being analyzed; Particularly, the basket of stocks exhibit different temporal dependence depending on how the portfolio is weighted. We notice that the equally weighted index (which puts more emphasis on small cap and penny stocks than the other two indexes) has consistently much higher Hurst index values than the value weighted and S&P

indexes. The S&P 500 and the value weighted index are very similar in how they are constructed, however the value weighted index has a subset of stocks with less liquidity and capitalization than the S&P 500. The value weighted index is consistently exhibiting higher Hurst index values over the S&P. These observations suggest that the capitalization or liquidity of an asset may have a direct relationship to the Hurst index, particularly the lower the liquidity, the higher the Hurst index.

We hypothesize that low liquidity causes greater price friction in the marketplace and therefore prices tend to exhibit higher auto-correlation and consequently a higher Hurst index; just as higher friction (or high Reynolds numbers) in fluids causes movements in turbulent flow to exhibit higher auto-correlation and Hurst index. Additionally, as liquidity and the availability of information has increased, the Hurst index has historically trended toward $H = 0.5$. Lastly, as information has become easier to access and trading quicker to execute, the Hurst index has decreased.

In summary, the analysis of the Hurst index leads to three primary observations:

1. The Hurst index is not constant over time.
2. The Hurst index is greater than $H = 0.5$ (on a 95% confidence interval) for most of time period analyzed.
3. Capitalization/Volume/Liquidity may influence the Hurst index.

One could also hypothesize that since the equity and bond markets are connected then long range dependence in the stock market could also mean long range dependence in the bond market. Given this assumption and Cheridito's [11] work on a fractional Ornstein-Uhlenbeck process, a short rate model could be developed to estimate the Hurst index for the bond market and analyze its significance in bond prices.

Part II

Gaussian Markov Processes

Chapter 5

Introduction and a Brief Literature Survey

In Part I we introduced new techniques and insights in modeling with fractional Brownian motion. We showed that our new techniques are fast and competitive to top methods. We also find that many of the properties of fractional Brownian motion make it a better candidate for modeling in finance than Brownian motion. However, in the preceding section we also discussed that the popularity of fBm in finance has been limited by a few barriers in theory and implementation. The largest restriction of the model is that Option Pricing Theory cannot be derived using stochastic calculus. FBm is not a semi-martingale, and therefore computations are limited to the use of Wick Calculus. Moreover, Rogers [44] shows that the fBm model exhibits arbitrage strategies. The ability to make riskless profit violates the "First Fundamental Theorem of Asset Pricing". (Shreve [46], Theorem 5.4.7) Additionally, Sottinen [47] shows that replication of the option price through a portfolio of stocks and bonds and self-financing strategies have no economic meaning in the Wick calculus case. This combined with the arbitrage violation limits the formation of proper risk management and hedging techniques.

Another restriction in the application of fBm is in regards to pricing of American style and barrier options. Tree representations of fBm are difficult to construct and they do not recombine. (Konstantopoulos and Skhanenko [26]) Without a recombining tree, construction and computations of the tree grows exponentially as the number of time periods increases (The tree has $2^{n+1} - 1$ nodes for periods $n = 0, 1, \dots$). This makes pricing many types of options computa-

tionally infeasible and inaccurate since the tree structure limits tree generation to a small number of periods. Lastly, while the market has shown signs of long-memory, there may be other dependence structures that can better model the process. In fBm, once a Hurst index is determined, the structure of fBm is fixed. The fBm model definition restricts incorporating heteroscedasticity, non-stationarity, or different dependence structures which have been observed in the markets. (Cont [12])

In an attempt to expand beyond the limitations of fBm, Part II of this thesis focuses on continuous sample path Gaussian processes of the Markov type. General diffusion models in physics, thermodynamics, fluid mechanics, biology and finance are all Markovian and many are Gaussian. Gaussian Markov (GM) processes are widely used to model many stochastic dynamic systems. In these fields, parameter estimation, discrete representations and stochastic calculus methodologies are used for decision making under uncertainty.

One of the most prevalent applications of GM processes is in financial mathematics, where pricing strategies for derivatives on equities, commodities, foreign exchange rates, and interest rate products are largely based on GM models of the underlying asset. With the ever increasing speed of trading algorithms and movement of prices, the need for fast and accurate methodologies for proper pricing and risk management are needed. The goal of this part of the thesis is to develop methods for quick and accurate modeling and decision making. In finance, commodity prices, foreign exchange rates and stock prices are commonly evaluated by models such as

- Bachelier: $dX_t = \mu dt + \sigma dB_t$ (Note that this was the first model in mathematical finance)
- Black-Scholes: $dX_t = \mu X_t dt + \sigma X_t dB_t$,

while the term structure of interest rates (known as short-rate models) and some foreign exchange rates are commonly evaluated with models by

- Ornstein-Uhlenbeck: $dY_t = \beta X_t dt + \sigma dB_t$
- Vasicek: $dX_t = \theta(\mu - X_t)\,dt + \sigma dB_t$ (Note that Vasicek is a type of Ornstein-Uhlenbeck model)
- Dothan: $dX_t = (\alpha + \beta X_t)\,dt + \sigma X_t dB_t$
- Black-Derman-Toy: $dX_t = \beta(t)\,X_t dt + \sigma(t)\,X_t dB_t$
- Black-Karasinski: $d(\log X_t) = (\alpha(t) + \beta(t)\log X_t)\,dt + \sigma(t)\,dB_t$

- Ho-Lee: $dX_t = \alpha(t)\,dt + \sigma dB_t$
- Hull-White (Extended Vasicek): $dX_t = (\alpha(t) + \beta(t) X_t)\,dt + \sigma(t)\,dB_t$.

In these models $(B_t)_{t\in[0,T]}$ is a standard Brownian motion process. Each of the mentioned models can be represented either as a GM process or a simple transformation of a GM process. These models incorporate a wide range of properties including stationary or non-stationary transition probabilities, homoscedasticity or heteroscedasticity, and long-range dependence, short-range dependence or memoryless. Furthermore, GM process models, like these, can be shown to be arbitrage free under some weak restrictions. (Shreve [46])

In this section we present definitions and key theorems for Gaussian Markov processes that will be needed in the construction of binomial trees, quadratic variation and stochastic calculus. Some of the results are taken from *Gaussian Processes* by Hida and Hitsuda [20] , Ojeda [34] and Hida [19].

Definition 5.1. (Gaussian Process)

A process $(X_t)_{t\in T}$ where T is any index set (for example a sphere or real line) is a Gaussian process if for any finite set $\{t_1, t_2, \ldots, t_n\}$, the random variable $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is an n dimensional normal.

A sub-class of Gaussian process indexed by $\mathbb{R}$ is described below.

Given a function $\{F(t,s), s \le t\}$ such that $\int\int |F(t,s)|\,dt\,ds < \infty$, a Gaussian processes $(X_t)_{t\in\mathbb{R}}$ can be defined by

$$X_t = \int_0^t F(t,s)\,dB_s, \text{ (Wiener integral)}$$

where $(B_t)_{t\in\mathbb{R}}$ is a standard Brownian motion process. If X_t can be written in the above form, it is said to be *represented canonically*, and $F(t,s)$ is called a *canonical kernel*. (Hida [19])

If $\int\int F^2(t,s)\,dt\,ds < \infty$ then $(X_t)_{t\in\mathbb{R}}$can be decomposed as

$$X_t = \int_0^s F(t,u)\,dB_u + \int_s^t F(t,s)\,dB_s$$

and therefore

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = \int_0^s F(t,u)\,dB_u,$$

where $\mathcal{F}_t = \sigma(X_r \mid r \le t)$.

Definition 5.2. (Markov Process)

The process $(X_t)_{t\in T}$, where $T \subseteqq \mathbb{R}$, is a Markov process if

$$\mathbb{E}\left[f\left(X_t\right) \mid \mathcal{F}_s\right] = \mathbb{E}\left[f\left(X_t\right) \mid X_s\right], \forall t > s,\, t, s \in T$$

where $\mathcal{F}_t = \sigma\left(X_{t,r\in T} \mid s \leq t\right)$, f is a bounded Borel function.

Markov processes provide a framework to describe the relationship between past observations and future values of a stochastic process.

In 1960 Hida [19] established the following theorem.

Theorem 5.3. *(Gaussian Markov Process)*

Let $(X_t)_{t\in I}$ be a non-degenerate centered Gaussian process on the interval $I = [0,\infty)$ or $[0,T]$ with continuous non-zero covariance matrix on I^2. Then the process $(X_t)_{t\in I}$ is a Markov process if and only if there exists a continuous function Ψ, $\Psi(t) \neq 0, \forall t \in I$, so that the process $(M_t)_{t\in I}$ defined by

$$M_t = \Psi^{-1}(t)\, X_t$$

is a Gaussian martingale.

Remark 5.4. If we set $\nu(t) \equiv \mathbb{E}\left[M_t^2\right]$, then by Jensen's inequality, ν is a monotone increasing function on I.

The following corollary is a straightforward consequence of Theorem 5.3 which is proved in Hida and Hitsuda [20] on page 30.

Corollary 5.5. *(GM Covariance Structure)*

A non-degenerate Gaussian process is Markov if and only if the covariance is expressed in the form

$$Cov\left(X_t, X_s\right) = \Psi(t)\,\Psi(s)\,\nu(s), \;\; t \geq s.$$

Moreover, the underlying Gaussian martingale $(M_t)_{t\in\mathbb{R}}$ has covariance

$$Cov\left(M_t, M_s\right) = \nu(t \wedge s),$$

where $x \wedge y$ denotes the minimum of x and y.

Remark 5.6. The theorem and corollary above will be extensively used throughout this dissertation.

Theorem 5.7. *(The Itô Itegral is a Martingale Process)*

Let $\{f(s) : s \in [0,T]\}$ *be an adapted process such that* $\int_0^T \mathbb{E}\left[f^2(s)\right] ds < \infty$. *Then every Itô integral*

$$I_t = \int_0^t f(s)\, dB_s,\ t \in [0,T]$$

is a martingale with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$, *that is*

$$\mathbb{E}\left[I_t \mid \mathcal{F}_s\right] = \int_0^s f(u)\, dB_u = I_s.$$

A proof is provided in Theorem 4.2.1 in Shreve [46] using elementary processes.

Moreover if f *is a deterministic function then* $(I_t)_{t\in[0,T]}$ *is a Gaussian martingale.*

Remark 5.8. Many Gaussian martingales are "time changes of Brownian motion", notice that the theorem above can be used to establish this statement.

The following two results will be extensively used in the dissertation and are provided here for the reader's convenience.

Theorem 5.9. *(Doob's Martingale Maximal Inequality)*

Let $(X_t)_{t\in\mathbb{R}}$ *be a continuous or discrete non-negative real valued sub-martingale. Then Doob establishes that, for any constant* $C > 0$ *and* $p \geq 1$,

$$P\left(\sup_{0\leq t\leq T} X_t \geq C\right) \leq \frac{\mathbb{E}\left[X_T^p\right]}{C^p}.$$

Lemma 5.10. *(Levy's Lemma)*

Let ξ_i be symmetric, that is $P(\xi \leq x) = P(-\xi \leq x)$, then

$$P\left(\max_{1\leq k\leq K}\left|\sum_{i=1}^{K}\xi_i\right| > x\right) \quad \leq \quad 4P\left(\left|\sum_{i=1}^{K}\xi_i\right| > x\right).$$

Proof in Appendix V.

The following definition comes from Steele [50] on page 126.

Definition 5.11. (Itô and Diffusion Processes)

Let $(B_t)_{t\in[0,T]}$be a standard Brownian motion process and $\mathcal{F}_t = \sigma(B_r \mid r \leq t)$. An Itô process $(X_t)_{t\in[0,T]}$ is a stochastic process of the form

$$X_t = X_0 + \int_0^t \mu(s,\omega)\, ds + \int_0^t \sigma(s,\omega)\, dB_s,$$

where X_0 is non-random and $(\mu(t,\omega) \mid \omega \in [0,1], t \in [0,T])$ and $(\sigma(t,\omega) \mid \omega \in [0,1], t \in [0,T])$ are adapted stochastic processes such that

$$P\left(\int_0^T |\mu(s,\omega)|\, ds < \infty\right) = 1 \text{ and } P\left(\int_0^T |\sigma(s,\omega)|^2\, ds < \infty\right) = 1.$$

Furthermore, the Itô process $(X_t)_{t\in[0,T]}$ is called a *diffusion process* if

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

that is

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s,$$

where functions μ and σ are called the *drift* and *diffusion*, respectively.

Remark 5.12. Diffusion processes are widely used as models in physics and financial engineering. In this thesis, we focus on understanding the scaling function and underlying martingales and their affect on modeling with dependence for Gaussian processes.

Remark 5.13. In this dissertation, we are interested in diffusion processes because they are Markovian. The next corollary shows when diffusion processes are Gaussian.

Corollary 5.14. *(Gaussian Itô Processes)*

Let Ψ be a continuously differentiable function on $[0,T]$. Then, a centered continuous sample path Gaussian Markov process satisfying Hida's representation (Theorem 5.3), is represented as

$$X_t = \Psi(t)\, M_t,$$

is a centered diffusion process of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t,$$

when

$$\mu(t, X_t) = \frac{X_t}{\Psi(t)}\Psi'(t) \text{ and } \sigma(t, X_t) = \Psi(t)\sqrt{\nu'(t)}.$$

Recall, that in Remark 5.4,we assume that $\nu(t) \equiv \mathbb{E}\left[M_t^2\right]$.

Furthermore, non-centered diffusion processes take the form

$$X_t = \int_0^t \mu(t, s) X_s ds + \Psi(t) \int_0^t \sqrt{\nu'(s)} dW_s, \tag{5.1}$$

where μ *is a continuous integrable function.*

Proof. Since $X_t = \Psi(t) M_t$, then the stochastic differential equation takes form

$$dX_t = M_t d\Psi(t) + \Psi(t) dM_t.$$

Using Theorem 5.7,

$$dM_t = \sqrt{\nu'(t)} dB_t,$$

then the stochastic differential equation for a Gaussian Markov process takes the form

$$dX_t = M_t \Psi'(t) dt + \Psi(t) \sqrt{\nu'(t)} dB_t.$$

□

Remark 5.15. Note that the list of Gaussian Markov models in finance at the beginning of this section are all diffusion processes that can be written as equation 5.1 or a simple transformation. However, not all Gaussian Markov processes can be written as a diffusion process. There are martingales that *cannot* be represented as an Itô integral.

Remark 5.16. This part of the dissertation will exclusively deal with Gaussian Markov (GM) processes and the development of new techniques based on Hida's [19] underlying martingale representation (Theorem 5.3).

Proposition 5.17. *(Long/Short-Range Dependence of GM Processes)*

The function Ψ *controls the dependence structure of a Gaussian Markov processes.*

Proof. Let

$$\gamma_n = \mathbb{E}\left[(X_1 - X_0)(X_{n+1} - X_n)\right],\ n \geq 1.$$

Substituting the definition of the Gaussian Markov we obtain

$$\begin{aligned}\gamma_n &= \mathbb{E}\left[(\Psi_1 M_1 - \Psi_0 M_0)(\Psi_{n+1} M_{n+1} - \Psi_n M_n)\right] \\ &= (\Psi_{n+1} - \Psi_n)(\Psi_1 \nu_1 - \Psi_0 \nu_0).\end{aligned}$$

The sum of the equation above is telescoping and therefore,

$$\sum_{n=1}^{N} \gamma_n = (\Psi_{N+1} - \Psi_1)(\Psi_1 \nu_1 - \Psi_0 \nu_0).$$

Therefore if:

1. $\Psi_{N+1} = \Psi_1, \forall N$ then the process X has *independent increments.*
2. $\lim_{N\to\infty} \Psi_N = \infty$, then the process X has *long-range dependence.*
3. $\lim_{N\to\infty} \Psi_N = c < \infty$,then the process X has *short-range dependence.* □

Example 5.18. Determination of Ψ and ν in the Vasicek/Ornstein-Uhlenbeck Model

The Ornstein-Uhlenbeck process $(X_t)_{t\in[0,T]}$ is a common model for modeling mean-reversion in interest rate derivative pricing and the length dynamics of over-damped springs under thermal fluctuations. The Ornstein-Uhlenbeck process

$$dX_t = \theta(\mu - X_t)\,dt + \sigma dB_t,$$

where $(B_t)_{t\in[0,T]}$ is a standard Brownian motion process and where constants $\mu > 0$ is the long term mean reversion, $\theta > 0$ is the dampening constant (or strength of reversion), and $\sigma > 0$ is the volatility or diffusion parameter.

The solution to the stochastic differential equation can be found using stochastic calculus and is given by

$$X_t = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma e^{-\theta t} \int_0^t e^{\theta s} dB_s.$$

Using the Itô isometry on the stochastic term, yields

$$\begin{aligned} Var\left(\sigma e^{-\theta t}\int_0^t e^{\theta s}dB_s\right) &= \sigma^2 e^{-2\theta t}\int_0^t e^{2\theta s}ds \\ &= \frac{1-e^{-2\theta t}}{2\theta}. \end{aligned}$$

If $s \leq t$,

$$\begin{aligned} Cov\left(\sigma e^{-\theta s}\int_0^s e^{\theta u}dB_u, \sigma e^{-\theta t}\int_0^t e^{\theta u}dB_u\right) &= e^{-\theta(t-s)}\int_0^s e^{2\theta u}du \\ &= \frac{e^{-\theta(t-s)}}{2\theta}\left(e^{2\theta s}-1\right). \end{aligned}$$

In Corollary 5.5 we established that the covariance of a GM process is given by $\Psi(t)\Psi(s)\nu(s)$, therefore equating terms we obtain

$$\Psi(t) = \frac{\sigma e^{-\theta t}}{\sqrt{2\theta}} \text{ and } \nu(t) = e^{2\theta t} - 1.$$

Note that the functions Ψ and ν are unique up to a multiplicative constant and we could have placed 2θ into the denominator of ν-s formula instead.

Consequentially, the Ornstein-Uhlenbeck process takes the form

$$X_t = X_0 e^{-\theta t} + \mu\left(1-e^{-\theta t}\right) + \Psi(t) M_t.$$

Chapter 6

Parameter Estimation with the Quadratic Variation of a GM Processes on $[0, T]$

In this chapter, we study the quadratic variation of a Gaussian Markov process. We prove that the quadratic variation can be used as a consistent estimator of the diffusion parameter in a GM model. While the Markov property makes the process's structure seem simple, even in the Gaussian case, it allows for an infinite variety of temporal dependence structures and heteroscedasticity which can complicate computations and estimation of parameters. The results of this chapter also show how quadratic variation gives important insights into the behavior of a process.

The simplest Gaussian Markov process is Brownian motion, for which the quadratic variation evolves linearly with time. (Shreve [46], Theorem 3.4.3) We begin this chapter with a brief literature review of quadratic variation and established results for Brownian motion and fractional Brownian motion. In section 6.2 we use Theorem 5.3 (Hida's [19] representation for the Gaussian Markov process) to explicitly derive the quadratic variation for a Gaussian Markov process. This result is more general than Øksendal's [35] quadratic variation representation for Itô processes of the Gaussian type since it allows for GM processes that cannot be represented as diffusion processes. Additionally, since Hida's representation of GM processes decomposes the memory (controlled by Ψ) and the instantaneous variance (controlled by the martingale's variance ν), our

quadratic variation representation allows for a better understanding of the influence of these two properties on the behavior of the process.

In subsection 6.2.3, we address the need for fast and accurate parameter estimation techniques which are consistent on a fixed interval. Current methods like the MLE, are not consistent on fixed intervals, but require an infinite time horizon. We show how to use the quadratic variation as an estimator of the diffusion parameter in a GM model. The diffusion of a process is the most important factor in determining strategy, price and risk in option pricing theory, and therefore a need for estimation of the diffusion parameter is essential to model implementation. Additionally, since modeling in finance is accomplished under the risk-neutral measure, Cameron-Martin-Girsanov theorem establishes that any drift parameters do not influence the optimal price and hedging strategies of a derivative.

Quadratic variation has been widely used as an estimator of the volatility in the Black-Scholes option pricing model (see Shreve [46] and §6.1.1), however convergence of this estimator is easily proved due to independence of Brownian increments. In the case of the Gaussian Markov process, the function Ψ results in a variety of memory structures from long-range to short-range dependence, while the variance of the martingale process, ν, destroys the identical distribution of increments (unless the martingale is Brownian motion).

Using the Borel-Cantelli lemma we find that almost sure convergence of the sample quadratic variation can be established when Ψ is a Lipschitz or monotone function. Furthermore, we explicitly find the sub-sequences that converge almost surely. This results in Theorem 6.4, that establishes a quadratic variation estimator of the diffusion which is consistent on a fixed interval $[0, T]$, fast and accurate.

We conclude this chapter with a discussion and comparison of our estimator to the MLE (§6.3) for the Ornstein-Uhlenbeck (O-U) process. Using the likelihood definition from Prakasa Rao [42], we derive closed form estimators for the O-U process, which is only consistent for samples in $[0, \infty]$. We compare the quadratic variation estimator to the MLE estimator of the diffusion parameter σ through a simulation study. Our results indicate that the quadratic variation diffusion estimator is simpler and faster to calculate since the MLE method requires estimation of the spring coefficient and long term mean parameters before estimation of the diffusion parameter. Furthermore, the simulation study shows that the quadratic variation diffusion estimator is more accurate than the MLE in terms of absolute deviation on small to moderate size sample sizes and equivalent on very large samples on a fixed interval.

The results of this chapter are then used in Chapter 7 to develop a more general model for Option Pricing Theory. We discuss the Itô-Doeblin formula 7.1 and stochastic calculus which establishes that the quadratic variation is the most important term in determining price and optimal policies in dynamic systems.

6.1 Quadratic Variation of a Function: A Brief Review

Definition 6.1. Let $f(t)$ be a function defined on the interval $[0, T]$. The quadratic variation of a function f up to time T, $[f, f]_T$, is defined as:

$$[f, f]_T \equiv \lim_{\|\Pi_n\| \to 0} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2$$

where $\Pi_n = \{t_0, t_1, \ldots, t_n\}$, $0 = t_0 < t_1 < \cdots < t_n = T$ and $\|\Pi_n\| = \max_{j=0,\ldots,n-1}(t_{j+1} - t_j)$.

The quadratic variation of a function is a measure of its oscillations.

According to Steele [50], if a process X_t can be written as a stochastic integral of the form:

$$X_t(\omega) = \int_0^t a(\omega, t)\, dt + \int_0^t b(\omega, t)\, dB_t,$$

then the quadratic variation of X_t exists and is given by:

$$[X, X]_T = \int_0^T b^2(\omega, s)\, ds.$$

"The quadratic variation of a standard process can be given explicitly as an integral of the instantaneous variance of the process." (Steele [50], pp.129)
Additionally, if $a(\omega, s)$ is a mensurable adapted process with

$$\int_0^t |a(\omega, s)|\ ds < \infty$$

then the quadratic variation of the process $A_t = \int_0^t a(\omega, s)\, ds$ exists and is equal to zero. (Steele [50])

6.1.1 Quadratic Variation of Brownian Motion

In section 3.4.3, Shreve [46], shows that the quadratic variation of a standard Brownian motion $(B_t)_{t\in[0,T]}$ evolves linearly with time:

$$[B,B]_T = T.$$

Additionally, the quadratic variation of Brownian motion can be generalized on any other partition $[T_1, T_2]$ as:

$$[B,B]_{T_2} - [B,B]_{T_1} = T_2 - T_1.$$

This result indicates that Brownian motion accumulates quadratic variation at a rate of one unit time, hence

$$d\,[B,B]_t = dt.$$

The derivation of the quadratic variation for Brownian motion is simple since the Brownian increments are independent identically distributed (i.i.d.) Gaussian random variables.
Furthermore, Shreve [46] demonstrates that the quadratic variation can be used as a uniformly consistent estimator of volatility in the Black-Scholes model, $dS_t = S_t\,(\mu t + \sigma dB_t)$. If we manipulate the stock process $(S_t)_{t\in[0,T]}$ so that $X_t = \ln(S_t/S_0)$, then $X_t = \mu t - \frac{1}{2}\sigma^2 t + \sigma dB_t$ and discrete observations of the process $(X_{t_i})_{i=0}^n$ are sampled on $[0,T]$, then since

$$[X,X]_T = \sigma^2 T,$$

an estimator of volatility take the form

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n}\left(X_{t_i} - X_{t_{i-1}}\right)^2, \text{ almost surely.}$$

6.1.2 Quadratic Variation of Fractional Brownian Motion

In Part I, we quoted literature that the quadratic variation of fractional Brownian motion process, $\left(B_t^H\right)_{t\in[0,T]}$, is given by

$$\left[B^H, B^H\right] = \begin{cases} \infty & \text{if } H < \frac{1}{2} \\ 0 & \text{if } H > \frac{1}{2}. \end{cases}$$

Since the quadratic variation of fBm is either zero or infinite, parameter estimation cannot be accomplished using this property.

Discrete observations of fractional Brownian motion, $\left(B^H_{t_j}\right)_{j=0}^{n-1}$, are mean zero Gaussian self-similar process with stationary increments. All of these properties are needed to derive its quadratic variation. Here is a sketch of the proof required to find the quadratic variation of fBm.

Proof. Assume we are on the partition $\Pi_n = \{t_0, t_1, \ldots, t_{n-1}, t_n\}$ and set

$$\begin{aligned} Q &\equiv \sum_{j=0}^{n-1} \mathbb{E}\left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2 \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[\left(B^H_{t_{j+1}}\right)^2 - 2B^H_{t_{j+1}} B^H_{t_j} + \left(B^H_{t_j}\right)^2\right]. \end{aligned}$$

Without loss of generality, assuming the standard fBm process, that is $\mathbb{E}\left[\left(B^H_{t_0}\right)^2\right] = 1$, then $\mathbb{E}\left[\left(B^H_{t_j}\right)^2\right] = t_j^{2H}$ and $\mathbb{E}\left[B^H_{t_{j+1}} B^H_{t_j}\right] = \frac{1}{2}\left[t_{j+1}^{2H} + t_j^{2H} - (t_{j+1} - t_j)^{2H}\right]$ and so

$$\begin{aligned} Q &= \sum_{j=0}^{n-1} (t_{j+1} - t_j)^{2H} \\ &= \sum_{j=0}^{n-1} (t_{j+1} - t_j)\,(t_{j+1} - t_j)^{2H-1}. \end{aligned}$$

Replace the terms $(t_{j+1} - t_j)^{2H-1}$ with $\max_j \left\{(t_{j+1} - t_j)^{2H-1}\right\}$ and notice that depending on $0 \leq H \leq 1$, there are three different situations; $H < \frac{1}{2}$ $H = \frac{1}{2}$and $H > \frac{1}{2}$. When $H = \frac{1}{2}$ the process is Brownian motion for which we know the quadratic variation, otherwise

$$Q \geq T\left[\max_j \{t_{j+1} - t_j\}\right]^{2H-1}, \; H < 1/2,$$

$$Q \leq T\left[\max_j \{t_{j+1} - t_j\}\right]^{2H-1}, \; H > 1/2.$$

Notice that when $\max_j \{t_{j+1} - t_j\} \to 0$, we obtain the expected value of the quadratic variation of fractional Brownian motion:

$$\mathbb{E}\left[B^H, B^H\right]_T = \begin{cases} \infty & \text{for } H < 1/2 \\ 0 & \text{for } H > 1/2. \end{cases}$$

To show that the sample quadratic variation of fBm converges in L^2 to its mean when $H > 1/2$ we need to compute the variance of the quadratic variation. (We are not concerned with proving convergence when $H < 1/2$ since the expected value of the quadratic variation is infinite.) Set,

$$\begin{aligned} V &\equiv \mathbb{E}\left[\sum_{j=0}^{n-1}\left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2 - \sum_{j=0}^{n-1}\mathbb{E}\left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2\right]^2. \\ &= \mathbb{E}\left[\sum_{j=0}^{n-1}\left\{\left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2 - (t_{j+1} - t_j)^{2H}\right\}\right]^2. \end{aligned}$$

Let $X_j \equiv \left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2 - (t_{j+1} - t_j)^{2H}$, and split the expected value into two parts, one for the sum of variances and one for the sum of covariances

$$\begin{aligned} V &= \mathbb{E}\left[\sum_{j=0}^{n-1} X_j\right]^2 \\ &= \mathbb{E}\left[\sum_{j=0}^{n-1} X_j^2\right] + \mathbb{E}\left[\sum_{j\neq i} X_i X_j\right]. \end{aligned}$$

Using the definition of X_j, coupled with the fact that $\left(B^H_{t_{j+1}} - B^H_{t_j}\right) \stackrel{d}{=} N_j(0, (t_{j+1} - t_j)^{2H})$, the fourth and second moments of the normal distribution yields:

$$\mathbb{E}\left[\sum_{j=0}^{n-1} X_j^2\right] = \sum_{j=0}^{n-1} 2(t_{j+1} - t_j)^{4H}.$$

The second term is more difficult to find, since it involves finding

$$\mathbb{E}\left[\left(B^H_{t_{j+1}} - B^H_{t_j}\right)^2\left(B^H_{t_{i+1}} - B^H_{t_i}\right)^2\right].$$

If X and Y are two normals, then $\mathbb{E}\left[X^2Y^2\right] = \sigma_X^2\sigma_Y^2 + 2\left(\sigma_{X,Y}\right)^2$. Using the fact that,

$$\mathbb{E}\left[\left(= B^H_{t_{j+1}} - B^H_{t_j}\right)\left(B^H_{t_{i+1}} - B^H_{t_i}\right)\right] = \frac{(t_{j+1} - t_j)^H (t_{i+1} - t_i)^H}{2}\left\{|i - j + 1|^{2H} + |i - j - 1|^{2H} - 2\,|i - j|^{2H}\right\},$$

$$\mathbb{E}\left[X^2Y^2\right] = 2\left(\mathbb{E}\left[\left(B^H_{t_{j+1}} - B^H_{t_j}\right)\left(B^H_{t_{i+1}} - B^H_{t_i}\right)\right]\right)^2.$$

.

Since $H > \frac{1}{2}$, if we let $\Delta t_i \equiv t_{i+1} - t_i$ and $\Delta t_i = \Delta t_j = \frac{T}{n}$ then

$$\lim_{n\to\infty} 2\left\{\sum_{j=0}^{n-1}\left(\frac{T}{n}\right)^{4H} + \sum_{j\neq i}\left(\frac{T}{n}\right)^{4H}\left\{|i-j+1|^{2H} + |i-j-1|^{2H} - 2|i-j|^{2H}\right\}^2\right\} = 0.$$

This proves that the sample quadratic variation converges to zero in L^2. □

6.2 Quadratic Variation of Gaussian Markov Processes

In the following subsections we explicitly derive a closed form representation of the quadratic variation of a Gaussian Markov process. We establish Theorem 6.2 , which shows that the expected value of the quadratic variation is represented by a Riemann-Stieltjes integral, while Theorem 6.3 imposes weak conditions on Ψ for the quadratic variation to convergence in L^2 to zero. The goal of this section is to establish that the quadratic variation can be used as an estimator for the diffusion coefficient in GM processes. In Theorem 6.4 we establish this result and provide weak restrictions for convergence almost surely. Confidence interval bounds along with the convergence rate of the quadratic variation estimator is also provided. Examples are provided for the Ornstein-Uhlenbeck process along with a comparison of the quadratic variation estimator to the MLE method.

6.2.1 Expected Value of the Quadratic Variation

In this subsection we explicitly derive the expected value of the quadratic variation and establish its representation as the Riemann–Stieltjes integral in Theorem 6.2.

Theorem 6.2. *Let $(X_t)_{t\in[0,T]}$ be a centered Gaussian Markov process such that*

$$X_t = \Psi_t M_t,$$

where $(M_t)_{t\in[0,T]}$ is a centered Gaussian martingale with $\mathbb{E}\left[M_t^2\right] = \nu(t)$, Ψ is a continuous function with bounded variation and $\Psi \neq 0$ for all $t \in [0,T]$.

Then the quadratic variation is of $(X_t)_{t\in[0,T]}$ *is given by*

$$\mathbb{E}\left[X,X\right]_T = \int_0^T \Psi^2\left(t\right) d\nu\left(t\right).$$

Proof. Let,

$$\begin{aligned} Q &\equiv \sum_{j=0}^{n-1} \mathbb{E}\left[\left(X(t_{j+1}) - X(t_j)\right)^2\right] \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[\left(\Psi(t_{j+1})M(t_{j+1}) - \Psi(t_j)M(t_j)\right)^2\right]. \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[\left(\Psi(t_{j+1})\left(M(t_{j+1}) - M(t_j)\right) - \left(\Psi(t_{j+1}) - \Psi(t_j)\right)M(t_j)\right)^2\right]. \end{aligned} \tag{6.1}$$

Utilizing the martingale property $\mathbb{E}\left[M_t M_s\right] = \nu(t\wedge s)$, equation 6.1 takes the following form:

$$Q = \sum_{j=0}^{n-1} \left[\Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right) - \Psi(t_j)\nu(t_j)\left(\Psi(t_{j+1}) - \Psi(t_j)\right)\right]. \tag{6.2}$$

Using summation by parts, the second term in equation 6.2 can be written as

$$\sum_{j=0}^{n-1} \Psi(t_j)\nu(t_j)\left(\Psi(t_{j+1}) - \Psi(t_j)\right) = \Psi(t_n)\Psi(t_{n-1})\nu(t_{n-1}) - \sum_{j=0}^{n-2} \Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right). \tag{6.3}$$

The second term in equation 6.3 is the same as the first term in equation 6.2. Substituting equation 6.3 into equation 6.2 and readjusting the summations by adding the term

$$\Psi(t_n)\left(\Psi(t_n)\nu(t_n) - \Psi(t_{n-1})\nu(t_{n-1})\right),$$

we obtain

$$\begin{aligned} Q &= \Psi(t_n)\left(\Psi(t_n)\nu(t_n) - \Psi(t_{n-1})\nu(t_{n-1})\right) - \Psi(t_n)\Psi(t_{n-1})\nu(t_{n-1}) \\ &\quad +2\sum_{j=0}^{n-1} \Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right). \end{aligned}$$

Notice that

$$\lim_{n\to\infty}\sum_{j=0}^{n-1}\Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1})-\Psi(t_j)\nu(t_j)\right)=\int_0^T\Psi(t)\,d\left(\Psi(t)\nu(t)\right).$$

The expected value of the quadratic variation of a Gaussian Markov process is then given by

$$[X,X]_T=2\int_0^T\Psi(t)\,d\left(\Psi(t)\nu(t)\right)-\Psi^2(T)\nu(T).$$

Applying integration by parts, we obtain

$$[X,X]_T=\int_0^T\Psi^2(t)\,d\nu(t). \tag{6.4}$$

□

6.2.2 L^2 Convergence of the Quadratic Variation

In this subsection we prove that the expected value of the quadratic variation of the GM process is constant.

Theorem 6.3. *Let $(X_t)_{t\in[0,T]}$ be a centered Gaussian Markov process such that*

$$X_t=\Psi_t M_t,$$

where $(M_t)_{t\in[0,T]}$ is a centered Gaussian martingale with $\mathbb{E}\left[M_t^2\right]=\nu(t)$, Ψ is continuous with bounded variation and $\Psi\neq 0$ for all $t\in[0,T]$.

Then the quadratic variation is of $(X_t)_{t\in[0,T]}$ is given by

$$[X,X]_T=\int_0^T\Psi^2(t)\,d\nu(t),\ \textit{almost surely.}$$

To prove Theorem 6.3 we analyze the variance of the quadratic variation and show that converges to zero in L^2.

Proof. Set,

$$N_j\equiv\Psi(t_{j+1})M(t_{j+1})-\Psi(t_j)M(t_j).$$

In equation 6.2 we determined that

$$\sigma_j^2 \equiv \mathbb{E}\left[N_j^2\right] = \Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right) - \Psi(t_j)\nu(t_j)\left(\Psi(t_{j+1}) - \Psi(t_j)\right).$$

Therefore the variance of the quadratic variation is given by

$$\begin{aligned}
V_n &\equiv \mathbb{E}\left[\sum_{j=0}^{n-1} N_j^2 - \sum_{j=0}^{n-1}\mathbb{E}\left[N_j^2\right]\right]^2 \\
&= \mathbb{E}\left[\sum_{j=0}^{n-1} N_j\right]^2 - \left(\sum_{j=0}^{n-1}\sigma_j^2\right)^2 \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\mathbb{E}\left[N_i^2 N_j^2\right] - \left(\sum_{j=0}^{n-1}\sigma_j^2\right)^2 \\
&= \sum\sum_{i\neq j}\mathbb{E}\left[N_i^2 N_j^2\right] + \sum_{j=0}^{n-1}\mathbb{E}\left[N_j^4\right] - \left(\sum_{j=0}^{n-1}\sigma_j^2\right)^2
\end{aligned}$$

Since $N_j \stackrel{d}{=} N(0,\sigma_j^2)$, then

$$\begin{aligned}
\mathbb{E}\left[N_i^2 N_j^2\right] &= \sigma_i^2\sigma_j^2 + 2\mathbb{E}^2\left[N_i N_j\right] \\
\mathbb{E}\left[N_j^4\right] &= 3\sigma_j^4.
\end{aligned}$$

Therefore,

$$V_n = \sum\sum_{i\neq j}\sigma_i^2\sigma_j^2 + 2\sum\sum_{i\neq j}\mathbb{E}^2\left[N_i N_j\right] + 3\sum_{j=0}^{n-1}\sigma_j^4 - \left(\sum_{j=0}^{n-1}\sigma_j^2\right)^2.$$

Since,

$$\sum\sum_{i\neq j}\sigma_i^2\sigma_j^2 + \sum_{j=0}^{n-1}\sigma_j^4 = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sigma_i^2\sigma_j^2 = \left(\sum_{j=0}^{n-1}\sigma_j^2\right)^2,$$

then,

$$\begin{aligned}
V_n &= 2\sum\sum_{i\neq j}\mathbb{E}^2\left[N_i N_j\right] + 2\sum_{j=0}^{n-1}\sigma_j^4 \\
&= 4\sum\sum_{i<j}\mathbb{E}^2\left[N_i N_j\right] + 2\sum_{j=0}^{n-1}\sigma_j^4.
\end{aligned}$$

Notice,

$$\begin{aligned}\mathbb{E}\left[N_i N_j\right] &= \Psi(t_{i+1})\Psi(t_{j+1})\nu(t_{i+1}\wedge t_{j+1}) - \Psi(t_{i+1})\Psi(t_j)\nu(t_{i+1}\wedge t_j)\\ &\quad -\Psi(t_i)\Psi(t_{j+1})\nu(t_i\wedge t_{j+1}) + \Psi(t_i)\Psi(t_j)\nu(t_i\wedge t_j).\end{aligned}$$

Consequentially for $i<j$,

$$\begin{aligned}V_n &= 2\sum_{j=0}^{n-1}\left\{\Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right) - \Psi(t_j)\nu(t_j)\left(\Psi(t_{j+1}) - \Psi(t_j)\right)\right\}^2\\ &\quad +4\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1}\left(\Psi(t_{j+1}) - \Psi(t_j)\right)^2\left(\Psi(t_{i+1})\nu(t_{i+1}) - \Psi(t_i)\nu(t_i)\right)^2\\ &\equiv V_{1,n} + V_{2,n}.\end{aligned}$$

To prove that the variance of the quadratic variation converges the Gaussian Markov process must have finite expected quadratic variation; that is, using the result from equation 6.4, since Ψ is continuous and ν has bounded variation,

$$\int_0^T \Psi^2(t)\, d\nu(t) < \infty,$$

which is equivalent to

$$\sum_{j=0}^{n-1}\Psi^2(t_{j+1})\left(\nu(t_{j+1}) - \nu(t_j)\right) < \infty \Leftrightarrow 2\sum_{j=0}^{n-1}\Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right) - \Psi(t_n)\nu(t_n) < \infty.$$

Additionally, let us restrict Ψ to the family of continuous functions with bounded variation; that is, there exists an $M<\infty$, such that

$$\sum_{j=0}^{n-1}\left|\Psi(t_{j+1}) - \Psi(t_j)\right| \le M, \text{ for any partition of } [0,T].$$

Notice that

$$\begin{aligned}M &\ge \sum_{j=0}^{n-1}\left|\Psi(t_{j+1}) - \Psi(t_j)\right|\\ M\delta &\ge \sum_{j=0}^{n-1}\left(\Psi(t_{j+1}) - \Psi(t_j)\right)^2,\end{aligned}$$

where $\delta \equiv \max_{0\leq j\leq n-1} |\Psi(t_{j+1}) - \Psi(t_j)|$,which proves that if a continuous function of Ψ has bounded variation, it also has bounded quadratic variation. The term $V_{1,n}$ can be rewritten as

$$\begin{aligned}
V_{1,n} &= 2\sum_{j=0}^{n-1}\left\{\Psi^2(t_{j+1})\left(\nu(t_{j+1})-\nu(t_j)\right)+\left(\Psi(t_{j+1})-\Psi(t_j)\right)^2\nu(t_j)\right\}^2 \\
&= 2\sum_{j=0}^{n-1}\Big\{\Psi^4(t_{j+1})\left(\nu(t_{j+1})-\nu(t_j)\right)^2 \\
&\quad +2\Psi^2(t_{j+1})\left(\nu(t_{j+1})-\nu(t_j)\right)\left(\Psi(t_{j+1})-\Psi(t_j)\right)^2\nu(t_j)+\left(\Psi(t_{j+1})-\Psi(t_j)\right)^4\nu^2(t_j)\Big\} \\
&\leq 2\Bigg\{\sup_{x\in[0,T]}\Psi^4(x)\max_{0\leq j\leq n-1}\left(\nu(t_{j+1})-\nu(t_j)\right)\nu(T) \\
&\quad +2\nu^2(T)\sup_{x\in[0,T]}\Psi^2(x)\max_{0\leq j\leq n-1}\left(\Psi(t_{j+1})-\Psi(t_j)\right)^2 \\
&\quad +\max_{0\leq j\leq n-1}\left(\Psi(t_{j+1})-\Psi(t_j)\right)^2 M\delta\nu^2(T)\Bigg\}. \qquad (6.5)
\end{aligned}$$

Since ν and Ψ are continuous then

$$\lim_{n\to\infty} V_{1,n} \leq 0.$$

Analyzing the second term in the variance

$$\begin{aligned}
V_{2,n} &= 4\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1}\left(\Psi(t_{j+1})-\Psi(t_j)\right)^2\left(\Psi(t_{i+1})\nu(t_{i+1})-\Psi(t_i)\nu(t_i)\right)^2 \\
&\leq 4M\delta\sum_{i=0}^{n-1}\left(\Psi(t_{i+1})\left(\nu(t_{i+1})-\nu(t_i)\right)+\left(\Psi(t_{i+1})-\Psi(t_i)\right)\nu(t_i)\right)^2 \\
&\leq 4M\delta\Bigg\{\max_{x\in[0,T]}|\Psi(x)|^2\max_{0\leq i\leq n-1}\left(\nu(t_{i+1})-\nu(t_i)\right)\nu(T) \\
&\quad +2\max_{x\in[0,T]}|\Psi(x)|^2\max_{0\leq i\leq n-1}\left(\nu(t_{i+1})-\nu(t_i)\right)\max_{0\leq i\leq n-1}|\Psi(t_{i+1})-\Psi(t_i)|\nu^2(T) \\
&\quad +\max_{0\leq i\leq n-1}|\Psi(t_{i+1})-\Psi(t_i)|M\nu(T)\Bigg\}. \qquad (6.6)
\end{aligned}$$

Since Ψ and ν are a continuous functions, then

$$\lim_{n\to\infty} V_{2,n} \leq 0.$$

Since both $V_{1,n}$ and $V_{2,n}$ converge to zero, this establishes that the quadratic variation is a constant, almost surely. □

6.2.3 Almost Sure Convergence and Estimation with Quadratic Variation

In the previous subsection, we proved that the quadratic variation of a Gaussian Markov process converges in L^2, so there is a sub-sequence that converges almost surely. In this subsection we establish almost sure convergence of the quadratic variation for GM processes by identifying those sub-sequences explicitly. This result establishes that the quadratic variation can be used as a consistent estimator.

Theorem 6.4. *Let $(X_t)_{t\in[0,T]}$ be a centered Gaussian Markov process such that*

$$X_t = \Psi_t M_t,$$

where $(M_t)_{t\in[0,T]}$ is a centered Gaussian martingale with $\mathbb{E}\left[M_t^2\right] = \nu(t)$, Ψ and ν are Lipschitz, and $\Psi \neq 0$ for all $t \in [0,T]$. If $\Pi_n = \{t_0, t_1, \ldots, t_{n-1}, t_n\}$ is a partition on $[0,T]$ so that $\|\Pi_n\| \leq k\frac{T}{n}$, for some $k \geq 1$, and $\{n_k\}_{k=1}^{\infty}$ is any sub-sequence $\{n\}_{n=1}^{\infty}$ so that $\sum_{k=1}^{\infty} 1/n_k < \infty$ then,

$$[X,X]_{T,n} \equiv \sum_{j=0}^{n-1}\left[\left(X(t_{j+1}) - X(t_j)\right)^2\right] \overset{n\to\infty}{\to} \int_0^T \Psi^2(t)\, d\nu(t), \text{ almost surely.}$$

Proof. Since Ψ and ν are Lipschitz,

$$\max_{0\leq j\leq n-1} |\Psi(t_{j+1}) - \Psi(t_j)| \leq C_1 \|\Pi_n\| \leq C_1 k\frac{T}{n} \tag{6.7}$$

and

$$\max_{0\leq j\leq n-1} |\nu(t_{j+1}) - \nu(t_j)| \leq C_2 \|\Pi_n\| \leq C_2 k\frac{T}{n}, \tag{6.8}$$

where C_1, and C_2 are Lipschitz constants for Ψ and ν, respectively, and $\Pi = \{t_0, t_1, \ldots, t_n\}$ is any partition of $[0,T]$ so that $\|\Pi_n\| = \max_{j=0,\ldots,n-1}(t_{j+1} - t_j) \leq k\frac{T}{n}$.

Set,

$$\Lambda_{\tilde{\epsilon},n} \equiv \left\{\sum_{j=0}^{n-1} N_j^2 - \sum_{j=0}^{n-1}\mathbb{E}\left[N_j^2\right] > \tilde{\epsilon}\right\}.$$

Then by Chebyshev's inequality,

$$P\left(\Lambda_{\tilde{\epsilon},n}\right) \leq \frac{V_n}{\tilde{\epsilon}^2}.$$

By 6.5 and 6.6 , inequalities 6.7 and 6.8 imply

$$V_n \leq C\frac{1}{n}.$$

If $\{n_k\}_{k=1}^{\infty}$ is a sub-sequence of $\{n\}_{n=1}^{\infty}$ so that $\sum_{k=1}^{\infty} 1/n_k < \infty$, then

$$\sum_{k=1}^{\infty} V_{n_k} < \infty$$

and by Borel-Cantelli lemma,

$$P\left(\Lambda_{\tilde{\epsilon},n_k} \text{ i.o.}\right) = 0,$$

where "$\Lambda_{\tilde{\epsilon},n_k}$ i.o." denotes the event $\Lambda_{\tilde{\epsilon},n_k}$ occurring k *infinitely often*, which means

$$P\left(\limsup_{k\to\infty} \Lambda_{\tilde{\epsilon},n_k}\right) = 0,$$

which is equivalent to

$$\Lambda_{\tilde{\epsilon},n_k} \to 0, \quad a.s.$$

Therefore, the expected quadratic variation converges to the true quadratic variation almost surely (i.e. with probability 1). For example, sampling with the speed $n_k = (1+k)^{1+\epsilon}$ will suffice. □

Example 6.5. Vasicek and the Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process $(X_t)_{t\in[0,T]}$ is a common model for modeling mean-reversion in interest rate derivative pricing and the length dynamics of over-damped springs under thermal fluctuations. The Ornstein-Uhlenbeck process is a Gaussian Markov process defined by

$$X_t = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma e^{-\theta t}\int_0^t e^{\theta s} dB_s,$$

where $(B_t)_{t\in[0,T]}$ is a standard Brownian motion process and $\mu > 0$, $\theta > 0$, and $\sigma > 0$ are constants. In example 5.18 we used the Itô isometry, to show that $\sigma e^{-\theta t}\int_0^t e^{\theta s} dB_s$, has functions

$$\Psi(t) = \frac{\sigma e^{-\theta t}}{\sqrt{2\theta}} \text{ and } \nu(t) = e^{2\theta t} - 1.$$

This yields a centered GM process plus a deterministic term in the O-U process representation,

$$X_t = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \Psi(t) M_t.$$

The deterministic term $X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right)$ in the above equation is continuous and differentiable and therefore has zero quadratic variation. Additionally, the cross variation with the deterministic term is also zero. (Steele [50]). Therefore, applying Theorem 6.3, the quadratic variation of the Ornstein-Uhlenbeck process is

$$\begin{aligned} [X, X]_T &= \int_0^T \Psi^2(t)\, d\nu(t) \\ &= \sigma^2 T. \end{aligned}$$

The functions Ψ and ν are continuously differentiable, and therefore satisfy all properties needed for almost sure convergence of the quadratic variation.

Since Ψ and ν are continuously differentiable on $[0, T]$, they are Lipschitz. Let us find the Lipschitz constants. Now,

$$\begin{aligned} \max_i |\Psi(t_{i+1}) - \Psi(t_i)| &= \max_i \left| \frac{\sigma e^{-\theta t_{i+1}}}{\sqrt{2\theta}} - \frac{\sigma e^{-\theta t_i}}{\sqrt{2\theta}} \right| \\ &= \max_i \left(\left| e^{-\theta(t_{i+1} - t_i)} - 1 \right| \left| \frac{\sigma e^{-\theta t_i}}{\sqrt{2\theta}} \right| \right) \\ &\leq \frac{\sigma}{\sqrt{2\theta}} \max_i \left| e^{-\theta(t_{i+1} - t_i)} - 1 \right|. \end{aligned}$$

Using Mean Value Theorem with $x_0 = 0$ and $x = t_{i+1} - t_i$,

$$e^{-\theta x_0} = 1 - \theta(t_{i+1} - t_i) e^{-\theta c}, \text{ where } c \in [0, (t_{i+1} - t_i)].$$

Therefore

$$\begin{aligned} \max_i |\Psi(t_{i+1}) - \Psi(t_i)| &\leq \frac{\sigma}{\sqrt{2\theta}} \max_i \left| \theta(t_{i+1} - t_i) e^{-\theta c} \right| \\ &\leq \sigma \sqrt{\frac{\theta}{2}} \max_i (t_{i+1} - t_i) \leq C_1 \max_i (t_{i+1} - t_i) \leq C_1 k \frac{T}{n}. \end{aligned}$$

Finding the Lipschitz constant for ν yields,

$$\begin{aligned}\max_i |\nu(t_{i+1}) - \nu(t_i)| &= \max_i \left| e^{2\theta t_{i+1}} - e^{2\theta t_i} \right| \\ &= \max_i \left(\left| 1 - e^{-2\theta(t_{i+1}-t_i)} \right| \left| e^{2\theta t_{i+1}} \right| \right) \\ &\leq e^{2\theta T} \max_i \left| 1 - e^{-2\theta(t_{i+1}-t_i)} \right|\end{aligned}$$

Using Mean Value Theorem with $x_0 = 0$ and $x = t_{i+1} - t_i$,

$$e^{-\theta x_0} = 1 - 2\theta(t_{i+1} - t_i) e^{-2\theta c}, \text{ where } c \in [0, (t_{i+1} - t_i)].$$

Therefore,

$$\begin{aligned}\max_i |\nu(t_{i+1}) - \nu(t_i)| &= e^{2\theta T} \max_i \left| 2\theta(t_{i+1} - t_i) e^{-2\theta c} \right| \\ &\leq 2\theta e^{2\theta T} \max_i (t_{i+1} - t_i) \leq C_2 \max_i (t_{i+1} - t_i) \leq C_2 k \frac{T}{n}.\end{aligned}$$

Consequentially, the Ornstein-Uhlenbeck process has the property

$$\lim_{k\to\infty} \left(\sum_{j=0}^{n_k - 1} (X(t_{j+1}) - X(t_j))^2 \right) = [X, X]_T, \text{ almost surely,}$$

for any sequence $\{n_k\}_{k=1}^{\infty}$ so that $\sum_{k=1}^{\infty} 1/n_k < \infty$.

6.2.4 Confidence Interval Bounds of the Sample Quadratic Variation for a Gaussian Markov Process

In this section, we use the prior results to find the confidence interval bounds for the sample quadratic variation $[X, X]_{T,n}$.

To find the confidence interval (and convergence rate) we need to show that

$$A_{\epsilon,n_k} \equiv P\left(\left| [X, X]_{T,n} - [X, X]_T \right| > \epsilon \right) < \alpha.$$

By Chebyshev's inequality,

$$
\begin{aligned}
A_{\epsilon,n} &\leq \frac{\mathbb{E}\left[[X,X]_{T,n} - [X,X]_T\right]^2}{\epsilon^2} \\
&\leq 2\left(\frac{\mathbb{E}\left([X,X]_{T,n} - \mathbb{E}\left[[X,X]_{T,n}\right]\right)^2}{\epsilon^2} + \frac{\left(\mathbb{E}\left[[X,X]_{T,n}\right] - [X,X]_T\right)^2}{\epsilon^2}\right) \\
&= 2\left(\frac{V_n}{\epsilon^2} + \frac{\left(\mathbb{E}\left[[X,X]_{T,n}\right] - [X,X]_T\right)^2}{\epsilon^2}\right),
\end{aligned}
$$

where V_n was derived in §6.2.2. Using the results that appear in computing V_n in §6.2.2, we have

$$
\mathbb{E}\left[X,X\right]_{T,n} - [X,X]_T = \sum_{j=0}^{n-1}\sigma_j^2 - \sum_{j=0}^{n-1}\int_j^{j+1}\Psi^2\left(t\right)d\nu\left(t\right).
$$

Let, $B_{T,n} \equiv \left|\mathbb{E}\left[X,X\right]_{T,n} - [X,X]_T\right|$, then by Mean Value Theorem, there are $t_j^* \in [t_j, t_{j+1}]$ such that

$$
\begin{aligned}
B_{T,n} &= \left|\sum_{j=0}^{n-1}\sigma_j^2 - \sum_{j=0}^{n-1}\Psi\left(t_j^*\right)\left(\nu\left(t_{j+1}\right) - \nu\left(t_j\right)\right)\right| \\
&= \Bigg|\sum_{j=0}^{n-1}\{\Psi(t_{j+1})\left(\Psi(t_{j+1})\nu(t_{j+1}) - \Psi(t_j)\nu(t_j)\right) - \Psi(t_j)\nu(t_j)\left(\Psi(t_{j+1}) - \Psi(t_j)\right) \\
&\quad -\Psi^2\left(t_j^*\right)\left(\nu\left(t_{j+1}\right) - \nu\left(t_j\right)\right)\}\Bigg| \\
&= \left|\sum_{j=0}^{n-1}\left\{\left(\Psi(t_{j+1}) - \Psi(t_j)\right)^2\nu(t_j) - \left(\Psi^2(t_{j+1}) - \Psi^2(t_j^*)\right)\left(\nu(t_{j+1}) - \nu(t_j)\right)\right\}\right| \\
&\leq M\nu\left(T\right)\max_{0\leq j\leq n_k-1}\left|\Psi(t_{j+1}) - \Psi(t_j)\right| + 2\nu\left(T\right)\max_{x\in[0,T]}\left|\Psi\left(x\right)\right|\max_{0\leq j\leq n_k-1}\left|\Psi(t_{j+1}) - \Psi(t_j^*)\right|.
\end{aligned}
$$

If Ψ is Lipschitz or a monotone function then,

$$
\left|\mathbb{E}\left[X,X\right]_{T,n} - [X,X]_T\right|^2 \leq \frac{C_4}{n^2}.
$$

This yields a bound on the confidence interval such that

$$
P\left(\left|[X,X]_{T,n} - [X,X]_T\right| > \epsilon\right) \leq \frac{2}{\epsilon^2}\left(\frac{C_3}{n} + \frac{C_4}{n^2}\right).
$$

This establishes that the sample quadratic variation converges to the true quadratic variation at a rate of $1/\sqrt{n}$ in probability. Notice that the computation above indicates that the primary driver of errors in the estimator is the standard deviation, while there is relatively small bias.

Furthermore, if

$$\frac{2}{\epsilon^2}\left(\frac{C_3}{n}+\frac{C_4}{n^2}\right)\leq\alpha,$$

then

$$\epsilon\leq\sqrt{\frac{2}{\alpha}\left(\frac{C_3}{n}+\frac{C_4}{n^2}\right)}.$$

In general, Chebyshev's inequality is a very crude bound, and therefore the error bounds on the quadratic variation estimator is much more conservative than the actual error bound. This will be demonstrated in the example below.

Example 6.6. Estimation of Volatility in an Ornstein-Uhlenbeck Process

In this example we demonstrate the efficacy of the quadratic variation as an estimator of the volatility in Gaussian Markov processes. We show how to use this result to estimate the diffusion parameter of an Ornstein-Uhlenbeck process.

We begin by computing the confidence interval on the sample quadratic variation estimator. Recall,

$$\Psi(t)=\frac{\sigma e^{-\theta t}}{\sqrt{2\theta}}\text{ and }\nu(t)=e^{2\theta t}-1.$$

Let the partition Π_n be defined as $t_i=i\frac{T}{n}$, $i=0,1,\ldots,n$. Since Ψ is monotone decreasing, the following summation is telescoping:

$$M\geq\sum_{j=0}^{n-1}\left|\Psi(t_{j+1})-\Psi(t_j)\right|=\Psi(0)-\Psi(T)=\frac{\sigma}{\sqrt{2\theta}}\left(1-e^{-\theta T}\right).$$

Since we assume n is large, the Taylor series expansion yields

$$\begin{aligned}\max_{0\leq j\leq n-1}\left|\Psi(t_{j+1})-\Psi(t_j)\right|^p &= \left|\Psi(0)-\Psi\left(\frac{1}{n}\right)\right|^p=\frac{\sigma^p}{(2\theta)^{p/2}}\left(1-e^{-\theta/n}\right)^p\\ &\leq \left(\frac{\sigma^p}{(2\theta)^{p/2}}\right)\left(\frac{\theta}{n}\right)^p,\forall p=1,2,\ldots\end{aligned}$$

and

$$\delta=\max_{0\leq j\leq n-1}\left|\Psi(t_{j+1})-\Psi(t_j)\right|\leq\left(\frac{\sigma}{\sqrt{2\theta}}\right)\left(\frac{\theta}{n}\right).$$

The function Ψ has its maximum occurring at $x = 0$. Therefore,

$$\sup_{x\in[0,T]} \Psi^p(x) = \frac{\sigma^p}{(2\theta)^{p/2}}, \forall p = 1, 2, \ldots$$

It is easy to verify that

$$\max_{0\le j\le n-1} (\nu(t_{j+1}) - \nu(t_j)) = \left(\nu(T) - \nu\left(T - \frac{1}{n}\right)\right) = e^{2\theta T}\left(1 - e^{-2\theta/n}\right).$$

By Taylor series expansion,

$$e^{2\theta T}\left(1 - e^{-2\theta/n}\right) \le 2\left(\frac{\theta}{n}\right)e^{2\theta T}.$$

Let us use the above to find the variance of the diffusion estimator. Since,

$$\begin{aligned} V_{1,n} \le\ & 2\left\{\sup_{x\in[0,T]} \Psi^4(x) \max_{0\le j\le n-1} (\nu(t_{j+1}) - \nu(t_j))\,\nu(T)\right. \\ & +2\nu^2(T) \sup_{x\in[0,T]} \Psi^2(x) \max_{0\le j\le n-1} (\Psi(t_{j+1}) - \Psi(t_j))^2 \\ & \left. + \max_{0\le j\le n-1} (\Psi(t_{j+1}) - \Psi(t_j))^2 M\delta\nu^2(T)\right\}, \end{aligned}$$

which yields

$$\begin{aligned} V_{1,n} \le\ & 2\frac{\sigma^4}{(2\theta)^2}\left\{2\left(\frac{\theta}{n}\right)\left(e^{2\theta T}\right)\left(e^{2\theta T} - 1\right) + 2\left(\frac{\theta}{n}\right)^2\left(e^{2\theta T} - 1\right)^2\right. \\ & \left. + \left(\frac{\theta}{n}\right)^3\left(1 - e^{-\theta T}\right)\left(e^{2\theta T} - 1\right)^2\right\}, \end{aligned}$$

and since,

$$\begin{aligned} V_{2,n} \le\ & 4M\delta\left\{\max_{x\in[0,T]} |\Psi(x)|^2 \max_{0\le i\le n-1} (\nu(t_{i+1}) - \nu(t_i))\,\nu(T)\right. \\ & +2\max_{x\in[0,T]} |\Psi(x)|^2 \max_{0\le i\le n-1} (\nu(t_{i+1}) - \nu(t_i)) \max_{0\le i\le n-1} |\Psi(t_{i+1}) - \Psi(t_i)|\,\nu^2(T) \\ & \left. + \max_{0\le i\le n-1} |\Psi(t_{i+1}) - \Psi(t_i)|\,M\nu(T)\right\}, \end{aligned}$$

assuring that

$$V_{2,n} \le \sigma^4\left(1 - e^{-\theta T}\right)\left(e^{2\theta T} - 1\right)\left(\frac{1}{n^2}\right)\left\{2e^{2\theta T} + 4\left(\frac{\theta}{n}\right)\frac{\sigma}{\sqrt{2\theta}}e^{2\theta T}\left(e^{2\theta T} - 1\right) + \left(1 - e^{-\theta T}\right)\right\}.$$

Recall that the variance is given by $V_n = V_{1,n} + V_{2,n}$.

The absolute bias term takes the form,

$$\begin{aligned}\left|\mathbb{E}\left[X,X\right]_{T,n} - \left[X,X\right]_T\right| &\leq M\nu\left(T\right)\max_{0\leq j\leq n-1}\left|\Psi(t_{j+1}) - \Psi(t_j)\right| \\ &\quad +2\nu\left(T\right)\max_{x\in[0,T]}\left|\Psi\left(x\right)\right|\max_{0\leq j\leq n-1}\left|\Psi(t_{j+1}) - \Psi(t_j^*)\right| \\ \left|\mathbb{E}\left[X,X\right]_{T,n} - \left[X,X\right]_T\right| &\leq \left(\frac{1}{n}\right)\frac{\sigma^2}{2}\left(e^{2\theta T}-1\right)\left(3-e^{-\theta T}\right).\end{aligned}$$

Therefore, the quadratic variation estimator has

$$\epsilon \leq \sqrt{\frac{2}{\alpha}\left(\frac{C_3}{n} + \frac{C_4}{n^2}\right)}, \tag{6.9}$$

where,

$$\begin{aligned}C_3 &= \frac{\sigma^4}{\theta^2}\left(e^{2\theta T}\right)\left(e^{2\theta T}-1\right) \\ C_4 &= \frac{\sigma^4}{4}\left(e^{2\theta T}-1\right)^2\left(3-e^{-\theta T}\right)^2.\end{aligned}$$

To demonstrate the efficacy of this method, we use Monte Carlo simulation to generate 1000 independent sample paths of the Ornstein-Uhlenbeck process with $\Delta t = 1/n$ and $t \in [0,1]$. In the simulations we set the initial value $X_0 = 0.08$, the long term mean $\mu = 0.05$, the spring constant $\theta = 1$, and the instantaneous volatility $\sigma = 0.02$. The Ornstein-Uhlenbeck process has quadratic variation

$$\left[X,X\right]_T = \sigma^2 T.$$

Therefore, the quadratic variation estimator of volatility is given by

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{j=0}^{n-1}\left[\left(X(t_{j+1}) - X(t_j)\right)^2\right]}.$$

Implementing the estimator on the 1000 sample paths with $n = 128$ observations, we obtain the histogram of the volatility estimates in Figure 6.1 below.

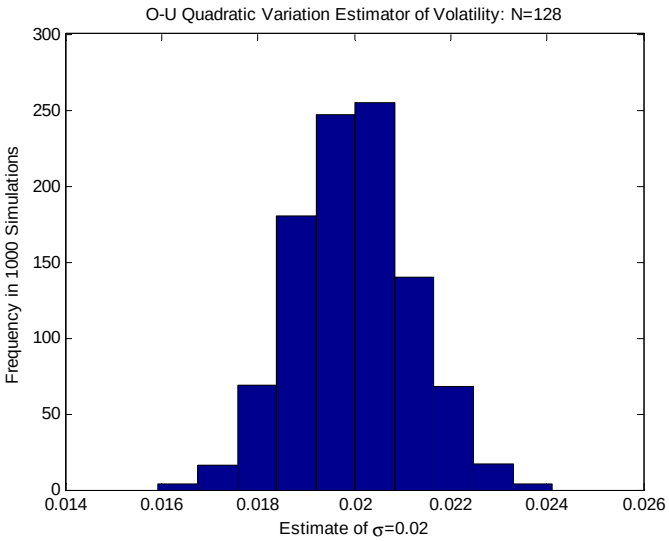


Figure 6.1: Quadratic Variation Estimates of Volatility for an O-U Process

If we use the error function results derived in inequality 6.9 and set $\alpha = 0.05$, Figure 6.2 shows that the derived ϵ is a very conservative bound for the true error.

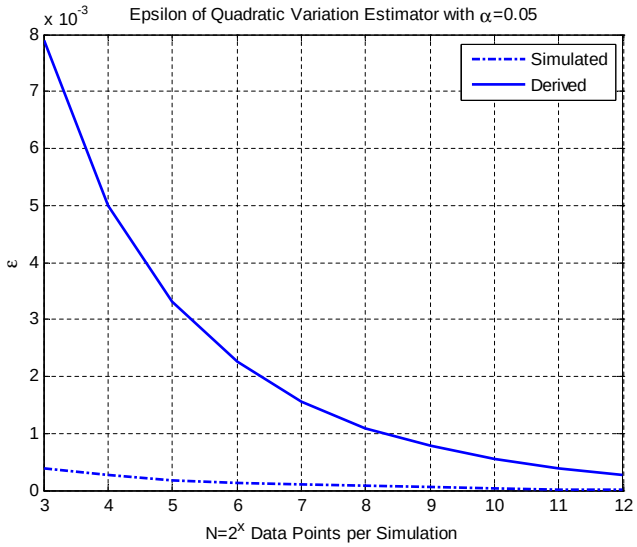


Figure 6.2: Error Bound for Quadratic Variation Estimates of Volatility of an O-U Process

To further understand the estimator's errors, the Root Mean Square Error (RMSE), bias and standard deviation are shown for different n in Figure 6.3 $n_k = 2^k$.

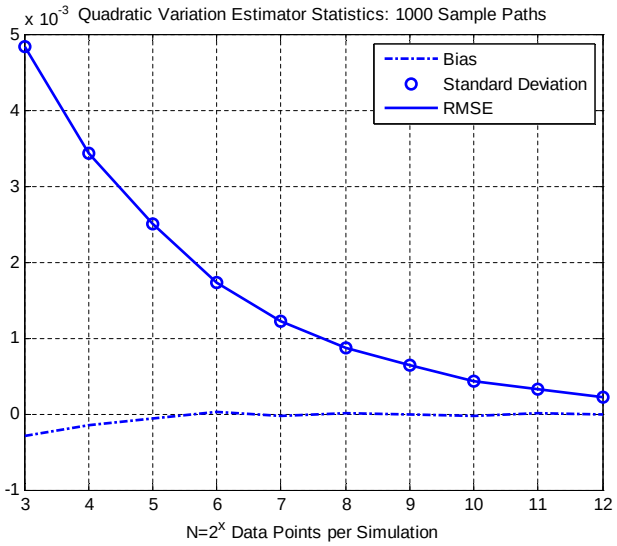


Figure 6.3: Statistics on the Quadratic Variation Estimator of Volatility of an O-U Process

Figure 6.3 above confirms that the estimator's error is primarily driven by the variance and that the bias has very little effect, especially after $n = 32$ data points.

6.3 Discussion and Comparison to the MLE

As mentioned in the introduction, derivative pricing with stochastic models is done under a change of probability measure, the new measure is known as the risk neutral measure. The change of measure eliminates the drift parameters. Therefore, typical financial applications do not need estimation of drift parameters. The primary exception is short rate bond models which are all Markovian and developed under the risk neutral probability measure. Estimation of the drift parameters is impeded since the model's parameters and the observed prices are under different probability measures. Therefore optimization techniques are typically used to "calibrate" all the model parameters. (Shreve [46])

In financial mathematics, the change of measure is accomplished by the Radon-Nikodym derivative, which preserves the quadratic variation (and diffusion) properties of the process. This means that our estimator of the quadratic variation and the diffusion parameter are obtained on historical price data. While calibration methods are still needed to find the drift parameters, our proposed estimator could be used to reduce the number of variables in the optimization. Additionally, since calibration of interest rate models is typically done on a snapshot of the

forward rate curve, and the quadratic variation estimator uses a series of historical data, our estimator of diffusion may provide better estimates of the diffusion parameter in practice.

Modeling outside of finance typically requires estimation of all model parameters. Estimation of parameters in stochastic differential equations is a well studied subject. Girsanov's theorem (which uses the Radon-Nikodym derivative) establishes that the change of measure provides weights (or likelihood) on different outcomes of the stochastic process. Given an Itô process,

$$dX_t = b\left(X_t, \Theta\right) dt + \sigma\left(X_t, \Theta\right) dB_t,$$

where $\Theta \subset \mathbb{R}^p$ is a multidimensional parameter, Prakasa Rao [42] shows that using the likelihood function $L_T\left(\Theta\right)$ is equivalent to the Radon-Nikodym derivative:

$$L_T\left(\Theta\right) = \exp\left(\int_0^T \frac{b\left(X_s, \Theta\right)}{\sigma^2\left(X_s, \Theta\right)} dX_s - \frac{1}{2}\int_0^T \frac{b^2\left(X_s, \Theta\right)}{\sigma^2\left(X_s, \Theta\right)} ds\right).$$

The discrete counterpart of the likelihood function is given by

$$L_n\left(\Theta\right) = \prod_{i=1}^{n} p_\Theta\left(\Delta_i, X_{t_i} \mid X_{t_{i-1}}\right) p_\Theta\left(X_0\right),$$

where $p_\Theta\left(\Delta_i, X_{t_i} \mid X_{t_{i-1}}\right)$ is the transition density from value $X_{t_{i-1}}$ at time $t_{i-1}$to value X_{t_i} at time t_i, where $t_i = i\Delta_i$, $i = 0, 1, \ldots, n$ and $T = n\Delta_n$. In some cases, the sampling rate Δ_i can be constant, while in others it takes the form Δ_n varies and it is assumed that $n\Delta_n^k \to 0$ for some power $k \geq 2$. (Iacus [24], pp. 111) The choice of Δ_i is dependent on the behavior of the process. For each model, a specific sampling rate Δ_i is needed to guarantee almost sure convergence of the Maximum Likelihood Estimation (MLE). Additionally, it should be noted that closed form representations of the transition density is not always available and numerical methods are sometimes needed.

For consistency of the MLE estimators the asymptotics require $n \to \infty$, which is equivalent to $T \to \infty$. While estimation with the MLE on discrete data can be accurate on finite time intervals, optimization of the likelihood function can often fail because of the requirement for a time series of infinite length. (Iacus [24], pp. 116)

Example 6.7. One popular example of parameter estimation using the MLE is on the Ornstein-Uhlenbeck process. Iacus [24] simulates an Ornstein-Uhlenbeck process and uses numerical optimization to maximize the likelihood function and find the three parameters of the model. To

the best of our knowledge, the only closed form MLE representation of estimators of the O-U process parameters is given by Florens-Zmirou [16], where he assumes that the long term mean reversion parameter $\mu = 0$. Under this assumption, the O-U process takes the form

$$dX_t = -\theta X_t dt + \sigma dB_t.$$

Florens-Zmirou [16] finds that if $\Delta_i = 1/n = \Delta t$, then the MLE estimators are

$$\hat{\theta} = -\frac{1}{\Delta t} \ln \left(\frac{\sum_{i=1}^{n} X_{i-1} X_i}{\sum_{i=1}^{n} X_{i-1}^2} \right)$$

and

$$\hat{\sigma}^2 = \frac{2\hat{\theta}}{n\left(1 - e^{-2\Delta t \hat{\theta}}\right)} \sum_{i=1}^{n} \left(X_i - X_{i-1} e^{-\Delta t \hat{\theta}} \right)^2.$$

Notice that a major difference in Florens-Zmirou's [16] estimators and our quadratic variation estimator is the dependence on $\hat{\theta}$.

Motivated by Florens-Zmirou [16], we decided to attempt to derive the MLE for the full O-U model assuming that we already have an estimate of σ from the quadratic variation. Given the model

$$dX_t = \theta \left(\mu - X_t \right) dt + \sigma dB_t$$

its solution is

$$X_{t_i} = X_0 e^{-\theta t_i} + \mu \left(1 - e^{-\theta t_i} \right) + \sigma e^{-\theta t_i} \int_0^{t_i} e^{\theta u} dB_u$$

or

$$X_{t_i} = \mu \left(1 - e^{-\theta \Delta t} \right) + e^{-\theta \Delta t} X_{t_{i-1}} + \sigma \int_{t_{i-1}}^{t_i} e^{-\theta (t_i - u)} dB_u,$$

therefore,

$$p_\Theta \left(\Delta t, X_{t_i} \mid X_{t_{i-1}} \right) \stackrel{d}{=} N \left(\mu \left(1 - e^{-\theta \Delta t} \right) + e^{-\theta \Delta t} X_{t_{i-1}}, \frac{\sigma^2 \left(1 - e^{-2\theta \Delta t} \right)}{2\theta} \right).$$

Notice that the O-U process has transition probabilities that are invariant for fixed Δt. This

is used to form the likelihood function

$$\begin{aligned} L_n(\Theta) &= \prod_{i=1}^{n}\left(\frac{\theta}{\pi\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)^{\frac{1}{2}}\exp\left\{-\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2\right\} \\ &= \left(\frac{\theta}{\pi\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)^{\frac{n}{2}}\exp\left\{-\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2\right\}. \end{aligned}$$

The log-likelihood can be written as

$$l_n(\Theta)=\frac{n}{2}\log\left(\frac{\theta}{\pi\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)-\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2.$$

The maximum likelihood can be found by taking partials with respect to μ and θ:

$$\frac{\partial l_n(\Theta)}{\partial\mu}=\left(\frac{2\theta\left(1-e^{-\theta\Delta t}\right)}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)=0. \tag{6.10}$$

Therefore,

$$\hat{\mu}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}\right)}{\left(1-e^{-\theta\Delta t}\right)}. \tag{6.11}$$

Even though we assume that we have an estimate of σ^2, let us see if we can get more information out of the optimal σ^2:

$$\frac{\partial l_n(\Theta)}{\partial\sigma^2}=-\frac{n}{2}\frac{1}{\sigma^2}+\frac{\theta}{\sigma^4\left(1-e^{-2\theta\Delta t}\right)}\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2=0$$

$$\hat{\sigma}^2=\frac{2\theta}{n\left(1-e^{-2\theta\Delta t}\right)}\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2.$$

This gives an equation for the conditional variance of the Ornstein-Uhlenbeck process,

$$\frac{\hat{\sigma}^2\left(1-e^{-2\theta\Delta t}\right)}{2\theta}=\frac{1}{n}\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2. \tag{6.12}$$

The partial with respect to θ gives,

$$\begin{aligned}\frac{\partial l_n(\Theta)}{\partial\theta} &= \frac{n}{2}\frac{\left(\left(1-e^{-2\theta\Delta t}\right)-2\theta\Delta t e^{-2\theta\Delta t}\right)}{\theta\left(1-e^{-2\theta\Delta t}\right)} \\ &\quad -\frac{\left(1-e^{-2\theta\Delta t}\right)-2\theta\Delta t e^{-2\theta\Delta t}}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)^2}\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)^2 \\ &\quad -2\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)\left(\Delta t e^{-\theta\Delta t}X_{t_{i-1}}-\mu\Delta t e^{-\theta\Delta t}\right)\end{aligned}$$

Substituting equation 6.12 into the above results in a cancellation of the first and second terms of $\frac{\delta l_n(\Theta)}{\delta\theta}$:

$$\frac{\partial l_n(\Theta)}{\partial\theta} = -2\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)\left(\Delta t e^{-\theta\Delta t}X_{t_{i-1}}-\mu\Delta t e^{-\theta\Delta t}\right).$$

In equation 6.10, the partial with respect to μ established that

$$\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)=0,$$

and therefore

$$\frac{\partial l_n(\Theta)}{\partial\theta} = -2\left(\frac{\theta}{\sigma^2\left(1-e^{-2\theta\Delta t}\right)}\right)\sum_{i=1}^{n}\left(X_{t_i}-e^{-\theta\Delta t}X_{t_{i-1}}-\mu\left(1-e^{-\theta\Delta t}\right)\right)\left(\Delta t e^{-\theta\Delta t}X_{t_{i-1}}\right)=0.$$

Substituting the estimator of $\mu\left(1-e^{-\theta\Delta t}\right)$ which was established in equation 6.11, yields

$$0 = \left(\sum_{i=1}^{n}X_{t_i}X_{t_{i-1}}-e^{-\theta\Delta t}\sum_{i=1}^{n}X_{t_{i-1}}^2-\frac{1}{n}\sum_{i=1}^{n}X_{t_{i-1}}\sum_{j=1}^{n}X_{t_j}+\frac{1}{n}e^{-\theta\Delta t}\sum_{i=1}^{n}X_{t_{i-1}}\sum_{j=1}^{n}X_{t_{j-1}}\right).$$

Therefore the maximum likelihood estimator for θ takes the form

$$\hat{\theta}=\frac{-1}{\Delta t}\ln\left\{\frac{\frac{1}{n}\sum_{i=1}^{n}X_{t_{i-1}}\sum_{j=1}^{n}X_{t_j}-\sum_{i=1}^{n}X_{t_i}X_{t_{i-1}}}{\frac{1}{n}\left(\sum_{i=1}^{n}X_{t_{i-1}}\right)^2-\sum_{i=1}^{n}X_{t_{i-1}}^2}\right\}.$$

The above results establish a MLE estimator for all three parameters of the O-U process. We notice that when μ is unknown, the estimator for σ^2 and θ are similar to the Florens-Zmirou's [16], however extra terms are needed to account for the influence of μ. These extra terms cause an interesting effect on the estimator for θ. The numerator and denominator in the log function are both very close to each other, in fact, both approach zero as n tends to infinity. This is one

reason MLE consistency proofs require Δ fixed and $T \to \infty$. The delicate balance between the ratio makes this method very sensitive to perturbations in the process and can cause $\hat{\theta}$ estimates less than zero. Additionally, when $\hat{\theta}$ approaches zero, the denominator in equation 6.11 tends to zero and $\hat{\mu}$ explodes. In this situation the estimator $\hat{\sigma}^2$ is relatively unaffected for small Δt.

This result gives further insights into cases that cause a failure in the numerical maximization of Iacus [24]. If the optimal $\hat{\theta}$ is some value less than or equal to zero, the model requirements that $\theta > 0$ are violated and since the objective function can be very flat, the optimization fails.

Interestingly, while the MLE estimators for μ and σ^2 are functions of θ, as long as $\hat{\theta}$ is positive, $\hat{\mu}$ and $\hat{\sigma}^2$ are relatively insensitive to explosions of $\hat{\theta}$. The bias of these estimators is only slightly affected in the same direction as the bias in $\hat{\theta}$.

The major advantage of the quadratic variation estimator of volatility is the ability to prove consistency on a fixed and finite interval $[0, T]$, while the MLE estimator is not consistent unless $T \to \infty$. This difference brings into question which estimator is more accurate. In the remainder of this example we investigate the performance of the two estimators.

In applications, parameter estimates are found on a finite time interval and therefore we analyze the estimators performance for various N-s on the fixed interval $[0, 1]$. Simulating 10000 paths of the O-U process with the same parameters as in Example 6.6 and using the MLE and quadratic variation estimators, we obtain the following statistics.

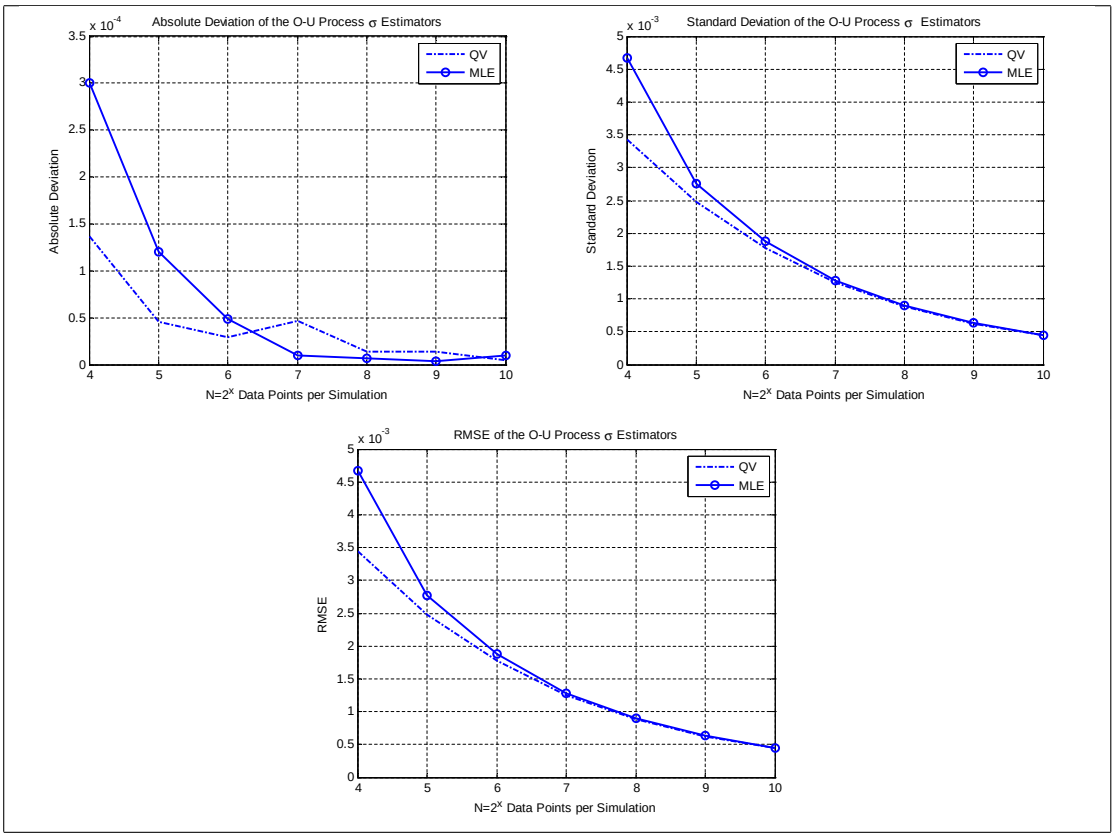


Figure 6.4: Comparison of Volatility Estimators of an O-U Process: Fixed Interval

The statistics above suggest that the quadratic variation estimator of σ performs better on smaller sample sizes than the MLE. To determine if there is a significant difference in the accuracy of the estimators, let us compare the estimator's absolute deviation using the paired t-test. Let,

$$D = average\left[|\hat{\sigma}_1 - \sigma_{Actual}| - |\hat{\sigma}_2 - \sigma_{Actual}|\right]$$

$$\sigma_D^2 = Var\left[|\hat{\sigma}_1 - \sigma_{Actual}| - |\hat{\sigma}_2 - \sigma_{Actual}|\right]$$

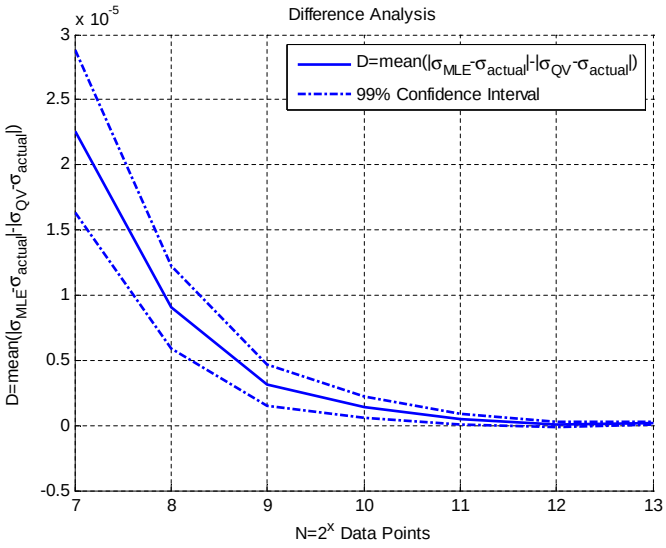
The confidence interval on the statistic D can be shown to be approximately,

$$D \pm \sigma_D t_{\frac{\alpha}{2},N-1}. \tag{6.13}$$

We use equation 6.13 to construct 99% confidence interval for testing the hypothesis

$$H_0 : |\hat{\sigma}_1 - \sigma| = |\hat{\sigma}_2 - \sigma|$$
$$H_1 : |\hat{\sigma}_1 - \sigma| \neq |\hat{\sigma}_2 - \sigma|$$

The results of the analysis can be seen below in Figure 6.5. Inclusion of zero in the confidence interval indicates that there is no significant difference in the estimators. The 99% upper and lower confidence interval bounds on D are denoted by D_U and D_L, respectively.



N =	**128**	**256**	**512**	**1024**	**2048**	4096
D	**2.26E-05**	**9.06E-06**	**3.12E-06**	**1.39E-06**	**4.52E-07**	1.28E-07
D_L	**1.63E-05**	**5.88E-06**	**1.53E-06**	**6.05E-07**	**5.89E-08**	-7.00E-08
D_U	**2.88E-05**	**1.22E-05**	**4.70E-06**	**2.18E-06**	**8.45E-07**	3.26E-07

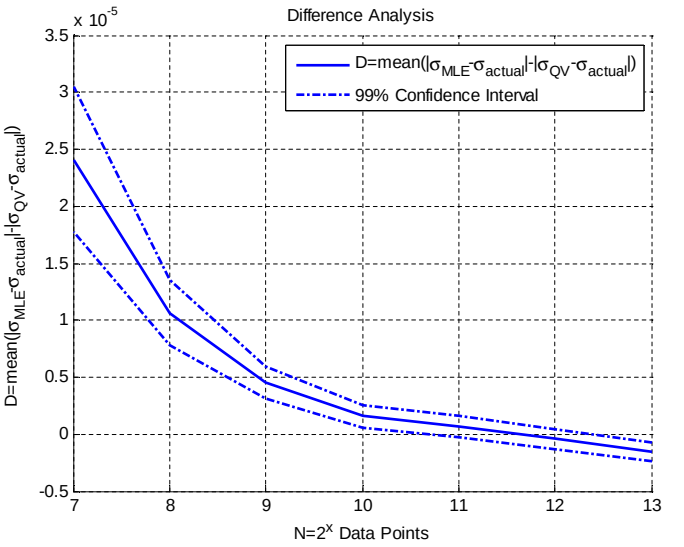
Figure 6.5: Difference Analysis on Estimators of Volatility of an O-U Process: Fixed Interval

Given the definition of D, if D is significantly great than zero, then the quadratic variation estimator of σ has less absolute deviation and vice-versa. The bold numbers in the table of Figure 6.5 highlight the values for N for which the analysis indicated that the quadratic variation estimator is more accurate. Further refinement indicates that the quadratic variation estimator is significantly more accurate for samples up to approximately 4000 data points. After this point, the analysis indicates that there is no significant difference in the accuracy of the estimators.

The prior analysis tested the performance of the estimators on a fixed interval as $\Delta t \to 0$. The quadratic variation estimator's consistency proof is derived under this assumption. However, the MLE requires Δt to remain fixed, while $T \to \infty$. Therefore, let us test the performance of the estimators under the MLE's consistency requirements to see which performs better.

Fixing $\Delta t = 50/2^{13}$ and simulating 10000 paths of the O-U process each with $t \in [0, 50]$ and the same parameters as in Example 6.6, the paths were sampled from $t_0 = 0$ to $T = 50/2^{13-i}$, for $i = 0, \ldots, 6$. The MLE and quadratic variation estimators were then applied to each sample and

analyzed using the difference analysis described above. The analysis gave the following results.



N =	**256**	**512**	**1024**	2048	4096	***8192***
D	**1.06E-05**	**4.52E-06**	**1.55E-06**	6.51E-07	-4.09E-07	***-1.56E-06***
D_L	**7.77E-06**	**3.15E-06**	**5.50E-07**	-2.64E-07	-1.26E-06	***-2.36E-06***
D_U	**1.35E-05**	**5.89E-06**	**2.55E-06**	1.57E-06	4.46E-07	***-7.48E-07***

Figure 6.6: Difference Analysis on Estimators of Volatility of an O-U Process: Fixed Delta t

The results of the analysis show that the quadratic variation estimator is significantly more accurate for samples up to approximately 2000 data points and the MLE estimator is significantly more accurate once approximately 8000 data points are exceeded. In between, there is no significant difference in the accuracy of the estimators. It should be noted that further analysis shows that choosing different fixed Δt-s change the cross point for the switch in out-performance. However, each analysis indicated that the quadratic variation estimator is significantly more accurate on smaller to moderate samples and the MLE is significantly more accurate on very large samples.

In the markets, conditions can change quickly. If the simulations of the O-U process with fixed Δt was a financial instrument and if σ was an annualized volatility (which is typical of financial applications), this implies that the MLE method would not significantly out-perform the quadratic variation method until you obtained 50 years of data that was sampled every 1.5 days. Accurate and fast analysis of smaller data sets on small fixed intervals is critical to decision making in finance. For these reasons, the quadratic variation estimator may be better suited for application in real world scenarios.

6.4 Summary of Results

Through the use of Hida's [19] underlying martingale representation we were able to show that the quadratic variation for any continuous GM process can be represented as a Riemann–Stieltjes integral for which we integrate the square of the scaling function with respect to the martingale's variance. This result gives insight into the influence of the dependence structure Ψ and variance ν on the range properties of a GM process. Using the discrete representation of the quadratic variation and its convergence properties, we were able to establish that the quadratic variation can be used as a statistical estimator of diffusion parameter which is consistent on a fixed interval $[0, T]$ and has a convergence rate of $1/\sqrt{n}$. A numerical study of the quadratic variation estimator of diffusion for an Ornstein-Uhlenbeck process provides support that the method is fast and accurate. Additionally, error (measured in terms of RMSE) of these estimators is primarily driven by the variance while the bias is relatively small.

A simulation analysis of the MLE and quadratic variation diffusion estimators of an Ornstein-Uhlenbeck process indicate that the quadratic variation estimator is significantly more accurate on small to moderate size samples. Analysis also indicated that out-performance may be due to the consistency requirements; the MLE estimator requires an infinite time horizon of data points, while the quadratic variation estimator requires an infinite number of data points on a fixed time interval. The results also show that the MLE estimator of diffusion σ requires estimation of both the spring dampening constant θ and the long term mean μ first, while the quadratic variation estimator of diffusion is independent of the other parameters. This makes the quadratic variation estimator both faster and more accurate or equivalent for samples on fixed finite time intervals.

Chapter 7

Derivative Pricing Theory and Itô Calculus with GM Processes

In this chapter we review the role of quadratic variation in stochastic calculus and demonstrate the importance of accurate estimation of the diffusion parameter in application. In sections 7.2 and 7.3 we explicitly derive the call option pricing formula and the Kolmogorov (diffusion) backward equation under the assumption that an asset's price process follows a geometric Gaussian Markov process. Our results go beyond the typical i.i.d. modeling assumptions to extend the Black-Scholes-Merton option pricing formula and Black-Scholes pde to accommodate for various structures of long/short dependence and scedasticity. The Raydon-Nikodym derivative is explicitly derived to find the proper change of measure and arbitrage free conditions. Results in this chapter demonstrate how quadratic variation plays a key role in pricing, hedging and risk management.

7.1 Quadratic Variation in Stochastic Calculus: A Brief Review

Let $(x,t) \mapsto f(t,x)$ be a function for which the partial derivatives f_t, f_x, and f_{xx} are continuous. If $X = (X_t)_{t\in\mathbb{R}}$ is a continuous semi-martingale process, then the Itô-Doeblin formula (also known as Itô's lemma) states that

$$f(T,X(T))-f(0,X(0))=\int_0^T f_t(t,X(t))\,dt+\int_0^T f_x(t,X(t))\,dX(t)+\frac{1}{2}\int_0^T f_{xx}(t,X(t))\,d\left[X,X\right]_t\,, \tag{7.1}$$

where the quadratic variation of the process X, $[X,X]_t$, is of bounded variation. Alternatively, the Itô-Doeblin formula can be written as

$$df(t,X(t))=f_t(t,X(t))\,dt+f_x(t,X(t))\,dX(t)+\frac{1}{2}f_{xx}(t,X(t))\,d\left[X,X\right]_t\,. \tag{7.2}$$

Remark 7.1. In Chapter 6 we derived the quadratic variation of the Gaussian Markov process $(X_t)_{t\in[0,T]}$ and established that $d\left[X,X\right]_t=\Psi_t^2 d\nu_t$. As long as Ψ is continuous and has bounded variation, we can use the Itô-Doeblin formula with the Gaussian Markov process since we can write

$$dX_t=\Psi_t dM_t+M_t d\Psi_t.$$

For a proof of the Itô-Doeblin formula see Chapter 8 in Steele [50].

7.2 Derivation of the Call Option Pricing Model

In this section we explicitly derive the call option pricing formula under the assumption of a stock process following a geometric Gaussian Markov process. This derivation adapts the standard techniques used to derive the Black-Scholes-Merton call option pricing formula to accommodate for a Gaussian Markov process.

Let $(S_t)_{t\in[0,T]}$ be a stock process model so that

$$S_t=S_0e^{\mu t-\frac{1}{2}\sigma^2\int_0^t\Psi_s^2 d\nu_s+\sigma X_t}. \tag{7.3}$$

where $(X_t)_{t\in[0,T]}$ is the Gaussian Markov process such that $X_t=\Psi_t M_t$, and σ and μ are diffusion and drift constants, respectively.

In Chapter 6 we derived the quadratic variation of the Gaussian Markov process and established that $d\left[X,X\right]_t=\Psi_t^2 d\nu_t$. Applying the Itô-Doeblin formula 7.2 to S_t, it can be seen that this model is a generalization of the Black-Scholes model so that

$$dS_t = \sigma S_t dX_t + \mu S_t dt.$$

Note that if the scaling function $\Psi_t = 1$ and the underlying martingale's variance $\nu_t = t$, this model is exactly the Black-Scholes model. However if $\Psi \neq C$, this model can incorporate long or short range dependence, while if ν can be chosen to make the log returns of the process be highly non-stationary.

Additionally, assume a risk free bond defined by

$$B_t = B_0 e^{rt},$$

where r is a constant for the risk free rate.

Let $\mathcal{F}_t = \sigma\left(X_r \mid r \leq t\right)$ and define

$$Z_t \equiv B_t^{-1} S_t = S_0 e^{(\mu - r)t - \frac{1}{2}\sigma^2 \int_0^T \Psi_t^2 d\nu_t + \sigma \Psi_t M_t}. \tag{7.4}$$

By the Itô-Doeblin formula 7.1

$$\begin{aligned} dZ_t &= Z_t\left(\sigma d\left(\Psi_t M_t\right) + \left(\mu - r\right) dt\right) \\ &= \sigma Z_t \left(d\left(\Psi_t M_t\right) + \frac{\mu - r}{\sigma} dt\right) \\ &= \sigma \Psi_t Z_t \left(dM_t + \frac{M_t}{\Psi_t} d\Psi_t + \frac{\mu - r}{\sigma \Psi_t} dt\right). \end{aligned} \tag{7.5}$$

Recall that $\left(M_t\right)_{t \in [0,T]}$ is a $\mathbb{P}$ martingale process and let

$$d\gamma_t = \frac{M_t}{\Psi_t} d\Psi_t + \frac{\mu - r}{\sigma \Psi_t} dt,$$

then $\left(\gamma_t\right)_{t \in [0,T]}$ is an adapted (or previsible) process to $\left\{\mathcal{F}_t\right\}_{t \in [0,T]}$. Define

$$\tilde{M}_t = M_t + \int_0^t \gamma_s d\nu_s.$$

Set $Y_t \equiv -\int_0^t \gamma_s dM_s - \frac{1}{2}\int_0^t \gamma_s^2 d\nu_s$, $\Lambda_t = e^{Y_t}$ and assume $\mathbb{E}\left[e^{\frac{1}{2}\int_0^T \gamma_s^2 d\nu_s}\right] < \infty$. Since,

$$\begin{aligned} d\Lambda_t &= e^{Y_t} dY_t + \frac{1}{2} e^{Y_t} d\left[Y,Y\right]_t \\ &= e^{Y_t}\left(-\int_0^t \gamma_s dM_s - \frac{1}{2}\int_0^t \gamma_s^2 d\nu_s\right) + \frac{1}{2} e^{Y_t}\gamma_t^2 d\nu_t \\ &= -\gamma_t \Lambda_t dM_t, \end{aligned}$$

$(\Lambda_t)_{t\in[0,T]}$ is a stochastic integral and therefore it is also a martingale process. Let the measure $\mathbb{Q}$ be defined by the Raydon-Nikodym derivative $\Lambda_T = d\mathbb{Q}/d\mathbb{P}$. Then $(\Lambda_t)_{t\in[0,T]}$ is the Raydon-Nikodym process of Λ_T. Note that

$$\begin{aligned} d\left(\tilde{M}_t\Lambda_t\right) &= \tilde{M}_t d\Lambda_t + \Lambda_t d\tilde{M}_t + d\Lambda_t d\tilde{M}_t \\ &= \left(1-\gamma_t\tilde{M}_t\right)\Lambda_t dM_t, \end{aligned}$$

which assures that the process $\left(\tilde{M}_t\Lambda_t\right)_{t\in[0,T]}$ is a $\mathbb{P}$ martingale since, for $t > s$,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}\left[\tilde{M}_t \mid \mathcal{F}_s\right] &= \frac{1}{\Lambda_s}\mathbb{E}_{\mathbb{P}}\left[\tilde{M}_t\Lambda_t \mid \mathcal{F}_s\right] \\ &= \tilde{M}_s. \end{aligned}$$

This establishes that $\left(\tilde{M}_t\right)_{t\in[0,T]}$ is a $\mathbb{Q}$ Gaussian martingale process with $\mathbb{E}_{\mathbb{Q}}\left[\tilde{M}_t^2\right] = \nu_t$. Therefore, the risk neutral measure $\mathbb{Q}$ of the discounted stock process is given by

$$dZ_t = \sigma Z_t \Psi_t d\tilde{M}_t,$$

consequentially the solution to this spde is given by

$$Z_t = Z_0 e^{\sigma\int_0^t \Psi_s d\tilde{M}_s - \frac{1}{2}\sigma^2\int_0^t \Psi_s^2 d\nu_s},$$

which is a $\mathbb{Q}$ martingale.

Remark 7.2. In Shreve [46], Theorem 5.4.7: *The First Fundamental Theorem of Asset Pricing* states that if a market model has a risk-neutral probability measure, then it does not admit arbitrage. Therefore, as long as $\mathbb{E}\left[e^{\frac{1}{2}\int_0^T \gamma_s^2 d\nu_s}\right] < \infty$, the geometric GM stock model will be arbitrage free.

Let F_T be the payoff of an option on the stock at time T. Define

$$E_t = \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1} F_T \mid \mathcal{F}_t\right]. \tag{7.6}$$

Therefore, the discounted option $(E_t)_{t\in[0,T]}$ is a $\mathbb{Q}$ martingale. If we proceed with standard Option Pricing Theory, according to the Martingale Representation Theorem, since $(E_t)_{t\in[0,T]}$ and $(Z_t)_{t\in[0,T]}$ are two $\mathbb{Q}$ martingales, then there exists a previsible ($\mathcal{F}_t$- measurable) process $(\phi_t)_{t\in[0,T]}$ so that

$$dE_t = \phi_t dZ_t. \tag{7.7}$$

Let $(\Pi_t)_{t\in[0,T]}$ be a portfolio consisting of $(\phi_t)_{t\in[0,T]}$ shares of the underlying stock and $(\psi_t)_{t\in[0,T]}$ shares of a risk free bond, where

$$\psi_t = E_t - \phi_t Z_t. \tag{7.8}$$

Then, the value of the portfolio is at time t is given by

$$V_t = \phi_t S_t + \psi_t B_t = E_t B_t. \tag{7.9}$$

Using the standard approach in Option Pricing Theory, it can be shown that the portfolio $(V_t)_{t\in[0,T]}$ replicates the option payoff at time T. (Shreve [46]) Additionally, the portfolio $(\Pi_t)_{t\in[0,T]}$ is *self financing.* (Shreve [46]) Since we construct the model so it does not admit arbitrage, the value of the option F_t at time $t \in [0,T]$, equals the value the portfolio V_t. Therefore,

$$F_t = B_t \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1} F_T \mid \mathcal{F}_t\right], \tag{7.10}$$

which is the fundamental formula in Option Pricing Theory. Let us use fundamental formula to derive the price of a call option for the geometric Gaussian Markov stock model.

Definition 7.3. A *call option* is a financial contract between two parties that gives its buyer the right, but not the obligation to buy a financial instrument (the underlying security S) from the seller of the option at a predefined time T, known as the expiration date, for an agreed upon price K, known as the strike price.

The payoff of the call option at the expiration time, T, is given by the formula

$$X_T = (S_T - K)^+ .$$

The derivation of the call option's value at time $t \in [0, T]$ is accomplished under the risk-neutral $\mathbb{Q}$-measure. Since the discounted stock price under the $\mathbb{Q}$-measure was given by

$$Z_t \quad = \quad Z_0 e^{\sigma \int_0^t \Psi_s d\tilde{M}_s - \frac{1}{2}\sigma^2 \int_0^t \Psi_s^2 d\nu_s}$$

then the stock process under measure $\mathbb{Q}$ is given by

$$S_t \quad = \quad S_0 e^{rt - \frac{1}{2}\sigma^2 \int_0^t \Psi_s^2 d\nu_s + \sigma \int_0^t \Psi_s d\tilde{M}_s}.$$

Recall, the value of a derivative is given by equation 7.10. Let $\tau \equiv T - t$ and $B_0 \equiv 1$, then the call option value is given by

$$\begin{aligned}
F_t &= B_t \mathbb{E}_{\mathbb{Q}} \left[e^{-rT} \left(S_0 e^{rT + \sigma \int_0^T \Psi_s d\tilde{M}_s - \frac{1}{2}\sigma^2 \int_0^T \Psi_s^2 d\nu_s} - K \right)^+ \mid \mathcal{F}_t \right] \\
&= e^{-r\tau} \mathbb{E}_{\mathbb{Q}} \left[\left(S_0 e^{rT + \sigma \int_0^T \Psi_s d\tilde{M}_s - \frac{1}{2}\sigma^2 \int_0^T \Psi_s^2 d\nu_s} - K \right)^+ \mid \mathcal{F}_t \right] \\
&= e^{-r\tau} \mathbb{E}_{\mathbb{Q}} \left[\left(S_0 e^{rt - \frac{1}{2}\sigma^2 \int_0^t \Psi_s^2 d\nu_s + \sigma \int_0^t \Psi_s d\tilde{M}_s} e^{r\tau - \frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s + \sigma \int_t^T \Psi_s d\tilde{M}_s} - K \right)^+ \mid \mathcal{F}_t \right] \\
&= e^{-r\tau} \mathbb{E}_{\mathbb{Q}} \left[\left(S_t e^{r\tau - \frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s + \sigma \int_t^T \Psi_s d\tilde{M}_s} - K \right)^+ \mid \mathcal{F}_t \right] \\
&= e^{-r\tau} \mathbb{E}_{\mathbb{Q}} \left[\left(S_t e^{r\tau - \frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s + \sigma \int_t^T \Psi_s d\tilde{M}_s} - K \right)^+ \right] \qquad (7.11) \\
&= e^{-r\tau} \mathbb{E}_{\mathbb{Q}} \left[\left(\hat{S}_t e^{\sigma \int_t^T \Psi_s d\tilde{M}_s} - K \right)^+ \right] \\
&= e^{-r\tau} \mathbb{E} \left[\left(\hat{S}_t e^{\sigma Z \sqrt{\int_t^T \Psi_s^2 d\nu_s}} - K \right)^+ \right] \qquad (7.12)
\end{aligned}$$

where

$$\hat{S}_t = S_t e^{r\tau - \frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s}$$

and Z is a standard normal random variable. Note that equation 7.11 deduces from the fact that $S_t e^{r\tau - \frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s}$ is deterministic and $\sigma \int_t^T \Psi_s d\tilde{M}_s$ is a Gaussian process that is independent of the filtration $\mathcal{F}_t$. Since the Gaussian martingale $d\tilde{M}$ has quadratic variation $d\nu_t$, equation 7.12

is a consequence of the Itô isometry

$$\begin{aligned}\mathbb{E}\left[\int_t^T f_s d\tilde{M}_s\right]^2 &= \mathbb{E}\left[\int_t^T f_s^2 d\left[X,X\right]_s\right] \\ &= \mathbb{E}\left[\int_t^T f_s^2 d\nu_t\right].\end{aligned}$$

Let $\Sigma \equiv \sigma^2 \int_t^T \Psi_s^2 d\nu_s$. Then

$$F_t = e^{-r\tau}\mathbb{E}\left[\left(\hat{S}_t e^{Z\sqrt{\Sigma}} - K\right)^+\right]$$

consequently this form allows us to use the Black-Scholes-Merton call option formula by substituting σ with Σ. Therefore, the value of the call option is given by

$$\begin{aligned}F_t &= e^{-r(T-t)}\left[\hat{S}_t e^{\frac{1}{2}\sigma^2 \int_t^T \Psi_s^2 d\nu_s}\Phi\left(d_1\right) - K\Phi\left(d_2\right)\right] \\ &= S_t\Phi\left(d_1\right) - e^{-r(T-t)}K\Phi\left(d_2\right). \qquad (7.13)\end{aligned}$$

where $\Phi\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{1}{2}x^2}dx$ is the standard cumulative normal distribution and

$$\begin{aligned}d_1 &= \frac{\ln\left(\frac{S_t}{K}\right) + r\left(T-t\right) + \frac{1}{2}\sigma^2\int_t^T \Psi_s^2 d\nu_s}{\sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}} \\ d_2 &= d_1 - \sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}.\end{aligned}$$

We notice that if $\Psi = 1$ and $\nu\left(t\right) = t$, this yields the quadratic variation of Brownian motion, $\int_t^T \Psi_s^2 d\nu_s = T - t$, and we obtain the Black-Scholes-Merton option pricing formula 4.3. Additionally, equation 7.13 demonstrates that choosing a GM process with the proper quadratic variation and having accurate estimation techniques for σ is essential since the quadratic variation and σ significantly contribute to the theoretical option price. Furthermore, since Ψ and ν can be chosen so that Σ can be non-linear over time, this model could accommodate for the term structure to volatility that is observed in the market. (Hull [23], pp.377)

7.3 The Kolmogorov Backward Equation for Pricing Derivatives

In this section we use fundamental principles of modern financial theory and the previous results to adapt the Black-Scholes partial differential equation (otherwise known as the Kolmogorov backward equations) to the Gaussian Markov process. The derivation of such a pde is essential for pricing derivatives and finding hedging strategies for risk management. These results are used in the next section to find the *Greeks* for a call option, which give the hedging strategies and dynamics of the replicating portfolio.

Let $(X_t)_{t\in\mathbb{R}}$ be a Gaussian Markov process and assume that the stock process $(S_t)_{t\in\mathbb{R}}$ is given by

$$dS_t = \sigma(t, S_t)\, dX_t + \mu(t, S_t)\, dt, \tag{7.14}$$

where functions σ and μ satisfy standard assumptions in definition 5.11. Additionally, let a risk free bond follow the model

$$dB_t = rB_t dt.$$

Since the option price F_t satisfies,

$$V_t = \phi_t S_t + \psi_t B_t = F_t, \tag{7.15}$$

then $F_t = f(t, S_t)$.

Furthermore, since the portfolio is *self-financing* it satisfies

$$\begin{aligned} df(t, S_t) &= \phi_t dS_t + \psi_t dB_t. \\ &= \sigma(t, S_t)\, \phi_t dX_t + (\mu(t, S_t)\, \phi_t + r\psi_t B_t)\, dt. \end{aligned} \tag{7.16}$$

On the other hand, applying the Itô-Doeblin formula 7.1 to the stock process 7.14 yields

$$df(t, S_t) = f_t(t, S_t)\, dt + f_x(t, S_t)\mu(t, S_t)\ dt + f_x(t, S_t)\sigma(t, S_t)\ dX_t + \frac{1}{2} f_{xx}(t, S_t)\sigma^2(t, S_t)\ d[X, X]_t. \tag{7.17}$$

Equating 7.17 to 7.16 yields

$$\begin{aligned}(\sigma(t,S_t)\,\phi_t - f_x(t,S_t)\sigma(t,S_t))\,dX_t &= (f_t(t,S_t) + f_x(t,S_t)\mu(t,S_t) - \mu(t,S_t)\,\phi_t - r\psi_t B_t)\,dt \\ &\quad + \frac{1}{2} f_{xx}(t,S_t)\sigma^2(t,S_t)\; d[X,X]_t\,. \end{aligned} \tag{7.18}$$

Since

$$dX_t = \Psi_t dM_t + M_t d\Psi_t,$$

then

$$\begin{aligned}(\sigma(t,S_t)\,\phi_t - f_x(t,S_t)\sigma(t,S_t))\,\Psi_t dM_t &= (f_t(t,S_t) + f_x(t,S_t)\mu(t,S_t) - \mu(t,S_t)\,\phi_t - r\psi_t B_t)\,dt \\ &\quad - M_t d\Psi_t\,(\sigma(t,S_t)\,\phi_t - f_x(t,S_t)\sigma(t,S_t)) \\ &\quad + \frac{1}{2} f_{xx}(t,S_t)\sigma^2(t,S_t)\,\Psi_t^2 d\nu_t. \end{aligned}$$

If

$$a_t = \Psi_t\,(\sigma(t,S_t)\,\phi_t - f_x(t,S_t)\sigma(t,S_t)) \neq 0,$$

then we can divide the rhs by a_t, and the lhs dM_t is independent of $\mathcal{F}_t$ while the rhs is $\mathcal{F}_t$-measurable. This can only happen if M_t is zero, which violates the definition of $\nu_t \neq 0$ for all $t > 0$. Therefore since $\Psi_t \neq 0$,

$$\sigma(t,S_t)\,\phi_t - f_x(t,S_t)\sigma(t,S_t) = 0.$$

Thus the shares of stock in the replicating portfolio

$$\phi_t = f_x(t,S_t). \tag{7.19}$$

Substituting this result into equation 7.18, then

$$(f_t(t,S_t) - r\psi_t B_t)\,dt + \frac{1}{2} f_{xx}(t,S_t)\sigma^2(t,S_t)\,\Psi_t^2 d\nu_t = 0. \tag{7.20}$$

Rearranging equation 7.15

$$\begin{aligned} \psi_t B_t &= f(t, S_t) - \phi_t S_t. \\ &= f(t, S_t) - f_x(t, S_t) S_t, \end{aligned} \tag{7.21}$$

and substituting into equation 7.20, we obtain the Kolmogorov derivative pricing partial differential equation

$$rf(t, S_t)\, dt = (f_t(t, S_t) + rS_t f_x(t, S_t))\, dt + \frac{1}{2}\sigma^2 S_t^2 f_{xx}(t, S_t)\, \Psi_t^2 d\nu_t. \tag{7.22}$$

7.4 The Greeks and Sensitivity of the Call Option Pricing Equation

The previous derivations form a replicating portfolio of a derivative in continuous time. We have seen how the assumption of continuous time processes is essential for simplifying the large sums that would occur if we assumed we were in discrete time. Basically, the power of the Itô integral (among properties of Gaussian martingales) enables the ease of derivation of the call option pricing model. However, even with the microsecond scaled speeds in which electronic trading currently takes place, these security prices evolve as a discrete process. Perfect replication of a derivative would require continuous re-balancing, something that is not realizable (not to mention infeasible when transaction costs are present). The use of a continuous time model to approximate a discrete time system means an introduction of modeling errors. Additionally, implicit assumptions pertaining to constant volatility, σ, drift, μ and dependence structure (modeled in $\Psi$and ν) may change as information is reflected in prices.

In the remainder of this section we focus on deriving the sensitivity of the derived model to changes. In finance, the sensitivity analysis of a pricing model to model factors is known as the *Greeks*. Furthermore, hedging strategies are needed to reduce exposures to risks in the market. The Greeks allow for the hedging strategies that are needed to mitigate these risks.

Using the Kolmogorov pde 7.22 we can find the shares in the $(\phi_t, \psi_t)_{t\in[0,T]}$ replicating portfolio for the call option and the *Greeks*. In the following, we find the terms needed in the Kolmogorov pde to gain insights into hedging strategies and sensitivities. Recall the call option pricing formula 7.13:

$$f(t,S) = S\Phi(d_1) - e^{-r(T-t)}K\Phi(d_2),$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{S}{K}\right) + r(T-t) + \frac{1}{2}\sigma^2\int_t^T \Psi_s^2 d\nu_s}{\sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}} \\ d_2 &= d_1 - \sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}. \end{aligned}$$

As mentioned in the derivation of equation 7.13, if we let $\Sigma \equiv \sigma^2\int_t^T \Psi_s^2 d\nu_s$, then equation 7.13 is the Black-Scholes-Merton call option pricing model with σ substituted by Σ. Therefore, we can use the traditional Greeks (see Hull [23] Chapter 15) with the substitution of Σ to find the following:

$$\begin{aligned} \Delta &\equiv \frac{\partial f}{\partial S} = \phi_t = \Phi(d_1) \\ \Gamma &\equiv \frac{\partial^2 f}{\partial S^2} = \frac{\Phi'(d_1)}{\sigma S\sqrt{\int_t^T \Psi_s^2 d\nu_s}} \\ \rho &\equiv \frac{\partial f}{\partial r} = \tau e^{-r\tau}K\Phi(d_2). \end{aligned}$$

The Greeks Theta and Vega are not the same as in the Black-Scholes-Merton call option Greeks since the quadratic variation of a Gaussian Markov process may not evolve linearly with time (as it does in the Brownian motion case). In the derivation that follows we find Theta. Let,

$$\Theta \equiv \frac{\partial f}{\partial t} = S\Phi'(d_1)\frac{\partial d_1}{\partial t} - rKe^{-(T-t)}\Phi(d_2) - Ke^{-r(T-t)}\Phi'(d_2)\frac{\partial d_2}{\partial t},$$

notice that

$$\frac{\partial d_2}{\partial t} = \frac{\partial d_1}{\partial t} - \sigma\frac{\delta\sqrt{\int_t^T \Psi_s^2 d\nu_s}}{\delta t} = \frac{\partial d_1}{\partial t} + \frac{\sigma}{2\sqrt{\int_t^T \Psi_s^2 d\nu_s}}\frac{\Psi_t^2 d\nu_t}{dt},$$

and therefore

$$\begin{aligned} \Theta &= \frac{\partial d_1}{\partial t}\left(S\Phi'(d_1) - Ke^{-r(T-t)}\Phi'(d_2)\right) - rKe^{-(T-t)}\Phi(d_2) \\ &\quad -Ke^{-r(T-t)}\Phi'(d_2)\left(\frac{\sigma}{2\sqrt{\int_t^T \Psi_s^2 d\nu_s}}\frac{\Psi_t^2 d\nu_t}{dt}\right). \end{aligned}$$

Notice that

$$\begin{aligned} Ke^{-r(T-t)}\Phi'(d_2) &= \frac{K}{\sqrt{2\pi}}e^{-r(T-t)-\frac{1}{2}\left(d_1-\sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}\right)^2} \\ &= \frac{K}{\sqrt{2\pi}}e^{-r(T-t)-\frac{1}{2}\left(d_1-\sigma\sqrt{\int_t^T \Psi_s^2 d\nu_s}\right)^2} \\ &= \frac{K}{\sqrt{2\pi}}e^{\ln\left(\frac{S}{K}\right)-\frac{1}{2}d_1^2} \\ &= S\Phi'(d_1). \end{aligned}$$

In the derivation of Δit is easy to show that the term $S\Phi'(d_1) - Ke^{-r(T-t)}\Phi'(d_2) = 0$, and therefore we obtain,

$$\Theta = -rKe^{-r\tau}\Phi(d_2) - \frac{\sigma S}{2\sqrt{\int_t^T \Psi_s^2 d\nu_s}}\Psi_t^2 d\nu_t \Phi'(d_1), \tag{7.23}$$

where $\tau = T - t$.

The derivation of Vega is similar to that of the Black-Scholes-Merton model's calculation with minor modifications for the quadratic variation of the GM process. It is easy to show that:

$$\mathcal{V} \equiv \frac{\partial f}{\partial \sigma} = S\Phi'(d_1)\sqrt{\int_t^T \Psi_s^2 d\nu_s}.$$

Note that each Greek is highly dependent on the quadratic variation of the GM process. This means that proper risk management and hedging strategies require choosing a GM process with the proper quadratic variation and accurate estimation of σ.

7.5 Discussion of Results

In this chapter, we demonstrated how the pricing equations and hedging strategies are modified for general forms of a continuous path Gaussian Markov processes. Particularly we observe that proper risk management and pricing is extremely sensitive to the quadratic variation and estimates of σ. These results confirm the importance of accurate estimators of the diffusion parameter for pricing and hedging strategies, like the one introduced in Chapter 6. Most importantly, the pricing formulas indicate that the theoretical price can be significantly misvalued if modeling is done with a process that does not have the market's true quadratic variation. Additionally, this will result in an inability to completely hedge exposures to different risks. In

this regard, we reviewed how stochastic calculus and Itô-Doeblin formula are highly influenced by the quadratic variation. The results show that the GM process has an advantage over the fractional Brownian motion model, which admits arbitrage, lacks exact hedging strategies, and does not have economical interpretations of replication and self-financing portfolios. We saw how the Radon-Nikodym derivative is constructed for the geometric GM stock model along with the condition required for the change of measure, which, if satisfied, guarantees an arbitrage free model.

Most importantly, the results of this chapter demonstrate how memory and heteroscedasticity *independently* influence Option Pricing Theory and the management of risk. We showed that we can expand the assumptions of the Black-Scholes model by replacing Brownian motion with a Gaussian Markov process. This modification allows for the flexibility to model many observed phenomenon, like long-range dependence and non-stationarity. The change does not significantly complicate derivations since solutions only replace the linear quadratic variation of Brownian motion with the quadratic variation of the Gaussian Markov process.

All of the results in this chapter show that the quadratic variation is important term in determining price, optimal policy and hedging strategies. The dependence structure and or non-linear scaling of time can result in a non-linear representation of quadratic variation. In this scenario, the results of this chapter show that the process (and it's increments) may have marginal Gaussian distribution that constantly change over time. Just as in the case of fractional Brownian motion, this would result in a term-structure to volatility and misvaluations of risk if the standard Brownian motion model was used on such a GM process. However, more properties than fBm could be incorporated that could complicate statistical analysis and give a false impression of process characteristics. Since many GM processes are non-stationary, in general, if these processes were analyzed under the assumption of i.d.d. increments or stationarity (like in the Brownian motion model), the processes would not appear Gaussian (and would fail tests for normality). This is because the convolution of the Gaussian random variables with different variances is not Gaussian. While there are many observed market properties that GM processes could explain, Option Pricing Theory suggests that Gaussian Markov process cannot explain the volatility smile that has been observed in the market since the 1987 market crash. On the other hand, the theory makes one point clear; choosing a model with an improper quadratic variation will result in improper risk management and exposures to unforeseen risks that could have significant consequences.

Chapter 8

Replication and Simulation of a GM Process

In this chapter we address the need for discrete representations of stochastic processes. In §8.2, we present a general method to generate sample paths of continuous sample path GM processes using Monte-Carlo simulation, while in §8.3 we present a general method for generating recombining binomial trees of any continuous path GM process and prove that the trees converge, as stochastic processes, to their corresponding GM processes.

8.1 Introduction

Derivative contracts come with a variety of clauses stipulating when the contract can be exercised by the buyer. In the case of the *European* style derivatives, execution of the contract can only occur at the expiration date, while *American* style derivatives give the buyer the right to exercise the contract anytime before expiration. The vast majority of options are European and American style and are referred to as *vanilla options*. Non-vanilla option styles like Bermudan, Canary, compound, shout and swing options have a discrete number of exercise dates with varying terms.

For many European and non-vanilla payoff functions, f, the price and hedging strategies can be found by analysis at each of the discrete exercise times. The discrete set of execution times often allows for closed form solutions to be found using stochastic calculus.

In the case of American style derivatives, the maximization of the expected payoff can require exercising the option before the expiration date. According to Theorem 8.5.2 in Shreve [46], if

$f(x)$, $x \geq 0$, is a non-negative and convex payoff function of a derivative such that $f(0) = 0$, then the American style derivative expiring at time T is optimal to be exercised at T. This means that the American style and European style derivative have the same price and we can evaluate the American derivative as if it is of European style. Furthermore, using Jensen's inequality, it can be shown that the optimal policy for maximizing expected payoff is to hold the derivative to expiration. For example, given an expiration date and a strike price, the buyer of an American call option should hold the contract until expiration and pay the same price as a European call option on the same security. However, in the case of an American style put option with strike price K, the payoff function, $f(x) = (K - x)^+$, is not zero when the underlying is zero. Since Theorem 8.5.2 in Shreve [46] is not satisfied, it may be optimal to exercise the put option before the expiration date.

For derivative payoffs where Theorem 8.5.2 in Shreve [46] are not satisfied, the optimal execution time of the contract is a stopping time random variable with range $[0, T]$. Stochastic calculus gives the formula in form of maximum over all stopping times, for these types of American style derivatives closed form solutions are not known. The most popular example of a contract that does not satisfy the theorem's requirements is the American put option. This creates a need to have a discrete representation of the stochastic process to find the optimal price, policies and hedging strategies that maximize expected payoff. One such approach is Monte-Carlo simulation which typically requires the generation and analysis of millions of sample paths to have the accuracy required by the market place. Another popular method is to numerically solve the Kolmogorov backwards equations by performing finite difference analysis. In this type of analysis, errors can occur when there are non-smooth terminal boundary conditions, which are typical of almost all financial derivative products with the exception of futures contracts (forward contracts can be solved explicitly and therefore do not need discretization methods). (Shreve [46], Exercises 6.8 & 6.9) Another approach is the application of dynamic programming on binomial trees. Given a recombining tree, the procedure to obtain optimal policies is well understood, fast, and without approximations. (See Luenberger [30]) The accuracy of the solution is determined by the number of periods n in the tree. In the Cox-Ross-Rubinstein (CRR) model, as $n \to \infty$, the tree converges to the Brownian motion process at a rate of $1/\sqrt{n}$. The CRR model is often used in the valuation of American style derivatives because of the speed and accuracy of dynamic programming methodologies.

8.2 Simulation of GM Processes

Even though Monte-Carlo simulation can be an inefficient brute force method of determining policy and price, it is often needed to price exotic derivatives for which other methods are intractable. One such example would be the pricing of an American style Asian option. Asian options determine payoff by averaging prices over the contract time period. In this case, a recombining binomial tree would have a payoff tree which is non-recombining at each node, making the payoff binomial tree grow exponentially as the number of periods are increased (the tree has $2^{n+1}-1$ nodes for periods $n=0,1,\dots$). In order to determine price and policy we derive a general approach to simulate the paths of a Gaussian Markov process.

Let $(X_t)_{t\in[0,T]}$ be a continuous path GM process. Then by Theorem 5.3, $X_t=\Psi_t M_t$ so that $\mathbb{E}\left[M_t^2\right]=\nu(t)$. Note that when $t>s$ then

$$\begin{aligned}
P\left(X_t\le x\mid X_s=y\right) &= P\left(\Psi_t M_t\le x\mid \Psi_s M_s=y\right)\\
&= P\left(M_t\le \frac{x}{\Psi_t}\mid M_s=\frac{y}{\Psi_s}\right) \qquad (8.1)\\
&= P\left(M_t-M_s\le \frac{x}{\Psi_t}-M_s\mid M_s=\frac{y}{\Psi_s}\right)\\
&= P\left(\frac{M_t-M_s}{\sqrt{\nu(t)-\nu(s)}}\le \frac{\frac{x}{\Psi_t}-M_s}{\sqrt{\nu(t)-\nu(s)}}\right) \qquad (8.2)\\
&= P\left(Z\le \frac{\frac{x}{\Psi_t}-M_s}{\sqrt{\nu(t)-\nu(s)}}\right),
\end{aligned}$$

where equation 8.2 is a result of the Gaussian martingale property that (M_t-M_s) is independent of M_s, for all $t>s$. Additionally, if $t>s$ then $(M_t-M_s)\overset{d}{\sim}N(0,\nu(t)-\nu(s))$, since $\mathbb{E}\left[M_tM_s\right]=\min\left\{\nu(t),\nu(s)\right\}=\nu(s)$.

The derivation above can be used to find the one-step predictor (forecast) and its error function as well as recursively generate an exact simulation the Gaussian Markov process. To simulate the Gaussian Markov process at time t given time $s<t$ we need only generate a standard normal Z and perform the transformation

$$X_t=\Psi_t\left(\frac{X_s}{\Psi_s}+Z\sqrt{\nu(t)-\nu(s)}\right).$$

Example 8.1. To generate a sample paths of the Ornstein-Uhlenbeck (O-U) process $(X_{t_i})_{i=0}^n$, such that $t_i=i/n$, the algorithm above implies that we first generate $(Z_i)_{i=1}^n$ i.i.d. standard

normal random variables. Using the functions

$$\Psi(t) = \frac{\sigma e^{-\theta t}}{\sqrt{2\theta}} \text{ and } \nu(t) = e^{2\theta t} - 1,$$

the algorithm builds on itself inductively for $i = 1, \ldots, n$, so that

$$X_{t_i} = e^{-\theta/n} X_{t_{i-1}} + Z_i \frac{\sigma e^{-\theta \frac{i}{n}}}{\sqrt{2\theta}} \sqrt{e^{2\theta \frac{i}{n}} - e^{2\theta \frac{i-1}{n}}}.$$

Generating 500 replications of each sample path with the following parameters

$$\begin{aligned} X_0 &= 0.025 \\ \mu &= 0.05 \\ \theta &= 1.3 \\ \sigma &= 0.01, \end{aligned}$$

we obtain the following result.

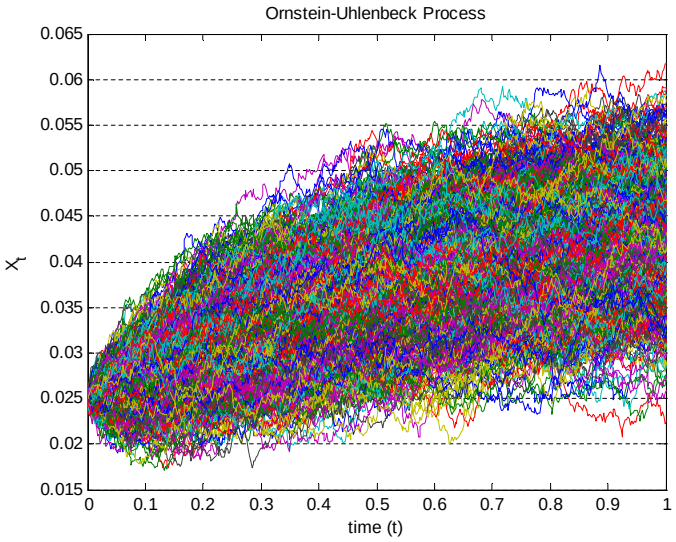


Figure 8.1: Simulations of an O-U Process

8.3 A Binomial Tree Representation of a Gaussian Markov Process

Binomial tree models have been individually and specifically developed for different stochastic processes. These trees are commonly used for determining optimal policies and valuations. To prove that a tree convergences to a continuous stochastic process is a demanding task. Like the derivations in Pollard [41], most convergence proofs require working on the metric space $D([0,T])$, the Skorohod topology. Furthermore, it is not uncommon for models to be proposed or used in practice without a proof of convergence to a specific stochastic process or when convergence proofs are provided, many derivations incorrectly use theorems like the Central Limit Theorem. In this section, we create a binomial tree representation that weakly converges to a continuous path Gaussian Markov process. To prove weak convergence, and to avoid working in a Skorohod topology, we adapt the Stochastic Central Limit Theorem by Anderson and Dobrić [2] to Hida's [19] definition of Gaussian Markov processes (Theorem 5.3).

The use of dynamic programming on trees to find optimal policies for options is well studied and applied. Given a tree, the value of an American type option is typically found by using the backward induction algorithm. As the name implies, this algorithm starts at the terminal node values of the tree and inductively works its way back to the first node while it evaluates the times for which the expected payoff is maximized. Dynamic programming on stochastic tree representations are also wildly used by economists and engineers in areas such as resource, inventory, operational management and capital budgeting. (Luenberger [30])

The most common tree representation is the random walk, which weakly converges to Brownian motion. Brownian motion is the most basic of all GM processes and the proof of weak convergence takes advantage of the independent and identically distributed (i.i.d.) Gaussian increments. (Pollard [41], Theorem 19) Pollard [41] shows that if a process has independent increments with weak regularity, then the tree can be shown to weakly convergence to a specified process.

8.3.1 Gaussian Markov Binomial Trees

In this section, we propose a binomial tree representation for GM processes. In Theorem 8.2 we state the results of the Anderson and Dobrić [2] Stochastic Central Limit Theorem, which establishes the three requirements which are needed to prove that the tree weakly converges to

a continuous path stochastic process. Proofs of these requirements are derived in three lemmas. Lemma 8.3 establishes that in each finite interval the random variables on a tree are eventually bounded in the limit. Lemma 8.4 provides a proof of weak convergence of the process. Lemma 8.5 is needed to assure that the tree process converges to a continuous path stochastic process. Lastly, Lemma 8.6 proves that the random processes of the tree converge to a specified continuous path Gaussian Markov process.

To begin, let $(X_t)_{t\in I}$ be a continuous path stochastic process, where I is an index set. For our purposes, $I = [0,\infty)$ or $[0,T]$. We say that the sequence $(X_{n,t})_{t\in I}$ converges weakly to $(X_t)_{t\in I}$ and we write

$$(X_{n,t})_{t\in I} \overset{w}{\Longrightarrow} (X_t)_{t\in I}$$

if

$$\int f\left((X_{n,t})_{t\in I}\right) dP \to \int f\left((X_t)_{t\in I}\right) dP$$

for all f bounded continuous on $C(I)$, where $C(I)$ is equipped with the usual $\|\quad\|_\infty$ norm.

Next, if $t_i = \nu^{-1}\left(\frac{i}{n}\right)$, this implies that $\nu(t_{i+1}) - \nu(t_i) = \frac{i}{n}$. We can define a binomial tree such that,

$$X_n \equiv (X_{n,t})_{t\in[0,T]} = \left(\frac{\Psi_n(t)}{\sqrt{n}} \sum_{i=1}^{[n\nu(t)]} \xi_i\right)_{t\in[0,T]}, \tag{8.3}$$

where $X_{n,0} = 0$, $\{\xi_i\}_{i=1}^{\infty}$ are i.i.d. with $P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}$, $\Psi_n(t) = \Psi\left(\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right)\right)$. We could have also chosen $\Psi_n(t) = \Psi\left(\frac{[nt]}{n}\right)$, however then time changes would not evolve at the same times as $n\nu(t)$, but the theorem below would still hold. Note that,

$$\|X_n\|_T = \sup_{t\in[0,T]} \left|\frac{\Psi_n(t)}{\sqrt{n}} \sum_{i=1}^{[n\nu(t)]} \xi_i\right| \leq \frac{M[n\nu(T)]}{\sqrt{n}} \leq M\nu(T)\sqrt{n} < \infty,$$

therefore $(X_{n,t})_{t\in[0,T]} \in \mathbb{B}([0,T], \|\quad\|_I)$, where $\mathbb{B}([0,T], \|\quad\|_I)$ is the set of all bounded functions on $[0,T]$ with respect to $\|\quad\|_I$. In what follows we assume that Ψ, ν are continuous on $[0,T]$.

The following theorem is an application of Theorem 5.5 and Theorem 4.2 from Anderson and Dobrić [2].

Theorem 8.2. *(Gaussian Markov Functional Central Limit Theorem (CLT))*

A process $(X_{n,t})_{t\in[0,T]}$ *(given in equation 8.3), converges weakly to* $(X_t)_{t\in[0,T]}$*, where* $(X_t)_{t\in[0,T]}$ *is a continuous path Gaussian Markov process with* $X_t = \Psi(t) M_t$*, and* $(M_t)_{t\in[0,T]}$ *is a Gaussian martingale with* $\mathbb{E}\left[M_t^2\right] = \nu(t)$*.*

By Theorem 5.5 in Anderson and Dobrić [2], to prove that

$$(X_{n,t})_{t\in[0,T]} \overset{w}{\Longrightarrow} (X_t)_{t\in[0,T]},$$

we must show:

a) Eventual boundedness, that is, (X_n) is *eventually bounded* if

$$\lim_{a\to\infty} \limsup_{n\to\infty} P\left(\|X_n\|_T > a\right) = 0, \tag{8.4}$$

where $\|\ \|_T = ess\sup_{s\in S} |g(s)|$.

b) Eventual Totally Boundedness, that is if $\Gamma(T)$ is the set of all finite partitions of the set T, then (X_n) is *eventually totally bounded* if

$$\forall \epsilon > 0, \quad \exists a \in \Gamma(T) : \limsup_{n\to\infty} P\left(\max_{A\in a} \omega(X_n, A) > \epsilon\right) < \epsilon, \tag{8.5}$$

where $\omega(\phi, A) = \sup\{|\Delta\phi(u,v)| \,|u,v \in A\}$, $A \subseteqq T$, $\phi \in \mathbb{R}^T$ and $\Delta\phi$ is the difference function defined by $\Delta\phi(u,v) = \phi(u) - \phi(v)$, $\forall u, v \in T$.

To prove that $(X_t)_{t\in[0,T]}$ has continuous sample paths by Theorem 4.2 in Anderson and Dobrić [2], we have to show Eventually Uniformly $\|\quad\|_\infty$-equicontinuity of the process (X_t). We say that (X_n) is *eventually uniformly ρ-equicontinuous,* where ρ is a pseudometric on I, if

$$\lim_{a\to 0} \limsup_{n\to\infty} P\left(\omega_\rho(X_n, a) > \epsilon\right) = 0, \ \forall \epsilon > 0, \tag{8.6}$$

where $\omega_\rho(\phi, a) = \sup\{|\Delta\phi(u,v)| \,|u,v \in T : \rho(u,v) < a\}$, $a > 0$. In our case,

$$\rho(f,g) \equiv \sup_{x\in[0,T]} |f(x) - g(x)| = \|f - g\|_\infty.$$

Throughout the proofs we will use

$$\sup_{t\in[0,T]}\left|\Psi\left(\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right)\right)\right| \leq \sup_{t\in[0,T]}|\Psi(t)| \equiv M,$$

which by continuity of Ψ implies $M < \infty$.

For easier reading the proof of Theorem 8.2 is broken into three Lemmas.

Lemma 8.3. *The process $(X_{n,t})_{t\in[0,T]}$ is eventually bounded.*

Proof. According to equation 8.4 we need to show that

$$\lim_{a\to 0}\limsup_{n\to\infty} P\left(\|X_n\|_T > a\right) = 0.$$

Since $M < \infty$,

$$\begin{aligned} P\left(\|X_n\|_T > a\right) &\leq P\left(\sup_{t\in[0,T]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{[n\nu(t)]}\xi_i\right| > \frac{a}{M}\right) \\ &\leq \frac{M\mathbb{E}\left|\sum_{i=1}^{[n\nu(T)]}\xi_i\right|}{a\sqrt{n}} \\ &\leq \frac{M\sqrt{\mathbb{E}\left(\sum_{i=1}^{[n\nu(T)]}\xi_i\right)^2}}{a\sqrt{n}} \\ &\leq \frac{M\sqrt{\nu(T)}}{a}, \end{aligned} \tag{8.7}$$

where inequality 8.7 follows from Doob's (Martingale) Maximal Inequality (see Theorem 5.9), and so

$$\lim_{a\to\infty}\limsup_{n\to\infty}\left\{\frac{M\sqrt{\nu(T)}}{a}\right\} = 0.$$

□

Lemma 8.4. *The process $(X_{n,t})_{t\in[0,T]}$ is eventually totally bounded.*

Proof. Let us start by dividing the interval $[0,T]$ into N equal sub-intervals

$$I_j = \left[\nu^{-1}\left((j-1)\frac{\nu(T)}{N}\right), \nu^{-1}\left(j\frac{\nu(T)}{N}\right)\right), j = 1,\ldots,N,$$

where N will be determined later, and assume $t, s \in [0,T]$ so that $s \leq t$. According to inequality 8.5 we need to show that for all $\epsilon > 0$,

$$\limsup_{n\to\infty} P\left(\max_j \sup_{t,s\in I_j} |X_{n,t} - X_{n,s}| > \epsilon\right) < \epsilon.$$

If

$$B_{\epsilon,n} \equiv \left(\max_{1\leq j\leq N} \sup_{t,s\in I_j} |X_{n,t} - X_{n,s}| > \epsilon\right),$$

then

$$\begin{aligned}
P\left(B_{\epsilon,n}\right) &= P\left(\max_{1\leq j\leq N}\max_{t,s\in I_j}\left|\left(\Psi_n(t) - \Psi_n(s)\right)\sum_{i=1}^{[n\nu(s)]}\xi_i + \Psi_n(t)\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \epsilon\sqrt{n}\right) \\
&\leq P\left(\max_{1\leq j\leq N}\max_{t,s\in I_j}\left|\left(\Psi_n(t) - \Psi_n(s)\right)\sum_{i=1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&\quad + P\left(\max_{1\leq j\leq N}\max_{t,s\in I_j}\left|\Psi_n(t)\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \qquad (8.8) \\
&\equiv I_n + II_n,
\end{aligned}$$

where to obtain inequality 8.8 we used $P\left(|X| + |Y| > \epsilon\right) \leq P\left(|X| > \frac{\epsilon}{2}\right) + P\left(|Y| > \frac{\epsilon}{2}\right)$. Set

$$D_n \equiv \max_{1\leq j\leq N}\max_{t,s\in I_j} |\Psi_n(t) - \Psi_n(s)|,$$

and observe

$$\begin{aligned}
I_n &= P\left(\max_{0\leq j\leq N}\max_{t,s\in I_j}\left|\left(\Psi_n(t) - \Psi_n(s)\right)\sum_{i=1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&\leq P\left(\sup_{s\in[0,T]}\left|\sum_{i=1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2D_n}\right) \\
&\leq \frac{2D_n\sqrt{\nu(T)}}{\epsilon}, \qquad (8.9)
\end{aligned}$$

where inequality 8.9 is a result of Doob's (Martingale) Maximal Inequality (see Theorem 5.9). Our goal is to show that D_n is small enough to yield a small I_n. Since Ψ is uniformly continuous on $[0,T]$, for all $\epsilon' > 0$ there exists $\delta > 0$ such that if $|x - y| < \delta$ then $|\Psi(x) - \Psi(y)| < \epsilon'$. By uniformly continuity and monotonicity of ν^{-1} on $[0, \nu(T)]$ for $\delta/2$, there is an n_0 so that if

$n \geq n_0$ then

$$\nu^{-1}\left(\nu\left(s\right)\right) - \nu^{-1}\left(\nu\left(s\right) - \frac{1}{n}\right) < \frac{\delta}{2},$$

which by monotonicity of ν is equivalent to $s - \frac{\delta}{2} < \nu^{-1}\left(\nu\left(s\right) - \frac{1}{n}\right)$ and therefore,

$$\begin{aligned} \nu^{-1}\left(\frac{[n\nu\left(t\right)]}{n}\right) - \nu^{-1}\left(\frac{[n\nu\left(s\right)]}{n}\right) &\leq t - \nu^{-1}\left(\nu\left(s\right) - \frac{1}{n}\right) \\ &< t - s + \frac{\delta}{2}. \end{aligned}$$

Consequently, if we set $\epsilon' = \epsilon^2/4\sqrt{\nu\left(T\right)}$ and choose N so that $\frac{\nu(T)}{N} < \delta/2$, then if $|x-y| < \frac{\nu(T)}{N} < \delta/2$ and $n \geq n_0$ we have, for $s < t$,

$$\nu^{-1}\left(\frac{[n\nu\left(t\right)]}{n}\right) - \nu^{-1}\left(\frac{[n\nu\left(s\right)]}{n}\right) < t - s + \frac{\delta}{2} < \delta, \tag{8.10}$$

and therefore

$$D_n = \max_j \max_{t,s \in I_j} \left|\Psi\left(\nu^{-1}\left(\frac{[n\nu\left(t\right)]}{n}\right)\right) - \Psi\left(\nu^{-1}\left(\frac{[n\nu\left(s\right)]}{n}\right)\right)\right| < \epsilon' = \frac{\epsilon^2}{4\sqrt{\nu\left(T\right)}},$$

yielding

$$\limsup_{n\to\infty} I_n \leq \limsup_{n\to\infty} \frac{2D_n\sqrt{\nu\left(T\right)}}{\epsilon} \leq \frac{\epsilon}{2}. \tag{8.11}$$

Consider II_n in inequality 8.8. Since

$$\max_{0\leq j\leq N} \max_{t\in I_j} |\Psi_n\left(t\right)| = \max_{t\in[0,T]} |\Psi_n\left(t\right)| \leq M,$$

then

$$
\begin{aligned}
II_n &= P\left(\max_{0\le j\le N}\max_{t,s\in I_j}\left|\Psi_n(t)\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&\le P\left(\max_{0\le j\le N}\max_{t,s\in I_j}\left|\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2M}\right) \equiv III_n.
\end{aligned}
$$

Set $Y_j \equiv \max_{t,s\in I_j}\left|\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right|$ and observe that $Y_j - s$ are i.i.d. Our goal is to transform III_n into a form which is suitable for the application of the Central Limit Theorem (CLT).

$$
\begin{aligned}
III_n &= 1 - P^N\left(\max_{t,s\in I_j}\left|\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| \le \frac{\epsilon\sqrt{n}}{2M}\right) \\
&= 1 - P^N\left(\max_{t,s\in I_j}\left|\sum_{i=\left[j\frac{n\nu(T)}{N}\right]+1}^{[n\nu(t)]}\xi_i - \sum_{i=\left[j\frac{n\nu(T)}{N}\right]+1}^{[n\nu(s)]}\xi_i\right| \le \frac{\epsilon\sqrt{n}}{2M}\right) \\
&= 1 - \left(1 - P\left(\max_{t,s\in I_j}\left|\sum_{i=\left[j\frac{n\nu(T)}{N}\right]+1}^{[n\nu(t)]}\xi_i - \sum_{i=\left[j\frac{n\nu(T)}{N}\right]+1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2M}\right)\right)^N \\
&\le 1 - \left(1 - P\left(\max_{t\in I_j}\left|\sum_{i=\left[j\frac{n\nu(T)}{N}\right]+1}^{[n\nu(t)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{4M}\right)\right)^N \\
&\le 1 - \left(1 - 2P\left(\left|\sum_{i=1}^{\left[\frac{n\nu(T)}{N}\right]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{4M}\right)\right)^N \qquad (8.12) \\
&\le 1 - \left(1 - 4P\left(\sum_{i=1}^{\left[\frac{n\nu(T)}{N}\right]}\xi_i > \frac{\epsilon\sqrt{n}}{4M}\right)\right)^N \qquad (8.13) \\
&= 1 - \left(1 - 4P\left(\frac{1}{\sqrt{\left[\frac{n\nu(T)}{N}\right]}}\sum_{i=1}^{\left[\frac{n\nu(T)}{N}\right]}\xi_i > \frac{\epsilon\sqrt{n}}{4M\sqrt{\left[\frac{n\nu(T)}{N}\right]}}\right)\right)^N \\
&\le 1 - \left(1 - 4P\left(\frac{1}{\sqrt{\left[\frac{n\nu(T)}{N}\right]}}\sum_{i=1}^{\left[\frac{n\nu(T)}{N}\right]}\xi_i > \frac{\epsilon}{4M}\sqrt{\frac{N}{\nu(T)}}\right)\right)^N,
\end{aligned}
$$

where inequality 8.12 is a result of Levy's Lemma (see Theorem 5.10) and inequality 8.13 is by the symmetry of $\{\xi_i\}_{i=1}^{\infty}$. By the Central Limit Theorem, we have

$$\limsup_{n\to\infty} II_n \quad \leq \quad 1-\left(1-4P\left(N\left(0,1\right)>\frac{\epsilon}{4M}\sqrt{\frac{N}{\nu\left(T\right)}}\right)\right)^N.$$

Since

$$P\left(N\left(0,1\right)>x\right)<\frac{1}{x\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ x>0 \tag{8.14}$$

and using the Taylor series expansion, $1-\frac{x}{2}\geq e^{-x}$ if $0\leq x\leq 1$, for N large enough, we obtain

$$\begin{aligned}\limsup_{n\to\infty} II_n &< 1-\left(1-\frac{16M}{\epsilon\sqrt{2\pi}}\sqrt{\frac{\nu\left(T\right)}{N}}\exp\left\{-\frac{N\epsilon^2}{32M^2\nu\left(T\right)}\right\}\right)^N \\ &< 1-\exp\left(\frac{-32M}{\epsilon\sqrt{2\pi}}\sqrt{N\nu\left(T\right)}\exp\left\{-\frac{N\epsilon^2}{32M^2\nu\left(T\right)}\right\}\right).\end{aligned}$$

Penultimately, since $\sqrt{N}\exp\left(-Nc\right)\to 0$, when $c>0$ as $N\to\infty$, therefore there exists an N such that $1-\exp\left(\sqrt{N}\exp\left\{-Nc\right\}\right)\leq\epsilon/2$,

$$\limsup_{n\to\infty} II_n<\frac{\epsilon}{2}. \tag{8.15}$$

Finally, combining inequalities 8.11 and 8.15 we see that

$$\limsup_{n\to\infty} P\left(B_{\epsilon,n}\right) \quad < \quad \epsilon.$$

So we have proved that

$$\left(X_{n,t}\right)_{t\in[0,T]} \overset{w}{\Longrightarrow} \left(X_t\right)_{t\in[0,T]}.$$

□

Lemma 8.5. *The process $\left(X_{n,t}\right)_{t\in[0,T]}$ is eventually uniformly $\|\quad\|_\infty$-equicontinuous.*

Proof. According to equation 8.6 we need to prove that for any $\epsilon>0$,

$$\lim_{a\to 0}\limsup_{n\to\infty} P\left(\sup_{|t-s|<a}|X_{n,t}-X_{n,s}|>\epsilon\right) \quad = \quad 0.$$

Without loss of generality, let $s < t \leq T$ and set

$$\omega(a) \equiv \sup_{|t-s|<a} |\nu(t) - \nu(s)|.$$

Our goal is to rewrite the supremum in a form which allows for analysis similar to the proof of eventually totally bounded (Lemma 8.4). To achieve our goal, we need to partition $[0,T]$ into appropriate sub-intervals.

Let j_0 be chosen so that $[nj_0\omega(a)] = [n\nu(T)]$. Since

$$\left[n\omega(a)\left[\frac{\nu(T)}{\omega(a)}\right]\right] \leq [n\nu(T)] = [nj_0\omega(a)]$$

we have $j_0 \geq \left[\frac{\nu(T)}{\omega(a)}\right]$ and since

$$\left[n\omega(a)\left(\left[\frac{\nu(T)}{\omega(a)}\right]+1\right)\right] \geq [n\nu(T)] = [n\omega(a)\, j_0]$$

we have $j_0 \leq \left[\frac{\nu(T)}{\omega(a)}\right] + 1$. Now we can define our sub-intervals by

$$I_j \equiv \{t \in [0,T] \mid [nj\omega(a)] < [n\nu(t)] \leq [n(j+1)\omega(a)]\},$$

and, for a fixed $\epsilon > 0$, we have

$$B_{a,n} \equiv \left(\sup_{|t-s|<a} |X_{n,t} - X_{n,s}| > \epsilon\right) = \left(\max_j \max_{s\in I_j, |t-s|<a} |X_{n,t} - X_{n,s}| > \epsilon\right). \quad (8.16)$$

Similarly to Lemma 8.4,

$$\begin{aligned}
P(B_{a,n}) &= P\left(\max_j \max_{s\in I_j,|t-s|<a} \left|(\Psi_n(t) - \Psi_n(s)) \sum_{i=1}^{[n\nu(s)]} \xi_i + \Psi_n(t) \sum_{i=[n\nu(s)]+1}^{[n\nu(t)]} \xi_i\right| > \epsilon\sqrt{n}\right) \\
&\leq P\left(\max_j \max_{s\in I_j,|t-s|<a} \left|(\Psi_n(t) - \Psi_n(s)) \sum_{i=1}^{[n\nu(s)]} \xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&\quad + P\left(\max_j \max_{s\in I_j,|t-s|<a} \left|\Psi_n(t) \sum_{i=[n\nu(s)]+1}^{[n\nu(t)]} \xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&= P(\Xi_n) + P(\Theta_n).
\end{aligned}$$

We will first compute $P(\Xi_n)$. Recall

$$\max_j \max_{s\in I_j,|t-s|<a} \Psi_n(t) = \sup_{t\in[0,T]} \Psi_n(t) \le M,$$

and set

$$D_n \equiv \sup_{|t-s|<a} |\Psi_n(t) - \Psi_n(s)|.$$

Let $\epsilon'' > 0$ and set $\epsilon' = \frac{\epsilon}{2\sqrt{\nu(T)}}\epsilon''$. Since Ψ is uniformly continuous on $[0,T]$, there exists $\delta > 0$ so that if $|x-y| < \delta$, then $|\Psi(x) - \Psi(y)| < \epsilon'$. If $a < \delta/2$, then by inequality 8.10,

$$\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right) - \nu^{-1}\left(\frac{[n\nu(s)]}{n}\right) < t - s + \frac{\delta}{2} < a + \frac{\delta}{2} < \delta,$$

and consequently $D_n < \epsilon'$, implying

$$\begin{aligned}
P(\Xi_n) &= P\left(\sup_{|t-s|<a}\left|(\Psi_n(t) - \Psi_n(s))\sum_{i=1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right) \\
&\le P\left(\sup_{s\in[0,T]}\left|\sum_{i=1}^{[n\nu(s)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2D_n}\right) \\
&\le \frac{2D_n\sqrt{\nu(T)}}{\epsilon} \qquad (8.17) \\
&< \frac{2\epsilon'\sqrt{\nu(T)}}{\epsilon} \\
&= \epsilon'',
\end{aligned}$$

where inequality 8.17 is a result of Doob's (Martingale) Maximal Inequality (see Theorem 5.9), yielding

$$\lim_{a\to 0}\limsup_{n\to\infty} P(\Xi_n) = 0. \qquad (8.18)$$

Set

$$\eta_j \equiv [nj\omega(a)],\ j = 0, 1, \ldots, j_0$$

and

$$\Theta_{n,j} \equiv \left(\max_{s\in I_j,|t-s|<a}\left|\Psi_n(t)\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \frac{\epsilon\sqrt{n}}{2}\right).$$

Observe that

$$\Theta_n = \bigcup_{j=1}^{j_0} \Theta_{n,j}.$$

If $t \in I_j$, then

$$\begin{aligned}
\Theta_{n,j} &\subseteq \left(\max_{s \in I_j, |t-s|<a} \left| \sum_{i=\eta_j+1}^{[n\nu(t)]} \xi_i - \sum_{i=\eta_j+1}^{[n\nu(s)]} \xi_i \right| > \frac{\epsilon\sqrt{n}}{2M} \right) \\
&\subseteq \left(\max_{s \in I_j, |t-s|<a} \left\{ \left| \sum_{i=\eta_j+1}^{[n\nu(t)]} \xi_i \right| + \left| \sum_{i=\eta_j+1}^{[n\nu(s)]} \xi_i \right| \right\} > \frac{\epsilon\sqrt{n}}{2M} \right) \\
&\subseteq \left(\max_{\eta_j < k \leq \eta_{j+1}} \left| \sum_{i=\eta_j+1}^{k} \xi_i \right| > \frac{\epsilon\sqrt{n}}{4M} \right), \qquad (8.19)
\end{aligned}$$

and if $t \in I_{j+1}$ (and if $j+2 \leq j_0$), then

$$\begin{aligned}
\Theta_{n,j} &\subseteq \left(\max_{s \in I_j, |t-s|<a} \left| \sum_{i=[ns]+1}^{\eta_{j+1}} \xi_i + \sum_{i=\eta_{j+1}+1}^{[n\nu(t)]} \xi_i \right| > \frac{\epsilon\sqrt{n}}{2M} \right) \\
&\subseteq \left(\max_{s \in I_j, |t-s|<a} \left\{ \left| \sum_{i=[n\nu(s)]+1}^{\eta_{j+1}} \xi_i \right| + \left| \sum_{i=\eta_{j+1}+1}^{[n\nu(t)]} \xi_i \right| \right\} > \frac{\epsilon\sqrt{n}}{2M} \right) \\
&\subseteq \left(\max_{\eta_j < k \leq \eta_{j+1}} \left| \sum_{i=k}^{\eta_{j+1}} \xi_i \right| + \max_{\eta_{j+1} < k \leq \eta_{j+2}} \left| \sum_{i=\eta_{j+1}+1}^{k} \xi_i \right| > \frac{\epsilon\sqrt{n}}{2M} \right) \\
&= \left(\left(\max_{\eta_j < k \leq \eta_{j+1}} \left| \sum_{i=\eta_j+1}^{\eta_{j+1}} \xi_i - \sum_{i=\eta_j+1}^{k} \xi_i \right| \right. \right. \\
&\quad \left. \left. \vee \max_{\eta_{j+1} < k \leq \eta_{j+2}} \left| \sum_{i=\eta_{j+1}+1}^{k} \xi_i \right| \right) > \frac{\epsilon\sqrt{n}}{4M} \right) \\
&\subseteq \left(\left(\max_{\eta_j < k \leq \eta_{j+1}} \left(\left| \sum_{i=\eta_j+1}^{\eta_{j+1}} \xi_i \right| + \left| \sum_{i=\eta_j+1}^{k} \xi_i \right| \right) \right. \right. \\
&\quad \left. \left. \vee \max_{\eta_{j+1} < k \leq \eta_{j+2}} \left| \sum_{i=\eta_{j+1}+1}^{k} \xi_i \right| \right) > \frac{\epsilon\sqrt{n}}{4M} \right) \\
&\subseteq \left(\left(2 \max_{\eta_j < k \leq \eta_{j+1}} \left| \sum_{i=\eta_j+1}^{k} \xi_i \right| \vee \max_{\eta_{j+1} < k \leq \eta_{j+2}} \left| \sum_{i=\eta_{j+1}+1}^{k} \xi_i \right| \right) > \frac{\epsilon\sqrt{n}}{4M} \right). \qquad (8.20)
\end{aligned}$$

As before our goal is to bound $P(\Theta_n)$ by an expression for which we can apply CLT. Combining [8.19] and [8.20], it follows that

$$
\begin{aligned}
P(\Theta_n) &\leq P\left(\max_j\left(2\left(\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|\vee\max_{\eta_{j+1}<k\leq\eta_{j+2\wedge j_0}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|\right)\right.\right.\\
&\quad\left.\left.\vee\max_{\eta_{j+1}<k\leq\eta_{j+2\wedge j_0}}\left|\sum_{i=\eta_{j+1}+1}^{k}\xi_i\right|\right)>\frac{\epsilon\sqrt{n}}{4M}\right)\\
&\leq P\left(\max_j\left(\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|\vee\max_{\eta_{j+1}<k\leq\eta_{j+2\wedge j_0}}\left|\sum_{i=\eta_{j+1}+1}^{k}\xi_i\right|\right)>\frac{\epsilon\sqrt{n}}{8M}\right)\\
&\leq P\left(\max_j\left(\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|\vee\max_j\max_{\eta_{j+1}<k\leq\eta_{j+2\wedge j_0}}\left|\sum_{i=\eta_{j+1}+1}^{k}\xi_i\right|\right)>\frac{\epsilon\sqrt{n}}{8M}\right)\\
&= P\left(\max_j\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|>\frac{\epsilon\sqrt{n}}{8M}\right)\\
&= 1-P\left(\max_j\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|\leq\frac{\epsilon\sqrt{n}}{8M}\right).
\end{aligned}
$$

For fixed n, $Z_j\equiv\max_{\eta_j<k\leq\eta_{j+1}}\left|\sum_{i=\eta_j+1}^{k}\xi_i\right|, j=0,\ldots,j_0$ are i.i.d., and so we have

$$
\begin{aligned}
P(\Theta_n) &\leq 1-P^{j_0}\left(\max_{0<k\leq\eta_1,|t-s|<a}\left|\sum_{i=1}^{k}\xi_i\right|\leq\frac{\epsilon\sqrt{n}}{8M}\right)\\
&= 1-\left(1-P\left(\max_{0<k\leq\eta_1,|t-s|<a}\left|\sum_{i=1}^{k}\xi_i\right|>\frac{\epsilon\sqrt{n}}{8M}\right)\right)^{j_0}\\
&\leq 1-\left(1-4P\left(\sum_{i=1}^{\eta_1}\xi_i>\frac{\epsilon\sqrt{n}}{4M}\right)\right)^{j_0}, \qquad (8.21)
\end{aligned}
$$

where inequality 8.21 is a result of Levy's lemma and symmetry (see Theorem 5.10). Since $\eta_1=[n\nu(a)]$ and $j_0\leq\left[\frac{\nu(T)}{\omega(a)}\right]+1$,

$$
\begin{aligned}
P(\Theta_n) &\leq 1-\left(1-4P\left(\frac{1}{\sqrt{[n\nu(a)]}}\sum_{i=1}^{[n\nu(a)]}\xi_i>\frac{\epsilon\sqrt{n}}{8M\sqrt{[n\nu(a)]}}\right)\right)^{\left[\frac{\nu(T)}{\nu(a)}\right]+1}\\
&\leq 1-\left(1-4P\left(\frac{1}{\sqrt{[n\nu(a)]}}\sum_{i=1}^{[n\nu(a)]}\xi_i>\frac{\epsilon}{8M\sqrt{\nu(a)}}\right)\right)^{\left[\frac{\nu(T)}{\nu(a)}\right]+1}.
\end{aligned}
$$

By CLT and inequality 8.14

$$\begin{aligned}
\lim_{n\to\infty} P(\Theta_n) &\leq 1 - \left(1 - 4P(N(0,1)) > \frac{\epsilon}{8M\sqrt{\nu(a)}}\right)^{\left[\frac{\nu(T)}{\nu(a)}\right]+1} \\
&\leq 1 - \left(1 - \frac{32M\sqrt{\nu(a)}}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{\epsilon^2}{128M^2\nu(a)}\right)\right)^{\left[\frac{\nu(T)}{\nu(a)}\right]+1} \\
&< 1 - \exp\left\{-\left(\frac{64M\sqrt{\nu(a)}}{\epsilon\sqrt{2\pi}}\right)\left(\left[\frac{\nu(T)}{\nu(a)}\right]+1\right)\exp\left(-\frac{\epsilon^2}{128M^2\nu(a)}\right)\right\}.
\end{aligned}$$

Finally since $\lim_{a\to 0}\nu(a) = 0$,

$$\lim_{a\to 0}\lim_{n\to\infty} P(\Theta_n) = 1 - e^0 = 0. \tag{8.22}$$

From equations 8.18 and 8.22

$$\lim_{a\to 0}\limsup_{n\to\infty} P\left(\sup_{|t-s|<a}\left|\frac{1}{\sqrt{n}}\sum_{i=[n\nu(s)]+1}^{[n\nu(t)]}\xi_i\right| > \epsilon\right) \leq 0,$$

establishing that $(X_t)_{t\in[0,T]}$ has continuous sample paths and thereby concluding the proof of Theorem 8.2. □

Lemma 8.6. *The process $(X_{n,t})_{t\in[0,T]}$ convergences to the GM process $(\Psi(t) M_t)_{t\in[0,T]}$.*

Proof. Now that we have shown that the sequence converges to a Gaussian stochastic process, we need to show that both the mean and covariance of the discrete process $(X_{n,t})_{t\in[0,T]}$ converges to that of the continuous time Gaussian Markov process. Since $P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}$ it is trivial to see that the mean of the process $(X_{n,t})_{t\in[0,T]}$ is zero and therefore

$$\lim_{n\to\infty}\mathbb{E}[X_{n,t}] = 0.$$

Therefore we only need to prove that

$$\lim_{n\to\infty}\mathbb{E}[X_{n,t}X_{n,s}] = \Psi(t)\Psi(s)\nu(t\wedge s).$$

Without loss of generality let $s < t$. Then

$$\begin{aligned}
\mathbb{E}\left[X_{n,t}X_{n,s}\right] &= \mathbb{E}\left[\frac{\Psi_n(t)\Psi_n(s)}{n}\sum_{i=1}^{[n\nu(t)]}\xi_i\sum_{i=1}^{[n\nu(s)]}\xi_i\right] \\
&= \frac{\Psi_n(t)\Psi_n(s)}{n}\sum_{i=1}^{[n\nu(t)]}\sum_{j=1}^{[n\nu(s)]}\mathbb{E}\left[\xi_i\xi_j\right] \\
&= \frac{\Psi_n(t)\Psi_n(s)}{n}\sum_{i=1}^{[n\nu(s)]}\mathbb{E}\left[\xi_i^2\right] \qquad (8.23) \\
&= \left(\Psi_n(t)\Psi_n(s)\right)\frac{[n\nu(s)]}{n}.
\end{aligned}$$

Recall that $\Psi_n(t) = \Psi\left(\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right)\right)$ and therefore

$$\begin{aligned}
\lim_{n\to\infty}\Psi\left(\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right)\right) &= \lim_{n\to\infty}\Psi\left(\nu^{-1}\left(\frac{[n\nu(t)]}{n}\right)\right) \\
&= \Psi(t).
\end{aligned}$$

Additionally,

$$\lim_{n\to\infty}\frac{[n\nu(s)]}{n} = \nu(s).$$

Combining these results, we obtain the covariance of the Gaussian Markov process

$$\lim_{n\to\infty}\left(\Psi_n(t)\Psi_n(s)\right)\frac{[n\nu(s)]}{n} = \Psi(t)\Psi(s)\nu(s).$$

□

8.3.2 Example:A Binomial Tree of an Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck (O-U) process $(X_t)_{t\in[0,T]}$ is a common model in many applications for modeling mean-reversion. For example, in finance it is widely used to for interest rate derivative pricing (where it is called the Vasicek model) and in physics it governs the length dynamics of over-damped springs under thermal fluctuations. The O-U process satisfies the stochastic differential equation (sde)

$$dX_t = \theta(\mu - X_t)\,dt + \sigma dB_t,$$

where $(B_t)_{t\in[0,T]}$ is a standard Brownian motion process and $\mu > 0$, $\theta > 0$, and $\sigma > 0$ are

constants.

The solution to the sde is the process

$$X_t = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma e^{-\theta t} \int_0^t e^{\theta s} dB_s.$$

Using the Itô isometry, example 5.18 established that the stochastic term, $\sigma e^{-\theta t} \int_0^t e^{\theta s} dB_s$, is a GM process with

$$\Psi(t) = \frac{\sigma e^{-\theta t}}{\sqrt{2\theta}} \text{ and } \nu(t) = e^{2\theta t} - 1.$$

Let us assume that interest rates evolve according to the Vasicek model with the following parameters

$$\begin{aligned} X_0 &= 0.025 \\ \mu &= 0.05 \\ \theta &= 1.3 \\ \sigma &= 0.01. \end{aligned}$$

We can implement equation 8.3, in stages to show how each component of the Gaussian Markov process is influencing the properties of the tree. In the following example, we inductively generate a recombining tree with $n = 25$ stages.

By the definition of Brownian motion

$$\Psi_n(t) = 1, \forall t \text{ and } \nu(t) = t.$$

Plugging into equation 8.3, this yields the familiar random walk which converges to Brownian motion:

$$X_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \xi_i.$$

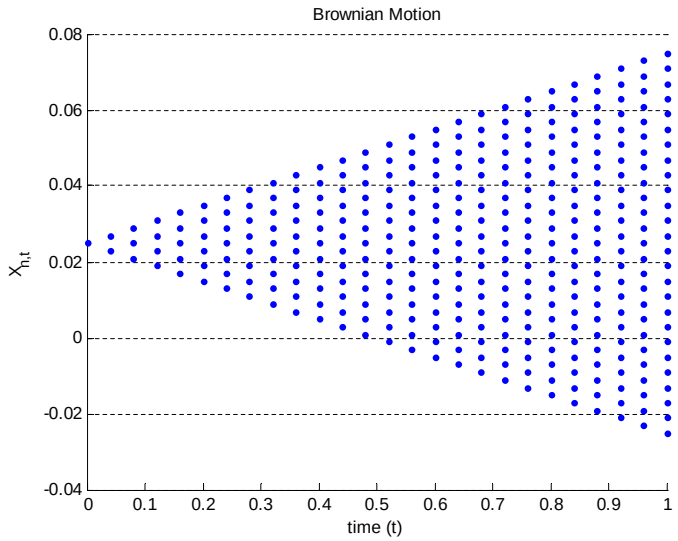


Figure 8.2: Random Walk

In Chapter 5 we saw that all Gaussian martingales are time changes of Brownian motion (Lemma 5.8) . Equation 8.3 establishes that time evolves such that $t_i = \nu^{-1}\left(\frac{i}{n}\right)$. In the Ornstein-Uhlenbeck process, the underlying martingale has variance $\nu(t) = e^{2\theta t} - 1$ and therefore time evolves as

$$t_i = \frac{\ln\left(\frac{i}{n} + 1\right)}{2\theta},\ i = 0, 1, \ldots, n.$$

This gives the underlying martingale the representation:

$$X_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\nu(t)]} \xi_i.$$

Notice that it is the non-linear scaling of time that results in the recombination of nodes in the tree. Without scaling time as an inverse of the martingale's variance, the process would not have an equal probability to go up or down for each n. This would destroy the recombining structure which is critical to computation and application of the tree.

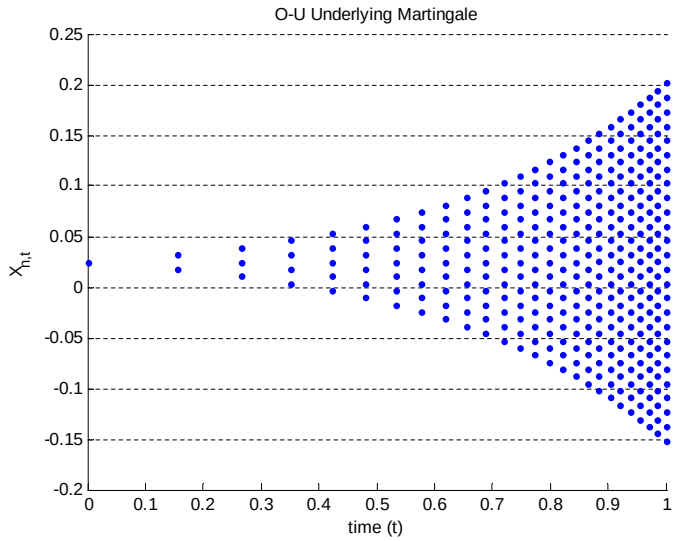


Figure 8.3: Binomial Tree for the Underlying Martingale of an O-U Process

The addition of the dependence structure, controlled by Ψ, gives the tree of a centered Gaussian Markov process:

$$X_{n,t} = \frac{\Psi_n(t)}{\sqrt{n}} \sum_{i=1}^{[n\nu(t)]} \xi_i,$$

such that

$$\Psi_n(t) = \frac{\sigma e^{-\theta t}}{\sqrt{2\theta}}.$$

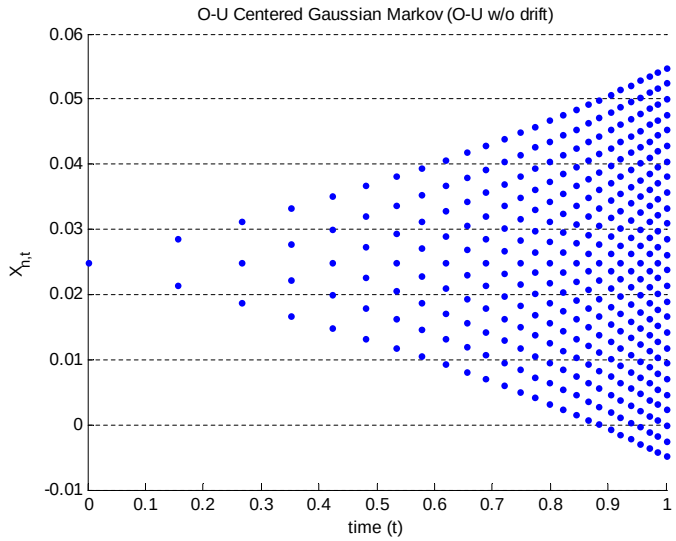


Figure 8.4: Binomial Tree for the Centered Gaussian Markov of the O-U Process

Lastly, the deterministic drift, which skews the tree toward its long term mean is needed to obtain the final representation of the binomial tree for the Ornstein-Uhlenbeck process:

$$X_{n,t} = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \frac{\Psi_n(t)}{\sqrt{n}} \sum_{i=1}^{[n\nu(t)]} \xi_i.$$

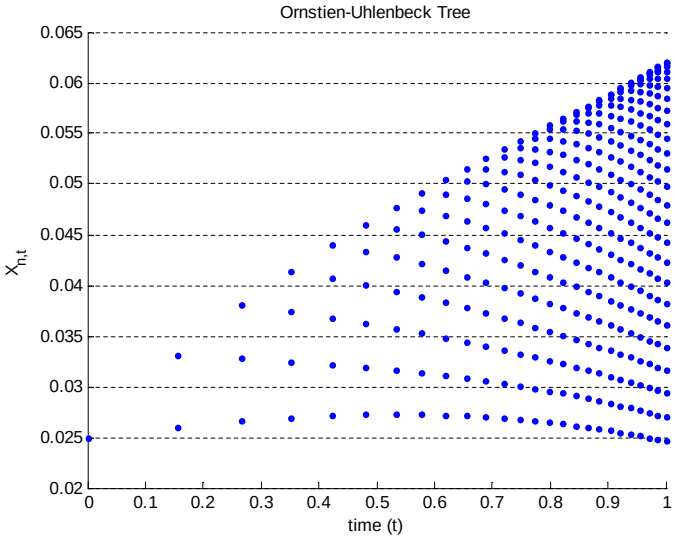


Figure 8.5: Binomial Tree Representation of an O-U Process

The O-U random walk above can be seen to have similar shape and characteristics to figure

8.1, which was generated with the same parameters.

Mining and harvesting are areas where the stochastic fluctuations of prices can have a significant effect on a company's bottom line. The determination of the optimal times to extract a resource or harvest a crop is an example of a real option. While many commodities like corn, oil or gold are typically modeled by a geometric Brownian motion (or the Bachelier model), commodities tend to have long-range dependence which make Brownian motion a poor candidate. (Mandelbrot [31]) Since Gaussian Markov processes can adjust their dependence structure through different Ψ functions and allow for heteroscedasticity by choosing ν, the GM binomial tree may be able to be used to improve company profits.

One such example is given in Luenberger [30] in example 12.8, where a company wants to extract gold from a mine. In this example, extraction costs depend on the amount of gold mined and remaining. Since mining in one period affects future mining costs, to maximize profits, the stochastic fluctuations of the gold prices must be taken into account. While Luenberger [30] does not reveal the model for his gold price tree model, the evaluation of the optimal extraction amounts and times is explicitly found using backward induction. This example shows how a GM processes tree could be used to evaluate real options and find optimal risk management strategies.

8.3.3 Example:Derivative Pricing and Risk Management with the Gaussian Markov Tree

Let us assume that an institution wants to issue a \$100 par valued 1 year variable rate bond that is currently paying $r_0 = 4.0\%$. This institution believes that rates are going to tend to $\mu = 5\%$ in the long term, however rates have been volatile with $\sigma = 1\%$.

If the institution is afraid that rates will fall or remain below 4% and wants to protect the bond from low yields, the institution could protect itself by purchasing an American style *interest rate put option* with a strike price $K = 4\%$. If the institution believes that short rates, r, are evolving according to the Vasicek model (which is Ornstein-Uhlenbeck process) with say $\theta = 1.3$, the optimal execution times and price of the put option can be determined by finding

$$\max_{0 \le \tau \le 1} \$100 \mathbb{E} e^{-\int_0^\tau r_s ds} \left(K - r_\tau\right)^+,$$

where τ is the exercise time.

In this application, the determination of policy through Dynamic Programming on the payoff

tree has a restriction which is typically only present in interest rate products. Indeed, most equity models assume that interest rate, r, is fixed; this assumption makes the term $e^{-\int_0^\tau r_s ds}$ deterministic and therefore, it can be removed from the expected value. In this situation, the payoff at each node in the O-U tree, and therefore the derivative payoff tree will remain recombining. However since interest rates are stochastic in this example, the discount factor, $e^{-\int_0^\tau r_s ds}$, is a random variable which is path dependent. The path dependency causes the derivative payoff tree to be non-recombining. To address this issue, we can use the Itô isometry to find the distribution of the discount factor using the continuous model. Over the time period $[0, T]$, the discount factor has distribution

$$\begin{aligned}\bar{r} &\equiv \int_0^\tau r_s ds \\ &\stackrel{d}{=} N\left(\frac{1}{T}\left[\frac{r_0}{\theta}\left(1-e^{-\theta T}\right)+\mu T+\frac{\mu}{\theta}\left(e^{-\theta T}-1\right), \frac{\sigma^2}{\theta^2 T^2}\left[T-\frac{2}{\theta}\left(1-e^{-\theta T}\right)+\frac{1}{2\theta}\left(1-e^{-2\theta T}\right)\right]\right]\right).\end{aligned}$$

Substituting in the parameters and assuming a contract length of $T = 1$,

$$\bar{r} \stackrel{d}{=} N\left(4.44\%, 0.0014\%\right).$$

The above indicates that the discounting factor plays has a small effect on the price. Therefore, we can keep our recombining tree by implementing the approximation

$$\max_{0\leq\tau\leq 1} \$100 e^{-\bar{r}\tau}\mathbb{E}\left(K-r_\tau\right)^+.$$

Application of the backward induction algorithm requires a comparison of the payoff at each node, $f\left(r_{t_i}\right) = \left(K - r_{t_i}\right)^+$ to the value of the option in the future, $V_{t_{i+1}}$. To start the backwards induction algorithm we need to compute, $V_{t_n} = f\left(r_{t_n}\right)$ for each of the $n+1$ nodes. Since the Vasicek model is constructed under the risk-neutral measure, the risk neutral probabilities, q, of moving up or down are fixed, by construction, as $q = 1/2$ for each branch in the tree. This means the expected value of the option if held from time t_i to the next time t_{i+1} is $V^*_{t_{i+1}} = \frac{1}{2}V^{up}_{t_{i+1}} + \frac{1}{2}V^{down}_{t_{i+1}}$ for each node. The value of the option at any node is determined by comparing the discounted expected value of the two successive nodes to the value of the option at the current node. If $f\left(r_{t_i}\right)$ is greater than $\left(1+\bar{r}\Delta t_i\right)^{-1} V^*_{t_{i+1}}$, then we exercise the option at time t_i, otherwise we hold onto the option. This means the optimal policy and value of the option is given by $V_{t_i} = \max\left(f\left(r_{t_i}\right), \left(1+\bar{r}\Delta t_i\right)^{-1} V^*_{t_{i+1}}\right)$. The algorithm is repeated inductively for all

$i = n-1, n-2, \ldots, 1, 0$ to determine value of the option at time zero, V_{t_0} (see Luenberger [30], Chapter 12 for details on the backwards induction algorithm and option pricing).

Implementation of the backwards induction algorithm with $n = 500$ gives an expected price of \$0.08 and a 99% confidence interval of $(0.0848, 0.0849)$. In figure 8.6, we show the tree for $n = 100$ periods. In this figure, the analysis shows the optimal interest rates (in green) that the institution should exercise the put option. The analysis provides the institution with both the price to pay for the insurance policy as well as the optimal exercise barrier of rates and times to execute the put option.

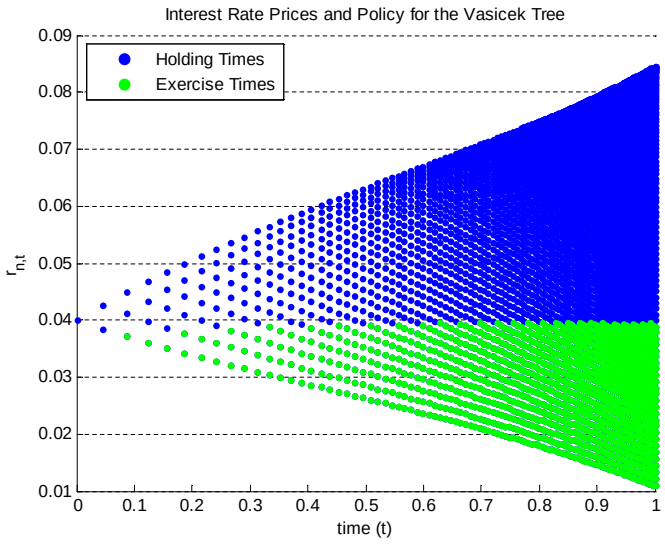


Figure 8.6: Binomial Tree Interest Rate Put Option Policy

The figure below shows the corresponding payoff for the "Exercise Times" along with the shares of stocks and bonds that replicate the payoff of the put option. Option Pricing Theory establishes that the replicating portfolio for the option value takes the form,

$$V_{t_i} = \phi_{t_i} S_{t_i} + \psi_{t_i} B_{t_i},$$

where

$$\phi_{t_i} = \frac{V^{up}_{t_{i+1}} - V^{down}_{t_{i+1}}}{r^{up}_{t_{i+1}} - r^{down}_{t_{i+1}}}$$

and

$$\psi_{t_i} = (1 + \bar{r}\Delta t_i)^{-1} \left(V\left(r^{up}_{t_{i+1}}\right) - \phi_{t_i} r_{t_i} \right).$$

Computation of the shares for the replicating portfolio is important for hedging risk and capturing any market inefficiencies.

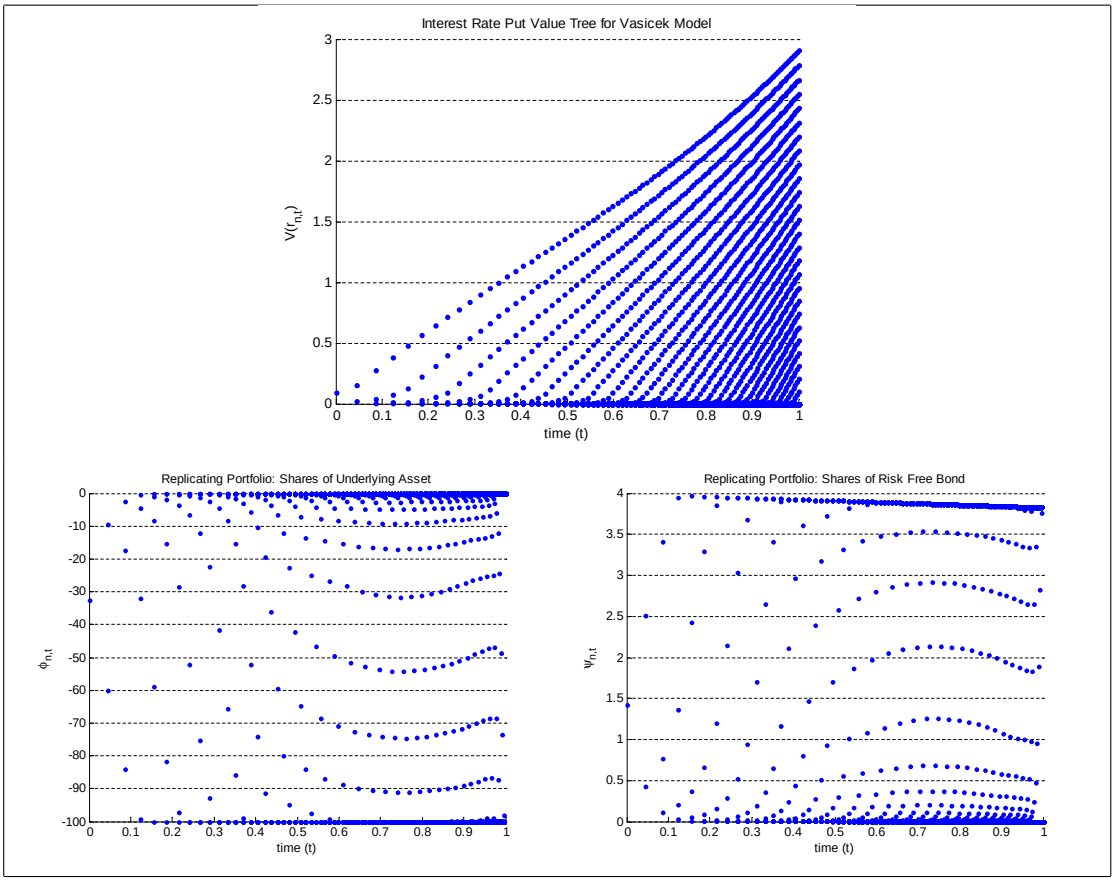


Figure 8.7: Replicating Portfolio/Policy for Interest Rate Put

8.4 Summary of Results

In this chapter we introduced methodologies to discretize GM processes. These methods allow for the evaluation of complex contracts through simulation or binomial tree construction. The most significant contribution of this chapter is the derivation of a recombining binomial tree that converges weakly to a continuous path stochastic Gaussian Markov process. The use of the Anderson and Dobrić [2] Stochastic Central Limit Theorem, coupled with Levy's lemma and Gaussian martingale properties enabled simpler convergence proofs than working in the Skorohod topology. We saw that the total boundedness and the ρ-equicontinuity requirements of the Stochastic Central Limit Theorem have a delicate balance which makes the choice of the intervals to be analyzed critical to the derivation of the convergence proofs.

Unlike the trees of fractional Brownian motion, we have shown that the GM tree has a general form for all continuous path GM processes that always recombines, making it fast and easy to generate since the number of points at each time step increases linearly with n. The recombining tree is a result of the proper time scaling, due to the variance of the underlying martingale. The recombining tree is critical to real world applications, since trees that do not recombine grow exponentially. Additionally, the application of dynamic programming to recombining binomial trees is well understood, exact and fast.

In fiance, Gaussian Markov processes are popular tools in modeling, however, without tree representations that converge weakly, many applications require approximate and/or slow methods to determine price and risk management policies. Our result accommodates these models and allows for accurate and fast analysis of many complex contracts or systems where no closed form solutions can be explicitly derived. For example, the binomial tree models of GM processes are crucial for dynamic optimal policies in pumping and valuing mines and wells, resource management and pricing American put options on financial products. With these results, decisions (and policies) for optimal pricing, risk management and hedging can be obtained by analyzing the proposed discrete binomial tree model, while being confident that the tree converges to a particular continuous path Gaussian Markov process.

Chapter 9

Conclusions

In this dissertation we focused on developing modeling techniques to handle the complex and fast paced world of finance. In 2009, trading algorithms had enabled execution of trades in under 128 microseconds. Applications in finance to modeling and decision making requires quick, efficient methods that are accurate on small sample sizes. In this motivation we developed two new estimators for the parameters of a fractional Wiener process. We showed that our Ratio method estimator of the Hurst index is significantly more accurate than Whittle's approximate MLE on small sample sizes ($n = 128$ data points or less) for $0.4 < H < 0.8$. Additionally, analysis showed that the Ratio method estimator shows no significant difference in accuracy for large sample sizes. To address the need for robust estimation of parameters, our second estimator of the Hurst index, which we call the Quadrant method, ignores the magnitude of movements, but measures the correlation structure of series of two dimensional Gaussian random variables. This method, while not as accurate as Whittle's MLE or the Ratio method, outperforms the second best estimator (according to Taqqu [51]), known as the Variance of Residuals method. Both of our methods are of the order of 10^4 times faster than Whittle's approximate MLE, and are approximately 500 times faster than the Variance of Residuals method. Our new methods address the disproportionate trade-off in accuracy and computation time that is present in current methods.

Keeping in mind that models in finance are typically Markovian, and many are Gaussian, the second part of this thesis focused on expanding modeling techniques for continuous path Gaussian Markov processes. We showed that the quadratic variation is a Riemann-Stieltjes integral which can be used as a new estimator of the diffusion parameter. We proved that the

quadratic variation estimator of diffusion is consistent on a fixed time interval, while the MLE method requires an infinite time horizon. The convergence rate and confidence interval bounds for the estimator are also presented. We proposed a model for extending the Black-Scholes equations to Gaussian Markov processes. This extension goes beyond the typical Black-Scholes model assumptions by allowing for the inclusion of observed market properties like long-range dependence, non-stationarity and heteroscedasticity. Unlike fractional Brownian motion, these models allow for the use of Itô calculus and are typically arbitrage free. In our derivations we explicitly show the arbitrage free condition (and change of measure) for a geometric Gaussian Markov stock model. Using our representation for the quadratic variation we demonstrated the importance of accurate estimators of quadratic variation and diffusion parameters in Option Pricing Theory and risk management techniques.

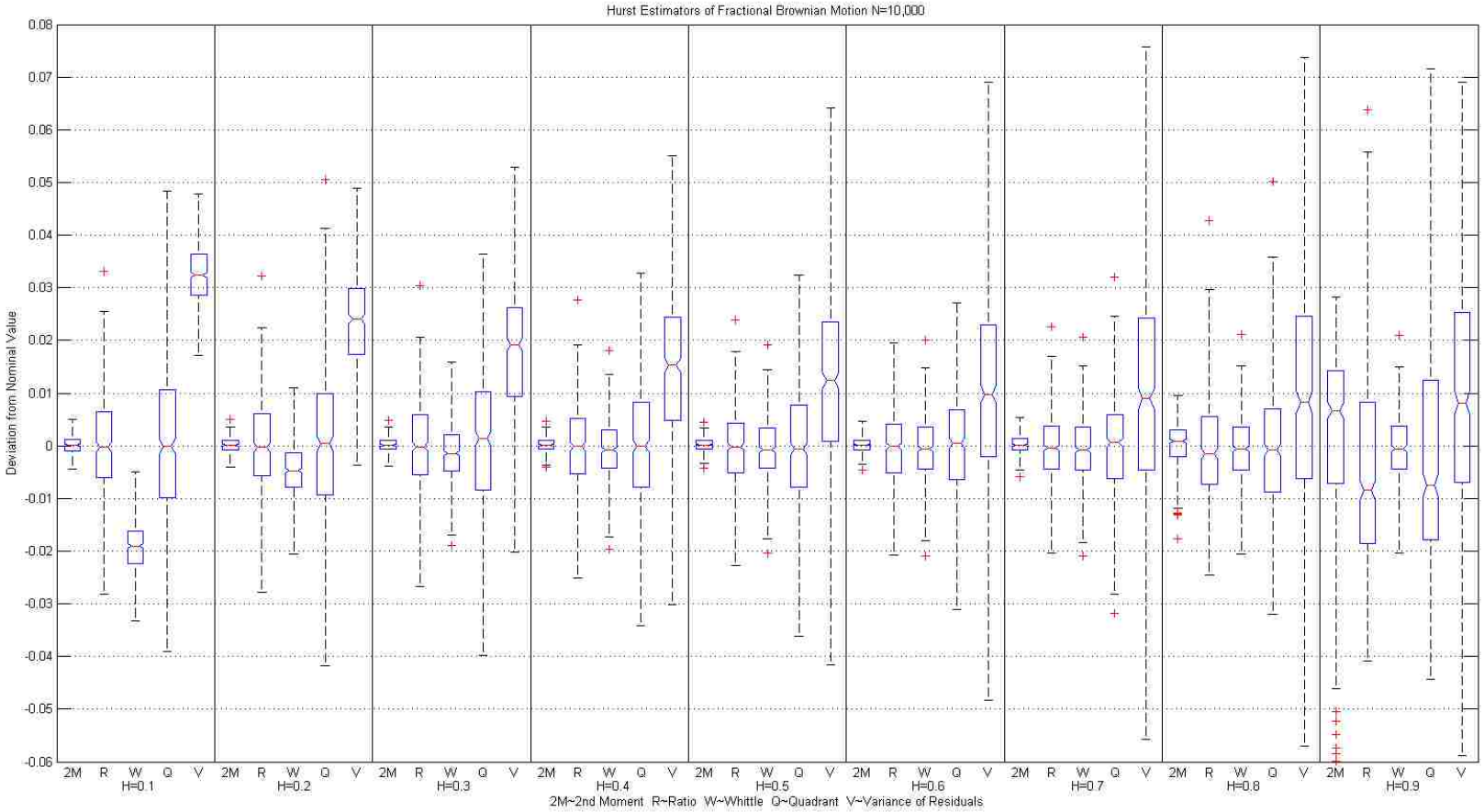
We also showed that for the Ornstein-Uhlenbeck process, three parameter numerical optimization of the likelihood function is not needed to find the Maximum Likelihood Estimators (MLE). We were able to derive closed form estimators of all model parameters and apply these estimators to simulated data to test the accuracy of the MLE. Analysis showed that our new quadratic variation method is computationally faster, since it does not require estimation of other parameters to find the diffusion parameter. More importantly, the quadratic variation estimator of diffusion is significantly more accurate on small to moderate size sample sizes. While consistency of the MLE requires an infinite time horizon for the data sample, consistency of the quadratic variation method is given on a fixed time interval, which explains the results. These properties establish that the quadratic variation method has a clear advantage in financial applications since speed and accuracy on small data sets is becoming ever more important with the increased speed of trading algorithms.

Lastly, we addressed the need for methods to price many American style derivatives by deriving a consistent discrete representation of general continuous path Gaussian Markov processes. We used the Stochastic Central Limit Theorem to prove that the binomial tree converges weakly to the continuous path Gaussian Markov process on a fixed interval. Given the need for accurate and fast methods in finance to determine optimal execution times, we show that the method always gives a recombining tree for the GM process. We apply the tree to price an American put option. Our example shows the optimal strategies for managing risk and maximizing the expected payoff of the option.

Using our results, we established that the time scaling and dependence characteristics of Gaus-

sian Markov processes and fractional Brownian motion may be able to explain many phenomenon observed in markets, such as the term structure to volatility and underestimation of risk. The infinite dependence and heteroscedasticity structures that Gaussian Markov processes can allow, along with the quick and accurate methods presented in this dissertation, enable practitioners to formulate models that could improve risk management, pricing and hedging in the marketplace.

Appendix I: Box Plot Comparisons of Hurst Estimators



Appendix II: Difference Analysis of Hurst Estimators

Difference Analysis

Whittle vs Ratio — Ratio Method Significantly Better Estimator — Statistically Equivalent Estimators — 99% Confidence Interval

N	H=0.1	H=0.2	H=0.3	H=0.4	H=0.5
10000	0.0117 (0.0108, 0.0126)	-0.0014 (-0.0021, -0.0006)	-0.0020 (-0.0026, -0.0014)	-0.0014 (-0.0019, -0.0009)	-0.0008 (-0.0011, -0.0004)
5000	0.0083 (0.0071, 0.0095)	-0.0029 (-0.0038, -0.0020)	-0.0028 (-0.0036, -0.0020)	-0.0018 (-0.0025, -0.0012)	-0.0008 (-0.0013, -0.0003)
2500	0.0021 (0.0003, 0.0038)	-0.0056 (-0.0070, -0.0043)	-0.0046 (-0.0057, -0.0035)	-0.0031 (-0.0040, -0.0021)	-0.0016 (-0.0023, -0.0009)
1250	-0.0043 (-0.0066, -0.0020)	-0.0086 (-0.0105, -0.0066)	-0.0065 (-0.0082, -0.0048)	-0.0041 (-0.0055, -0.0027)	-0.0018 (-0.0028, -0.0007)
625	-0.0107 (-0.0139, -0.0074)	-0.0124 (-0.0150, -0.0097)	-0.0093 (-0.0116, -0.0071)	-0.0061 (-0.0080, -0.0042)	-0.0030 (-0.0044, -0.0015)
312	-0.0246 (-0.0288, -0.0204)	-0.0196 (-0.0233, -0.0158)	-0.0142 (-0.0174, -0.0109)	-0.0092 (-0.0118, -0.0065)	-0.0047 (-0.0067, -0.0026)
156	-0.0308 (-0.0367, -0.0249)	-0.0206 (-0.0261, -0.0151)	-0.0131 (-0.0179, -0.0082)	-0.0064 (-0.0105, -0.0023)	-0.0009 (-0.0041, 0.0023)
78	-0.0463 (-0.0542, -0.0385)	-0.0306 (-0.0378, -0.0233)	-0.0206 (-0.0269, -0.0143)	-0.0120 (-0.0172, -0.0067)	-0.0041 (-0.0083, 0.0001)
39	-0.0636 (-0.0758, -0.0515)	-0.0382 (-0.0492, -0.0273)	-0.0213 (-0.0311, -0.0115)	-0.0077 (-0.0162, 0.0008)	0.0041 (-0.0032, 0.0114)

N	H=0.6	H=0.7	H=0.8	H=0.9
10000	-0.0002 (-0.0004, 0.0000)	-0.0001 (-0.0004, 0.0001)	-0.0030 (-0.0038, -0.0023)	-0.0116 (-0.0129, -0.0103)
5000	-0.0001 (-0.0004, 0.0002)	-0.0003 (-0.0006, 0.0001)	-0.0033 (-0.0042, -0.0023)	-0.0120 (-0.0135, -0.0104)
2500	-0.0003 (-0.0007, 0.0001)	0.0003 (-0.0002, 0.0007)	-0.0020 (-0.0032, -0.0007)	-0.0101 (-0.0120, -0.0082)
1250	0.0001 (-0.0006, 0.0008)	0.0004 (-0.0003, 0.0011)	-0.0031 (-0.0046, -0.0017)	-0.0119 (-0.0140, -0.0097)
625	-0.0002 (-0.0012, 0.0008)	0.0014 (0.0002, 0.0026)	0.0001 (-0.0021, 0.0024)	-0.0069 (-0.0096, -0.0042)
312	-0.0009 (-0.0024, 0.0006)	0.0010 (-0.0007, 0.0027)	-0.0002 (-0.0030, 0.0025)	-0.0072 (-0.0104, -0.0040)
156	0.0027 (0.0002, 0.0052)	0.0039 (0.0012, 0.0066)	0.0015 (-0.0021, 0.0050)	-0.0008 (-0.0049, 0.0032)
78	0.0022 (-0.0014, 0.0059)	0.0066 (0.0025, 0.0107)	0.0071 (0.0023, 0.0119)	0.0052 (-0.0005, 0.0109)
39	0.0150 (0.0081, 0.0219)	0.0231 (0.0156, 0.0305)	0.0245 (0.0163, 0.0326)	0.0231 (0.0136, 0.0326)

Variance of Residuals vs Quadrant — Quadrant Significantly Better Estimator — Statistically Equivalent Estimators — 99% Confidence Interval

N	H=0.1	H=0.2	H=0.3	H=0.4	H=0.5
10000	0.0199 (0.0187, 0.0212)	0.0123 (0.0108, 0.0138)	0.0081 (0.0065, 0.0097)	0.0076 (0.0060, 0.0092)	0.0079 (0.0063, 0.0095)
5000	0.0159 (0.0142, 0.0175)	0.0087 (0.0067, 0.0106)	0.0063 (0.0042, 0.0084)	0.0060 (0.0038, 0.0082)	0.0086 (0.0064, 0.0109)
2500	0.0083 (0.0059, 0.0108)	0.0028 (-0.0001, 0.0056)	0.0026 (-0.0002, 0.0055)	0.0049 (0.0020, 0.0078)	0.0093 (0.0063, 0.0123)
1250	-0.0023 (-0.0059, 0.0014)	-0.0043 (-0.0080, -0.0006)	0.0014 (-0.0025, 0.0053)	0.0090 (0.0051, 0.0130)	0.0168 (0.0128, 0.0208)
625	-0.0156 (-0.0205, -0.0106)	-0.0123 (-0.0178, -0.0068)	-0.0037 (-0.0093, 0.0018)	0.0092 (0.0034, 0.0151)	0.0192 (0.0134, 0.0251)
312	-0.0228 (-0.0296, -0.0159)	-0.0208 (-0.0281, -0.0135)	-0.0109 (-0.0181, -0.0038)	0.0050 (-0.0023, 0.0124)	0.0153 (0.0078, 0.0227)
156	-0.0218 (-0.0317, -0.0119)	-0.0228 (-0.0325, -0.0132)	-0.0118 (-0.0218, -0.0019)	0.0025 (-0.0076, 0.0127)	0.0173 (0.0072, 0.0274)
78	-0.0352 (-0.0508, -0.0195)	-0.0313 (-0.0468, -0.0158)	-0.0181 (-0.0333, -0.0030)	-0.0031 (-0.0185, 0.0123)	0.0199 (0.0058, 0.0340)
39	-0.0204 (-0.0468, 0.0059)	-0.0132 (-0.0402, 0.0138)	0.0052 (-0.0192, 0.0296)	0.0143 (-0.0094, 0.0380)	0.0481 (0.0258, 0.0705)

N	H=0.6	H=0.7	H=0.8	H=0.9
10000	0.0094 (0.0077, 0.0110)	0.0106 (0.0089, 0.0123)	0.0092 (0.0073, 0.0111)	0.0005 (-0.0020, 0.0030)
5000	0.0115 (0.0092, 0.0138)	0.0134 (0.0110, 0.0158)	0.0135 (0.0108, 0.0161)	0.0056 (0.0023, 0.0089)
2500	0.0135 (0.0103, 0.0166)	0.0176 (0.0143, 0.0209)	0.0184 (0.0149, 0.0220)	0.0126 (0.0085, 0.0168)
1250	0.0227 (0.0185, 0.0269)	0.0274 (0.0230, 0.0317)	0.0295 (0.0247, 0.0343)	0.0246 (0.0192, 0.0300)
625	0.0296 (0.0235, 0.0357)	0.0365 (0.0303, 0.0428)	0.0433 (0.0365, 0.0501)	0.0454 (0.0378, 0.0530)
312	0.0287 (0.0209, 0.0364)	0.0405 (0.0326, 0.0484)	0.0486 (0.0404, 0.0568)	0.0504 (0.0413, 0.0595)
156	0.0339 (0.0238, 0.0440)	0.0499 (0.0396, 0.0602)	0.0628 (0.0518, 0.0738)	0.0696 (0.0580, 0.0812)
78	0.0405 (0.0268, 0.0543)	0.0597 (0.0456, 0.0739)	0.0733 (0.0589, 0.0876)	-0.0026 (-0.0918, 0.0866)
39	0.0683 (0.0469, 0.0897)	0.0979 (0.0763, 0.1196)	0.0995 (0.0547, 0.1443)	-0.1892 (-0.3421, -0.0363)

Appendix III: Bias, Standard Deviation and RMSE of Hurst Estimators

Estimated Root Mean Square Error								Paths=500	
Technique	**H=0.1**	**H=0.2**	**H=0.3**	**H=0.4**	**H=0.5**	**H=0.6**	**H=0.7**	**H=0.8**	**H=0.9**
N=10,000									
Var of Res	0.033	0.026	0.022	0.021	0.021	0.022	0.023	0.024	0.025
Whittle	0.020	0.007	0.006	0.006	0.006	0.006	0.007	0.007	0.007
Second Moment	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.004	0.018
Ratio	0.010	0.009	0.008	0.008	0.007	0.007	0.007	0.011	0.020
Quadrant	0.016	0.015	0.014	0.012	0.011	0.010	0.010	0.013	0.024
N=5,000									
Var of Res	0.034	0.027	0.025	0.026	0.027	0.029	0.031	0.033	0.035
Whittle	0.020	0.009	0.008	0.008	0.009	0.009	0.009	0.009	0.009
Second Moment	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.006	0.021
Ratio	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.014	0.023
Quadrant	0.021	0.020	0.018	0.017	0.015	0.014	0.014	0.016	0.028
N=2,500									
Var of Res	0.035	0.030	0.030	0.032	0.035	0.038	0.041	0.045	0.047
Whittle	0.020	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.014
Second Moment	0.003	0.003	0.003	0.002	0.002	0.003	0.003	0.007	0.021
Ratio	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.017	0.026
Quadrant	0.031	0.029	0.028	0.026	0.023	0.021	0.019	0.021	0.030
N=1,250									
Var of Res	0.037	0.036	0.040	0.045	0.050	0.054	0.059	0.063	0.068
Whittle	0.021	0.015	0.015	0.016	0.017	0.018	0.018	0.018	0.018
Second Moment	0.004	0.004	0.004	0.003	0.003	0.004	0.004	0.009	0.026
Ratio	0.027	0.025	0.023	0.021	0.019	0.018	0.018	0.022	0.031
Quadrant	0.045	0.042	0.038	0.034	0.029	0.026	0.025	0.028	0.037
N=625									
Var of Res	0.041	0.045	0.054	0.062	0.071	0.078	0.086	0.093	0.101
Whittle	0.026	0.023	0.024	0.026	0.027	0.028	0.029	0.029	0.029
Second Moment	0.006	0.006	0.005	0.005	0.005	0.005	0.007	0.012	0.030
Ratio	0.040	0.038	0.036	0.033	0.031	0.028	0.027	0.030	0.037
Quadrant	0.063	0.060	0.056	0.051	0.046	0.041	0.040	0.040	0.044
N=312									
Var of Res	0.057	0.061	0.068	0.077	0.086	0.094	0.102	0.109	0.116
Whittle	0.029	0.031	0.034	0.036	0.038	0.039	0.040	0.040	0.038
Second Moment	0.009	0.008	0.008	0.008	0.007	0.008	0.010	0.016	0.034
Ratio	0.058	0.055	0.051	0.047	0.043	0.040	0.039	0.040	0.045
Quadrant	0.092	0.088	0.080	0.070	0.065	0.057	0.050	0.049	0.053
N=156									
Var of Res	0.077	0.080	0.088	0.098	0.109	0.121	0.133	0.144	0.153
Whittle	0.037	0.044	0.048	0.051	0.053	0.055	0.057	0.057	0.056
Second Moment	0.012	0.011	0.011	0.010	0.010	0.010	0.012	0.019	0.039
Ratio	0.076	0.071	0.066	0.061	0.056	0.052	0.051	0.053	0.056
Quadrant	0.118	0.109	0.103	0.097	0.089	0.080	0.072	0.067	0.066
N=78									
Var of Res	0.128	0.126	0.132	0.141	0.153	0.165	0.176	0.184	0.193
Whittle	0.048	0.061	0.068	0.073	0.077	0.079	0.080	0.079	0.080
Second Moment	0.016	0.015	0.015	0.014	0.014	0.015	0.018	0.026	0.047
Ratio	0.106	0.100	0.094	0.088	0.082	0.076	0.072	0.070	0.071
Quadrant	0.194	0.174	0.159	0.151	0.130	0.113	0.100	0.089	0.774
N=39									
Var of Res	0.232	0.229	0.231	0.237	0.246	0.258	0.268	0.279	0.290
Whittle	0.084	0.102	0.115	0.124	0.131	0.135	0.136	0.133	0.133
Second Moment	0.023	0.022	0.021	0.020	0.019	0.020	0.024	0.033	0.054
Ratio	0.161	0.152	0.143	0.134	0.124	0.114	0.106	0.101	0.096
Quadrant	0.289	0.271	0.238	0.225	0.190	0.169	0.147	0.364	1.365

Estimated Standard Deviation **Paths=500**

Technique	H=0.1	H=0.2	H=0.3	H=0.4	H=0.5	H=0.6	H=0.7	H=0.8	H=0.9
N=10,000									
Var of Res	0.006	0.009	0.012	0.015	0.017	0.019	0.021	0.022	0.023
Whittle	0.005	0.005	0.006	0.006	0.006	0.006	0.007	0.007	0.007
Second Moment	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.004	0.018
Ratio	0.010	0.009	0.008	0.008	0.007	0.007	0.007	0.011	0.020
Quadrant	0.016	0.014	0.014	0.012	0.011	0.010	0.010	0.013	0.024
N=5,000									
Var of Res	0.007	0.012	0.017	0.020	0.024	0.027	0.030	0.032	0.034
Whittle	0.007	0.007	0.008	0.008	0.009	0.009	0.009	0.009	0.009
Second Moment	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.006	0.021
Ratio	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.014	0.022
Quadrant	0.021	0.020	0.018	0.017	0.015	0.014	0.014	0.016	0.028
N=2,500									
Var of Res	0.011	0.018	0.024	0.029	0.033	0.037	0.040	0.043	0.046
Whittle	0.010	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.014
Second Moment	0.003	0.003	0.003	0.002	0.002	0.003	0.003	0.007	0.020
Ratio	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.016	0.024
Quadrant	0.031	0.029	0.028	0.026	0.023	0.021	0.019	0.021	0.030
N=1,250									
Var of Res	0.015	0.026	0.034	0.042	0.048	0.053	0.057	0.062	0.066
Whittle	0.014	0.014	0.015	0.016	0.017	0.018	0.018	0.018	0.018
Second Moment	0.004	0.004	0.004	0.003	0.003	0.004	0.004	0.009	0.025
Ratio	0.027	0.025	0.023	0.021	0.019	0.018	0.018	0.022	0.029
Quadrant	0.045	0.042	0.038	0.034	0.029	0.026	0.025	0.028	0.036
N=625									
Var of Res	0.022	0.036	0.049	0.059	0.068	0.076	0.083	0.091	0.098
Whittle	0.021	0.022	0.024	0.026	0.027	0.028	0.028	0.029	0.028
Second Moment	0.006	0.006	0.005	0.005	0.005	0.005	0.007	0.012	0.030
Ratio	0.040	0.038	0.036	0.033	0.031	0.028	0.027	0.029	0.034
Quadrant	0.063	0.060	0.056	0.050	0.046	0.041	0.040	0.040	0.044
N=312									
Var of Res	0.027	0.043	0.056	0.068	0.078	0.087	0.095	0.102	0.109
Whittle	0.028	0.031	0.034	0.036	0.038	0.039	0.040	0.040	0.037
Second Moment	0.009	0.008	0.008	0.008	0.007	0.008	0.010	0.016	0.033
Ratio	0.058	0.055	0.051	0.047	0.043	0.040	0.038	0.039	0.041
Quadrant	0.092	0.088	0.080	0.070	0.065	0.057	0.050	0.049	0.052
N=156									
Var of Res	0.038	0.056	0.073	0.087	0.101	0.114	0.126	0.137	0.146
Whittle	0.037	0.044	0.048	0.051	0.053	0.055	0.056	0.055	0.048
Second Moment	0.012	0.011	0.011	0.010	0.010	0.010	0.012	0.019	0.038
Ratio	0.076	0.071	0.066	0.061	0.056	0.052	0.051	0.052	0.051
Quadrant	0.118	0.109	0.103	0.097	0.089	0.080	0.072	0.067	0.065
N=78									
Var of Res	0.055	0.075	0.094	0.111	0.129	0.144	0.156	0.165	0.174
Whittle	0.046	0.061	0.068	0.073	0.077	0.078	0.078	0.074	0.063
Second Moment	0.016	0.015	0.015	0.014	0.014	0.015	0.018	0.026	0.045
Ratio	0.105	0.100	0.094	0.088	0.081	0.076	0.072	0.068	0.063
Quadrant	0.189	0.172	0.158	0.150	0.130	0.113	0.099	0.089	0.766
N=39									
Var of Res	0.120	0.142	0.163	0.184	0.204	0.223	0.239	0.254	0.268
Whittle	0.079	0.101	0.115	0.124	0.130	0.133	0.132	0.125	0.110
Second Moment	0.023	0.022	0.021	0.020	0.019	0.020	0.024	0.033	0.053
Ratio	0.161	0.152	0.143	0.134	0.124	0.114	0.105	0.097	0.086
Quadrant	0.287	0.270	0.237	0.224	0.190	0.169	0.147	0.362	1.315

Estimated Bias **Paths=500**

Technique	H=0.1	H=0.2	H=0.3	H=0.4	H=0.5	H=0.6	H=0.7	H=0.8	H=0.9
N=10,000									
Var of Res	0.032	0.024	0.018	0.015	0.012	0.010	0.009	0.009	0.008
Whittle	-0.019	-0.005	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
Ratio	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.004
Quadrant	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	-0.001
N=5,000									
Var of Res	0.033	0.024	0.019	0.015	0.013	0.012	0.010	0.010	0.009
Whittle	-0.019	-0.005	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
Ratio	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.002	-0.006
Quadrant	0.000	0.000	0.000	0.000	-0.001	-0.002	-0.001	-0.001	-0.002
N=2,500									
Var of Res	0.033	0.025	0.019	0.015	0.013	0.011	0.010	0.010	0.009
Whittle	-0.017	-0.004	-0.001	0.000	0.000	0.000	0.000	0.000	-0.001
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.006
Ratio	0.001	0.000	0.000	0.000	0.000	0.000	-0.001	-0.003	-0.010
Quadrant	-0.001	-0.001	0.000	-0.001	-0.001	0.000	-0.001	-0.002	-0.006
N=1,250									
Var of Res	0.034	0.026	0.021	0.017	0.015	0.013	0.013	0.013	0.014
Whittle	-0.015	-0.003	0.000	0.000	0.000	0.000	0.000	0.000	-0.001
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Ratio	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.003	-0.011
Quadrant	-0.001	0.000	0.000	0.000	0.001	0.002	0.001	-0.001	-0.006
N=625									
Var of Res	0.034	0.027	0.023	0.020	0.019	0.019	0.020	0.022	0.026
Whittle	-0.017	-0.005	-0.003	-0.003	-0.004	-0.004	-0.004	-0.004	-0.005
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Ratio	0.001	0.000	-0.001	-0.001	-0.002	-0.003	-0.004	-0.005	-0.013
Quadrant	-0.002	-0.002	-0.004	-0.005	-0.006	-0.003	-0.005	-0.005	-0.004
N=312									
Var of Res	0.050	0.043	0.039	0.036	0.035	0.035	0.036	0.038	0.039
Whittle	-0.007	0.001	0.002	0.002	0.001	0.000	-0.001	-0.003	-0.010
Second Moment	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.009
Ratio	0.005	0.004	0.003	0.003	0.002	0.000	-0.002	-0.008	-0.019
Quadrant	0.000	-0.002	-0.001	0.000	0.000	-0.001	0.000	-0.004	-0.009
N=156									
Var of Res	0.067	0.056	0.049	0.045	0.043	0.042	0.043	0.044	0.045
Whittle	-0.001	0.000	-0.001	-0.003	-0.006	-0.008	-0.011	-0.016	-0.029
Second Moment	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.009
Ratio	0.002	0.000	-0.001	-0.002	-0.004	-0.005	-0.007	-0.013	-0.024
Quadrant	-0.005	0.000	-0.002	-0.005	-0.003	0.000	-0.004	-0.006	-0.012
N=78									
Var of Res	0.115	0.102	0.092	0.086	0.083	0.081	0.082	0.082	0.083
Whittle	0.012	0.006	0.002	-0.002	-0.007	-0.011	-0.018	-0.028	-0.049
Second Moment	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.003	0.012
Ratio	-0.011	-0.009	-0.008	-0.007	-0.006	-0.006	-0.009	-0.016	-0.031
Quadrant	-0.044	-0.029	-0.018	-0.012	-0.006	-0.007	-0.007	-0.005	-0.106
N=39									
Var of Res	0.199	0.180	0.163	0.149	0.139	0.130	0.121	0.114	0.112
Whittle	0.029	0.014	0.005	-0.003	-0.010	-0.019	-0.030	-0.047	-0.075
Second Moment	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	0.001	0.010
Ratio	0.004	0.003	0.000	-0.002	-0.006	-0.010	-0.016	-0.026	-0.041
Quadrant	-0.036	-0.029	-0.018	-0.013	-0.005	-0.007	-0.006	-0.034	-0.369

Appendix IV: Review of Simulation and Linear Prediction Algorithms

Choleski (Lower-Upper Triangular) Decomposition

The simulation of any zero mean finite variance Gaussian process, stationary or non-stationary, can always be accomplished using Choleski decomposition. This is because the auto-covariance matrix is always symmetric and non-negative definite. Choleski requires a Hermitian and positive definite matrix in order to decompose the matrix into a product of a lower triangular matrix. Now let us derive the value of a call option and it conjugate transpose. The decomposition the non-negative definite Hermitian matrix Σ_{NxN} is found by solving the system linear equations

$$\Sigma_{NxN} = LL^*,$$

where L is a lower triangular matrix and $L^*$is its conjugate transpose. L is called the square root of the matrix. In the case of a covariance matrix, $\Sigma_{NxN} = LL^T$. In order to generate simulations of the the Gaussian process, $X = (X_i)_{i=1}^N$, the matrix L is applied to a vector of independent standard normal random variables, z_i:

$$X = L\,Z,$$

where $Z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}$. Unfortunately, in order to compute simulations, the algorithm requires a good deal of virtual memory in order to calculation and storage of all elements in the matrix, Σ_{NxN} , and calculate and store L which have a $NxN/2 + N/2$ elements. Furthermore, the algorithm can is quite time consuming, having complexity $\mathcal{O}\left(N^3\right)$ (in actuality the LU decomposition algorithm takes $N^3/3$ Floating point Operations). (Doukhan [13], pp.581)

Example 9.1. Simulation of Brownian Motion

Choleski Decomposition is trial for simulating the stationary increments of Brownian motion. Recall that by the definition of Brownian motion 2.6, the increments (white noise) are independent identically distributed as $N\left(0, \Delta t\right)$. Therefore the auto-covariance matrix for white noise is

$$\Sigma_{NxN} = \Delta t \, I_{NxN},$$

where I_{NxN} is an NxN dimensional identity matrix. The decomposition of this matrix is trivially

$$L = \sqrt{\Delta t} \, I_{NxN}.$$

Therefore, to simulate a discrete realization of a Brownian motion $(B_{t_i})_{i=1}^{N}$, all that is required is the generation of $N-1$ independent normal random variable realizations $Z = \{z_i\}_{i=1}^{N-1}$ with mean zero and variance Δt. The sum of the random variables Z generates a realization of the Brownian motion process, where

$$B_{t_{n+1}} = \sum_{i=1}^{n} z_i, \; n = 0, \ldots, N-1, \; t_{n+1} = n\Delta t.$$

As $\Delta t \to 0$, $(B_{t_i})_{i=1}^{N}$ converges by the Functional Central Limit Theorem to the Brownian motion process $(B_t)_{t \in \mathbb{R}}$.

Linear Prediction Algorithms

In this section, we do a literature review in which we sketch the proof of the formulation of the linear prediction algorithms. This section relies on the formulations presented in [7] *"Introduction to Time Series and Forecasting"* and [17] *"Time Series Analysis"* . The main advantage of Prediction algorithms (like the Durbin-Levinson algorithm and Innovations algorithm) over the Choleski decomposition algorithm is the ability to recursively compute the next observation in a time series. This is accomplished by recursively computing one-step linear predictors (a prediction of the next observation conditioned on all previous observations) and the corresponding one-step Mean Square Error of the predictors. The theory for these algorithms is formed by minimizing the Mean Square Error of the one-step predictor.

Let the time series $\{X_i\}_{i=1}^{N}$ have an auto-covariance $\mathbb{E}[X_i X_j] = k(i,j)$ with finite variance. Let us find the "best" predictor as a function f in terms of the minimal Mean Square Error. If we choose $\hat{X}_{n+1} = \mathbb{E}[X_{n+1} | X_1, X_2, \ldots, X_n]$ it can be shown that
$\mathbb{E}\left(X_{n+1} - f(X_1, X_2, \ldots, X_n)\right)^2 =$

$$\mathbb{E}\left\{X_{n+1} - \mathbb{E}\left(X_{n+1}|X_1, X_2, \ldots, X_n\right)\right\}^2 + \mathbb{E}\left\{\left(\mathbb{E}\left(X_{n+1}|X_1, X_2, \ldots, X_n\right) - f(X_1, X_2, \ldots, X_n)\right)^2\right\}$$

By the law of total expectation:

$$\mathbb{E}\left(X_{n+1} - \mathbb{E}\left(X_{n+1}|X_1, X_2, \ldots, X_n\right)\right)\mathbb{E}\left(X_{n+1} - f(X_1, X_2, \ldots, X_n)\right) = 0.$$

Therefore, the best estimator in terms of MSE has to be

$$f(X_1, X_2, \ldots, X_n) = \mathbb{E}\left[X_{n+1}|X_1, X_2, \ldots, X_n\right].$$

If the process X is jointly Gaussian, then the linear combination of these random variables are also Gaussian. Additionally, the conditional expectation $\mathbb{E}\left[X_{n+1}|X_n\right]$ is Gaussian. This is the motivation behind choosing a linear prediction operator, P_n, such that

$$\hat{X}_{n+1} = P_n X_{n+1} = \alpha_1 X_n + \cdots + \alpha_n X_1 = X^T \alpha.$$

Notice that this is a linear projection of X_{n+1} on all history of X_n, just like the conditional expectation. When choosing α we want the forecast error $\left(X_{n+1} - X^T\alpha\right)$ to be uncorrelated to X; $\mathbb{E}\left[\left(X_{n+1} - X^T\alpha\right)X\right] = 0$. Therefore,

$$\mathbb{E}\left[XX^T\right]\alpha = \mathbb{E}\left[X_{n+1}X\right].$$

If v_n is the one-step MSE of the forecasts X_n then $v_n = \mathbb{E}\left[X_{n+1} - P_n X_{n+1}\right]^2$. This how the Durbin Levinson algorithm is formed and establishes the best linear predictor of the n^{th} observation of the process.

Durbin Levinson Algorithm

The Durbin Levinson Algorithm uses the previous n observations of a mean zero stationary time series to generate the $n+1$ one-step predictor. The algorithm is recursive in this way. Since we have a stationary process we only need the one row (or column) of the auto-covariance function $k(i,j)$. Let $\gamma_n = [k(1,1), k(1,2), \ldots, k(1,n)]$, then

$$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] \nu_{n-1}^{-1},$$

$$\phi_{n,i} = \phi_{n-1,i} - \phi_{n,n}\phi_{n-1,n-i}, \quad for\ i < n,$$

$$\nu_n = \nu_{n-1}\left[1 - \phi_{n,n}^2\right],$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $\nu_0 = \gamma(0)$. To generate the time series $(X_i : i = 1, \ldots, N)$, we need to generate a sequence of i.i.d. standard normal random variables, $\{z_i : i = 1, \ldots, N\}$. Then we use the lower triangular matrix ϕ to obtain

$$X_{n+1} = \phi_{n,1}X_n + \cdots + \phi_{n,n}X_1 + \nu_{n-1}^{1/2} z_{n+1}.$$

Note that $\nu^{1/2}$ is the one-step RMSE on the one-step predictor.

Definition 9.2. Toeplitz matrix is defined as a diagonal-constant matrix, with each descending diagonal from left to right is constant. For example,

$$T = \begin{bmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{bmatrix}.$$

Remark 9.3. Stationary Gaussian process have auto-covariance matrices which meet the definition of a Toeplitz matrix, plus an additional property of symmetry and non-negative definiteness. All information in the auto-covariance function is represented in the first row (or column) of the auto-covariance matrix. All other rows can be represented by a circular shift in the first row. The Toeplitz matrix can be used to construct discrete convolution operations as matrix multiplication. The Durbin-Levinson algorithm takes advantage of this property, it uses the convolution

property of any zero mean stationary process to compute one-step predictors for simulation purposes. These properties are important for the complexity, efficiency and speed of the Durbin-Levinson algorithm since it does not require the computation and storage of an NxN one-step predictor matrix to simulate N observations; it only requires the computation and storage of a $Nx1$ vector. The speed of the Durbin-Levinson algorithm is $\mathcal{O}\left(N^2\right)$,much faster $\left(\mathcal{O}\left(N^3\right)\right)$ and computationally efficient than Choleski decomposition methods. (Doukhan [13], pp.581)

Innovations Algorithm

The Innovation algorithm approaches prediction slightly differently; through linear combination of the forecast errors, $U_n = X_n - \hat{X}_n$. Here instead of forming predictions from the time series values itself, prediction is done on the one-step prediction errors, U. The difference function $U_i = X_i - \hat{X}_i$ is called an innovation. Notice that there exists a nonsigular matrix A such that

$$U = A\,X, \tag{9.1}$$

where A is a lower triangular matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{11} & 1 & 0 & \ddots & 0 \\ a_{22} & a_{21} & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & 1 & \vdots \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \dots & 1 \end{bmatrix}$$

If X is a stationary process then $a_{i,j} = -\alpha_j$, where $\alpha = \left(\mathbb{E}\left[XX^T\right]\right)^{-1}\mathbb{E}\left[X_{n+1}X\right]$. This allows for the computation of the vector of one-step predictors, $\hat{X}_n = (X_1, P_1X_2, \dots, P_{n-1}X_n)^T$:

$$\hat{X} = X - U = A^{-1}U - U = \Theta(X - \hat{X}),$$

where

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \theta_{11} & 0 & 0 & \ddots & 0 \\ \theta_{22} & \theta_{21} & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \dots & 0 \end{bmatrix}.$$

Therefore,

$$\hat{X}_{n+1} = \begin{cases} 0 & if\ n = 0 \\ \sum_{i=1}^{n} \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}) & if\ n = 1, 2, \dots \end{cases}$$

The Innovations algorithm can be written recursively as:

$$v_0 = k(1,1),$$

$$\theta_{n,n-k} = v_k^{-1}\left(k(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right), \quad 0 \le k < n,$$

and

$$v_n = k(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

(Brockwell [6], pp.73)

The Innovations algorithm makes no assumption on the stationarity of the Gaussian process, unlike the Durbin Levinson algorithm. This algorithm is based on conditional expectations and utilizes the lower triangle of the auto-covariance (not just the first row) matrix to compute the one-step predictors of forecast errors. In order to simulate a Gaussian process, we generate N standard normal random variables, $\{z_n\}_{n=0}^{N}$, and recursively compute for all $n \in N$

$$X_n = \sum_{i=1}^{n} \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}) + v_n^{1/2}\, z_n.$$

Innovations requires the storage of only the most recent element of the auto-covariance func-

tion. Additionally, the algorithm is extremely efficient in terms of memory since it at any given integration only needs to store $2n-3$ values of Θ and $n-1$ values of v_n and the previous X_{n-1} values to generate the predictor $\hat{X}_n$. This algorithm is must faster than Choleski decomposition, with a complexity of $\mathcal{O}(N^2)$.(Doukhan et al.[13],pp. 581) Note that the main difference in the speed of this algorithm verses the speed of the Durbin Levinson algorithm comes from the computation of the one-step MSE, v_n. In Innovations algorithm there is a sum adding an additional $N(N+1)/2$ operations, while there is not sum in the Durbin Levinson algorithm for this computation.

Appendix V: Levy's Lemma

Let ξ_i be symmetric, that is $P(\xi \leq x) = P(-\xi \leq x)$, then

$$P\left(\max_{1\leq k\leq K}\left|\sum_{i=1}^{K}\xi_i\right| > x\right) \leq 4P\left(\left|\sum_{i=1}^{K}\xi_i\right| > x\right).$$

Proof. Let j be the first k so that $\sum_{i=1}^{j}\xi_i > x$. Define T as a random variable with values j. If there exists no such j, set $T = k+1$. Then,

$$\begin{aligned}
P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x,\ \sum_{i=1}^{K}\xi_i < x\right) &= \sum_{j=1}^{K}P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x,\ T=j,\ \sum_{i=1}^{K}\xi_i < x\right)\\
&= \sum_{j=1}^{K-1}P\left(\sum_{i=1}^{j}\xi_i > x,\ T=j,\ \sum_{i=1}^{j}\xi_i + \sum_{i=j+1}^{K}\xi_i < x\right)\\
&= \sum_{j=1}^{K-1}P\left(\sum_{i=1}^{j}\xi_i > x,\ T=j,\ \sum_{i=j+1}^{K}\xi_i < 0\right).
\end{aligned}$$

Since the random sums are disjoint, independent, and symmetric,

$$\begin{aligned}
P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x,\ \sum_{i=1}^{K}\xi_i < x\right) &= \sum_{j=1}^{K-1}P\left(\sum_{i=1}^{j}\xi_i > x,\ T=j\right)P\left(\sum_{i=j+1}^{K}\xi_i > 0\right)\\
&\leq \sum_{j=1}^{K}P\left(\sum_{i=1}^{K}\xi_i > x,\ T=j\right)\\
&= P\left(\sum_{i=1}^{K}\xi_i > x\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x\right) &= P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x,\ \sum_{i=1}^{K}\xi_i > x\right) + P\left(\max_{1\leq k\leq K}\sum_{i=1}^{k}\xi_i > x,\ \sum_{i=1}^{K}\xi_i < x\right)\\
&\leq 2P\left(\sum_{i=1}^{K}\xi_i > x\right) - P\left(\sum_{i=1}^{K}\xi_i = x\right)\\
&\leq 2P\left(\sum_{i=1}^{K}\xi_i > x\right).
\end{aligned}$$

By symmetry,

$$P\left(\max_{1\leq k\leq K}\left|\sum_{i=1}^{k}\xi_i\right| > x\right) \leq 4P\left(\sum_{i=1}^{K}\xi_i > x\right).$$

□

Bibliography

[1] Abry, P. and Veitch, D. (1998). Wavelet Analysis of Long-Range-Dependent Traffic. IEEE Transactions on Information Theory, Vol. 44, No. 1, 2-15.

[2] Andersen, N.T, Dobrić, V. (1987). The Central Limit Theorem for Stochastic Processes, *The Annals of Probability*, Vol. 15, No. 1, 164-177.

[3] Bayraktar, E., Poor, H.V. and Sicar, K.R. (2003). Efficient Estimation of the Hurst Parameter in High Frequency Financial Data with Seasonalities Using Wavelets. Proceedings of 2003 International Conference on Computational Intelligence for Financial Engineering (CIFEr2003), Hong-Kong, March 21-25, 2003.

[4] Beran, J. (1994). Statistics for Long-Memory Processes, Chapman & Hall, New York.

[5] Bishwal.J. (2000). *Parameter Estimation in Stochastic Differential Equations*, Lecture Notes in Mathematics, Springer, New York.

[6] Brockwell, P.J. and Davis, R.A. (1987). Time Series: Theory and Methods. Springer, New York.

[7] Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting, Second Edition.* Springer, New York.

[8] Dobrić, V. and Ojeda, F. (2009). Conditional Expectations and Martingales in the fractional Brownian Field. Preprint.

[9] Dobrić, V. and Ojeda, F. (2006). Fractional Brownian Fields, Duality, and Martingales. High Dimensional Probability, Vol. 51, 77-95.

[10] Cheridito, P. (2002). Arbitrage in Fractional Brownian Motion. Finance and Stochastics, 7(4), 533-553.

[11] Cheridito, P., Kawaguchi, H. and Maejima, M. (2000). Fractional Ornstein-Uhlenbeck Processes. Electronic Journal of Probability, 8(3), 1-14.

[12] Cont R. (2001). Empirical properties of asset returns—stylized facts and statistical issues. Quant Finance, I(2), 223-236.

[13] Doukhan, P., Oppenheim, G. and Taqqu, M.S. (2003). Theory and Applications of Long-Range Dependence, Springer, New York.

[14] Embrechts, P. and Maejima, M. (2002). Self-Similar Processes. Princeton University Press, New Jersey.

[15] Flandrin, P. (1989). On the Spectrum of Fractional Brownian Motions, IEEE Transactions on Information Theory, Vol. 35, No. 1, 197-199.

[16] Florens-Zmirou, D. (1989). Approximate Discrete-Time Schemes for Statistics of Diffusion Processes. *Statistics 20*, 547-557.

[17] Hamilton, J. (1994). *Time Series Analysis.* Princeton University Press, Princeton, NJ.

[18] Hampel, F. (2001). Robust Statistics: A Brief Introduction and Overview. *Robust Statistics and Fuzzy Techniques in Geodesy and GIS Symposium*, ETH Zurich, March 12-16.

[19] Hida,T. (1960). Canonical Representation of Gaussian Processes and their Applications. *Memoirs of the College of Science, University of Kyoto*, 33(1), 109-155.

[20] Hida, T. and Hitsuda M. (1976). *Gaussian Processes, Translations of Mathematical Monographs*, *Volume 120*, The American Mathematical Society, United States of America.

[21] Hu, Y. and Oksendal, B. (2003). Fractional White Noise Calculus and Applications to Finance. Infinite Dimensional Analysis, Quantum Probability and Related Topics, Vol. 6, No. 1, 1-32.

[22] Huber, P.J. Ronchetti, E. (2009). *Robust Statistics, Second Edition.* John Wiley & Sons, Hoboken, NJ.

[23] Hull, J. (2005). Fundamentals of Futures and Option Markets, Pearson, New Jersey.

[24] Iacus, S. (2008). *Simulation and Inference for Stochastic Differential Equations*, Springer Series in Statistics, Springer, New York.

[25] Glasserman, P. (2000). Monte Carlo Methods in Financial Engineering, Springer, New York.

[26] Konstantopoulos, T. and Skhanenko, A. (2004). Convergence and Convergence Rate to Fractional Brownian Motion for Weighted Random Sums. Siberian Electronic Mathematical Reports, Tom. 1, cmp. 57-63.

[27] Ledesma, S. and Liu, D. (2000). A Fast Method for Generating Self-Similar Network Traffic. IEEE, 54-61.

[28] Lindstrom, T. (2007). A Weighted Random Walk Approximation to Fractional Brownian Motion. e-print, Department of Mathematics/CMA, University of Oslo.

[29] Lo, A. (1991). Long-Term Memory in Stock Market Prices. Econometrica, Vol. 59, No. 5, 1279-1313.

[30] Luenberger, D. (1997). *Investment Science*, Oxford, New York.

[31] Mandelbrot, B. (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer, New York.

[32] Mandelbrot, B. and Hudson, R.L. (2004). *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward.* Basic Books, New York, New York.

[33] Mandelbrot, B. and Van Ness, J.W. (1968). Fractional Brownian Motions, Fractional Noises and Applications. SIAM Review, Vol. 10, No. 4, 422-437.

[34] Ojeda, F. (2005). Orthonormal Expansions for Gaussian Processes, Ph.D. Thesis, Department of Mathematics, Lehigh University.

[35] Øksendal, B. (1998). *Stochastic Differential Equations: An Introduction with Applications.* 5th edition, Springer-Verlag, Berlin.

[36] Papanicolaou, G. and Solna, K. (2001). Wavelet Based Estimation of Local Kolmogorov Turbulence, in 'Long-Range Dependence Theory and Applications', edited by P. Doukhan, G. Oppenheim, M.S. Taqqu, Birkhauser, Boston.

[37] Paxson, V. (1997). Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic. Computer Communication Review, 27(5), 5-18.

[38] Peltier, R. and Levy Vehel, J. (1994). A New Method for Estimating the Parameter of fractional Brownian Motion. Institut National de Recherche en Informatique et en Automatique, No. 2396, Programme 4, 1-27.

[39] Peng, C.K, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L Golberger. (1994). Mosaic Organization of DNA Nucleotides, Physical Review E, Vol. 24, No. 2, 1685-1689.

[40] Pedersen, R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, Vol. 22, No. 1, 55-71.

[41] Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.

[42] Prakasa Rao, B.L.S. (1999). *Statistical Inference for Diffusion Type Processes*, Oxford University Press, New York.

[43] Risken, H. (1989). The Fokker–Planck Equation: Method of Solution and Applications, Springer-Verlag, New York.

[44] Rogers, L. (1997). Arbitrage with Fractional Brownian Motion. Mathematical Finance, Vol. 7, No. 1, 95-105.

[45] Shiryaev, A.N. (1998). On Arbitrage and Replication for Fractal Models. Research Report No. 2, MaPhySto, University of Aarhus.

[46] Shreve, S. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models.* Springer, New York, NY.

[47] Sottinen,T. and Valkeila, E. (2001). Fractional Brownian Motion as a Model in Finance. University of Helsinki, Department of Mathematics, Preprint 302.

[48] Sottinen,T. (2001). Fractional Brownian Motion, Random Walks and Binary Market Models. Finance and Stochastics, 5, 343-355.

[49] Sottinen,T. and Valkeila, E. (2003). On Arbitrage and Replication in the Fractional Black-Scholes Pricing Model. Statistics & Decisions, 21: 137-151.

[50] Steele, J.M. (2001). *Stochastic Calculus and Financial Applications.* Springer, New York, NY.

[51] Taqqu, M.S., Teverovsky, V. and Willinger,W. (1995). Estimators for long-range dependence: an empirical study. Fractals, 3(4):785-789.

[52] Taqqu, M.S. and Teverovsky, V. (1997). On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. In: A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Robert J. Adler, Raisa E. Feldman and Murad S. Taqqu, editors. Birkhäuser, Boston (1998), pages 177-217.

[53] Taqqu, M.S. (1997). Robustness of Whittle-Type Estimators for Time Series with Long-Range Dependence. *Stochastic Models*, Vol. 13, Issue 4, 723-757.

[54] Uhlenbeck, G.E. and Ornstein,L.S. (1930). "On the theory of Brownian Motion", Phys.Rev., 36:823–841.

[55] Varadhan, S.R.S. (2001). Courant Institute of Mathematical Sciences: Volume 7 Probability Theory, American Mathematical Society, Rhode Island.

[56] Whittle, P. (1961). Gaussian Estimation in Stationary Time Series. Bulletin de l'Institut International de Statistique,Vol. 39, 105-130.

[57] Willinger, W., Taqqu, M.S. and Teverovsky, V. (1999). Stock Market Prices and Long-Range Dependence. Finance and Stochastics, Vol. 3, 1-13.

Vita

Daniel Jonathan Scansaroli was born on April 24, 1982 in New Jersey, United States of America to John Monti Scansaroli and Cathleen Josephine (Vasta) Scansaroli. He attended Lehigh University for his undergraduate and graduate studies from August 2001 to January 2012. In May 2005, he graduated Magna Cum Laude receiving the degree of Bachelor of Science in Mechanical Engineering. Awarded Lehigh University's Presidential Scholarship, he received the degree of Master of Science in Applied Mathematics in May 2006. In August 2006, he accepted admission into the Lehigh University Ph.D. program in Industrial Engineering under the Dean Doctoral Assistantship, receiving a Master of Science in Management Science in January 2009. For his tenure as a Ph.D. candidate, Dan was employed as a Teaching Assistant for Lehigh's Integrated Business and Engineering honors program. In 2008, Dan accepted part time employment as a quantitative analyst for Lehigh University's endowment. His research and doctoral studies were a result of a unique collaboration between the Lehigh University Department of Mathematics and Industrial Engineering & Systems Department. Dan was guided under co-advisors Vladimir Dobrić, Ph.D. of the Lehigh University Department of Mathematics and Robert H. Storer, Ph.D. of the Lehigh University Industrial & Systems Engineering Department until receiving the degree of Doctor of Philosophy in Industrial Engineering in January 2012.