# Maximizing Expected Logarithmic Utility in a Regime-Switching Model with Inside Information

Kenneth Tay
Advisor: Professor Ramon van Handel

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Arts
Department of Mathematics
Princeton University

May 2010

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Kenneth Tay

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Kenneth Tay

# Acknowledgements

This page is possibly the most important page in this thesis. Without the the love and support of those around me, this work would not have been completed. I would like to dedicate this thesis to the following people:

Mum and Dad, for always being there for me. For pouring your heart and soul into my life, and for encouraging me to pursue my passions.

Professor van Handel, my thesis advisor. Thank you for guiding me through this entire process, for your patience in explaining ideas, for challenging me to give my very best. Professor Calderbank, my JP advisor and second reader for my thesis, your dedication to the field of applied mathematics has inspired me to explore this area.

Professor Xu Jiagu, the professors at SIMO and SIMO X-Men, for nurturing and honing my skills in mathematics in my earlier years. I am deeply honored to have been taught by so many brilliant minds.

Blake, David, Dave, Jon and Zach, for being my spiritual mentors over the past four years. Your words of wisdom have deeply influenced my perspective on life.

Karen, for your untiring support and love. Thank you for always being there when I needed a friend.

The Manna senior class, especially the guys: Alan, Andrew, Becker, Ben, Dan and Gyeong-Sik. Thank you for loving and accepting me for who I am, flaws and all. And for four years of shenanigans.

The 321 prayer group, for helping me to view the thesis from the right perspective.

Chris, David and Ryan, for being the most helpful freshmen ever.

Calvin, Ernest and Yoonju, thank you for taking precious time out of your busy schedules to proofread my thesis. Special thanks to Yoonju for help with LaTeX.

Last but not least, my Lord and Savior Jesus Christ, the rock of my strength and my refuge.

To God Be The Glory
Great Things He Has Done

# Contents

# Chapter 1

# Introduction

When faced with a financial market model, two questions naturally arise in the mind of an investor. The first is what the greatest possible expected utility that he can obtain at some terminal time $T$ is, and the second is what trading strategy, or portfolio selection, would allow him to achieve that maximum. A common model that is used as a framework in answering this question is a financial market consisting of a riskless asset (usually a bond assumed to have a fixed interest rate), and $n$ financial assets (which we assume to be stocks), whose price process $(\mathbf{S}_t)_{t\in[0,T]}$ is given by a system of stochastic differential equations (SDEs)

$$d\mathbf{S}_t = \mathrm{diag}(\mathbf{S}_t)\left[\boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{W}_t\right].$$

Here, $\mathbf{W}$ is an $n$-dimensional Wiener process. $\boldsymbol{\mu}$ is known as the ***drift rate*** or ***mean return***, while $\boldsymbol{\sigma}$ is known as the ***volatility matrix***.

## The Case of Full Information

When the model above was first introduced by Merton (38) in relation to the utility maximization problem, it was assumed that the investor had ***full information***, i.e. he could observe the drift rate and volatilities. Merton (38), (39) solved the utility maximization problem via a dynamic programming approach for the case where the parameters of the model were constant with time. The next approach that was used to tackle this problem was martingale theory, introduced by Harrison & Pliska (24). By relating the absence of arbitrage and the completeness of the financial market model with the existence and uniqueness of an equivalent martingale measure, martingale theory could be used to solve the utility maximization problem in the case of complete markets (see Cox & Huang (14) and Karatzas et al. (29)). For the case of incomplete markets, duality methods were used to solve the utility maximization problem (see He & Pearson (27) and Karatzas et al. (30)). With Kramkov & Schachermayer (33) proving the existence of a utility-maximizing trading strategy under weak conditions in a general setting, some work has been done to find explicit solutions for these optimal trading strategies with specific utility functions (see Goll & Kallsen (19), (20) for work on the logarithmic utility function).

## The Case of Partial Information

The case of full information is not particularly realistic as it assumes that investors are able to see the underlying drift and volatility processes that drive the stock price process. As such, the case of ***partial information*** was introduced, in which investors could not see the underlying drift processes, but only the historical stock prices. Mathematically, this meant that the trading strategies which could be considered by the investor had to be adapted to a smaller filtration (the filtration generated by the stock price process) compared to the filtration used in the case of full information.

Lakner (35), (36) provides a general theory for the utility maximization problem in the case of partial information. The idea of modeling the drift rate process as a continuous-time finite state Markov chain was first introduced by Elliott & Rishel (17), making the financial market model a hidden Markov model (HMM). When the drift rate is modeled as such, the financial market model is called a ***regime-switching model***. Sass & Haussmann (45) derived an explicit representation of the utility-maximizing trading strategy for a regime-switching model in the case of constant volatility using HMM filtering results and Malliavin calculus. This result was extended by Haussmann & Sass (26) to the case of stochastic volatility and more general stochastic interest rates for the bond. (See Björk et al. (9) for a more extensive literature review for the case of partial information.)

## The Case of Inside Information

Recent work has attempted to focus on investors who do not fall in the case of complete information or partial information. While it is unreasonable to assume that investors have complete knowledge of the drift rate, it is also not realistic for an investor to make trading decisions based solely on historical stock prices. It is more likely that an investor would attempt to obtain a more accurate estimate of the drift rate than that given by the historical stock prices before executing his trading strategy. There are several ways to obtain estimates of the drift rate, including being an "insider" of the company. For this reason, this case is known as the case of ***inside information***. There are a number of ways to model the additional information that the insider has (see Corcuera et al. (13)). A natural way to model an insider's information is to enlarge the filtration generated by the stock prices to incorporate any additional information on the drift rate.

The case of inside information was first studied by Pikovsky & Karatzas (43) for the case of complete financial market models, where the additional information was modeled as a random variable dependent on the future value of some underlying process. (An example of such a random variable is a noise-corrupted value of the stock at the terminal time.) These results were then extended by Amendinger et al. (2) to the case of incomplete financial market models, who also noted that the additional utility gained from the insider's random variable is the relative entropy of that random variable with respect to the original probability measure. These two works deal with the "initial enlargement" setting, where the extra information the insider gets is some fixed extra information at the beginning of the trading interval. (See Ankirchner et al. (4) for a comprehensive list of works that have analyzed the utility maximization problem in the "initial enlargement" setting.)

## Problem Formulation & Thesis Outline

This thesis explores the notion of inside information from different perspective in the context of a regime-switching model. We will assume that the financial market consists of a riskless bond with fixed interest rate $r$, and a risky stock whose price process $(S_t)_{t\in[0,T]}$ is given by the SDE

$$dS_t = \mu_t S_t dt + \sigma S_t dW_t,$$

where $\mu$ is a continuous-time Markov chain with two states, $\sigma$ is constant, and $W$ is a one-dimensional Wiener process. Instead of assuming that the investor has some knowledge of future prices, we assume that the investor is able to see a noise-corrupted signal which is dependent on the historical drift rate. In particular, we assume that apart from the stock price process, the investor observes another process given by the SDE

$$dY_t = \mu_t dt + \varepsilon dV_t,$$

where $V$ is a Wiener process independent of $W$, and $\varepsilon$ is a parameter that controls the amount of noise present. (In this thesis, $\varepsilon$ is assumed to be an exogenous variable.) In addition, we assume that the investor's utility at time $t$ is given by the log of the discounted value of his portfolio at time $t$.

Using tools from stochastic filtering theory, we will derive explicit expressions for the greatest possible expected utility the investor can achieve, the long-run discounted growth rate of this wealth process, along with the trading strategy that achieves this maximum. These expressions will be obtained for the cases of complete, partial and inside information. Finally, we will show that in some sense, by varying the parameter $\varepsilon$ in the case of inside information, we approach the cases of complete and partial information.

This thesis is organized as follows: In Chapter 2, we outline background material on stochastic processes, continuous-time Markov chains, ergodic theory and Bessel functions which will be used in future chapters. In Chapter 3, we develop the theory of stochastic filtering from the reference measure approach. Of particular importance is the Shiryaev-Wonham filter, which is presented in Section 3.4. Chapter 4 sets up the regime-switching model which will be used for the rest of the thesis. In particular, the cases of complete, partial and inside information within the context of the regime-switching model are defined in Section 4.2. We begin Chapter 5 by developing the concepts of the log-optimal wealth process and the long-run discounted growth rate. In Section 5.2, we derive explicit expressions for the greatest achievable expected log utility, the long-run discounted growth rate of this wealth process, along with the trading strategy that attains this maximum. Chapter 6 explores the relationship between the cases of complete, partial and inside information through the long-run discounted growth rate of the wealth process that maximizes expected log utility. Section 6.2 gives an explicit expression for this growth rate in the case of complete information, while in Sections 6.3 & 6.4 we derive the stationary distributions of a quantity from which we can calculate the growth rate in the cases of partial and inside information. In Section 6.6, we show that from the point of view of the measures defined by the stationary distributions in the previous sections, the cases of complete and partial information can be viewed as the case of inside information with $\varepsilon = 0$ and $\varepsilon = \infty$ respectively. Finally, in Section 6.7 we provide an explicit expression for the long-run discounted growth rate of the log-optimal wealth process, along with its asymptotic as $\varepsilon$ goes to zero, in the case where the the rates of entering each state of the Markov chain are equal. For the sake of readability, proofs of certain lemmas and propositions which are more technical in nature are in the Appendix.

# Chapter 2

# Background Material

This chapter covers the mathematical background needed for this thesis. Before we begin, let us first clear up some notational matters:

- $\mathbb{N} := \{0, 1, 2, \dots\}$.
- $\mathbb{R}_+ := [0, \infty)$.
- $x$ positive $\Leftrightarrow x > 0$, $x$ non-negative $\Leftrightarrow x \geq 0$.
- $x \vee y := \max(x, y)$, $x \wedge y := \min(x, y)$.
- For a topological set $X$, $\mathcal{B}(X) :=$ the Borel $\sigma$-algebra on $X$.
- To distinguish $\sigma$-algebras from filtrations, we will always denote a $\sigma$-algebra by a single script letter (e.g. $\mathcal{F}$), and a filtration by a letter with a time subscript, enclosed in braces (e.g. $\{\mathcal{F}_t\}$).
- When referring to a stochastic process, we will use a single letter (e.g. $X$), unless we wish to make the time dependency more explicit, in which case we will use either $(X_t)$ or $(X_t)_{t\in\mathbb{T}}$ (where $\mathbb{T}$ is the time set). While this notation is similar to that for $\sigma$-algebras and filtrations, the object we are referring to should be clear from the context.
- The script letter will serve two notation purposes: for a measurable space $(\Omega, \mathcal{F})$, $\mathcal{F}$ will denote both the $\sigma$-algebra on $\Omega$ and the set of all real-valued random variables which are measurable with respect to $(\Omega, \mathcal{F})$. Again, no confusion should arise because of the context.

Throughout this thesis, we will also make the following assumptions:

1. Unless otherwise stated, we will assume that the underlying probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. In addition, this probability space admits a filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

2. Any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ mentioned in this thesis is complete, i.e. if we define

$$\mathcal{N} := \{A : A \subset B \text{ for some } B \in \mathcal{F} \text{ with } \mathbb{P}(B) = 0\},$$

then $\mathcal{N} \subset \mathcal{F}$.

3. For any underlying filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, we will assume that $\{\mathcal{F}_t\}$ satisfies the "usual conditions", i.e. $\{\mathcal{F}_t\}$ is:

   - augmented: $\mathcal{N} \subset \mathcal{F}_0$, and
   - right-continuous: for all $t \geq 0$, $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$.

   These technical conditions are often included in the literature in order to ensure that no technicalities can result from the behavior of the filtration. As we will be considering relatively simple objects in probability theory for this thesis, we will not concern ourselves with these conditions.

## 2.1 Basic Probability Theory

**Definition 2.1.1.** *Given a set $\Omega$, a **$\sigma$-algebra $\mathcal{F}$** is a collection of subsets of $\Omega$ which contains $\Omega$, and is closed under complements and countable unions.*
*Given a collection $\mathcal{C}$ of subsets of $\Omega$, the **$\sigma$-algebra generated by $\mathcal{C}$**, denoted by $\sigma\mathcal{C}$, is the smallest $\sigma$-algebra which contains $\mathcal{C}$.*
*Given a real-valued random variable $X$ on $\Omega$, the **$\sigma$-algebra generated by $X$** is defined as*

$$\sigma X := \sigma\{X^{-1}B : B \in \mathcal{B}(\mathbb{R})\},$$

*where $X^{-1}B := \{\omega : \omega \in \Omega, X(\omega) \in B\}$.*
*Given a collection of real-valued random variables $\{X_i : i \in I\}$ on $\Omega$ indexed by some arbitrary set $I$, the **$\sigma$-algebra generated by $\{X_i : i \in I\}$**, denoted by $\sigma\{X_i : i \in I\}$, is the smallest $\sigma$-algebra that contains $\sigma X_i$ for all $i \in I$.*

**Definition 2.1.2.** *Given a stochastic process $X = (X_t)_{t\geq 0}$, the **filtration generated by $X$**, denoted by $\{\mathcal{F}_t^X\}_{t\geq 0}$, is given by $\mathcal{F}_t^X := \sigma\{X_s : s \leq t\}$ for all $t$.*

**Definition 2.1.3.** *Let $\mathcal{F}$ be some $\sigma$-algebra on $\Omega$. A real-valued random variable $X$ on $\Omega$ is said to be **measurable w.r.t. $\mathcal{F}$** if $X^{-1}B \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$. If $X$ is measurable w.r.t. $\mathcal{F}$, we write $X \in \mathcal{F}$.*

**Definition 2.1.4.** *Let $\{\mathcal{F}_t\}$ be a filtration on $(\Omega, \mathcal{F})$, and let $X$ be a real-valued stochastic process on $\Omega$. $X$ is **adapted to $\{\mathcal{F}_t\}$** if for all $t$, $X_t$ is measurable w.r.t. $\mathcal{F}_t$.*

**Definition 2.1.5.** (Wiener Processes.) *Let $\{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration on $(\Omega, \mathcal{F})$, and let $W = (W_t)_{t\in[0,T]}$ be a continuous stochastic process (i.e. the paths $t \mapsto W_t(\omega)$ are continuous for* *almost all $\omega \in \Omega$). $W$ is a **Wiener process/Brownian Motion on $[0,T)$ w.r.t. $\{\mathcal{F}_t\}$** (or an **$\{\mathcal{F}_t\}$-Wiener process**) if:*

1. *$W$ is adapted to $\{\mathcal{F}_t\}$,*
2. *$W_0 = 0$,*
3. *For any $0 \leq s \leq t < T$, the increment $W_t - W_s$ is independent of $\mathcal{F}_s$, and*
4. *For any $0 \leq s \leq t < T$, the increment $W_t - W_s$ has the Gaussian distribution with mean $0$ and variance $t - s$.*

*For $n \in \mathbb{N} \setminus \{0\}$, a vector process $\mathbf{W}$ is said to be an **$n$-dimensional Wiener process on $[0,T)$ w.r.t. $\{\mathcal{F}_t\}$** if each compenent of $\mathbf{W}$, denoted by $W^{(i)}$ for $i = 1, \ldots, n$, is a Wiener process on $[0,T)$ w.r.t $\{\mathcal{F}_t\}$, and if for all $i \neq j$, $W^{(i)}$ is independent of $W^{(j)}$.*

## 2.2 The Stochastic Integral

The fundamental object of the field of stochastic calculus is the stochastic integral, also known as the Itô integral. Following the introduction of stochastic processes, one might consider what it means to integrate a function with respect to a stochastic process. The concept of the Stieltjes integral does not apply to general stochastic processes as many of these processes have infinite variation. As such, the concept of the integral has to be extended so as to allow a stochastic process to be the integrator. This is the role that the stochastic integral plays.

The stochastic integral is so fundamental to the theory of stochastic processes that almost every book on stochastic processes contains at least a short introduction the stochastic integrals and their construction. For the construction of the stochastic integral with respect to a general semimartingale, refer to Métivier (40). In this thesis we will work almost exclusively with the stochastic integral where the integrator is an Itô process. Here, we provide a condensed version that will give the reader a brief overview of the stochastic integral necessary for the development of ideas in subsequent chapters. The ideas presented here are loosely based on Kallianpur (28), Steele (51) and van Handel (54).

Let $W$ be a Wiener process on $[0,T)$ w.r.t. some filtration $\{\mathcal{F}_t\}$. To construct the stochastic integral of a function $f$ with respect to $W$, it is necessary to restrict the class of functions for which the integral makes sense.

**Definition 2.2.1.** *Let $\mathcal{B} := \mathcal{B}([0,T])$, i.e. the Borel $\sigma$-algebra of $[0,T]$. For each $t \in [0,T]$, define $\mathcal{F}_t \times \mathcal{B} := \sigma\{A \times B : A \in \mathcal{F}_t, B \in \mathcal{B}\}$. Let $f$ be a real-valued stochastic process with time index $[0,T]$. Then,*

- *$f$ is **measurable** if the mapping $(\omega, t) \mapsto f_t(\omega)$ is measurable w.r.t. $\mathcal{F}_T \times \mathcal{B}$.*
- *$f$ is **adapted** if $f_t \in \mathcal{F}_t$ for all $t \in [0,T]$.*

*Let $\boldsymbol{\mathcal{H}^2[0,T]}$ denote the collection of all measurable, adapted mappings $f : \Omega \times [0,T] \mapsto \mathbb{R}$ which satisfy the integrability constraint*

$$\mathbb{E}\left[\int_0^T f_t^2 dt\right] < \infty.$$

**Theorem 2.2.2.** (Stochastic integral w.r.t. a Wiener process). *For every $f \in \mathcal{H}^2[0,T]$, the* ***stochastic integral of $f$ w.r.t. $W$*** *exists, and is denoted by $\int_0^T f_t dW_t$. In fact, the stochastic integral can be defined in a manner similar to that of a Riemann sum:*

$$\int_0^T f_t dW_t := \lim_{n\to\infty} \sum_{k=1}^{n} f_{\frac{T(k-1)}{n}} \left(W_{\frac{Tk}{n}} - W_{\frac{T(k-1)}{n}}\right),$$

*where the limit is understood in the sense of limit in $L^2$-norm, that is, if $Y = \int_0^T f_t dW_t$, then*

$$\lim_{n\to\infty} \mathbb{E}\left[\left| Y - \sum_{k=1}^{n} f_{\frac{T(k-1)}{n}} \left(W_{\frac{Tk}{n}} - W_{\frac{T(k-1)}{n}}\right)\right|^2\right] = 0.$$

The following is a well-known proposition, which follows directly from the fact that when the integrand is in $\mathcal{H}^2[0,T]$, the stochastic integral is a martingale.

**Proposition 2.2.3.** *Let $f = (f_t)_{t\in[0,T]}$ be an element of $\mathcal{H}^2[0,T]$. Then for all $0 \le t \le T$,*

$$\mathbb{E}\left[\int_0^t f_u dW_u\right] = 0.$$

The notion of the stochastic integral with respect to a Wiener process on $[0,T)$ can be extended to a class of functions bigger than $\mathcal{H}^2[0,T]$.

**Definition 2.2.4.** *Fix a terminal time $T$. Define $\boldsymbol{\mathcal{H}^1_{loc}[0,T]}$ as the collection of all measurable, adapted mappings $f : \Omega \times [0,T] \mapsto \mathbb{R}$ such that*

$$\mathbb{P}\left\{\int_0^T |f_t| dt < \infty\right\} = 1.$$

*Define $\boldsymbol{\mathcal{H}^2_{loc}[0,T]}$ as the collection of all measurable, adapted mappings $f : \Omega \times [0,T] \mapsto \mathbb{R}$ such that*

$$\mathbb{P}\left\{\int_0^T f_t^2 dt < \infty\right\} = 1.$$

For any stochastic process $f$ in $\mathcal{H}^2_{loc}[0,T]$, the stochastic integral $\int_0^T f_t dW_t$ can be defined through a process known as localization (see Chapter 7 of Steele (51) for more details). It should be noted that Proposition 2.2.3 does not hold in this general context— the condition that the integrand is in $\mathcal{H}^2[0,T]$ cannot be dropped.

Next, we will define Itô processes and the stochastic integral with respect to an Itô process.

**Definition 2.2.5.** (Itô processes). *An **Itô process** is a stochastic process $I = (I_t)_{t\in[0,T]}$ such that for each $t \in [0,T]$,*

$$I_t = I_0 + \int_0^t X_s ds + \int_0^t Z_s dW_s,$$

*where $I_0$ is a constant, $X \in H^1_{loc}[0,T]$ and $Z \in \mathcal{H}^2_{loc}[0,T]$.*

**Theorem 2.2.6.** (Stochastic integral w.r.t. an Itô process.) *Let $I$ be an Itô process of the form above. Let $Y = (Y_t)_{t\in[0,T]}$ be a process adapted to $\{\mathcal{F}_t\}$. Then, the stochastic integral*

$$\int_0^T Y_t dI_t := \lim_{n\to\infty} \sum_{k=1}^n Y_{\frac{T(k-1)}{n}} \left( I_{\frac{Tk}{n}} - I_{\frac{T(k-1)}{n}} \right)$$

*exists and the following equation holds:*

$$\int_0^T Y_t dI_t = \int_0^T Y_t X_t dt + \int_0^T Y_t Z_t dW_t,$$

*provided the processes $YX$ and $YZ$ are such that the integrals in the equation above are well-defined.*

It would be remiss to conclude an introduction to the stochastic integral without talking about ***Itô's formula*** (also known as ***Itô's lemma***). Itô's formula is the foundation of stochastic calculus, which allows us to manipulate stochastic integrals. Here we present the most general form of Itô's formula, along with Itô's formula in one dimension as a corollary.

**Theorem 2.2.7.** (Itô's Formula.) *Let $(W_t^{(k)})_{t\geq 0}$, $k = 1,\ldots,m$, be $m$ independent Wiener processes. Let $I_0^{(i)}$, $i = 1,\ldots,n$, be random variables which are independent of the $W^{(k)}$'s. Let $(X_t^{(i)})_{t\geq 0}$, $i = 1,\ldots,n$ and $(Z_t^{(ik)})_{t\geq 0}$, $i = 1,\ldots,n$, $k = 1,\ldots,m$, be adapted processes such that for each $i \in \{1,\ldots,n\}$, the process given by*

$$I_t^{(i)} = I_0^{(i)} + \int_0^t X_u^{(i)} du + \int_0^t Z_u^{(i1)} dW_u^{(1)} + \cdots + \int_0^t Z_u^{(im)} dW_u^{(m)}$$

*exists. Suppose that $f(t, x_1, \ldots, x_n)$ is a function that is differentiable in $t$ and twice differentiable in each of the $x_i$'s. Then*

$$\begin{aligned} f(t, I_t^{(1)}, \ldots, I_t^{(n)}) = f(0, I_0^{(1)}, \ldots, I_0^{(n)}) &+ \int_0^t \frac{\partial f}{\partial t}(u, I_u^{(1)}, \ldots, I_u^{(n)}) du \\ &+ \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(u, I_u^{(1)}, \ldots, I_u^{(n)}) dI_u^{(i)} \\ &+ \frac{1}{2} \sum_{i,j=1}^n \sum_{k=1}^m \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(u, I_u^{(1)}, \ldots, I_u^{(n)}) Z_u^{(ik)} Z_u^{(jk)} du. \end{aligned}$$

**Corollary 2.2.8.** *Let $W = (W_t)_{t\in[0,T]}$ be a Wiener process on $[0,T]$. Let $f = f(t,x)$ be continuously differentiable in $t$ and twice continuously differentiable in $x$. Then for any $t \in [0,T]$,*

$$\begin{aligned} f(t,W_t) = f(0,W_0) &+ \int_0^t \frac{\partial f}{\partial t}(u,W_u)du \\ &+ \int_0^t \frac{\partial f}{\partial x}(u,W_u)dW_u + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(u,W_u)du. \end{aligned} \tag{2.1}$$

*Here, the domain of $f$ is $\mathbb{R}^+ \times \mathbb{R}$.*

In working with stochastic processes, one often works with integral equations rather than differentiable equations. This is because stochastic processes are often non-differentiable. However, writing the integral signs can be cumbersome, and they do not add much to the discussion. As such, we introduce "differential notation" to make the notation a little lighter. In differential notation, the term $\int_0^t f_s dW_s$ is replaced by the "differential" $f_t dW_t$, and random variables independent of $t$ vanish. For example, in differential notation, an Itô process $I$ is of the form $dI_t = X_t dt + Z_t dW_t$.

Differentials obey the following "multiplication table":

| | $dt$ | $dW_t^{(i)}$ | $dW_t^{(j)}$ |
|---|---|---|---|
| $dt$ | 0 | 0 | 0 |
| $dW_t^{(i)}$ | 0 | $dt$ | 0 |
| $dW_t^{(j)}$ | 0 | 0 | $dt$ |

(Here, $W^{(i)}$ and $W^{(j)}$ are two independent Wiener processes.) For example,

$$\begin{aligned} (dt)(dW_t) = 0, \qquad (dW_t)^2 &= dt, \\ (dI_t)^2 &= (X_t dt + Z_t dW_t)^2 \\ &= X_t^2 (dt)^2 + 2X_t Z_t dt dW_t + Z_t^2 (dW_t)^2 \\ &= Z_t^2 dt, \end{aligned}$$

and Itô's formula for an Itô process becomes

$$df(t,I_t) = \frac{\partial f}{\partial t}(t,I_t)dt + \frac{\partial f}{\partial x}(t,I_t)dI_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,I_t)(dI_t)^2.$$

## 2.3 Continuous-Time Markov Chains

The theory of Markov chains is a very wide subject with many interesting results. Norris (41) provides a substantial introduction to both discrete-time and continuous-time Markov chains, along with some applications, while Dynkin (15) presents an in-depth treatment of the general theory of Markov processes. In this section, we will define what we mean by a Markov chain, along with a construction of a continuous-time Markov chain with finite state space, which will be used in future chapters. We then present an equation linking functions of a Markov chain with the generator of the Markov chain. For this thesis, we will assume that the Markov chain will not "explode" (see page 69 of Norris (41) for more details on explosion).

The following definition highlights the slight difference in the way that we use the terms "Markov process" and "Markov chain" in this thesis:

**Definition 2.3.1.** *A real-valued stochastic processs $X = (X_t)_{t \geq 0}$ is said to be a* ***continuous-time Markov process*** *if for all $0 < s < t$ and for all $B \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{P}\left\{X_t \in B | \mathcal{F}_s^X\right\} = \mathbb{P}\left\{X_t \in B | \sigma(X_s)\right\}.$$

*$X$ is said to be a* ***Markov chain*** *if $X$ is a Markov process whose state space is finite or countably infinite, i.e. the set $\{X_t(\omega) : t \in \mathbb{R}_+, \omega \in \Omega\}$ is finite or countably infinite.*

Intuitively, a Markov process is a process with no memory: the value of the process at a future time conditioned on its historical values up to the present time is the same as that conditioned only on its present value.

Norris (41) presents three possible ways to construct such a Markov chain. The construction below is based on the third construction presented there and in Chigansky (11).

Let $d \in \mathbb{N} \setminus \{0, 1\}$. Let $E = \{e_1, \ldots, e_d\}$ be the standard Euclidean basis for $\mathbb{R}^d$, and let $p_0$ be some probability distribution on $E$. (Note that $p_0$ can be identified with a $d$-dimensional vector with non-negative entries that sum up to 1.) Introduce a family of independent Poisson processes

$$\left\{\left(N_t^{ij}\right)_{t \geq 0} : i, j \in \{1, \ldots, d\}, i \neq j\right\},$$

such that $N^{ij}$ has rate $\lambda_{ij} \geq 0$. Let $Y = (Y_n)_{n \in \mathbb{N}}$ be a discrete-time stochastic process taking values in $\{1, \ldots, d\}$, with $\mathbb{P}\{Y_0 = i\} = p_0(i)$ for all $i$. (Here, $p_0(i)$ denotes the $i^{\text{th}}$ component of the vector $p_0$.) Let $J = (J_n)_{n \in \mathbb{N}}$ be a discrete-time stochastic process taking values in $\mathbb{R}_+$, with $J_0 = 0$. For $n = 0, 1, 2, \ldots$, define the processes $J$ and $Y$ inductively as follows:

$$J_{n+1} = \inf\left\{t > J_n : N_t^{Y_n j} \neq N_{J_n}^{Y_n j} \text{ for some } j \neq Y_n\right\},$$
$$Y_{n+1} = \begin{cases} j, & \text{if } J_{n+1} < \infty \text{ and } N_{J_{n+1}}^{Y_n j} \neq N_{J_n}^{Y_n j}, \\ Y_n, & \text{if } J_{n+1} = \infty. \end{cases}$$

(Note that if $\left\{t > J_n : N_t^{Y_n j} \neq N_{J_n}^{Y_n j} \text{ for some } j \neq Y_n\right\}$ is the empty set, we define $J_{n+1}$ to be $\infty$.) Now, define a vector process $I = (I_t)_{t\geq 0}$ taking values in $E$ such that for all $t \geq 0$,

$$I_t = e_{Y_n}, \text{ where } n \text{ is such that } J_n \leq t < J_{n+1}.$$

Under this set-up, we have the following theorem:

**Theorem 2.3.2.** (See Theorem 6.30 of Chigansky (11)). *The process $I$ is a continuous-time Markov chain with state space $E$, initial distribution $p_0$, and transition intensities matrix $\Lambda$, where $\Lambda$ is the $d \times d$ matrix defined by*

$$(\Lambda)_{ij} = \begin{cases} \lambda_{ij}, & \text{if } j \neq i, \\ -\sum_{j\neq i} \lambda_{ij}, & \text{if } j = i. \end{cases}$$

Because of this theorem, it is natural to define $\lambda_{ii} = -\sum_{j\neq i} \lambda_{ij}$ for $i = 1, \ldots, d$.

To construct a continuous-time Markov chain $X$ on the state space $\{a_1, \ldots, a_d\} \subset \mathbb{R}$ with initial distribution $p_0$ and transition intensities matrix $\Lambda$, it suffices to define $X$ by

$$X_t = \sum_{i=1}^{d} a_i I_t(i), \qquad \text{for all } t \geq 0, \tag{2.2}$$

where $I_t(i)$ is the $i^{\text{th}}$ component of the vector $I_t$. If $f$ is a function of $X$, it only has to be defined on the state space $\{a_1, \ldots, a_d\}$. As such, we can identify $f$ with a $d \times 1$ vector:

$$f = \begin{bmatrix} f(a_1) \\ \vdots \\ f(a_d) \end{bmatrix}. \tag{2.3}$$

It is clear that for any function $f$ on $X$, we have

$$f(X_t) = \sum_{i=1}^{d} f(a_i) I_t(i) = f^T I_t. \tag{2.4}$$

Any continuous-time Markov process has a generator associated with it (see Section 4.2 of Varadhan (56).) In the general setting, the generator, denoted by $L$, is a differential operator which acts on functions of the Markov process. In the special case where a Markov process $X$ is actually a finite state Markov chain, the generator is simply the transition intensities matrix $\Lambda$. More concretely, if $f$ is a function on $X$, it can be identified with the vector in equation (2.3). Then, $Lf$ is the function on $X$ identified with the vector $\Lambda f$. The following proposition gives a useful expression for functions of Markov chains involving its generator:

**Proposition 2.3.3.** (See equation (6.7) of Rogers (44).) *Let $f : \{a_1, \ldots, a_d\} \mapsto \mathbb{R}$ be some* *function. Then for any* $t \in [0, T]$,

$$f(X_t) = f(X_0) + \int_0^t (\Lambda f)(X_s)ds + M_t,$$

*where* $M$ *is an* $\{\mathcal{F}_t\}$*-martingale. In differential notation:*

$$df(X_t) = (\Lambda f)(X_t)dt + dM_t.$$

## 2.4 Ergodic Theory for Continuous-Time Markov Processes

Let $X = (X_t)_{t \geq 0}$ be a real-valued stochastic process which is the solution to the SDE

$$dX_t = f(X_t)dt + g(X_t)dW_t, \tag{2.5}$$

where $f$ and $g$ are continuous real-valued functions, and $W$ is a Wiener process. Assume that this SDE has a unique strong solution. (For some sufficient conditions for such a solution to exist, see pages 128–9 and Theorem 4.5.3 of Kloeden & Platen (32).) Note that $X$ is a homogeneous Markov process (see Theorem 4.21 of Xiong (57), page 141 of Kloeden & Platen (32)).

**Definition 2.4.1.** *A probability measure* $\nu$ *is an* ***invariant probability measure*** *(or a* ***stationary distribution****) for a Markov process* $X$ *if* $\mathbb{P}^\nu\{X_t \in B\} = \nu(B)$ *for all* $B \in \mathcal{B}(\mathbb{R}), t \geq 0$*, where* $\mathbb{P}^\nu$ *is is the law of the solution of equation* (2.5) *for the case where the initial condition* $X_0$ *has distribution* $\nu$*.*

Under certain conditions (which will be satisfied by all processes in this thesis), there exists an invariant probability measure for the process which is the solution of equation (2.5) (see page 90 of Has'minskiĭ (25) and page 120 of Soize (50)). In fact, this invariant probability measure is unique (see page 123 of Has'minskiĭ (25)), and under conditions which will be satisfied by all processes in this thesis, the invariant probability measure has a density (see page 24 of Skorokhod (48) and Theorem 1.1 of Kushner (34)).

The following theorem gives an explicit formula for the invariant probability measure under additional conditions. (For a more general version of the theorem, see Section 6.2.3 of Sobczyk (49).)

**Theorem 2.4.2.** *Let* $X$ *be a continuous-time stochastic process taking values in* $[0, 1]$*, such that* $X$ *is the solution to equation* (2.5)*. By the discussion above,* $X$ *has a unique invariant probability measure* $\nu$ *with density* $\pi$*, i.e.* $\nu(dx) = \pi(x)dx$*, where* $dx$ *is the Lebesgue measure. Assume that* $g(x) = 0$ *for* $x = 0, 1$*,* $g(x) \neq 0$ *for* $0 < x < 1$*, and that for some* $a \in (0, 1)$*,*

$$\int_a^1 \frac{2f(x')}{g^2(x')}dx' = -\infty, \quad \textit{and} \quad 0 < \int_0^1 \frac{1}{g^2(x)} \exp\left[\int_a^x \frac{2f(x')}{g^2(x')}dx'\right] dx < \infty.$$

*Then $\pi$ is given by*

$$\pi(x) = \frac{N}{g^2(x)} \exp\left[\int_a^x \frac{2f(x')}{g^2(x')} dx'\right],$$

*where $N$ is the normalization constant such that $\int_{-\infty}^{\infty} \pi(x)dx = 1$.*

*Proof.* For any continuous function $h : \mathbb{R} \mapsto \mathbb{R}$ with compact support and for any $t \geq 0$,

$$\begin{aligned} h(X_t) &= h(X_0) + \int_0^t \frac{dh}{dx}(X_u)dX_u + \int_0^t \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)du && \text{(Itô's formula)} \\ &= h(X_0) + \int_0^t \left\{\frac{dh}{dx}(X_u)f(X_u) + \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)\right\} du + \int_0^t \frac{dh}{dx}(X_u)g(X_u)dW_u. \end{aligned}$$

Taking expectations with respect to the invariant probability measure,

$$\begin{aligned} \mathbb{E}[h(X_t)] &= \mathbb{E}[h(X_0)] + \mathbb{E}\left[\int_0^t \left\{\frac{dh}{dx}(X_u)f(X_u) + \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)\right\} du\right] \\ &\quad + \mathbb{E}\left[\int_0^t \frac{dh}{dx}(X_u)g(X_u)dW_u\right] \\ &= \mathbb{E}[h(X_0)] + \int_0^t \mathbb{E}\left[\frac{dh}{dx}(X_u)f(X_u) + \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)\right] du. \\ & && \text{(by Prop 2.2.3 and Fubini)} \end{aligned}$$

Under the invariant measure, $\mathbb{E}[h(X_t)] = \mathbb{E}[h(X_0)]$. Hence,

$$\int_0^t \mathbb{E}\left[\frac{dh}{dx}(X_u)f(X_u) + \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)\right] du = 0.$$

As this holds for all $t \geq 0$, it means that

$$\begin{aligned} \mathbb{E}\left[\frac{dh}{dx}(X_u)f(X_u) + \frac{1}{2}\frac{d^2h}{dx^2}(X_u)g^2(X_u)\right] &= 0, \\ \int_{-\infty}^{\infty} \left\{\frac{dh}{dx}(x)f(x) + \frac{1}{2}\frac{d^2h}{dx^2}(x)g^2(x)\right\} \pi(x)dx &= 0. \end{aligned} \tag{2.6}$$

Using integration by parts and the fact that $f$ is compact,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dh}{dx}(x)f(x)\pi(x)dx &= -\int_{-\infty}^{\infty} h(x)\frac{d[f(x)\pi(x)]}{dx}dx, \\ \int_{-\infty}^{\infty} \frac{d^2h}{dx^2}g^2(x)\pi(x)dx &= -\int_{-\infty}^{\infty} \frac{dh}{dx}(x)\frac{d\left[g^2(x)\pi(x)\right]}{dx}dx \\ &= \int_{-\infty}^{\infty} h(x)\frac{d^2\left[g^2(x)\pi(x)\right]}{dx^2}dx. \end{aligned}$$

Substituting into equation (2.6):

$$\int_{-\infty}^{\infty}\left\{-\frac{d[f(x)\pi(x)]}{dx}+\frac{1}{2}\frac{d^2\left[g^2(x)\pi(x)\right]}{dx^2}\right\}h(x)dx=0.$$

As $h$ was arbitrary, it means that we must have

$$\frac{d^2\left[g^2(x)\pi(x)\right]}{dx^2}=2\frac{d[f(x)\pi(x)]}{dx}.$$

Integrating both sides with respect to $x$,

$$\frac{d\left[g^2(x)\pi(x)\right]}{dx}=2f(x)\pi(x)+c_1,$$

where $c_1$ is some constant. Writing $q(x)=g^2(x)\pi(x)$,

$$\begin{aligned}
\frac{d}{dx}q(x)-\frac{2f(x)}{g^2(x)}q(x)&=c_1,\\
\left[\frac{d}{dx}q(x)-\frac{2f(x)}{g^2(x)}q(x)\right]\exp\left[\int_a^x-\frac{2f(x')}{g^2(x')}dx'\right]&=c_1\exp\left[\int_a^x-\frac{2f(x')}{g^2(x')}dx'\right],\\
\frac{d}{dx}\left\{q(x)\exp\left[\int_a^x-\frac{2f(x')}{g^2(x')}dx'\right]\right\}&=c_1\exp\left[\int_a^x-\frac{2f(x')}{g^2(x')}dx'\right].
\end{aligned}$$

Integrating both sides from 0 to $x$,

$$q(x)\exp\left[\int_a^x-\frac{2f(x')}{g^2(x')}dx'\right]=c_1\int_0^x\exp\left[\int_a^{x''}-\frac{2f(x')}{g^2(x')}dx'\right]dx''+N,\qquad (N\text{ constant})$$

$$q(x)=N\exp\left[\int_a^x\frac{2f(x')}{g^2(x')}dx'\right]+c_1\int_0^x\exp\left[\int_{x''}^x\frac{2f(x')}{g^2(x')}dx'\right]dx''.$$

Substituting $x=1$ into the above equation:

$$0=c_1\int_0^1\exp\left[\int_{x''}^1\frac{2f(x')}{g^2(x')}dx'\right]dx''.$$

As $\frac{2f(x')}{g^2(x')}$ is finite for all $x'\in(0,1)$, $\int_0^1\exp\left[\int_{x''}^1\frac{2f(x')}{g^2(x')}dx'\right]dx''$ must be positive. This means that $c_1=0$. As such,

$$\pi(x)=\frac{N}{g^2(x)}\exp\left[\int_a^x\frac{2f(x')}{g^2(x')}dx'\right],$$

where

$$\frac{1}{N}=\int_0^1\frac{1}{g^2(x)}\exp\left[\int_a^x\frac{2f(x')}{g^2(x')}dx'\right]dx.$$

Now, the density $\pi$ is well-defined if and only if $N$ is finite and non-zero, but $N$ is indeed finite and non-zero by assumption. $\square$

Invariant probability measures are useful because they lie at the heart of ergodic theory. The following theorem shows that, in some sense, long-run time averages of functions of $X$ are determined by the invariant probability measure of $X$.

**Theorem 2.4.3.** *Let $X$ be the solution to equation (2.5). Assume that $X$ has a unique invariant probability measure that has density $\pi$. Let $h$ be a function integrable w.r.t. the invariant probability measure. Then, almost surely,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T h(X_t)dt = \int_{-\infty}^{\infty} h(x)\pi(x)dx.$$

*In particular,*

$$\mathbb{E}\left[\lim_{T\to\infty} \frac{1}{T}\int_0^T h(X_t)dt\right] = \int_{-\infty}^{\infty} h(x)\pi(x)dx.$$

*Proof.* See Theorem 5.1 of Chapter IV in Has'minskiĭ (25). □

Ergodic theory in the case where $X$ is a two-state continuous-time Markov chain is significantly easier. Let the state space of $X$ be $\{a, b\}$, and let the transition intensities matrix of $X$ be given by

$$\Lambda = \begin{bmatrix} -\lambda_b & \lambda_b \\ \lambda_a & -\lambda_a \end{bmatrix},$$

with $(\lambda_a, \lambda_b) \neq (0, 0)$. The assumption on the $\lambda$'s ensures that $X$ is dependent on $t$. While the labeling of the $\lambda$'s seems a little counterintuitive, they are labeled as such so that for each state $i$ with $i = a, b$, when the Markov chain enters the other state, the time taken for the Markov chain to return to state $i$ is an exponential random variable with rate $\lambda_i$.

In this case, invariant probability measures for $X$ can be defined in the following way (which is equivalent to Definition 2.4.1):

**Definition 2.4.4.** *Let $\pi = \begin{bmatrix}\pi_1 & \pi_2\end{bmatrix}$ be a $1 \times 2$ vector such that $\pi_1 + \pi_2 = 1$. Then $\pi$ is an* ***invariant probability measure*** *for $X$ if and only if*

$$\pi\Lambda = 0.$$

From the definition, it is clear that $X$ has a unique invariant probability measure given by

$$\pi_a = \frac{\lambda_a}{\lambda_a + \lambda_b}, \quad \pi_b = \frac{\lambda_b}{\lambda_a + \lambda_b},$$

and Theorem 2.4.3 reduces to the following:

**Theorem 2.4.5.** (See Theorem 3.8.1 of Norris (41)). *For any function $f : \{a, b\} \mapsto \mathbb{R}$,*

$$\mathbb{P}\left\{\lim_{T\to\infty} \frac{1}{T}\int_0^T f(X_s)ds = \frac{\lambda_a f(a) + \lambda_b f(b)}{\lambda_a + \lambda_b}\right\} = 1.$$

## 2.5 Modified Bessel Functions of the Second Kind

In this section we will present properties of the modified Bessel function of the second kind which we will use in Chapter 6.

For $\nu \in \mathbb{R}$, let $K_\nu(z)$ denote the modified Bessel function of the second kind of order $\nu$. By formula 8.486.16 of Gradshteyn & Ryzhik (21), for any $\nu$, $K_{-\nu}(z) = K_\nu(z)$. Thus, we only have to consider have to consider non-negative orders of the Bessel function. In this thesis, we will only consider $z \in \mathbb{R}_+$.

**Proposition 2.5.1.** (See page 435 of Olver (42).) *Let $\nu \geq 0$, $z > 0$. Then,*

1. *$K_\nu(z)$ is a continuous positive function of both $\nu$ and $z$.*
2. *For fixed $z$, $K_\nu(z)$ increases as $\nu$ increases.*

**Proposition 2.5.2.** (Asymptotic behavior of $K_\nu(z)$ as $z \to 0$, for fixed $\nu$.) *As $z \to 0$,*

$$K_\nu(z) \sim \frac{\Gamma(\nu)}{2\left(\frac{z}{2}\right)^\nu}, \quad \nu > 0,$$

*and*

$$K_0(z) \sim \log\left(\frac{1}{z}\right).$$

*Here, $f(z) \sim g(z)$ as $z \to 0$ means that $\lim_{z\to 0} \frac{f(z)}{g(z)} = 1$.*

The following proposition shows, in a precise manner, that as $z \to 0$, $K_\nu(z)$ approaches $\infty$ much faster for larger orders of $\nu$.

**Proposition 2.5.3.** *Fix $\nu_1 > \nu_2 \geq 0$. Then $\frac{K_{\nu_1}(z)}{K_{\nu_2}(z)} \to \infty$ as $z \to 0$.*

*Proof.* *Case 1:* $\nu_2 > 0$. Then by Proposition 2.5.2,

$$\begin{aligned}
\lim_{z\to 0} \frac{K_{\nu_1}(z)}{K_{\nu_2}(z)} &= \lim_{z\to 0} \frac{\Gamma(\nu_1)}{2\left(\frac{z}{2}\right)^{\nu_1}} \frac{2\left(\frac{z}{2}\right)^{\nu_2}}{\Gamma(\nu_2)} \\
&= C \lim_{z\to 0} z^{\nu_2-\nu_1} && (C \text{ some positive constant}) \\
&= \infty. && (\text{as } \nu_2 - \nu_1 < 0)
\end{aligned}$$

*Case 2:* $\nu_2 = 0$. Then by Proposition 2.5.2,

$$\begin{aligned}
\lim_{z\to 0} \frac{K_{\nu_1}(z)}{K_{\nu_2}(z)} &= \lim_{z\to 0} \frac{\Gamma(\nu_1)}{2\left(\frac{z}{2}\right)^{\nu_1}} \frac{1}{\log\left(\frac{1}{z}\right)} \\
&= C \lim_{z\to\infty} z^{\nu_1} \frac{1}{\log z} && (C \text{ some positive constant}) \\
&= \infty,
\end{aligned}$$

as the logarithm of $z$ grows more slowly than any positive power of $z$. □

The following is a technical lemma that will be used in Chapter 6.

**Lemma 2.5.4.** *Let $b$ be some real constant. Then,*

$$\lim_{z \to 0} \frac{K_{1+bz}(z)}{K_{bz}(z)} = \infty.$$

*Proof.* First, assume that $b \neq 0$. Pick some $\varepsilon_0$ such that $0 < \varepsilon_0 \ll \frac{1}{2}$. Then for all $z$ such that $0 < z \leq \frac{\varepsilon_0}{|b|}$,

$$|bz| \leq \varepsilon_0,$$
$$\Rightarrow 1 + bz \geq 1 - \varepsilon_0, \quad bz \leq \varepsilon_0.$$

Applying Proposition 2.5.1, for $0 < z \leq \frac{\varepsilon_0}{|b|}$,

$$\frac{K_{1+bz}(z)}{K_{bz}(z)} \geq \frac{K_{1-\varepsilon_0}(z)}{K_{\varepsilon_0}(z)}. \tag{2.7}$$

Taking limits on both sides and using the fact that $\varepsilon_0 \ll \frac{1}{2}$,

$$\lim_{z \to 0} \frac{K_{1+bz}(z)}{K_{bz}(z)} \geq \lim_{z \to 0} \frac{K_{1-\varepsilon_0}(z)}{K_{\varepsilon_0}(z)} = \infty. \qquad \text{(by Prop 2.5.3)}$$

Note that when $b = 0$, equation (2.7) still holds, and so the conclusion is true for $b = 0$ as well. □

We end this section with asymptotics for $K_0$ and $K_1$ as $z$ goes to zero, which are more precise compared to the asymptotics given in Proposition 2.5.2. This result will be used when calculating the asymptotic for the long-run discounted growth rate in Section 6.7. Before presenting the asymptotics for $K_0$ and $K_1$, we present asymptotics for $I_0$ and $I_1$, which are the modified Bessel functions of the first kind with orders 0 and 1 respectively.

**Lemma 2.5.5.** *As $z$ goes to zero,*

$$I_0(z) = 1 + O(z^2), \qquad \textit{and} \qquad I_1(z) = O(z).$$

*Here, $O$ represents the big-O notation, that is, if $f(z) = O(g(z))$ as $z$ goes to zero, then there exist constants $M$ and $\varepsilon > 0$ such that for all $z$ with $0 < z < \varepsilon$, $\left|\frac{f(z)}{g(z)}\right| < M$.*

*Proof.* From formula 9.6.12 of Abramowitz (1),

$$I_0(z) = 1 + \sum_{k=1}^{\infty} \frac{1}{(k!)^2} \left(\frac{z^2}{4}\right)^k.$$

As such, for $0 < z < \frac{1}{2}$,

$$\begin{aligned}\left|\frac{I_0(z)-1}{z^2}\right| &= \frac{1}{4}\sum_{k=1}^{\infty}\frac{1}{(k!)^2}\left(\frac{z^2}{4}\right)^{k-1}\\ &\leq \frac{1}{4}+\left(z^2+z^4+z^6+\dots\right)\\ &= \frac{1}{4}+\frac{z^2}{1-z^2}\\ &< \frac{1}{4}+\frac{1/4}{3/4}=\frac{7}{12}.\end{aligned}$$

Thus $I_0(z) = 1 + O(z^2)$ as $z$ goes to zero. From formula 9.6.10 of Abramowitz (1),

$$I_1(z) = \frac{z}{2}\sum_{k=0}^{\infty}\frac{1}{k!\,\Gamma(k+2)}\left(\frac{z^2}{4}\right)^k = \frac{z}{2}\sum_{k=0}^{\infty}\frac{1}{k!(k+1)!}\left(\frac{z^2}{4}\right)^k,$$

where $\Gamma$ is the gamma function. Thus, for $0 < z < \frac{1}{2}$,

$$\begin{aligned}\left|\frac{I_1(z)}{z}\right| &= \frac{1}{2}\left[1+\sum_{k=1}^{\infty}\frac{1}{k!(k+1)!}\left(\frac{z^2}{4}\right)^k\right]\\ &\leq \frac{1}{2}+\left(z^2+z^4+z^6+\dots\right)\\ &< \frac{1}{2}+\frac{1/4}{3/4}=\frac{5}{6}.\end{aligned}$$

Hence, $I_1(z) = O(z)$ as $z$ goes to zero. □

**Proposition 2.5.6.** *As $z$ goes to zero,*

$$K_0(z) = -\left(\log\frac{z}{2}+\gamma\right)+O(z^2\log z), \qquad \textit{and} \qquad K_1(z) = \frac{1}{z}+O(z\log z),$$

*where $\gamma$ is the Euler-Mascheroni constant.*

*Proof.* By equation 3.3.15 of Bender (6),

$$K_0(z) = -\left(\log\frac{z}{2}+\gamma\right)I_0(z)+\sum_{n=1}^{\infty}\frac{1}{(n!)^2}\left(1+\frac{1}{2}+\dots+\frac{1}{n}\right)\left(\frac{z^2}{4}\right)^n.$$

Note that for $0 < z < \frac{1}{2}$,

$$\begin{aligned}
\frac{1}{z^2}\sum_{n=1}^{\infty}\frac{1}{(n!)^2}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}\right)\left(\frac{z^2}{4}\right)^n &= \frac{1}{4}+\frac{1}{4}\sum_{n=2}^{\infty}\frac{1}{(n!)^2}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}\right)\left(\frac{z^2}{4}\right)^{n-1} \\
&< \frac{1}{4}+\frac{1}{4}\sum_{n=2}^{\infty}\frac{n}{(n!)^2}\left(\frac{z^2}{4}\right)^{n-1} \\
&< \frac{1}{4}+(z^2+z^4+\dots) \\
&< \frac{1}{4}+\frac{1/4}{3/4}=\frac{7}{12}.
\end{aligned}$$

Hence, as $z$ goes to zero,

$$\begin{aligned}
K_0(z) &= -\left(\log\frac{z}{2}+\gamma\right)\left[1+O(z^2)\right]+O(z^2) \\
&= -\left(\log\frac{z}{2}+\gamma\right)+O(z^2\log z).
\end{aligned}$$

By equation 3.3.21 of Bender (6),

$$K_1(z)=\left(\log\frac{z}{2}+\gamma\right)I_1(z)+\frac{1}{z}-\frac{z}{4}-\frac{z}{2}\sum_{n=1}^{\infty}\frac{\left(\frac{z^2}{4}\right)^n}{n!(n+1)!}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}+\frac{1}{2n+2}\right).$$

Note that for $0 < z < \frac{1}{2}$,

$$\begin{aligned}
\frac{1}{z^3}\frac{z}{2}\sum_{n=1}^{\infty}\frac{\left(\frac{z^2}{4}\right)^n}{n!(n+1)!}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}+\frac{1}{2n+2}\right) &< \frac{1}{2z^2}\sum_{n=1}^{\infty}\frac{n+1}{n!(n+1)!}\left(\frac{z^2}{4}\right)^n \\
&< \frac{1}{2z^2}\sum_{n=1}^{\infty}z^{2n} \\
&< \frac{1}{2}+(z^2+z^4+\dots),
\end{aligned}$$

which is finite. Hence, as $z$ goes to zero,

$$\begin{aligned}
K_1(z) &= \left(\log\frac{z}{2}+\gamma\right)O(z)+\frac{1}{z}-\frac{z}{4}+O(z^3) \\
&= \frac{1}{z}+O(z\log z),
\end{aligned}$$

as required. $\square$

# Chapter 3

# Filtering Theory

In this chapter, we first set up the filtering problem in continuous-time before presenting certain theorems which we will apply (in later chapters) to the logarithmic utility optimization problem in the regime-switching model. Several of the results here will be multi-dimensional versions of theorems presented in Chigansky (11). (For a deeper understanding of stochastic filtering theory, see Xiong (57).)

## 3.1 Setting up the Filtering Problem

Assume that there is an underlying complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Fix some terminal time $T \in \mathbb{R}_+$, and a parameter $n$, which is a positive integer. Let $X$ and $\mathbf{Y}$ be stochastic processes defined on the probability space with the following properties:

1. $X$ is a real-valued continuous-time Markov chain on the time interval [0,T] which is adapted to $\{\mathcal{F}_t\}$. Assume that $X$ has a finite state space. (Assume that $X$ is given by equation (2.2) and the associated construction presented in Section 2.3.)

2. $\mathbf{Y}$ takes values in $\mathbb{R}^n$.

3. Let the $i^{\text{th}}$ component of $\mathbf{Y}$ be the stochastic process denoted by $Y^{(i)}$. Then for each $i = 1, \ldots, n$, there exists a measurable function $g^{(i)} : \mathbb{R}^+ \times \mathbb{R} \mapsto \mathbb{R}$ such that the process $g^{(i)}(\cdot, X.)$ is in $\mathcal{H}^2[0,T]$, and there exists an $\{\mathcal{F}_t\}$-Wiener process $W^{(i)}$ on $[0,T]$ which is independent of $X$ such that

$$Y_t^{(i)} = \int_0^t g^{(i)}(s, X_s)ds + BW_t^{(i)},$$

   for all $t \in [0,T]$. Here, $B$ is a positive constant. In differential notation, the above would be written as

$$dY_t^{(i)} = g^{(i)}(t, X_t)dt + BdW_t^{(i)}. \tag{3.1}$$

4. For $i \neq j$, $W^{(i)}$ is independent of $W^{(j)}$.

For notational convenience, let

$$\mathbf{g} = \begin{bmatrix} g^{(1)} \\ \vdots \\ g^{(n)} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} W^{(1)} \\ \vdots \\ W^{(n)} \end{bmatrix}.$$

We can then aggregate the equations in property 3 to get

$$\mathbf{Y}_t = \int_0^t \mathbf{g}(s, X_s) ds + B\mathbf{W}_t \tag{3.2}$$

for all $t \in [0, T]$. Note that $\mathbf{W}$ is an $n$-dimensional $\{\mathcal{F}_t\}$-Wiener process, and that $X$ and $\mathbf{Y}$ are both adapted to the filtration $\{\mathcal{F}_t\}$.

We can interpret the preceding set-up in the following way: $X$ is a ***signal process***, i.e. it represents, in some way, the state of the system we are interested in. We cannot see $X$ directly. Instead, we can see $n$ ***observation processes***, modeled by $\mathbf{Y}$, which depend on $X$ in some way (more precisely, $\mathbf{Y}$ depends on $X$ through equation (3.2)). Note that $X$ cannot be calculated deterministically given $\mathbf{Y}$: the Wiener process $\mathbf{W}$ in equation (3.2) models the fact that there might be noise present in the observation processes.

With such a set-up, we can ask ourselves the following question: given that we know the values of the observation processes up till time $t$ (time $t$ included), what is our best guess of the value of the signal process at time $t$?

Note that the set-up makes certain implicit assumptions. For example, the amount of noise present in the observation processes does not depend on the state of the system. In some cases such assumptions might be unrealistic. However, these assumptions simplify the model and are plausible for the problem that we are working with in this thesis.

## 3.2 Information Known at Time $t$

Let $\mathbf{Y} : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))^{\mathbb{R}_+}$ be a continuous-time stochastic process taking values in $\mathbb{R}^n$. Recalling Definition 2.1.2, $\{\mathcal{F}_t^{\mathbf{Y}}\}_{t \geq 0}$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$, where

$$\mathcal{F}_t^{\mathbf{Y}} = \sigma\{\mathbf{Y}_s : 0 \leq s \leq t\}.$$

Intuitively, $\mathcal{F}_t^{\mathbf{Y}}$ represents the information that we learn from the process $\mathbf{Y}$ from time 0 to time $t$. This notion of "information" can be made more concrete: see Section 4 in Chapter 2 of Cinlar (12).

The following proposition relates the filtration generated by an $n$-dimensional stochastic process with the filtrations generated by each of its components.

**Proposition 3.2.1.** *Let $\mathbf{Y}$ be a stochastic process taking values on $\mathbb{R}^n$, for some $n \in \mathbb{N}$.*

*Denote the components of* $\mathbf{Y}$ *by* $Y^{(i)}$*, for* $i = 1, \ldots, n$*. Then, for any time* $t$*,*

$$\mathcal{F}_t^{\mathbf{Y}} = \mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}} = \mathcal{H}_t,$$

*where* $\mathcal{H}_t$ *is the smallest* $\sigma$*-algebra containing* $\sigma Y_s^{(i)}$*, for all* $s$ *and* $i$ *such that* $0 \leq s \leq t$*,* $i = 1, \ldots, n$*.*

*Proof.* We will repeatedly use the fact that if $\mathcal{C}$ is a subset of a $\sigma$-algebra $\mathcal{G}$, then $\sigma\mathcal{C} \subset \mathcal{G}$.

For any fixed $i$, $\sigma\{Y_s^{(i)} : 0 \leq s \leq t\}$ is the smallest $\sigma$-algebra containing $\sigma Y_s^{(i)}$ for all $s \in [0, t]$. Since $\mathcal{H}_t$ is another $\sigma$-algebra containing $\sigma Y_s^{(i)}$ for all $s \leq t$,

$$\begin{aligned} \sigma\{Y_s^{(i)} : 0 \leq s \leq t\} &\subset \mathcal{H}_t \quad \forall i, \\ \Leftrightarrow \mathcal{F}_t^{Y^{(i)}} &\subset \mathcal{H}_t \quad \forall i, && \text{(by definition of } \mathcal{F}_t^{Y^{(i)}}) \\ \Rightarrow \mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}} &\subset \mathcal{H}_t. \end{aligned}$$

But it is clear that $\sigma Y_s^{(i)} \subset \mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}}$ for all $s$ and $i$. Thus, by definition of $\mathcal{H}_t$, $\mathcal{H}_t$ is contained in $\mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}}$. Hence, we conclude that $\mathcal{H}_t = \mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}}$.

For the other equation,

$$\begin{aligned} \sigma Y_s^{(1)}, \ldots, \sigma Y_s^{(n)} &\subset \mathcal{H}_t \quad \forall\, 0 \leq s \leq t, \\ \Rightarrow \sigma Y_s^{(1)} \vee \cdots \vee \sigma Y_s^{(n)} &\subset \mathcal{H}_t \quad \forall\, 0 \leq s \leq t, \\ \Rightarrow \bigvee_{0 \leq s \leq t} \sigma Y_s^{(1)} \vee \cdots \vee \sigma Y_s^{(n)} &\subset \mathcal{H}_t, \\ \Rightarrow \bigvee_{0 \leq s \leq t} \sigma \mathbf{Y}_s &\subset \mathcal{H}_t, && \text{(by Prop II.4.3 of Cinlar (12))} \\ \Rightarrow \mathcal{F}_t^{\mathbf{Y}} &\subset \mathcal{H}_t. && \text{(by definition of } \mathcal{F}_t^{\mathbf{Y}}) \end{aligned}$$

But it is clear that $\sigma Y_s^{(i)} \subset \bigvee_{0 \leq s \leq t} \left[\sigma Y_s^{(1)} \vee \cdots \vee \sigma Y_s^{(n)}\right] = \mathcal{F}_t^{\mathbf{Y}}$ for all $s$ and $i$, and so by the definition of $\mathcal{H}_t$, $\mathcal{H}_t \subset \mathcal{F}_t^{\mathbf{Y}}$. In conclusion, we have

$$\mathcal{F}_t^{Y^{(1)}} \vee \cdots \vee \mathcal{F}_t^{Y^{(n)}} = \mathcal{H}_t = \mathcal{F}_t^{\mathbf{Y}}. \qquad \square$$

Intuitively, the proposition above makes sense. If we had the means to get $n$ different observation signals at each time, the amount of "information" about the system that we have at a certain time should be the same, whether the observation signals are presented to us as $n$ seperate real signals or as one vector signal.

## 3.3 The Reference Measure Approach

The observation process $\mathbf{Y}$ consists of two components: the deterministic portion dependent on the signal process $X$, and noise, modeled by the Wiener process $\mathbf{W}$. The underlying idea of the reference measure approach is the use of the observation process for introducing a new reference probability. Under this reference probability measure, we are somehow able to remove the noise introduced by the Wiener process, thus gaining some idea of what the underlying signal process looks like.

Fundamental to this approach is Girsanov's theorem, which we present in its multi-dimensional form:

**Theorem 3.3.1.** (Multi-dimensional Girsanov Theorem, see Theorem 7.1.3 of Kallianpur (28)). *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be a complete filtered probability space. Let $\mathbf{W}$ be an $n$-dimensional $\{\mathcal{F}_t\}$-Wiener process. Let $\mathbf{f} = (\mathbf{f}_t)_{t\geq 0}$ be an $n$-dimensional measurable process adapted to $\{\mathcal{F}_t\}$ such that*

$$\mathbb{P}\left\{\int_0^T \|\mathbf{f}_t\|^2 dt < \infty\right\} = 1.$$

*(Here, $\|\cdot\|$ represents the Euclidean norm on $\mathbb{R}^n$.) Define*

$$M_t(\mathbf{f}) = \exp\left[\int_0^t \mathbf{f}_s \cdot d\mathbf{W}_s - \frac{1}{2}\int_0^t \|\mathbf{f}_s\|^2 ds\right],$$

*where $\int_0^t \mathbf{f}_s \cdot d\mathbf{W}_s = \sum_{i=1}^N \int_0^t f_s^{(i)} dW_s^{(i)}$. Assume that $\mathbb{E}M_T(\mathbf{f}) = 1$. Let $\tilde{\mathbb{P}}$ be the probability measure given by $d\tilde{\mathbb{P}} = M_T(\mathbf{f})d\mathbb{P}$. If $\tilde{\mathbf{W}}$ is given by*

$$\tilde{\mathbf{W}}_t = \mathbf{W}_t - \int_0^t \mathbf{f}_s ds,$$

*then $\tilde{\mathbf{W}}$ is an $n$-dimensional Wiener process w.r.t. $\{\mathcal{F}_t\}$ under the probability measure $\tilde{\mathbb{P}}$. In addition, $M(\mathbf{f})$ is a martingale w.r.t. $\mathbb{P}$.*

The Kallianpur-Striebel formula, which is the main result of this section, gives an explicit expression for the expected value of a function depending solely on $X_t$ given the information $\mathcal{F}_t^{\mathbf{Y}}$, i.e. $\mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right]$. Before stating the Kallianpur-Striebel formula, we will present a few lemmas which will be used in its proof.

**Lemma 3.3.2.** (Bayes formula, see pages 174–5 of van Handel (55)). *Let $(\Omega, \mathcal{F})$ be a measurable space on which two equivalent measures, $\mathbb{P}$ and $\tilde{\mathbb{P}}$, are defined. Let $\mathcal{G}$ be a sub $\sigma$-algebra of $F$. Then for any integrable $X$,*

$$\mathbb{E}[X|\mathcal{G}] = \frac{\tilde{\mathbb{E}}\left[X\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}|\mathcal{G}\right]}{\tilde{\mathbb{E}}\left[\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}|\mathcal{G}\right]}.$$

*Here, $\tilde{\mathbb{E}}$ denotes expectation w.r.t. the probability measure $\tilde{\mathbb{P}}$.*

**Lemma 3.3.3.** *Let $M = (M_t)_{t\in[0,T]}$ be a positive $\{\mathcal{F}_t\}$-martingale w.r.t. $\mathbb{P}$. Define a new measure $d\tilde{\mathbb{P}}$ by $d\tilde{\mathbb{P}} = M_T d\mathbb{P}$. Then $\frac{1}{M}$ is an $\{\mathcal{F}_t\}$-martingale w.r.t. $\tilde{\mathbb{P}}$.*

*Proof.* As $M_t > 0$ for all $t$, the process $\frac{1}{M}$ is well-defined. It is clear that $\frac{1}{M}$ is adapted to $\{\mathcal{F}_t\}$. For each $t$,

$$\begin{aligned}
\tilde{\mathbb{E}}\left[\left|\frac{1}{M_t}\right|\right] &= \mathbb{E}\left[\frac{M_T}{M_t}\right] && \text{(by definition of } \tilde{\mathbb{P}}\text{, and } M > 0\text{)}\\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{M_T}{M_t}\middle|\mathcal{F}_t\right]\right]\\
&= \mathbb{E}\left[\frac{1}{M_t}\mathbb{E}[M_T \mid \mathcal{F}_t]\right] && \text{(as } M_t \in \mathcal{F}_t\text{)}\\
&= \mathbb{E}\left[\frac{1}{M_t}M_t\right] && \text{(as } M \text{ is an } \{\mathcal{F}_t\}\text{-martingale)}\\
&= 1 < \infty,
\end{aligned}$$

hence $\frac{1}{M_t}$ is integrable for all $t$. Now, fix $0 \leq s \leq t$. Then

$$\begin{aligned}
\tilde{\mathbb{E}}\left[\frac{1}{M_t} \;\middle|\; \mathcal{F}_s\right] &= \frac{\mathbb{E}\left[\frac{1}{M_t}\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\middle|\mathcal{F}_s\right]}{\mathbb{E}\left[\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\middle|\mathcal{F}_s\right]} && \text{(by Bayes formula)}\\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[\frac{M_T}{M_t}\middle|\mathcal{F}_t\right]\middle|\mathcal{F}_s\right]}{\mathbb{E}\left[M_T \mid \mathcal{F}_s\right]}\\
&= \frac{\mathbb{E}\left[1 \mid \mathcal{F}_s\right]}{\mathbb{E}\left[M_T \mid \mathcal{F}_s\right]} && \text{(as above)}\\
&= \frac{1}{M_s}. && \text{(as } M \text{ is an } \{\mathcal{F}_t\}\text{-martingale)}
\end{aligned}$$

In conclusion, $\frac{1}{M}$ is an $\{\mathcal{F}_t\}$-martingale w.r.t. $\tilde{\mathbb{P}}$. □

**Lemma 3.3.4.** (See Lemma 11.3.1 of Kallianpur (28)). *Let $(\Omega, \mathcal{F}, \{F_t\}_{t\in[0,T]}, \mathbb{P})$ be a complete filtered probability space. Let $\mathbf{W}$ be an $n$-dimensional $\{\mathcal{F}_t\}$-Wiener process on $[0,T]$, and let $(\mathbf{g}_t)_{t\in[0,T]}$ be a measurable process (in the sense of Definition 2.2.1) taking values in $\mathbb{R}^n$. let $B$ be some positive real constant. Assume that $\mathbb{P}\left\{\omega : \int_0^T \mathbf{g}_t(\omega)^2 dt < \infty\right\} = 1$, and that processes $\mathbf{g}$ and $\mathbf{W}$ are independent of each other. Then*

$$\mathbb{E}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^T \|\mathbf{g}_s\|^2 ds\right] = 1. \tag{3.3}$$

*If, in addition, $\mathbf{g}$ is adapted to the filtration $\{\mathcal{F}_t\}$, then by defining the probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ by*

$$d\tilde{\mathbb{P}} = \exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^T \|\mathbf{g}_s\|^2 ds\right] d\mathbb{P},$$

*The process $\tilde{\mathbf{W}}$ given by*

$$\tilde{\mathbf{W}}_t = \mathbf{W}_t - \frac{1}{B}\int_0^t \mathbf{g}_s ds$$

*is an $\{\mathcal{F}_t\}$-Wiener process under the probability measure $\tilde{\mathbb{P}}$.*

*Proof.* See page 67 in the Appendix. $\square$

**Theorem 3.3.5.** (Multi-dimensional Kallianpur-Striebel Formula, see Chigansky (11)). *Consider the set-up in Section 3.1. Let $(\breve{\Omega}, \breve{\mathcal{F}}, \breve{\mathbb{P}})$ be a copy of $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for any bounded measurable $f : \mathbb{R} \mapsto \mathbb{R}$ and for any $t \in [0, T]$,*

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right](\omega) = \frac{\breve{\mathbb{E}} f\left(X_t(\breve{\omega})\right) \psi_t\left(X(\breve{\omega}), \mathbf{Y}(\omega)\right)}{\breve{\mathbb{E}} \psi_t\left(X(\breve{\omega}), \mathbf{Y}(\omega)\right)}, \quad \mathbb{P} - a.s.,$$

*where*

$$\psi_t(X, \mathbf{Y}) = \exp\left\{\frac{1}{B^2}\int_0^t \mathbf{g}(s, X_s) \cdot d\mathbf{Y}_s - \frac{1}{2B^2}\int_0^t \|\mathbf{g}(s, X_s)\|^2 ds\right\}.$$

*Proof.* For $t \in [0, T]$, let

$$z_t(X, \mathbf{W}) = \exp\left\{-\frac{1}{B}\int_0^t \mathbf{g}(s, X_s) \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^t \|\mathbf{g}(s, X_s)\|^2 ds\right\}.$$

Note that for any $t$,

$$\begin{aligned}
z_t(X, \mathbf{W})^{-1} &= \exp\left\{\frac{1}{B}\int_0^t \mathbf{g}(s, X_s) \cdot d\mathbf{W}_s + \frac{1}{2B^2}\int_0^t \|\mathbf{g}(s, X_s)\|^2 ds\right\} \\
&= \exp\left\{\frac{1}{B}\int_0^t \mathbf{g}(s, X_s) \cdot \left[\frac{1}{B}\Big(d\mathbf{Y}_s - \mathbf{g}(s, X_s)ds\Big)\right]\right. \\
&\qquad \left. + \frac{1}{2B^2}\int_0^t \|\mathbf{g}(s, X_s)\|^2 ds\right\} \qquad (\text{as } d\mathbf{Y}_s = \mathbf{g}(s, X_s)ds + Bd\mathbf{W}_s) \\
&= \exp\left\{\frac{1}{B^2}\int_0^t \mathbf{g}(s, X_s) \cdot d\mathbf{Y}_s - \frac{1}{2B^2}\int_0^t \|\mathbf{g}(s, X_s)\|^2 ds\right\} \\
&= \psi_t(X, \mathbf{Y}).
\end{aligned}$$

By applying Lemma 3.3.4 to $-\mathbf{g}$ and $\mathbf{W}$, we have $\mathbb{E} z_T(X, \mathbf{W}) = 1$, and the probability measure $\tilde{\mathbb{P}}$ defined by

$$d\tilde{\mathbb{P}} = z_T(X, \mathbf{W})d\mathbb{P}$$

is a probability measure on $(\Omega, \mathcal{F})$. Using the fact that $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent,

$$\begin{aligned}
\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}} &= \left(\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right)^{-1} \\
&= z_T(X, \mathbf{W})^{-1} = \psi_T(X, \mathbf{Y}).
\end{aligned}$$

By Girsanov's theorem, $(z_t(X, \mathbf{W}))$ is an $\{\mathcal{F}_t\}$-martingale w.r.t. $\mathbb{P}$. By Lemma 3.3.3, $(\psi_t(X, \mathbf{Y}))$ is an $\{\mathcal{F}_t\}$-martingale w.r.t. $\tilde{\mathbb{P}}$.

Let $D_{[0,T]}$ be the space of càdlàg paths on $[0, T]$ taking values in $\mathbb{R}$. (A càdlàg path is a path that is everywhere right-continuous and has left limits everywhere.) For each $x \in D_{[0,T]}$, define process $\mathbf{Y}^x$ by

$$\mathbf{Y}_t^x = \frac{1}{B}\int_0^t \mathbf{g}(s, x_s)ds + \mathbf{W}_t, \qquad t \in [0,T].$$

By Girsanov's theorem, $\mathbf{Y}^x$ is an $n$-dimensional Wiener process under $\tilde{\mathbb{P}}$. Thus, for any bounded measurable functional $\Psi : C^n_{[0,T]} \mapsto \mathbb{R}$, where $C^n_{[0,T]}$ is the space of continuous paths on $[0,T]$ taking values in $\mathbb{R}^n$,

$$\mathbb{E}\left[z_T(x, \mathbf{W})\Psi(\mathbf{Y}^x)\right] = \int_{C^n_{[0,T]}} \Psi(y)\mu^{\mathbf{W}}(dy), \quad \mu^X - a.s., \tag{3.4}$$

where $\mu^{\mathbf{W}}$ is the Wiener measure on $C^n_{[0,T]}$, and $\mu^X$ is the probability measure on $D_{[0,T]}$ induced by $X$. Thus, for any bounded measurable functionals $\Psi$ and $\Phi$,

$$\begin{aligned}
\tilde{\mathbb{E}}\left[\Psi(\mathbf{Y})\Phi(X)\right] &= \mathbb{E}\left[z_T(X, \mathbf{W})\Psi(\mathbf{Y})\Phi(X)\right] && \text{(by definition of } \tilde{\mathbb{P}}) \\
&= \mathbb{E}\left[\mathbb{E}\left[z_T(X, \mathbf{W})\Psi(\mathbf{Y})\Phi(X) \mid \mathcal{F}_T^X\right]\right] \\
&= \mathbb{E}\left[\Phi(X)\mathbb{E}\left[z_T(X, \mathbf{W})\Psi(\mathbf{Y}) \mid \mathcal{F}_T^X\right]\right] \\
&= \int_{D_{[0,T]}} \Phi(x)\mathbb{E}\left[z_T(x, \mathbf{W})\Psi(\mathbf{Y}^x)\right]\mu^X(dx).
\end{aligned}$$

By equation (3.4) and the fact that $X$ and $\mathbf{W}$ are independent under $\mathbb{P}$,

$$\tilde{\mathbb{E}}\left[\Psi(\mathbf{Y})\Phi(X)\right] = \left[\int_{C^n_{[0,T]}} \Psi(y)\mu^{\mathbf{W}}(dy)\right]\left[\int_{D_{[0,T]}} \Phi(x)\mu^X(dx)\right], \tag{3.5}$$

which implies that $X$ and $\mathbf{Y}$ are independent under $\tilde{\mathbb{P}}$. Setting $\Phi = 1$, $\Psi$ arbitrary in equation (3.5) shows that $\mathbf{Y}$ is an $n$-dimensional Wiener process under $\tilde{\mathbb{P}}$. Setting $\Psi = 1$, $\Phi$ arbitrary in equation (3.5) shows that $X$ has the same distribution under $\mathbb{P}$ and $\tilde{\mathbb{P}}$. As such,

$$\begin{aligned}
\mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right] &= \frac{\tilde{\mathbb{E}}\left[f(X_t)\psi_T(X, \mathbf{Y}) \mid \mathcal{F}_t^{\mathbf{Y}}\right]}{\tilde{\mathbb{E}}\left[\psi_T(X, \mathbf{Y}) \mid \mathcal{F}_t^{\mathbf{Y}}\right]} && \text{(by Bayes formula)} \\
&= \frac{\tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[f(X_t)\psi_T(X, \mathbf{Y}) \mid \mathcal{F}_t\right] \mid \mathcal{F}_t^{\mathbf{Y}}\right]}{\tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[\psi_T(X, \mathbf{Y}) \mid \mathcal{F}_t\right] \mid \mathcal{F}_t^{\mathbf{Y}}\right]} && (\text{as } \mathcal{F}_t^{\mathbf{Y}} \subset \mathcal{F}_t) \\
&= \frac{\tilde{\mathbb{E}}\left[f(X_t)\psi_t(X, \mathbf{Y}) \mid \mathcal{F}_t^{\mathbf{Y}}\right]}{\tilde{\mathbb{E}}\left[\psi_t(X, \mathbf{Y}) \mid \mathcal{F}_t^{\mathbf{Y}}\right]},
\end{aligned}$$

where the last equality above holds because $X$ is adapted to $\{\mathcal{F}_t\}$ and $(\psi_t(X, \mathbf{Y}))$ is an $\{\mathcal{F}_t\}$-martingale under $\tilde{\mathbb{P}}$. Using the independence of $X$ and $\mathbf{Y}$ under $\tilde{\mathbb{P}}$, and the fact that

$X$ has the same distribution under $\mathbb{P}$ and $\tilde{\mathbb{P}}$,

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right](\omega) = \frac{\breve{\mathbb{E}} f\left(X_t(\breve{\omega})\right)\psi_t\left(X(\breve{\omega}), \mathbf{Y}(\omega)\right)}{\breve{\mathbb{E}}\psi_t\left(X(\breve{\omega}), \mathbf{Y}(\omega)\right)}.$$

for all $\omega \in \Omega$. $\square$

Introduce the following notation: For a function $f : \mathbb{R} \mapsto \mathbb{R}$, let $\pi_t(f) := \mathbb{E}[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}]$, and let $\sigma_t(f) := \tilde{\mathbb{E}}\left[f(X_t)\psi_t|\mathcal{F}_t^{\mathbf{Y}}\right]$, where $\tilde{\mathbb{E}}$ and $\psi$ are as defined in Theorem 3.3.5. $\pi$ is known as the ***normalized filter***, while $\sigma$ is known as the ***unnormalized filter***. Note that for any bounded and measurable $f$, the Kallianpur-Striebel formula can be restated as

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}. \tag{3.6}$$

(Here, 1 is the function that maps everything to 1.)

Now, assume that $\mathbf{g}$ is homogeneous, i.e. for each $i$, $g^{(i)}(s, X_s) = g^{(i)}(X_s)$. Then the unnormalized filter must satisfy the following SDE:

**Theorem 3.3.6.** (Multi-dimensional Zakai equation). *Under the assumptions of Theorem 3.3.5, along with the assumption that* $\mathbf{g}$ *is homogeneous, for bounded measurable* $f : \mathbb{R} \mapsto \mathbb{R}$*,*

$$d\sigma_t(f) = \sigma_t(\Lambda f)dt + \frac{1}{B^2}\sigma_t(f\mathbf{g}) \cdot d\mathbf{Y}_t, \tag{3.7}$$

*with* $\sigma_0(f) = \mathbb{E}f(X_0)$*.*

*Proof.* For each $t$, let

$$\begin{aligned} I_t &:= \frac{1}{B^2}\int_0^t \mathbf{g}(X_s) \cdot d\mathbf{Y}_s - \frac{1}{2B^2}\int_0^t \|\mathbf{g}(X_s)\|^2 ds \\ &= \frac{1}{B^2}\sum_{i=1}^n \int_0^t g^{(i)}(X_s)dY_s^{(i)} - \frac{1}{2B^2}\sum_{i=1}^n \int_0^t g^{(i)}(X_s)^2 ds. \end{aligned}$$

In differential notation,

$$\begin{aligned} dI_t &= \frac{1}{B^2}\sum_{i=1}^n g^{(i)}(X_t)dY_t^{(i)} - \frac{1}{2B^2}\sum_{i=1}^n g^{(i)}(X_t)^2 dt \\ &= \frac{1}{B^2}\mathbf{g}(X_t) \cdot d\mathbf{Y}_t - \frac{1}{2B^2}\|\mathbf{g}(X_t)\|^2 dt. \end{aligned}$$

Using the fact that $dY_t^{(i)}dY_t^{(j)} = 0$ for all $i \neq j$,

$$\begin{aligned}(dI_t)^2 &= \frac{1}{B^4}\sum_{i=1}^n g^{(i)}(X_t)^2\left(dY_t^{(i)}\right)^2 \\ &= \frac{1}{B^4}\sum_{i=1}^n g^{(i)}(X_t)^2 B^2 dt \qquad \text{(using Eqn (3.1))} \\ &= \frac{1}{B^2}\sum_{i=1}^n g^{(i)}(X_t)^2 dt \\ &= \frac{1}{B^2}\|\mathbf{g}(X_t)\|^2 dt.\end{aligned}$$

As $\psi_t = e^{I_t}$, using Itô's formula (Theorem 2.2.7),

$$\begin{aligned}d\psi_t &= \psi_t dI_t + \frac{1}{2}\psi_t(dI_t)^2 \\ &= \psi_t\left[\frac{1}{B^2}\mathbf{g}(X_t)\cdot d\mathbf{Y}_t - \frac{1}{2B^2}\|\mathbf{g}(X_t)\|^2 dt\right] + \frac{1}{2B^2}\psi_t\|\mathbf{g}(X_t)\|^2 dt \\ &= \frac{1}{B^2}\psi_t\mathbf{g}(X_t)\cdot d\mathbf{Y}_t.\end{aligned}$$

By Itô's formula for general martingales and noting that the independence of $X$ and $W$ implies $\Rightarrow df(X_t)d\psi_t = 0$,

$$\begin{aligned}f(X_t)\psi_t &= f(X_0)\psi_0 + \int_0^t \psi_s df(X_s) + \int_0^t f(X_s)d\psi_s \\ &= f(X_0) + \int_0^t \psi_s\Lambda f(X_s)ds + \int_0^t \psi_s dM_s + \frac{1}{B^2}\int_0^t f(X_s)\mathbf{g}(X_s)\psi_s\cdot d\mathbf{Y}_s \\ &\qquad \text{(by Prop 2.3.3)} \\ &= f(X_0) + \int_0^t \psi_s\Lambda f(X_s)ds + \int_0^t \psi_s dM_s + \frac{1}{B^2}\sum_{i=1}^n\int_0^t f(X_s)g^{(i)}(X_s)\psi_s dY_s^{(i)}.\end{aligned}$$

From this point, the proof is exactly the same as in the case of the single-dimensional Zakai equation (see pages 106–7 of Chigansky (11), page 90 of Xiong (57)). $\square$

**Theorem 3.3.7.** (Fujisaki-Kallianpur-Kunita (FKK) equation). *Under the assumptions of Theorem 3.3.5, for bounded measurable $f : \mathbb{R} \mapsto \mathbb{R}$, $\pi_t(f)$ satisfies the equation*

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\Lambda f)ds + \int_0^t \frac{1}{B}\big[\pi_s(f\mathbf{g}) - \pi_s(f)\pi_s(\mathbf{g})\big]\cdot d\tilde{\mathbf{W}}_s, \qquad (3.8)$$

*where the process $\tilde{\mathbf{W}}$ is given by*

$$\tilde{\mathbf{W}}_t = \frac{1}{B}\left[\mathbf{Y}_t - \int_0^t \pi_s(\mathbf{g})ds\right].$$

*Proof.* Using the Zakai equation (equation (3.7)) for $f = 1$ and for general $f$, we have

$$d\sigma_t(1) = \frac{1}{B^2}\sigma_t(\mathbf{g}) \cdot d\mathbf{Y}_t, \quad \sigma_0(1) = 1,$$
$$d\sigma_t(f) = \sigma_t(\Lambda f)dt + \frac{1}{B^2}\sigma_t(f\mathbf{g}) \cdot d\mathbf{Y}_t, \quad \sigma_0(f) = \mathbb{E}f(X_0).$$

Recalling that $\left(dY_t^{(i)}\right)^2 = B^2 dt$ for all $i = 1, \ldots, n$, and that $dY_t^{(i)} dY_t^{(j)} = 0$ for all $i \neq j$,

$$\begin{aligned}
d\sigma_t(f)d\sigma_t(1) &= \left[\frac{1}{B^2}\sum_{i=1}^n \sigma_t(g^{(i)})dY_t^{(i)}\right]\left[\sigma_t(\Lambda f)dt + \frac{1}{B^2}\sum_{i=1}^n \sigma_t(fg^{(i)})dY_t^{(i)}\right] \\
&= \frac{1}{B^4}\sum_{i=1}^n \sigma_t(g^{(i)})\sigma_t(fg^{(i)})(B^2 dt) \\
&= \frac{1}{B^2}\sum_{i=1}^n \sigma_t(g^{(i)})\sigma_t(fg^{(i)})dt \\
&= \frac{1}{B^2}\sigma_t(\mathbf{g}) \cdot \sigma_t(f\mathbf{g})dt,
\end{aligned}$$

and

$$\begin{aligned}
(d\sigma_t(1))^2 &= \frac{1}{B^4}\sum_{i=1}^n \sigma_t(g^{(i)})^2 \left(dY_t^{(i)}\right)^2 \\
&= \frac{1}{B^2}\sum_{i=1}^n \sigma_t(g^{(i)})^2 dt \\
&= \frac{1}{B^2}\|\sigma_t(\mathbf{g})\|^2 dt.
\end{aligned}$$

From the Kallianpur-Striebel formula (equation (3.6)), $\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}$. Using Itô's formula for the function $f(x, y) = \frac{x}{y}$,

$$\begin{aligned}
d\pi_t(f) &= \frac{1}{\sigma_t(1)}d\sigma_t(f) - \frac{\sigma_t(f)}{\sigma_t(1)^2}d\sigma_t(1) - \frac{1}{\sigma_t(1)^2}d\sigma_t(f)d\sigma_t(1) + \frac{1}{2}\frac{2\sigma_t(f)}{\sigma_t(1)^3}(d\sigma_t(1))^2 \\
&= \frac{\sigma_t(\Lambda f)}{\sigma_t(1)}dt + \frac{\sigma_t(f\mathbf{g}) \cdot d\mathbf{Y}_t}{B^2\sigma_t(1)} - \frac{\sigma_t(f)\sigma_t(\mathbf{g}) \cdot d\mathbf{Y}_t}{B^2\sigma_t(1)^2} \\
&\qquad - \frac{\sigma_t(\mathbf{g}) \cdot \sigma_t(f\mathbf{g})}{B^2\sigma_t(1)^2}dt + \frac{\sigma_t(f)\|\sigma_t(\mathbf{g})\|^2}{B^2\sigma_t(1)^3}dt \\
&= \pi_t(\Lambda f)dt + \frac{\pi_t(f\mathbf{g}) \cdot d\mathbf{Y}_t}{B^2} - \frac{\pi_t(f)\pi_t(\mathbf{g}) \cdot d\mathbf{Y}_t}{B^2} \\
&\qquad - \frac{\pi_t(\mathbf{g}) \cdot \pi_t(f\mathbf{g})}{B^2}dt + \frac{\pi_t(f)\pi_t(\mathbf{g}) \cdot \pi_t(\mathbf{g})}{B^2}dt \qquad \text{(by Eqn (3.6))} \\
&= \pi_t(\Lambda f)dt + \frac{1}{B^2}\Big\{\big[\pi_t(f\mathbf{g}) - \pi_t(f)\pi_t(\mathbf{g})\big] \cdot d\mathbf{Y}_t \\
&\qquad - \big[\pi_t(f\mathbf{g}) - \pi_t(f)\pi_t(\mathbf{g})\big] \cdot \pi_t(\mathbf{g})dt\Big\}
\end{aligned}$$

$$= \pi_t(\Lambda f)dt + \frac{1}{B}\big[\pi_t(f\mathbf{g}) - \pi_t(f)\pi_t(\mathbf{g})\big] \cdot \left\{\frac{1}{B}\big[d\mathbf{Y}_t - \pi_t(\mathbf{g})dt\big]\right\}.$$

As such, we have

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\Lambda f)ds + \int_0^t \frac{1}{B}\big[\pi_s(f\mathbf{g}) - \pi_s(f)\pi_s(\mathbf{g})\big] \cdot d\tilde{\mathbf{W}}_s. \qquad \square$$

It is worth noting that the FKK equation could be derived via a different approach known as the "innovation approach". (See pages 97–103 in Chigansky (11) for more details.)

We conclude this section with the following proposition, which we will use in Sections 6.3 & 6.4 when determining the stationary distribution of a stochastic process.

**Proposition 3.3.8.** (See Lemma 2.2 of Fujisaki, Kallianpur & Kunita (18).) *$\tilde{\mathbf{W}}$ is an $n$-dimensional Wiener process on $[0, T]$ with respect to $\{\mathcal{F}_t^{\mathbf{Y}}\}$.*

## 3.4 The Shiryaev-Wonham Filter

As with the Zakai and FKK equations, set up the observation process $\mathbf{Y}$ such that the process $\mathbf{g}$ depends only on the state of the signal process:

$$\mathbf{Y}_t = \int_0^t \mathbf{g}(X_s)ds + B\mathbf{W}_t. \tag{3.9}$$

As the signal process takes on finitely many values (let these values be $a_1, \ldots, a_d$), we can view $\mathbf{g}(X)$ as a matrix

$$\mathbf{G} = \begin{bmatrix} g^{(1)}(a_1) & \ldots & g^{(n)}(a_1) \\ \vdots & \ddots & \vdots \\ g^{(1)}(a_d) & \ldots & g^{(n)}(a_d) \end{bmatrix}.$$

The fact that the signal process only takes on a finite number of values makes it easy to calculate expected values. For each time $t$, let $\mathbf{\Phi}_t$ be the $d \times 1$ vector such that its $i^{\text{th}}$ component is given by $\Phi_t(i) = \mathbb{P}\{X_t = a_i | \mathcal{F}_t^{\mathbf{Y}}\}$. Then for any function $f$,

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right] = \mathbb{E}\left[\sum_{i=1}^d f(a_i)1_{\{X_t = a_i\}} \,\middle|\, \mathcal{F}_t^{\mathbf{Y}}\right] = \sum_{i=1}^d f(a_i)\Phi_t(i).$$

Below is the filter for the vector $\mathbf{\Phi}$ (see page 117 of Chigansky (11) for the case where $\mathbf{Y}$ is one-dimensional):

**Theorem 3.4.1.** (Multi-dimensional Shiryaev-Wonham filter). *The vector $\mathbf{\Phi}_t$ satisfies the SDE*

$$d\mathbf{\Phi}_t = \Lambda^T\mathbf{\Phi}_t dt + \frac{1}{B^2}\left[\operatorname{diag}(\mathbf{\Phi}_t) - \mathbf{\Phi}_t\mathbf{\Phi}_t^T\right]\mathbf{G}\left[d\mathbf{Y}_t - \mathbf{G}^T\mathbf{\Phi}_t dt\right], \quad \mathbf{\Phi}_0 = p_0. \tag{3.10}$$

*Proof.* For each $j$, let $I_t(j)$ denote the $j^{\text{th}}$ component of the vector $I_t$. Fix $i \in \{1, \dots, d\}$. Consider the function $f$ given by

$$f(a_j) = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{if } j \neq i. \end{cases}$$

By the construction of $I$, for any time $t$, exactly one of $I_t(1), \dots, I_t(d)$ has value 1, and the rest have value 0. By equation (2.2), we have $f(X_t) = I_t(i)$ for all $t$. Note that as a vector,

$$\Lambda f = \begin{bmatrix} \lambda_{11} & \dots & \lambda_{1d} \\ \vdots & \ddots & \vdots \\ \lambda_{d1} & \dots & \lambda_{dd} \end{bmatrix} e_i = \begin{bmatrix} \lambda_{1i} \\ \vdots \\ \lambda_{di} \end{bmatrix},$$

where $e_i$ is the $i^{\text{th}}$ vector in the standard Euclidean basis for $\mathbb{R}^d$. As such, applying Proposition 2.3.3 and using equation (2.4),

$$I_t(i) = I_0(i) + \int_0^t \sum_{j=1}^d \lambda_{ji} I_s(j) ds + M_t(i),$$

where $M(i)$ is some martingale. Note that $I_t(i) = \mathbf{1}_{\{X_t = a_i\}}$, hence $\Phi_t(i) = \mathbb{E}\left[I_t(i)|\mathcal{F}_t^{\mathbf{Y}}\right] = \mathbb{E}\left[f(X_t)|\mathcal{F}_t^{\mathbf{Y}}\right]$. Applying Theorem 3.3.7,

$$\begin{aligned} \Phi_t(i) &= \Phi_0(i) + \int_0^t \mathbb{E}\left[\sum_{j=1}^d \lambda_{ji} I_s(j) \middle| \mathcal{F}_s^{\mathbf{Y}}\right] ds \\ &\quad + \frac{1}{B^2}\int_0^t \left\{\mathbb{E}\left[I_s(i)\mathbf{g}(X_s)|\mathcal{F}_s^{\mathbf{Y}}\right] - \Phi_s(i)\mathbb{E}\left[\mathbf{g}(X_s)|\mathcal{F}_s^{\mathbf{Y}}\right]\right\} \cdot \left[d\mathbf{Y}_t - \int_0^t \mathbb{E}\left[\mathbf{g}(X_s)|F_s^{\mathbf{Y}}\right] ds\right] \\ &= \Phi_0(i) + \int_0^t \left[\sum_{i=1}^d \lambda_{ji}\Phi_s(j)\right] ds \\ &\quad + \frac{1}{B^2}\int_0^t \left\{\Phi_s(i)\mathbf{g}(a_i) - \Phi_s(i)\sum_{j=1}^d \Phi_s(j)\mathbf{g}(a_j)\right\} \cdot \left[d\mathbf{Y}_t - \int_0^t \sum_{j=1}^d \Phi_s(j)\mathbf{g}(a_j) ds\right]. \end{aligned}$$

In differential notation,

$$\begin{aligned} d\Phi_t(i) &= \left[\sum_{i=1}^d \lambda_{ji}\Phi_t(j)\right] dt \\ &\qquad + \frac{1}{B^2}\left\{\Phi_t(i)\mathbf{g}(a_i) - \Phi_t(i)\sum_{j=1}^d \Phi_t(j)\mathbf{g}(a_j)\right\}^T \left[d\mathbf{Y}_t - \sum_{j=1}^d \Phi_t(j)\mathbf{g}(a_j)dt\right]. \end{aligned}$$

Note that $\sum_{i=1}^d \lambda_{ji}\Phi_t(j)$ is the $i^{\text{th}}$ component of $\Lambda^T \mathbf{\Phi}_t$, $[\Phi_t(i)\mathbf{g}(a_i)]^T$ is the $i^{\text{th}}$ row of

$\text{diag}(\mathbf{\Phi}_t)\mathbf{G}$, and $\left[\Phi_t(i)\sum_{j=1}^{d}\Phi_t(j)\mathbf{g}(a_j)\right]^T$ is the $i^{\text{th}}$ row of $\mathbf{\Phi}_t\mathbf{\Phi}_t^T G$. In addition,

$$\sum_{j=1}^{d}\Phi_t(j)\mathbf{g}(a_j)dt = \mathbf{G}^T\mathbf{\Phi}_t.$$

As such, by aggregating the above equality for $i = 1, \ldots, d$,

$$\begin{aligned} d\mathbf{\Phi}_t &= \Lambda^T\mathbf{\Phi}_t dt + \frac{1}{B^2}\left[\text{diag}(\mathbf{\Phi}_t)\mathbf{G} - \mathbf{\Phi}_t\mathbf{\Phi}_t^T\mathbf{G}\right]\left[d\mathbf{Y}_t - \mathbf{G}^T\mathbf{\Phi}_t dt\right], \\ \Rightarrow d\mathbf{\Phi}_t &= \Lambda^T\mathbf{\Phi}_t dt + \frac{1}{B^2}\left[\text{diag}(\mathbf{\Phi}_t) - \mathbf{\Phi}_t\mathbf{\Phi}_t^T\right]\mathbf{G}\left[d\mathbf{Y}_t - \mathbf{G}^T\mathbf{\Phi}_t dt\right], \end{aligned}$$

as required. $\square$

# Chapter 4

# The Regime-Switching Model with Inside Information

## 4.1 The Regime-Switching Model for Two Assets

Consider a financial market model consisting of a bank account and one financial asset. The bank account is assumed to pay a fixed interest rate $r \geq 0$. As such, if an amount $A$ is deposited in the bank at time $t_1$, the bank account will have a value of $Ae^{(t_2-t_1)}$ at time $t_2 \geq t_1$. The financial asset which an investor can invest in is a stock, whose price process we will denote by $(S_t)_{t\geq 0}$. We wish to analyze the dynamics of trading within this framework.

In order for this framework to be plausible, we assume that at any time $t$, the investor does not know what the stock price at time $t'$ is, for any $t' > t$. An implication of this assumption is that the paths of the stock price process must be non-differentiable. A common way to do this is by modeling $(S_t)$ as the solution of an SDE.

One of the simplest (and most widely used) models for a stock price process is geometric Brownian motion. More explicitly, $(S_t)$ is given as the solution of the SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \quad S_0 = x,$$

where $\mu$ is the mean return of the stock, $\sigma$ is its volatility, $x$ is the price of the stock at time 0 and $W$ is a Wiener process. (Here, $x$, $\mu$ and $\sigma$ are constant, $x > 0$ and $\sigma \geq 0$.) Geometric Brownian motion is commonly used because it fits historical data relatively well and it is also easy to manipulate (in fact, there is an explicit expression for $S_t$).

Note that while it is relatively easy to analyze the stock price dynamics when geometric Brownian motion is used, the geometric Brownian motion model makes certain assumptions which may not be realistic (see Marathe & Ryan (37) for an in-depth analysis on the validity of the geometric Brownian motion model.) One assumption that can be unrealistic is that the expected growth rate $\mu$ of the stock is constant with time, i.e. $\mathbb{E}S_t = e^{\mu t}S_0$ for all values of $t$. Often it is more plausible to assume that the average growth rate of a stock is tied to company-specific factors (e.g. which part of the business cycle the company is in), or macroeconomic indicators. For example, we would generally expect the average growth rate of the stock to be higher when the economy is doing well than when the economy is in a recession. The geometric Brownian motion model does not account for such macroeconomic conditions.

The regime-switching model is an adaptation of the geometric Brownian motion model that fills in this gap. The volatility of the stock, $\sigma$, is still assumed to be constant with time. However, the expected growth rate of the stock is no longer constant with time. Instead, it is replaced by a stochastic process denoted by $(\mu_t)_{t\geq 0}$. The SDE governing the stock price process in this model is

$$dS_t = \mu_t S_t dt + \sigma S_t dW_t, \quad S_0 = x,$$

where $x$ is again the price of the stock at time 0. It remains to define $(\mu_t)_{t\geq 0}$ to model the assumptions we are making. Assume that there are $d$ different "regimes", or states of the economy, which we label as $1, \ldots, d$ (without loss of generality). Assume that the expected growth rate of the stock when the economy is in the $i^{\text{th}}$ regime is $a_i$. Lastly, assume that the state in which the economy is in is modeled as a continuous-time Markov chain. As such, the process $(\mu_t)$ is a continuous-time Markov chain with state space $\{a_1, \ldots, a_d\}$.

Regime-switching models were introduced to model interest rates in Hamilton (22), where it was found that the business cycle was characterized well by discrete shifts between a recessionary state and a growth state. Ang & Bekaert (3) give a list of works which examined empirical models of regime switches in interest rates. Hardy (23) found that the regime-switching model provided a significantly better fit to the data from the Standard and Poor's 500 and the Toronto Stock Exchange 300 indices compared to other popular models. For more information on the empirical results on regime switching models, see Schaller & van Norden (46).

It is worth noting that the regime-switching model is an example of a continuous-time hidden Markov model (HMM) (see Elliott et al. (16) for the general theory of HMMs).

## 4.2 Modeling Information on the Mean Return

Assume that we are in the regime-switching model set up in the previous section. It is unrealistic to assume that an investor would choose how much to invest based solely on the historical stock prices (i.e. the values of $S_u$ for $u \leq t$). He would seek out as much news as he could on the company and on the relevant factors (e.g. the economy) in order to have a more accurate guess of $\mu_t$ before deciding on his portfolio at time $t$. (Note that the values of $\mu_u$ for $u < t$ are irrelevant as $\mu$ is a Markov chain.)

For the sake of simplicity, let us model this idea of added information by assuming that the investor reads a newspaper which gives him an estimate of the stock's mean return for every time $t$. From experience we know that no newspaper is able to give completely accurate guesses all the time. As such, we can model this additional observation by

$$Y'_t = \int_0^t \mu_s ds + \varepsilon V_t, \qquad \forall t \geq 0,$$

where $V$ is a Wiener process independent of $W$, and $\varepsilon \geq 0$. In differential notation, the equation above can be written as

$$dY'_t = \mu_t dt + \varepsilon dV_t, \qquad Y'_0 = 0.$$

Notice that we can model how accurate the newspaper's estimate of the stock's mean return is by varying $\varepsilon$. More specifically, when $\varepsilon = 0$, we can determine $\mu_t$ for every $t$ with complete certainty. On the other hand, as $\varepsilon \to \infty$, the $\mu$ component in the signal above is overpowered by the noise component, meaning that we approach the situation where the investor has no information other than the historical stock prices.

**Definition 4.2.1.** *An investor has* ***complete information on*** $\boldsymbol{\mu}$ *if at any time $t$, he can see the values of $\mu$ up till and including $t$.*

*An investor has* ***inside information on*** $\boldsymbol{\mu}$ *if at any time $t$, he can see the values of $S$ and $Y'$ up till and including $t$, and $0 < \varepsilon < \infty$.*

*An investor has* ***partial information on*** $\boldsymbol{\mu}$ *if at any time $t$, he can only see the values of $S$ up till and including $t$.*

It is worth noting that the case of "partial information on $\mu$" is named as such because we can estimate $\mu$ even if we only have the stock prices available to us. For instance, in the case where $\sigma = 0$, the stock prices alone are enough to tell us what $\mu$ is at every time. The case of "inside information on $\mu$" is named as such to reflect the fact that the investor is getting information on $\mu$ over and above what is available in the historical stock prices.

In some sense, the cases of complete information and partial information correspond to the case of inside information with parameter $\varepsilon = 0$ and $\varepsilon = \infty$ respectively. These links will be elaborated on for the special case of the regime-switching model limited to two states in Section 6.7.

## 4.3 Trading Strategies

Assume that there is an underlying filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ on which the stock price process $S$ is defined. Assume also that $S$ is adapted to $\{\mathcal{F}_t\}$.

Consider an investor with an initial endowment of $V_0$ (which is a constant). If he has the option of continuously trading in the financial market model set up above, what parameters do we need to define in order to completely characterize his actions? At any time $t$, the investor can rebalance his portfolio by determining how much money he is going to invest in the stock. He would then deposit the rest of his money in the bank account. One way to characterize this is to specify the fraction of his wealth at time $t$ which is invested in stock:

**Definition 4.3.1.** *A trading strategy is a stochastic process $p = (p_t)_{t\geq 0}$, where $p_t$ denotes the fraction of wealth invested in the stock at time $t$.*

Note that $p$ is not restricted to the interval $[0, 1]$. If $p_t < 0$, it means that the investor is selling the stock short. If $p_t > 1$, it means that the investor is borrowing money from the bank in order to buy an amount of stock that is worth more than his total wealth. (This implies that in analyzing this set-up in future chapters, we assume that the investor can borrow and lend at the same rate.) Notice also that we have defined a trading strategy as a stochastic process and not as a deterministic process. This is because we want to allow for dynamic strategies which depend not only on the data available to the investor and time 0, but also on the evolution of the financial market model which the investor sees before making his investment decision.

The definition given above for a trading strategy is too wide in the sense that not all stochastic processes would make sense as trading strategies. Take, for example, a stochastic process $p$ such that for some time $t \geq 0$, $p_t \in \mathcal{F}^S_{t+1}$, but $p_t \notin \mathcal{F}^S_t$. This means that at time $t$, in deciding how much to invest in the stock the investor uses information from the stock price in the time period $(t, t+1]$! This does not make sense: the future has not happened yet, and so we cannot look into the future to make our current investment decision. (This assumes that you do not have inside information of the type suggested by Pikovsky & Karatzas (43), in which case you might have some idea of how the stock price might move in the near future.)

In this light, valid trading strategies must have the property that for all times $t \geq 0$, $p_t$ can only depend on the information available to you during the time period $[0, t]$. As such, $p$ must be adapted to some filtration $\{\mathcal{G}_t\}$ that captures all the information up till the current time. (For an explanation on how filtrations are related to information, see Section 4 of Chapter II in Cinlar (12).) Such a stochastic process is known as a ***non-anticipating process***.

At this point, we will not specify the filtration $\{\mathcal{G}_t\}$ explicitly: in future chapters this filtration will change across different cases. We will define the filtration which $p$ is adapted to in those cases for clarity's sake. However, to ensure that the wealth of the portfolio does not "explode" to infinity at any time, we will impose the following condition: over a finite time horizon $[0, T]$, the investment strategy $p$ must be an element of $\mathcal{H}^2[0, T]$. We will only deal with finite time horizons in this thesis.

For the remaining chapters, we will use the following updated definition for trading strategies:

**Definition 4.3.2.** *Let $\{\mathcal{G}_t\}$ be a sub-filtration of $\mathcal{F}$ such that for all times $t$, $\mathcal{G}_t$ represents the information available to the investor at time $t$. Let $p = (p_t)_{t\geq 0}$ be a stochastic* *process such that $p_t$ denotes the fraction of wealth invested in the stock at time $t$. $p$ is a* ***trading strategy on $[0, T]$*** *if $p$ is measurable and adapted to $\{\mathcal{G}_t\}$, and $p \in \mathcal{H}^2[0, T]$.*

# Chapter 5

# Deriving the Utility-Maximizing Trading Strategy

In this chapter, we formulate the concept of the log-optimal wealth process. By deriving an expression for the log of the wealth process in terms of the stock price process and the trading strategy, we will obtain explicit expressions for the trading strategy that optimizes the expected log-utility of the wealth process. We will also derive expressions for the greatest possible expected log-utility of the wealth process and long-run discounted growth rate.

As before, assume that there is an underlying filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

## 5.1 The Wealth Process and Expected Logarithmic Utility

When trading in a financial market, an investor is primarily interested in his total wealth. A rational investor would seek to maximize the expected utility that he gains from this wealth. To this end, introduce the stochastic process $V = (V_t)_{t\geq 0}$, where $V_t$ denotes the investor's wealth, or the value of the portfolio, at time $t$. $V$ is called the ***wealth process***. The goal of an investor would then be to maximize $\mathbb{E}[U(V_T)]$ for some terminal time $T$, where $U$ is some utility function.

Often, it is useful to discount prices by the risk-free asset. One simple strategy for investing in the market is by placing the entire initial endowment $V_0$ in the bank account that pays interest at a fixed rate $r$ at time 0. At a terminal time $T$, the investor would have a wealth of $V_0 e^{rT}$ without incurring any risk. In some sense this strategy should be the benchmark against which all other strategies are measured: a strategy that consistently does worse than this riskless strategy should never be employed. As such, it is natural to measure how well a trading strategy does, not by looking at the nominal values of portfolios, but by comparing it to the riskless trading strategy just mentioned. With this in mind, we introduce the discounted version of a stochastic process:

**Definition 5.1.1.** *Let $X = (X_t)_{t\geq 0}$ be some price or wealth process. Assume that the bank* *account offers a fixed interest rate $r$. The* ***discounted price or wealth process****, denoted by $\bar{X}$, is given by $\bar{X}_t := X_t e^{-rt}$ for all $t$.*

In this thesis, we will consider the "normalized" expected utilities $\mathbb{E}\left[U(\bar{V}_T)\right]$ rather than the nominal expected utilities $\mathbb{E}\left[U(V_T)\right]$.

In economic theory, in order for the concept of utility to be realistic, the utility function $U$ must satisfy certain basic conditions (see Section 3.4 of Karatzas & Shreve (31)). In this thesis, we will not consider the general utility function, but rather logarithmic utility (log-utility), given by

$$U(\bar{V}_T) := \log \bar{V}_T.$$

The idea of using the logarithm as a utility function was first posited by Bernoulli in 1738 (Bernoulli (7)). Bernoulli argued that it is reasonable to expect that the utility an investor gains from an increase in his wealth is inversely proportional to his original wealth. This condition leads to the conclusion that utility is proportional to the log of wealth. The logarithmic utility function can also be seen as an example of the Weber-Fechner law, which states that a wide variety of stimuli are perceived on a logarithmic scale.

The use of the logarithmic utility function is supported not just by psychological evidence, but also by some of its mathematical properties. Thorp (53) showed that in a financial market model, by maximizing logarithmic utility, an investor could asymptotically maximize the growth rate of his assets. Bell & Cover (5) showed that in certain set-ups, the expected log-optimal portfolio is also game-theoretically optimal in a single play or multiple plays of the stock market. Goll & Kallsen (20) also note that when using the logarithmic utility function, optimal solutions can be explicitly calculated for general dynamic models.

**Definition 5.1.2.** *Let $T \geq 0$ be a given terminal time. A wealth process $V^*$ is called a* ***log-optimal wealth process*** *if it maximizes expected log-utility, i.e. for all possible wealth processes $V$,*

$$\mathbb{E} \log \bar{V}_T \leq \mathbb{E} \log \bar{V}_T^*.$$

*A trading strategy is said to be* ***log-optimal*** *if the wealth process associated with it is log-optimal.*

We end this section with a definition of the long-run discounted growth rate of a wealth process.

**Definition 5.1.3.** *Given a wealth process $V = (V_T)_{T \geq 0}$, define the* ***long-run discounted growth rate*** *of the wealth process as*

$$\gamma(V) := \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\log \bar{V}_T\right],$$

*if the limit exists.*

The intuition for the definition above is as follows: if a discounted wealth process has a long-run growth rate of $r$, then $\bar{V}_T$ should look like $V_0 e^{rT}$ for terminal times that are large enough. The definition above captures this idea in a precise manner.

## 5.2 The Relationship between the Wealth and Stock Price Processes

Chen (10) presents a similar derivation of the following proposition:

**Proposition 5.2.1.** *Let $S$ be the stock price process, $p$ a trading strategy. Let $V$ be the wealth process associated with $p$. Then for any terminal time $T$,*

$$\int_0^T \frac{1}{\bar{V}_t} d\bar{V}_t = \int_0^T \frac{p_t}{\bar{S}_t} d\bar{S}_t,$$

*if the integral on the right hand side exists. Equivalently, we can write*

$$\frac{d\bar{V}_T}{\bar{V}_T} = p_T \frac{d\bar{S}_T}{\bar{S}_T}.$$

*Proof.* Fix a terminal time $T \geq 0$, let $k \in \mathbb{N}$ be fixed. Partition the time interval $[0, T]$ into $N = 2^k$ $(k \in \mathbb{N})$ intervals of equal length: define $T_{N,i} = \frac{Ti}{N}$ for $i = 0, 1, \ldots, N$. Consider the discrete trading strategy where you rebalance your portfolio only at times $T_{N,i}$ for $i = 0, 1, \ldots, N$, according to the strategy $p$. For any $i = 1, \ldots, N$, the table below shows the amount invested in each asset at time $T_{N,i-1}$, and the value of each component of the portfolio at time $T_{N,i}$ before rebalancing:

| | Value of bank account | Value of stock component |
|---|---|---|
| After rebalancing at $T_{N,i-1}$ | $(1 - p_{T_{N,i-1}})V_{T_{N,i-1}}$ | $p_{T_{N,i-1}} V_{T_{N,i-1}}$ |
| Before rebalancing at $T_{N,i}$ | $(1 - p_{T_{N,i-1}})V_{T_{N,i-1}} e^{T_{N,i} - T_{N,i-1}}$ | $p_{T_{N,i-1}} V_{T_{N,i-1}} \frac{S_{T_{N,i}}}{S_{T_{N,i-1}}}$ |

As such, for this trading strategy,

$$\begin{aligned}
V_{T_{N,i}} &= (1 - p_{T_{N,i-1}})V_{T_{N,i-1}} e^{T_{N,i} - T_{N,i-1}} + p_{T_{N,i-1}} V_{T_{N,i-1}} \frac{S_{T_{N,i}}}{S_{T_{N,i-1}}}, \\
\frac{V_{T_{N,i}}}{V_{T_{N,i-1}}} &= (1 - p_{T_{N,i-1}}) e^{T_{N,i} - T_{N,i-1}} + p_{T_{N,i-1}} \frac{S_{T_{N,i}}}{S_{T_{N,i-1}}}, \\
\frac{\bar{V}_{T_{N,i}} e^{T_{N,i}}}{\bar{V}_{T_{N,i-1}} e^{T_{N,i-1}}} &= (1 - p_{T_{N,i-1}}) e^{T_{N,i} - T_{N,i-1}} + p_{T_{N,i-1}} \frac{\bar{S}_{T_{N,i}} e^{T_{N,i}}}{\bar{S}_{T_{N,i-1}} e^{T_{N,i-1}}}, \\
\frac{\bar{V}_{T_{N,i}}}{\bar{V}_{T_{N,i-1}}} &= 1 - p_{T_{N,i-1}} + p_{T_{N,i-1}} \frac{\bar{S}_{T_{N,i}}}{\bar{S}_{T_{N,i-1}}}, \\
\frac{\bar{V}_{T_{N,i}} - \bar{V}_{T_{N,i-1}}}{\bar{V}_{T_{N,i-1}}} &= p_{T_{N,i-1}} \frac{\bar{S}_{T_{N,i}} - \bar{S}_{T_{N,i-1}}}{\bar{S}_{T_{N,i-1}}}.
\end{aligned}$$

As this relation holds for all $i = 1, \dots, N$, summing them up, we get

$$\sum_{i=1}^{N} \frac{\bar{V}_{T_{N,i}} - \bar{V}_{T_{N,i-1}}}{\bar{V}_{T_{N,i-1}}} = \sum_{i=1}^{N} p_{T_{N,i-1}} \frac{\bar{S}_{T_{N,i}} - \bar{S}_{T_{N,i-1}}}{\bar{S}_{T_{N,i-1}}}.$$

Taking $N \to \infty$ and using Theorem 2.2.6,

$$\begin{aligned}
\int_0^T \frac{1}{\bar{V}_t} d\bar{V}_t &= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{\bar{V}_{T_{N,i-1}}} \left(\bar{V}_{T_{N,i}} - \bar{V}_{T_{N,i-1}}\right) \\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{p_{T_{N,i-1}}}{\bar{S}_{T_{N,i-1}}} \left(\bar{S}_{T_{N,i}} - \bar{S}_{T_{N,i-1}}\right) \\
&= \int_0^T \frac{p_t}{\bar{S}_t} d\bar{S}_t.
\end{aligned}$$

□

Note that Proposition 5.2.1 holds for all stock price processes. The following proposition gives us the expected log-utility for the wealth process associated with a general trading strategy under the regime-switching model.

**Proposition 5.2.2.** *Let the stock price process be given by $dS_t = \mu_t S_t dt + \sigma S_t dW_t$, with the assumption that $\mu$ is bounded, i.e. there exists a constant $M$ such that $|\mu_t(\omega)| < M$ for all $t \in \mathbb{R}_+, \omega \in \Omega$. Fix the terminal time $T$. Then, for any trading strategy $p \in \mathcal{H}^2[0,T]$,*

$$\mathbb{E} \log \bar{V}_T = \log \bar{V}_0 + \int_0^T \mathbb{E}\left[(\mu_t - r)p_t - \frac{1}{2}\sigma^2 p_t^2\right] dt. \tag{5.1}$$

*Proof.* Consider the discounted stock price process $(\bar{S}_t)_{t\geq 0}$. Using Itô's formula for the function $f(t,x) = e^{-rt}x$:

$$\begin{aligned}
d\bar{S}_t &= df(t, S_t) \\
&= -re^{-rt} S_t dt + e^{-rt} dS_t + 0 \\
&= -r\bar{S}_t dt + e^{-rt}(\mu_t S_t dt + \sigma S_t dW_t),
\end{aligned}$$

Hence,

$$d\bar{S}_t = (\mu_t - r)\bar{S}_t dt + \sigma \bar{S}_t dW_t. \tag{5.2}$$

By equation (5.2), $(d\bar{S}_t)^2 = \sigma^2 \bar{S}_t^2 dt$ for all $t$. Using Itô's formula for the function $f(t,x) = \log x$,

$$\begin{aligned}
d(\log \bar{V}_T) &= 0 + \frac{1}{\bar{V}_T} d\bar{V}_T + \frac{1}{2}\left(-\frac{1}{\bar{V}_T^2}\right)(d\bar{V}_T)^2 \\
&= \frac{d\bar{V}_T}{\bar{V}_T} - \frac{1}{2}\left(\frac{d\bar{V}_T}{\bar{V}_T}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= p_T \frac{d\bar{S}_T}{\bar{S}_T} - \frac{1}{2} p_T^2 \left( \frac{d\bar{S}_T}{\bar{S}_T} \right)^2 && \text{(by Prop 5.2.1)} \\
&= p_T [(\mu_T - r) dT + \sigma dW_T] - \frac{1}{2} p_T^2 \frac{\sigma^2 \bar{S}_T^2 dT}{\bar{S}_T^2} \\
&= \left[ p_T(\mu_T - r) - \frac{1}{2} p_T^2 \sigma^2 \right] dT + p_T \sigma dW_T,
\end{aligned}
$$

$$
\text{i.e.} \quad \int_0^T d(\log \bar{V}_t) = \int_0^T \left[ p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 \right] dt + \int_0^T p_t \sigma dW_t,
$$

$$
\log \bar{V}_T = \log \bar{V}_0 + \int_0^T \left[ p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 \right] dt + \int_0^T p_t \sigma dW_t.
$$

Note that $p \in \mathcal{H}^2[0,T] \Rightarrow \sigma p \in \mathcal{H}^2[0,T]$ for any constant $\sigma \in \mathbb{R}$. As such,

$$
\begin{aligned}
\mathbb{E} \log \bar{V}_T &= \log \bar{V}_0 + \mathbb{E}\left[ \int_0^T p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 dt \right] + \mathbb{E}\left[ \int_0^T p_t \sigma dW_t \right] \\
&= \log \bar{V}_0 + \mathbb{E}\left[ \int_0^T p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 dt \right]. && \text{(by Prop 2.2.3)}
\end{aligned}
$$

It remains to show that we can switch the integral sign and the expectation sign for the second term on the right hand side above. By Fubini's theorem, it suffices to show that

$$
\mathbb{E}\left[ \int_0^T \left| p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 \right| dt \right] < \infty.
$$

Consider $p$ as a function $p : (\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}[0,T], \mathbb{P} \times \mathbb{Q}) \mapsto \mathbb{R}$, where $\mathbb{Q}$ is the uniform measure on $[0,T]$. (Note that $p$ is a measurable function.) Then

$$
\begin{aligned}
\left[ \mathbb{E} \frac{1}{T} \int_0^T |p_t| dt \right]^2 &= \left\{ (\mathbb{P} \times \mathbb{Q}) |p| \right\}^2 \\
&\leq (\mathbb{P} \times \mathbb{Q}) \left[ p^2 \right] && \text{(by Jensen's inequality)} \\
&= \mathbb{E}\left[ \frac{1}{T} \int_0^T |p_t|^2 dt \right].
\end{aligned}
$$

As such,

$$
\begin{aligned}
\mathbb{E}\left[ \int_0^T \left| p_t(\mu_t - r) - \frac{1}{2} p_t^2 \sigma^2 \right| dt \right] &\leq \mathbb{E}\left[ \int_0^T |p_t(\mu_t - r)| dt \right] + \frac{\sigma^2}{2} \mathbb{E}\left[ \int_0^T |p_t|^2 dt \right] \\
&\leq \big( |M| + |r| \big) T \mathbb{E}\left[ \frac{1}{T} \int_0^T |p_t| dt \right] + \frac{\sigma^2}{2} \mathbb{E}\left[ \int_0^T |p_t|^2 dt \right] \\
& \qquad \text{(as } \mu \text{ is bounded)} \\
&\leq \big( |M| + |r| \big) T \left\{ \mathbb{E}\left[ \frac{1}{T} \int_0^T |p_t|^2 dt \right] \right\}^{\frac{1}{2}} + \frac{\sigma^2}{2} \mathbb{E}\left[ \int_0^T |p_t|^2 dt \right]
\end{aligned}
$$

$$= \big(|M| + |r|\big)\sqrt{T}\left\{\mathbb{E}\left[\int_0^T |p_t|^2 dt\right]\right\}^{\frac{1}{2}} + \frac{\sigma^2}{2}\mathbb{E}\left[\int_0^T |p_t|^2 dt\right]$$

$$< \infty. \qquad \text{(as } p \in \mathcal{H}^2[0,T]\text{)}$$

This concludes the proof. □

The proposition below gives an explicit expression for the trading strategy which maximizes the expected log-utility of the wealth process.

**Proposition 5.2.3.** *Consider the setup of Proposition 5.2.2. Let $\{\mathcal{G}_t\}$ be a filtration on $\Omega$ such that for every time $t$, $\mathcal{G}_t \subset \mathcal{F}_t$. Fix a terminal time $T$. Then, among admissible trading strategies which are adapted to $\{\mathcal{G}_t\}$, the trading strategy which maximizes the expected log-utility of the wealth process is given by*

$$p_t^* = \frac{\{\mathbb{E}\left[\mu_t|\mathcal{G}_t\right] - r\}}{\sigma^2}, \quad \forall t \in [0,T].$$

*Proof.* For each $\omega \in \Omega$, consider the function

$$f_{t,\omega}(x) = \{\mathbb{E}\left[\mu_t|\mathcal{G}_t\right](\omega) - r\}\, x - \frac{1}{2}\sigma^2 x^2, \quad x \in \mathbb{R}.$$

$f_{t,\omega}$ is an inverted parabola, and so if $x^*$ is the value in $\mathbb{R}$ that maximizes $f_{t,\omega}$, then

$$\begin{aligned} f'_{t,\omega}(x^*) &= 0,\\ \Rightarrow \{\mathbb{E}\left[\mu_t|\mathcal{G}_t\right](\omega) - r\} - \sigma^2 x^* &= 0,\\ \Rightarrow x^* &= \frac{\{\mathbb{E}\left[\mu_t|\mathcal{G}_t\right](\omega) - r\}}{\sigma^2}\\ &= p_t^*(\omega). \end{aligned}$$

Hence, $p_t^*(\omega)$ maximizes $f_{t,\omega}$. By equation (5.1), for all $p$ adapted to $\{\mathcal{G}_t\}$,

$$\begin{aligned} \mathbb{E}[U(\bar{V}_T)] - \log \bar{V}_0 &= \int_0^T \mathbb{E}\left[(\mu_t - r)p_t - \frac{1}{2}\sigma^2 p_t^2\right] dt\\ &= \int_0^T \mathbb{E}\left\{\mathbb{E}\left[(\mu_t - r)p_t - \frac{1}{2}\sigma^2 p_t^2 \Big| \mathcal{G}_t\right]\right\} dt\\ &= \int_0^T \int_\Omega \Big(\mathbb{E}[\mu_t|\mathcal{G}_t](\omega) - r\Big)p_t(\omega) - \frac{1}{2}\sigma^2 p_t(\omega)^2 d\mathbb{P}dt\\ &\qquad \text{(as } p_t \text{ is } \mathcal{G}_t\text{-measurable)}\\ &= \int_0^T \int_\Omega f_{t,\omega}[p_t(\omega)] d\mathbb{P}dt. \end{aligned}$$

For all $p$ adapted to $\{\mathcal{G}_t\}$,

$$
\begin{aligned}
f_{t,\omega}[p_t(\omega)] &\leq f_{t,\omega}[p_t^*(\omega)] && \forall \omega \in \Omega, t \in [0,T], \\
\Rightarrow \int_\Omega f_{t,\omega}[p_t(\omega)] d\mathbb{P} &\leq \int_\Omega f_{t,\omega}[p_t^*(\omega)] d\mathbb{P} && \forall t \in [0,T], \\
& && \text{(by the monotonicity of the integral)} \\
\Rightarrow \int_0^T \int_\Omega f_{t,\omega}[p_t(\omega)] d\mathbb{P} dt &\leq \int_0^T \int_\Omega f_{t,\omega}[p_t^*(\omega)] d\mathbb{P} dt. && \text{(same reasoning)}
\end{aligned}
$$

Thus, the trading strategy $p^*$ gives an expected log-utility which is greater than or equal than that given by any trading strategy which is adapted to $\mathcal{G}_t$. It remains to show that $p^*$ is adapted to $\{\mathcal{G}_t\}$, and that $p^* \in \mathcal{H}^2[0,T]$.

As $\mathbb{E}[\mu_t|\mathcal{G}_t]$ is in $\mathcal{G}_t$ by definition, and $r$ and $\sigma$ are constants, it follows that $p_t^*$ is $\mathcal{G}_t$-measurable for all $t$. From this, it is also clear that $p^*$ is measurable w.r.t. $\mathcal{F}$ and is adapted to $\{\mathcal{F}_t\}$. Finally, as $\mu$ is a bounded process,

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T (p_t^*)^2 dt\right] &= \mathbb{E}\left[\int_0^T \frac{\{\mathbb{E}[\mu_t|\mathcal{G}_t] - r\}^2}{\sigma^4} dt\right] \\
&\leq \mathbb{E}\left[\int_0^T \frac{\{|\mathbb{E}[\mu_t|\mathcal{G}_t]| + r\}^2}{\sigma^4} dt\right] \\
&\leq \mathbb{E}\left[\int_0^T \frac{\{M + r\}^2}{\sigma^4} dt\right] \\
&= \frac{T(M+r)^2}{\sigma^4} < \infty.
\end{aligned}
$$

Thus, $p^*$ is indeed an admissible strategy which is adapted to $\{\mathcal{G}_t\}$. □

Proposition 5.2.3 has two important implications. First, for a fixed terminal time, the trading strategy which maximizes the expected-log utility at the terminal time is unique. Second, note that for each time $t$, $p_t^*$ depends only on the estimate of $\mu_t$. For any two terminal times $0 \leq T_1 \leq T_2$, let $(p_t^*)_{t\in[0,T_1]}$ and $(q_t^*)_{t\in[0,T_2]}$ be the log-optimal trading strategies associated with $T_1$ and $T_2$ respectively. Then $p^*$ is the restriction of $q^*$ to the time interval $[0,T_1]$. This means that in maximizing his expected log-utility an investor can be "myopic": his trading strategy and any point in time does not depend on the terminal time.

The following corollary gives an explicit expression for the value and the long-run discounted growth rate of the log-optimal wealth process.

**Corollary 5.2.4.** *Under the set-up of Proposition 5.2.3, the value of the log-optimal wealth* *process, when the information at time $t$ is given by the $\sigma$-algebra $\mathcal{G}_t$, is*

$$\mathbb{E}\log \bar{V}_T^* = \log \bar{V}_0 + \frac{1}{2\sigma^2}\int_0^T \mathbb{E}\Big\{\big(\mathbb{E}[\mu_t|\mathcal{G}_t] - r\big)^2\Big\}dt, \tag{5.3}$$

*and the long-run discounted growth rate of this wealth process is*

$$\gamma(V^*) = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{2\sigma^2 T}\int_0^T \big(\mathbb{E}[\mu_t|\mathcal{G}_t] - r\big)^2 dt\right]. \tag{5.4}$$

*Proof.* The expression for $\mathbb{E}\log \bar{V}_T^*$ can be obtained directly by substituting $p^*$ into equation (5.1). The expression for $\gamma(V^*)$ can be derived by dividing both sides of equation (5.3) by $T$, applying the limit $T \to \infty$, and using Fubini's theorem to switch the integral and the expectation. (Fubini's theorem applies due to the boundedness of $\mu$.) $\square$

Proposition 5.2.3 and Corollary 5.2.4 imply that in order to find the value of the log-optimal wealth process along with the trading strategy that obtains this value when information at time $t$ is given by $\mathcal{G}_t$, it is sufficient to determine $\mathbb{E}[\mu_t|\mathcal{G}_t]$. This will be the subject of the next chapter.

# Chapter 6

# Analysis of the Regime-Switching Model with Two States

In this chapter, we will apply results of stochastic filtering theory to obtain a method for determining the investment strategy that optimizes expected log-utility, and an expression for the growth rate of the log-optimal wealth process. We will consider three cases: when we have complete information on $\mu$, partial information on $\mu$ and inside information on $\mu$. For each case, after discussing the method for determining the filter for $\mu$, we will calculate the long-run discounted growth rate of the log-optimal wealth process. In addition, we will show weak convergence of measures which demonstrate that the cases of complete and partial information can be viewed as special cases of the case of inside information. We will end this chapter with an analysis of the long-run discounted growth rate of the log-optimal wealth process.

Based on Proposition 5.2.3 and Corollary 5.2.4, determining the optimal investment strategy amounts to determining $\mathbb{E}[\mu_t|\mathcal{G}_t]$ for each time $t$, where $\mu_t$ is the expected growth rate of the stock at time $t$, and $\mathcal{G}_t$ is the filtration representing the information available at time $t$. It is clear that an application of the Shiryaev-Wonham filter (Theorem 3.4.1) gives a system of SDEs for $\mathbb{P}\{\mu_t = a_i|\mathcal{G}_t\}$ for each state $a_i$. From such a system of SDEs, given information $\mathcal{G}_t$, we can use numerical approximation techniques to obtain an estimate for $\mathbb{P}\{\mu_t = a_i|\mathcal{G}_t\}$ for all $i$, which in turn gives us an estimate for $\mathbb{E}[\mu_t \mid \mathcal{G}_t]$.

We do not know how to obtain closed-form solutions for any analysis on this general regime-switching model. However, when the expected growth rate of the stock ($\mu_t$) is limited to two states, we can obtain some explicit expressions.

## 6.1 Setting up the Model

To orient ourselves, we will now set up the parts of the regime-switching model that will be used throughout this chapter. Fix some terminal time $T$. Let the interest rate of the bank account be fixed at $r$. Let the stock price process $S_t$ be governed by the following SDE:

$$dS_t = \mu_t S_t dt + \sigma S_t dW_t, \qquad S_0 = x,$$

where $W$ is a Wiener process, and $x$ is the stock price at time 0. Let the expected growth rate of the stock $\mu$ be modeled as a two-state continuous-time Markov chain with states $\{a, b\}$ such that $a > b$, and transition intensities matrix

$$\Lambda = \begin{bmatrix} -\lambda_b & \lambda_b \\ \lambda_a & -\lambda_a \end{bmatrix}, \qquad \lambda_a, \lambda_b > 0.$$

We have assumed that $\lambda_a$ and $\lambda_b$ are both strictly positive so as to ensure that the Markov chain $\mu$ is irreducible and recurrent. We are not losing much by making this assumption: if both $\lambda$'s were 0, $\mu$ would be constant with time. If only one of the $\lambda$'s was non-zero, after an almost-surely finite amount of time $t$, we would have $\mu_s = \mu_t$ for all $s \geq t$. Both these cases are not very interesting in terms of ergodic theory— the solutions are obvious.

In order for the filtering results of Chapter 3 to be applicable in our model, instead of using the stock price process $S$ as the observation process, we will use $Y$ given by $Y_t := \log S_t$. Using Itô's formula, we have

$$\begin{aligned} Y_0 = \log x, \quad dY_t &= \frac{1}{S_t} dS_t - \frac{1}{2}\frac{1}{S_t^2}(dS_t)^2 \\ &= (\mu_t dt + \sigma dW_t) - \frac{1}{2}\frac{1}{S_t^2}\sigma^2 S_t^2 dt \\ &= \left(\mu_t - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \end{aligned}$$

As the Markov chain has only two states, it means that $\mathbb{P}\{\mu_t = a|\mathcal{G}_t\} + \mathbb{P}\{\mu_t = b|\mathcal{G}_t\} = 1$ for all $t$. As such, by defining $\phi_t = \mathbb{P}\{\mu_t = a|\mathcal{G}_t\}$, we will be able to determine the filter for $\mu$ w.r.t. $\{\mathcal{G}_t\}$ from the process $\phi$. More specifically, for all $t$,

$$\begin{aligned} \mathbb{E}\left[\mu_t|\mathcal{G}_t\right] &= a\mathbb{P}\{\mu_t = a|\mathcal{G}_t\} + b\mathbb{P}\{\mu_t = b|\mathcal{G}_t\} \\ &= a\phi_t + b(1 - \phi_t) \\ &= b + (a - b)\phi_t. \end{aligned}$$

Finally, note that in the notation of Section 3.4, we have

$$\mathbf{\Phi}_t := \begin{bmatrix} \mathbb{P}\{\mu_t = a \mid \mathcal{G}_t\} \\ \mathbb{P}\{\mu_t = b \mid \mathcal{G}_t\} \end{bmatrix} = \begin{bmatrix} \phi_t \\ 1 - \phi_t \end{bmatrix},$$

and

$$\begin{aligned}
\operatorname{diag}(\boldsymbol{\Phi}_t) - \boldsymbol{\Phi}_t\boldsymbol{\Phi}_t^T &= \begin{bmatrix} \phi_t & 0 \\ 0 & 1-\phi_t \end{bmatrix} - \begin{bmatrix} \phi_t^2 & \phi_t(1-\phi_t) \\ \phi_t(1-\phi_t) & (1-\phi_t)^2 \end{bmatrix} \\
&= \begin{bmatrix} \phi_t(1-\phi_t) & -\phi_t(1-\phi_t) \\ -\phi_t(1-\phi_t) & (1-\phi_t)\left[1-(1-\phi_t)\right] \end{bmatrix} \\
&= \phi_t(1-\phi_t) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.
\end{aligned}$$

## 6.2 Complete Information on $\mu$

In this case, at time $t$, we know exactly which state $\mu_s$ is in for every time $s \in [0,t]$. Using Proposition 5.2.3 and Corollary 5.2.4, we know that the log-optimal trading strategy is given by

$$p_t^* = \frac{\mu_t - r}{\sigma^2}, \qquad \forall t,$$

and that this trading strategy gives a log-optimal utility of

$$\mathbb{E}\log \bar{V}_T^* = \log \bar{V}_0 + \frac{1}{2\sigma^2}\int_0^T \mathbb{E}\left[(\mu_t - r)^2\right] dt.$$

Corollary 5.2.4 also gives an expression for the long-run discounted growth rate of the log-optimal wealth process:

$$\begin{aligned}
\gamma(V^*) &= \frac{1}{2\sigma^2}\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\int_0^T (\mu_t - r)^2 dt\right] \\
&= \frac{1}{2\sigma^2}\mathbb{E}\left[\lim_{T\to\infty}\frac{1}{T}\int_0^T (\mu_t - r)^2 dt\right] && \text{(by the bounded convergence theorem)} \\
&= \frac{1}{2\sigma^2}\frac{\lambda_a(a-r)^2 + \lambda_b(b-r)^2}{\lambda_a + \lambda_b}. && \text{(by Thm 2.4.5)}
\end{aligned}$$

## 6.3 Partial Information on $\mu$

In this case, the only information available at time $t$ is the stock prices up till time $t$ (time $t$ included). Applying the Shiryaev-Wonham filter (equation (3.10)) to the signal process $\mu$ with observation process $Y$:

$$\begin{aligned}
\begin{bmatrix} d\phi_t \\ 1 - d\phi_t \end{bmatrix} &= \begin{bmatrix} -\lambda_b & \lambda_a \\ \lambda_b & -\lambda_a \end{bmatrix} \begin{bmatrix} \phi_t \\ 1-\phi_t \end{bmatrix} dt + \frac{1}{\sigma^2}\phi_t(1-\phi_t)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \\
&\quad \times \begin{bmatrix} a - \frac{\sigma^2}{2} \\ b - \frac{\sigma^2}{2} \end{bmatrix} \left[dY_t - \left(a - \frac{\sigma^2}{2}\right)\phi_t dt - \left(b - \frac{\sigma^2}{2}\right)(1-\phi_t)dt\right],
\end{aligned}$$

$$
\begin{aligned}
d\phi_t = [-\lambda_b\phi_t + \lambda_a(1-\phi_t)]\,dt + \frac{\phi_t(1-\phi_t)}{\sigma^2}\left[\left(a-\frac{\sigma^2}{2}\right)-\left(b-\frac{\sigma^2}{2}\right)\right] \\
\times\left[dY_t - \left((a-b)\phi_t + b - \frac{\sigma^2}{2}\right)dt\right],
\end{aligned}
$$

Hence,

$$
\begin{aligned}
d\phi_t = [\lambda_a - (\lambda_a+\lambda_b)\phi_t]\,dt + \frac{(a-b)\phi_t(1-\phi_t)}{\sigma^2} \\
\times\left[dY_t - \left((a-b)\phi_t + b - \frac{\sigma^2}{2}\right)dt\right]. \qquad (6.1)
\end{aligned}
$$

In practice, one would obtain stock price data from the stock market. One can then use approximating techniques such as Euler's method to obtain an estimate of $\phi_t$, which would allow one to trade according to the strategy given in Proposition 5.2.3. To estimate the distribution $\phi_t$, one can use Monte Carlo simulation. For each iteration, one would use approximating techniques to obtain sample paths of $Y$ and $\phi$, from which we can obtain a value of $\phi_t$. Averaging over many sample paths, one would obtain an estimate for $\phi_t$'s distribution.

We now derive an expression for the invariant probability measure for $\phi$, which is needed to obtain an explicit expression for the long-run discounted growth rate of the log-optimal wealth process. By Proposition 3.3.8, $\left[Y_t - \int_0^t \left((a-b)\phi_s + b - \frac{\sigma^2}{2}\right)ds\right]$ is a Wiener process. As such, we can write equation (6.1) as

$$
d\phi_t = [\lambda_a - (\lambda_a+\lambda_b)\phi_t]\,dt + \frac{(a-b)\phi_t(1-\phi_t)}{\sigma^2}dB_t,
$$

where $B$ is a Wiener process.

**Claim 6.3.1.** *The process $\phi$ satisfies the conditions of Theorem 2.4.2.*

While the proof of Claim 6.3.1 is rather technical, we provide it here because the proof has significant overlap with the calculations for the invariant probability measure associated with $\phi$.

*Proof.* As $\phi_t = \mathbb{P}\{\mu_t = a|\mathcal{F}_t^Y\}$ for all $t$, it is clear that $\phi$ must take on values in $[0,1]$. Using the notation of Theorem 2.4.2, define functions $f$ and $g$ by

$$
\begin{aligned}
f(x) &= \lambda_a - (\lambda_a+\lambda_b)x, \\
g(x) &= \frac{(a-b)x(1-x)}{\sigma^2}.
\end{aligned}
$$

From these definitions it is clear that $f$ and $g$ are continuous, and that $g(x)=0$ if and only if $x = 0, 1$.

Let $a = 0.5$. Note that by partial fractions,

$$\frac{\lambda_a - (\lambda_a + \lambda_b)y}{y^2(1-y)^2} = \frac{\lambda_a - \lambda_b}{y} + \frac{\lambda_a}{y^2} + \frac{\lambda_a - \lambda_b}{1-y} - \frac{\lambda_b}{(1-y)^2}.$$

As such,

$$\begin{aligned}
\int_{0.5}^{x} \frac{2f(y)}{g^2(y)} dy &= \frac{2\sigma^4}{(a-b)^2} \int_{0.5}^{x} \frac{\lambda_a - (\lambda_a + \lambda_b)y}{y^2(1-y)^2} dy \\
&= \frac{2\sigma^4}{(a-b)^2} \left[ (\lambda_a - \lambda_b) \log y - \frac{\lambda_a}{y} - (\lambda_a - \lambda_b)\log(1-y) - \frac{\lambda_b}{1-y} \right]_{0.5}^{x} \\
&= \frac{2\sigma^4}{(a-b)^2} \left[ (\lambda_a - \lambda_b) \log\left(\frac{x}{1-x}\right) - \left(\frac{\lambda_a}{x} + \frac{\lambda_b}{1-x}\right) - c_2 \right],
\end{aligned}$$

where $c_2$ is some constant. In particular,

$$\begin{aligned}
\int_{0.5}^{1} \frac{2f(y)}{g^2(y)} dy &= \lim_{x\to 1} \frac{2\sigma^4}{(a-b)^2} \left[ (\lambda_a - \lambda_b) \log\left(\frac{x}{1-x}\right) - \left(\frac{\lambda_a}{x} + \frac{\lambda_b}{1-x}\right) - c_2 \right] \\
&= c_3 + \frac{2\sigma^4}{(a-b)^2} \lim_{x\to 1} \left[ (\lambda_a - \lambda_b) \log\left(\frac{x}{1-x}\right) - \frac{\lambda_b}{1-x} \right] \qquad (c_3 \text{ some constant}) \\
&= c_3 + \frac{2\sigma^4}{(a-b)^2} \lim_{x\to 0} \left[ (\lambda_a - \lambda_b) \log\left(\frac{1-x}{x}\right) - \frac{\lambda_b}{x} \right] \\
&= c_3 + \frac{2\sigma^4}{(a-b)^2} \lim_{x\to \infty} \left[ (\lambda_a - \lambda_b) \log(x-1) - \lambda_b x \right] = -\infty,
\end{aligned}$$

as required. It remains to show that

$$\begin{aligned}
&0 < \int_0^1 \frac{1}{g^2(x)} \exp\left\{ \int_{0.5}^{x} \frac{2f(y)}{g^2(y)} dy \right\} dx < \infty, \\
\Leftrightarrow\ &0 < \int_0^1 \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_a - \lambda_b)} \exp\left[ -2c\left(\frac{\lambda_a}{x} + \frac{\lambda_b}{1-x}\right) \right] dx < \infty \\
\Leftrightarrow\ &0 < \int_0^1 \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_a - \lambda_b)} \exp\left[ -\frac{2c\left[\lambda_a - (\lambda_a - \lambda_b)x\right]}{x(1-x)} \right] dx < \infty,
\end{aligned}$$

where $c = \frac{\sigma^4}{(a-b)^2}$. As the integrand is always non-negative, continuous on $(0,1)$ and strictly positive on $\left[\frac{1}{3}, \frac{2}{3}\right]$, the first inequality obviously holds. The rest of the proof will focus on the finiteness of the integral. Denoting the integrand by $h$, for $y = 0, 1$, define $h(y) = \lim_{x\to y} h(x)$. It then remains to show that $h(0)$ and $h(1)$ are finite. (Assuming that these two limits are finite, $h : [0,1] \mapsto \mathbb{R}$ is a continuous function on $[0,1]$, hence it maps to compact set $[0,1]$ to a bounded set in $\mathbb{R}$, which would imply the finiteness of the integral.) In fact, we will show that $h(0) = h(1) = 0$.

*Case 1:* $\lambda_a = \lambda_b$. By symmetry, $h(1) = h(0) = \lim_{x\to 0} \frac{c}{x^2(1-x)^2} \exp\left[-\frac{2c\lambda_a}{x(1-x)}\right]$. Then,

$$\begin{aligned}
0 &\le \lim_{x\to 0} \frac{c}{x^2(1-x)^2} \exp\left[-\frac{2c\lambda_a}{x(1-x)}\right] \\
&= \lim_{x\to 0} \frac{c}{(1-x)^2} \lim_{x\to 0} \frac{1}{x^2} \exp\left[-\frac{2c\lambda_a}{x(1-x)}\right] \\
&\le c \lim_{x\to 0} \frac{1}{x^2} \exp\left[-\frac{2c\lambda_a}{x}\right] \\
&= c \lim_{x\to\infty} x^2 \exp\left[-2c\lambda_a x\right] = 0.
\end{aligned}$$

*Case 2:* $\lambda_a > \lambda_b$.

$$\begin{aligned}
0 &\le \lim_{x\to 0} \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] \\
&= \lim_{x\to 0} \frac{c}{(1-x)^{2c(\lambda_a-\lambda_b)+2}} \lim_{x\to 0} x^{2c(\lambda_a-\lambda_b)-2} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] \\
&\le c \lim_{x\to 0} x^{2c(\lambda_a-\lambda_b)-2} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)x\right]}{x}\right] \\
&= c \exp\left[2c(\lambda_a-\lambda_b)\right] \lim_{x\to\infty} x^{2-2c(\lambda_a-\lambda_b)} \exp[-2c\lambda_a x] = 0,
\end{aligned}$$

and

$$\begin{aligned}
0 &\le \lim_{x\to 1} \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] \\
&= \lim_{x\to 0} \frac{c}{x^2(1-x)^2} \left(\frac{1-x}{x}\right)^{2c(\lambda_a-\lambda_b)} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)(1-x)\right]}{x(1-x)}\right] \\
&= \lim_{x\to 0} c(1-x)^{2c(\lambda_a-\lambda_b)-2} \lim_{x\to 0} x^{-2c(\lambda_a-\lambda_b)-2} \exp\left[-\frac{2c\left[\lambda_b + (\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] \\
&\le c \lim_{x\to 0} x^{-2c(\lambda_a-\lambda_b)-2} \exp\left[-\frac{2c\left[\lambda_b + (\lambda_a-\lambda_b)x\right]}{x}\right] \\
&= c \exp\left[2c(\lambda_b-\lambda_a)\right] \lim_{x\to\infty} x^{2c(\lambda_a-\lambda_b)+2} \exp[-2c\lambda_b x] = 0.
\end{aligned}$$

*Case 3:* $\lambda_a < \lambda_b$. By the relations

$$\begin{aligned}
&\lim_{x\to 0,1} \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)} \exp\left[-\frac{2c\left[\lambda_a - (\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] \\
&\qquad = \lim_{x\to 1,0} \frac{c}{x^2(1-x)^2} \left(\frac{x}{1-x}\right)^{2c(\lambda_b-\lambda_a)} \exp\left[-\frac{2c\left[\lambda_b - (\lambda_b-\lambda_a)x\right]}{x(1-x)}\right],
\end{aligned}$$

case 3 reduces to case 2. This completes the proof. □

With Claim 6.3.1, we can apply Theorem 2.4.2 to $\phi$: Let $\pi$ be the density associated with the invariant probability measure for $\phi$. Then

$$\pi(x) = \frac{c_1\sigma^4}{(a-b)^2x^2(1-x)^2}\exp\left[\int_{0.5}^{x}\frac{2\sigma^4\left[\lambda_a-(\lambda_a+\lambda_b)y\right]}{(a-b)^2y^2(1-y)^2}dy\right], \tag{6.2}$$

where $c_1$ is the normalization constant such that $\int \pi(x)dx = 1$. Using the working in Claim 6.3.1, we have the following:

**Proposition 6.3.2.** (Invariant probability measure for $\phi$.) *In the case of partial information on $\mu$, let $\phi$ be the process defined by $\phi_t = \mathbb{P}\{\mu_t = a|\mathcal{F}_t^Y\}$. Then the invariant probability measure of $\phi$ has a density $\pi$, which is given by*

$$\pi(x) = Nh(x) = \frac{Nc}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)}\exp\left[-\frac{2c\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right],$$

*where $c = \frac{\sigma^4}{(a-b)^2}$, and*

$$\frac{1}{N} = \int_0^1 \frac{c}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)}\exp\left[-\frac{2c\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right]dx.$$

As $h(0) = h(1) = 0$, $\pi$ is a continuous, bounded function on $[0,1]$ which takes on the value 0 at the endpoints.

We will not derive the long-run discounted growth rate of the log-optimal wealth process at this point. Instead, we will develop certain results that will allow us to obtain this value as a simple corollary in Section 6.7.

## 6.4 Inside Information on $\mu$

In this case, at time $t$, we know the values of $S_u$ and $Y'_u$ for $u \le t$, where $Y'$ is given by

$$dY'_t = \mu_t dt + \varepsilon dV_t, \qquad Y'_0 = 0,$$

where $V$ is a Wiener process which is independent of $W$, and $\varepsilon > 0$.

In order to use the filtering tools developed in Chapter 3, it will be more convenient to use

$$dY''_t = \frac{\sigma\mu_t}{\varepsilon}dt + \sigma dV_t, \qquad Y''_0 = 0,$$

as the second observation process available to us (this is just a scaled version of $Y'$). Using the notation of Chapter 3, we have

$$\mathbf{X}_t = \mu_t, \qquad \mathbf{Y}_t = \begin{bmatrix} Y_t \\ Y''_t \end{bmatrix},$$

$$\mathbf{g}(\mu_t) = \begin{bmatrix} \mu_t - \frac{\sigma^2}{2} \\ \frac{\sigma\mu_t}{\varepsilon} \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} a - \frac{\sigma^2}{2} & \frac{\sigma a}{\varepsilon} \\ b - \frac{\sigma^2}{2} & \frac{\sigma b}{\varepsilon} \end{bmatrix},$$

$$\mathbf{W}_t = \begin{bmatrix} W_t \\ V_t \end{bmatrix}, \qquad B = \sigma.$$

Let $\phi_t^\varepsilon = \mathbb{P}\{\mu_t = a | \mathcal{F}_t^{\mathbf{Y}}\}$. We can then apply the Shiryaev-Wonham filter:

$$\begin{bmatrix} d\phi_t^\varepsilon \\ 1 - d\phi_t^\varepsilon \end{bmatrix} = \begin{bmatrix} -\lambda_b & \lambda_a \\ \lambda_b & -\lambda_a \end{bmatrix} \begin{bmatrix} \phi_t^\varepsilon \\ 1 - \phi_t^\varepsilon \end{bmatrix} dt + \frac{1}{\sigma^2} \phi_t^\varepsilon (1 - \phi_t^\varepsilon) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$
$$\times \begin{bmatrix} a - \frac{\sigma^2}{2} & \frac{\sigma a}{\varepsilon} \\ b - \frac{\sigma^2}{2} & \frac{\sigma b}{\varepsilon} \end{bmatrix} \left\{ \begin{bmatrix} dY_t \\ dY_t'' \end{bmatrix} - \begin{bmatrix} a - \frac{\sigma^2}{2} & b - \frac{\sigma^2}{2} \\ \frac{\sigma a}{\varepsilon} & \frac{\sigma b}{\varepsilon} \end{bmatrix} \begin{bmatrix} \phi_t^\varepsilon \\ 1 - \phi_t^\varepsilon \end{bmatrix} dt \right\},$$

$$\begin{aligned} d\phi_t^\varepsilon &= [-\lambda_b \phi_t^\varepsilon + \lambda_a (1 - \phi_t^\varepsilon)]\, dt + \frac{\phi_t^\varepsilon (1 - \phi_t^\varepsilon)}{\sigma^2} \begin{bmatrix} a - b & \frac{\sigma(a-b)}{\varepsilon} \end{bmatrix} \\ &\quad \times \begin{bmatrix} dY_t - \left(a - \frac{\sigma^2}{2}\right) \phi_t^\varepsilon dt - \left(b - \frac{\sigma^2}{2}\right)(1 - \phi_t^\varepsilon) dt \\ dY_t'' - \frac{\sigma a}{\varepsilon} \phi_t^\varepsilon dt - \frac{\sigma b}{\varepsilon}(1 - \phi_t^\varepsilon) dt \end{bmatrix} \\ &= [\lambda_a - (\lambda_a + \lambda_b)\phi_t^\varepsilon]\, dt + \frac{(a-b)\phi_t^\varepsilon (1 - \phi_t^\varepsilon)}{\sigma^2} \\ &\quad \times \left\{ \left[ dY_t - \left( (a-b)\phi_t^\varepsilon + b - \frac{\sigma^2}{2} \right) dt \right] + \frac{\sigma}{\varepsilon} \left[ dY_t'' - \left( \frac{\sigma(a-b)}{\varepsilon} \phi_t^\varepsilon + \frac{\sigma b}{\varepsilon} \right) dt \right] \right\}. \end{aligned} \tag{6.3}$$

As in the previous section, in reality one would have one sample path of each of $Y$ and $Y''$. Running Euler's method would give one an estimate for $\phi^\varepsilon$. One could also use Monte Carlo simulation to obtain an estimate of the distribution of $\phi_t^\varepsilon$ for each $t$.

We will now derive an expression for the invariant probability measure for $\phi^\varepsilon$. By Proposition 3.3.8, $\left[Y_t - \int_0^t \left((a-b)\phi_s^\varepsilon + b - \frac{\sigma^2}{2}\right) ds\right]$ and $\left[Y_t'' - \int_0^t \left(\frac{\sigma(a-b)}{\varepsilon}\phi_s^\varepsilon + \frac{\sigma b}{\varepsilon}\right) ds\right]$ are independent Wiener processes. As such, we can write equation (6.3) as

$$d\phi_t^\varepsilon = [\lambda_a - (\lambda_a + \lambda_b)\phi_t^\varepsilon]\, dt + \frac{(a-b)\phi_t^\varepsilon (1 - \phi_t^\varepsilon)}{\sigma^2} \sqrt{1 + \frac{\sigma^2}{\varepsilon^2}} dB_t,$$

where $B$ is a Wiener process. Let $c = \frac{\sigma^4}{(a-b)^2}$ (as in the previous section), and let $d(\varepsilon) = \frac{\varepsilon^2}{\varepsilon^2 + \sigma^2}$. By modifying the proof of Claim 6.3.1 by replacing $c$ with $cd(\varepsilon)$, in the case of partial information on $\mu$, $\phi^\varepsilon$ still satisfies the conditions of Theorem 2.4.2. As such, letting $\pi_\varepsilon$ be the density associated by the invariant probability measure for $\phi^\varepsilon$, Theorem 2.4.2 gives

$$\pi_\varepsilon(x) = \frac{c_1' \varepsilon^2 \sigma^4}{(a-b)^2 x^2 (1-x)^2 (\varepsilon^2 + \sigma^2)} \exp\left[ \int_{0.5}^x \frac{2\varepsilon^2 \sigma^4 [\lambda_a - (\lambda_a + \lambda_b) y]}{(a-b)^2 y^2 (1-y)^2 (\varepsilon^2 + \sigma^2)} dy \right], \tag{6.4}$$

where $c_1'$ is the normalization constant. By calculations identical to those in Section 6.3, we have the following result:

**Proposition 6.4.1.** (Invariant probability measure for $\phi^\varepsilon$.) *In the case of inside information on $\mu$ with parameter $\varepsilon$, let $\phi^\varepsilon$ be the process defined as $\phi_t^\varepsilon = \mathbb{P}\{\mu_t = a|\mathcal{F}_t^{\mathbf{Y}}\}$. Then the invariant probability measure of $\phi^\varepsilon$ has a density $\pi_\varepsilon$, which is given by*

$$\pi_\varepsilon(x) = \frac{N_\varepsilon cd(\varepsilon)}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2cd(\varepsilon)(\lambda_a-\lambda_b)} \exp\left[-\frac{2cd(\varepsilon)\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right],$$

*where*

$$\frac{1}{N_\varepsilon} = \int_0^1 \frac{cd(\varepsilon)}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2cd(\varepsilon)(\lambda_a-\lambda_b)} \exp\left[-\frac{2cd(\varepsilon)\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] dx.$$

As with the previous section, $\pi_\varepsilon$ is a continuous, bounded function on $[0,1]$ which takes on the value 0 at the endpoints.

Note that $\pi_\varepsilon$ is very similar to $\pi$. In fact, it is $\pi$ with $c$ replaced with $cd(\varepsilon)$.

We will end this section with a result that, while somewhat technical, will be very useful in the next two sections.

**Proposition 6.4.2.** (See page 248 of Stratonovich (52).) *For any $\varepsilon > 0$,*

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx = \frac{K_q(\Lambda)}{2K_q(\Lambda) + \sqrt{\frac{\lambda_a}{\lambda_b}}K_{1+q}(\Lambda) + \sqrt{\frac{\lambda_b}{\lambda_a}}K_{1-q}(\Lambda)},$$

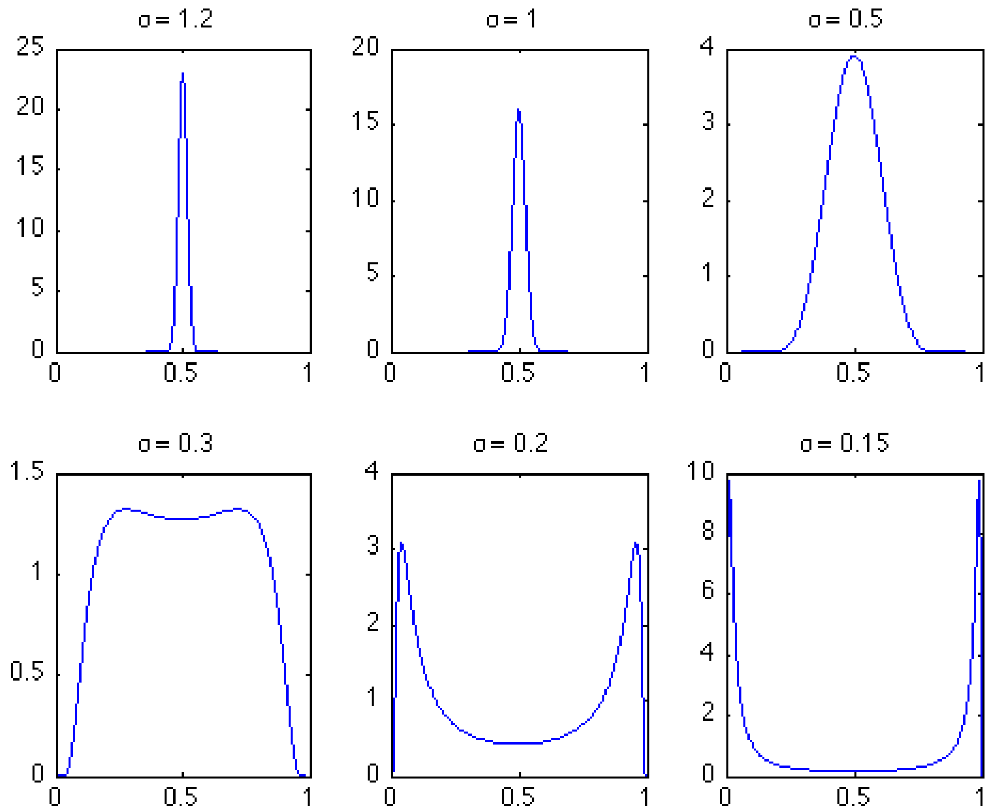
*where $N = 4cd(\varepsilon)$, $\Lambda = N\sqrt{\lambda_a\lambda_b}$, $q = \frac{N}{2}(\lambda_a - \lambda_b)$, and $K$ is the modified Bessel function of the second kind.*

*Proof.* The proof involves a few change of variables and some manipulation of integrals. For a complete proof, see page 69 in the Appendix. □

## 6.5 Intuition behind the Stationary Distributions

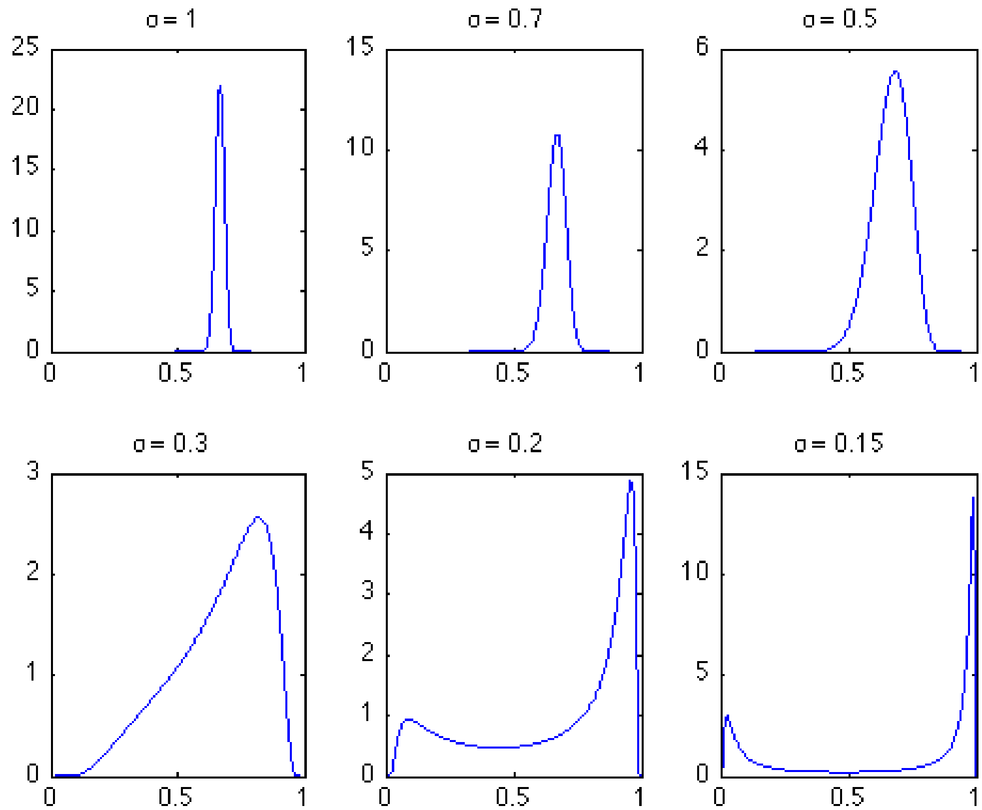
In this section, we present plots of the stationary distributions of $\phi$ and $\phi^\varepsilon$, and explain the intuition behind the shapes of plots we see.

First, we look at the stationary distribution of $\phi$ in the case of partial information on $\mu$. The plots on the next page show the stationary distribution of $\phi$ for different values of $\sigma$, when $a = 0.3$, $b = 0.1$, and $\lambda_a = \lambda_b = 1$:



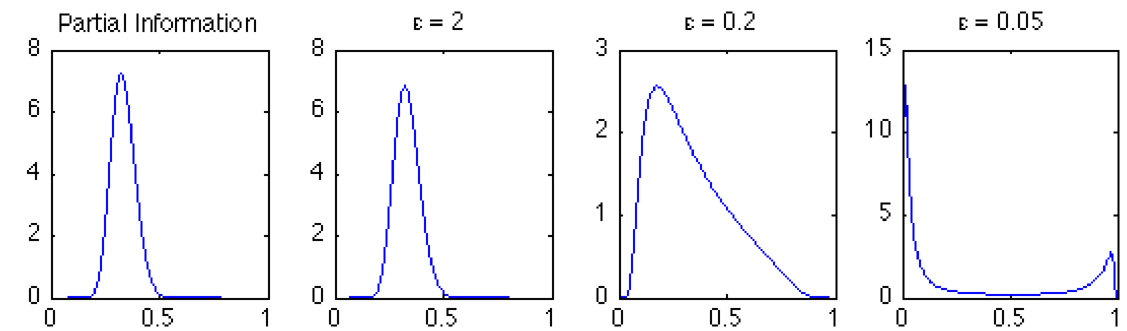

Recall that $\phi_t = \mathbb{P}\left\{\mu_t = a|\mathcal{F}_t^Y\right\}$. When $\sigma$ is big, it means that the volatility of the stock is so great that it becomes difficult to tell which state $\mu$ is in. As such, in the case where $\lambda_a = \lambda_b$, our best guess of the probability that $\mu$ is in state $a$ is 0.5. This can be seen in the plots: for bigger values of $\sigma$ the stationary distribution of $\phi$ becomes a spike centered at 0.5. As $\sigma$ becomes smaller, it means that the stock price process becomes less and less volatile. This means that at any point in time it should become easier to tell which state $\mu$ is in because a change in state will not be obscured by the volatility. We see this in the plots as well: as $\sigma$ becomes smaller, the stationary distribution becomes more and more like two spikes of equal height at 0 and 1.

In the case where $\lambda_a \neq \lambda_b$, the reasoning in the paragraph above still holds. The difference is that when $\sigma$ becomes big, our best guess of the probability that $\mu$ is in state $a$ is only informed by the rates of the Markov chain, i.e. $\mathbb{P}\left\{\mu_t = a|\mathcal{F}_t^Y\right\} = \frac{\lambda_a}{\lambda_a+\lambda_b}$. We can see this in the figures on the next page, where $a = 0.3$, $b = 0.1$, $\lambda_a = 2$ and $\lambda_b = 1$. When $\sigma$ is big, the spike is centered around $\frac{2}{3}$, as expected. When $\sigma$ is small, we have the two spikes at 0 and 1, with the spike at 0 being shorter than that at 1.

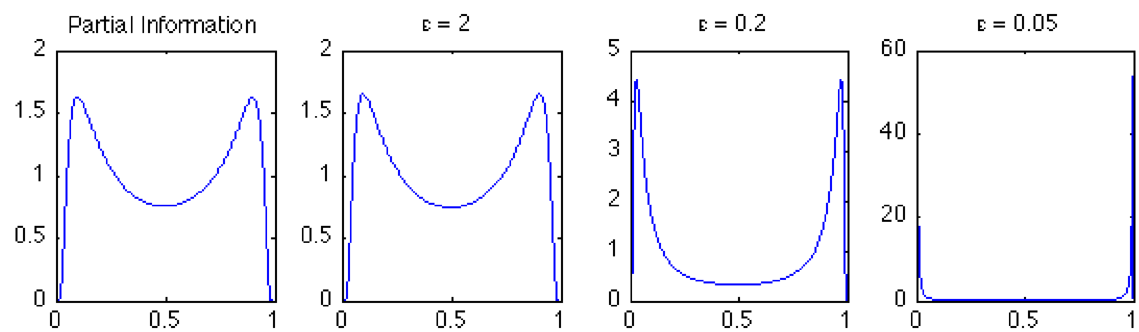



Next, we consider the case of inside information. When the parameter $\varepsilon$ is big, it means that the additional observation the insider has is basically just noise. As such, we would expect the stationary distribution of $\phi^{\varepsilon}$ to be much like that in the case of partial information. As $\varepsilon$ becomes smaller, the additional observation tells the insider which state $\mu$ is in with greater certainty. This implies that as $\varepsilon$ goes to zero, the stationary distribution of $\phi$ should become two spikes centered at 0 and 1, such that the ratio of their heights is given by ${}^{\lambda_b}/_{\lambda_a}$.

The plots confirm the intuition in the paragraph above. We demonstrate the properties above in two different cases. For the first row of figures on the next page, $a = 0.5$, $b = 0.2$, $\sigma$ is at a relatively high value of 0.7, $\lambda_a = 1$ and $\lambda_b = 2$. The first figure shows the stationary distribution of $\phi$ in the case of partial information on $\mu$, while the three remaining figures show the stationary distribution in the case of inside information for different values of $\varepsilon$.




The row of figures below are plots of the stationary distribution for $a = 0.5$, $b = 0.2$, $\sigma$ at a relatively low value of 0.3, $\lambda_a = 1$ and $\lambda_b = 1$. The intuition outlined earlier is clear in the case where $\sigma$ is relatively small and $\lambda_a = \lambda_b$ as well.




## 6.6 The Links Between the Three Cases

In Section 4.2, we made the claim that when $\varepsilon \to \infty$ we approach the case of partial information on $\mu$. The following proposition makes this claim more precise:

**Proposition 6.6.1.** (Inside information $\to$ Partial information as $\varepsilon \to \infty$.) *Define probability measures $\nu$ and $\nu_\varepsilon$ on $[0,1]$ by $\nu(dx) = \pi(x)dx$, and $\nu_\varepsilon(dx) = \pi_\varepsilon(x)dx$. Then as $\varepsilon \to \infty$, $\nu_\varepsilon$ converges weakly to $\nu$, i.e. for any bounded, continuous function $f$, $\int f d\nu_\varepsilon \to \int f d\nu$.*

The proof of Proposition 6.6.1 requires the following lemma:

**Lemma 6.6.2.** *For any $\varepsilon_0 > 0$, the family of functions $\left\{\frac{\pi_\varepsilon}{N_\varepsilon} : \varepsilon \geq \varepsilon_0\right\}$ is bounded above by an integrable function.*

*Proof.* See page 73 in the Appendix. □

*Proof of Proposition 6.6.1.* By Proposition III.5.7 of Cinlar (12), the proposition is equivalent to showing that the distribution functions associated with $\nu_\varepsilon$ converge pointwise to the distribution function associated with $\nu$. As both $\nu_\varepsilon$ and $\nu$ have densities, by Scheffé's theorem, it is sufficient to show that $\lim_{\varepsilon\to\infty} \pi_\varepsilon(x) = \pi(x)$ a.s.. Now, as $\lim_{\varepsilon\to\infty} d(\varepsilon) = 1$, for $x \in (0,1)$,

$$\lim_{\varepsilon\to\infty}\left(\frac{x}{1-x}\right)^{2cd(\varepsilon)(\lambda_a-\lambda_b)} = \left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)},$$
$$\lim_{\varepsilon\to\infty}\exp\left[-\frac{2cd(\varepsilon)\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right] = \exp\left[-\frac{2c\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right].$$

Hence, for all $x \in (0,1)$,

$$\lim_{\varepsilon\to\infty}\frac{\pi_\varepsilon(x)}{N_\varepsilon} = \frac{\pi(x)}{N}.$$

Integrating both sides of the equation over $[0,1]$,

$$\begin{aligned}
\frac{1}{N} &= \int_0^1 \frac{\pi(x)}{N}dx \\
&= \int_0^1 \lim_{\varepsilon\to\infty}\frac{\pi_\varepsilon(x)}{N_\varepsilon}dx \\
&= \lim_{\varepsilon\to\infty}\int_0^1 \frac{\pi_\varepsilon(x)}{N_\varepsilon}dx \qquad\qquad \text{(by the dominated convergence theorem)} \\
&= \lim_{\varepsilon\to\infty}\frac{1}{N_\varepsilon}.
\end{aligned}$$

By the finiteness of $N$ and $N_\varepsilon$ for all $\varepsilon > 0$, $\lim_{\varepsilon\to\infty} N_\varepsilon = N$. As such, for all $x \in (0,1)$,

$$\lim_{\varepsilon\to\infty}\pi_\varepsilon(x) = \lim_{\varepsilon\to\infty}N_\varepsilon\frac{\pi_\varepsilon(x)}{N_\varepsilon} = \lim_{\varepsilon\to\infty}N_\varepsilon\lim_{\varepsilon\to\infty}\frac{\pi_\varepsilon(x)}{N_\varepsilon} = N\frac{\pi(x)}{N} = \pi(x). \qquad \square$$

As such, we can identify the case of partial information on $\mu$ with the case of inside information on $\mu$ with $\varepsilon = \infty$. To make this connect more explicit in our notation, we may write $\pi$, the density of the invariant probability measure of $\phi$ in the case of partial information, as $\pi_\infty$.

In Section 4.2, we also claimed that when $\varepsilon \to 0$ we approach the case of complete information on $\mu$. As with the case of $\varepsilon \to \infty$, we can formulate this statement in terms of weak convergence of measures:

**Proposition 6.6.3.** (Inside information $\to$ Complete information as $\varepsilon \to 0$.) *As in Proposition 6.6.1, define probability measures $\nu_0$ and $\nu_\varepsilon$ on $[0,1]$ by $\nu_0\{0\} = \frac{\lambda_b}{\lambda_a+\lambda_b}$, $\nu_0\{1\} = \frac{\lambda_a}{\lambda_a+\lambda_b}$, and $\nu_\varepsilon(dx) = \pi_\varepsilon(x)dx$. Then as $\varepsilon \to 0$, $\nu_\varepsilon$ converges weakly to $\nu_0$, i.e. for any bounded, continuous function $f$, $\int f d\nu_\varepsilon \to \int f d\nu_0$.*

Before proving this proposition, we will first prove the following lemma:

**Lemma 6.6.4.**
$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

*Proof.* See page 74 in the Appendix. $\square$

*Proof of Proposition 6.6.3.* It is enough to show that for any subsequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\varepsilon_n = 0$, there exists a further subsequence $\{\varepsilon'_n\}_{n\in\mathbb{N}}$ of $\{\varepsilon_n\}$ such that $\nu_{\varepsilon'_n}$ converges weakly to $\nu_0$ as $n\to\infty$ (see Theorem 1.3.3 of Silvestrov (47)).

Let $\{\varepsilon_n\}$ be a subsequence such that $\varepsilon_n \to 0$ as $n\to\infty$. By Helly's Selection Principle (see Theorem 5.5.1 of Bhatia (8)), there exists some probability measure $\tilde{\nu}$ on $[0,1]$ and a subsequence $\{\varepsilon'_n\}$ of $\{\varepsilon_n\}$ such that $\nu_{\varepsilon'_n}$ converges weakly to $\tilde{\nu}$ as $n\to\infty$. As the function $f(x) = x(1-x)$ is continuous and bounded on $[0,1]$, the weak convergence implies that

$$\begin{aligned}\int_0^1 x(1-x)\tilde{\nu}(dx) &= \lim_{n\to\infty}\int_0^1 x(1-x)\pi_{\varepsilon'_n}(x)dx \\ &= 0, \qquad\qquad \text{(by Lemma 6.6.4)} \\ \Rightarrow \tilde{\nu}(0,1) &= 0,\end{aligned}$$

i.e. the measure of the open set $(0,1)$ under $\tilde{\nu}$ is zero. Now, note that for each $\varepsilon'_n$, as $\pi_{\varepsilon'_n}$ is the invariant probability measure for $\phi^{\varepsilon'_n}$,

$$\int_0^1 x\pi_{\varepsilon'_n}(x)dx = \lim_{t\to\infty}\mathbb{E}\left[\phi_t^{\varepsilon'_n}\right].$$

By the tower property of expectation,

$$\begin{aligned}\mathbb{E}\left[\phi_t^{\varepsilon'_n}\right] &= \mathbb{E}\left[\mathbb{P}\{\mu_t = a|\mathcal{F}_t^{\mathbf{Y}}\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[1_{\{\mu_t=a\}}|\mathcal{F}_t^{\mathbf{Y}}\right]\right] \\ &= \mathbb{E}\left[1_{\{\mu_t=a\}}\right] \\ &= \mathbb{P}\{\mu_t = a\}, \\ \Rightarrow \lim_{t\to\infty}\mathbb{E}\left[\phi_t^{\varepsilon'_n}\right] &= \lim_{t\to\infty}\mathbb{P}\{\mu_t = a\} \\ &= \frac{\lambda_a}{\lambda_a+\lambda_b}.\end{aligned}$$

As such,

$$\begin{aligned}0\times\tilde{\nu}\{0\} + 1\times\tilde{\nu}\{1\} &= \lim_{n\to\infty}\int_0^1 x\pi_{\varepsilon'_n}(x)dx \\ &= \frac{\lambda_a}{\lambda_a+\lambda_b}, \\ \Rightarrow \tilde{\nu}\{1\} = \frac{\lambda_a}{\lambda_a+\lambda_b}, &\qquad \tilde{\nu}\{0\} = \frac{\lambda_b}{\lambda_a+\lambda_b}, \\ \Rightarrow \tilde{\nu} &= \nu_0.\end{aligned}$$

As such, the sequence of measures $\{\nu_{\varepsilon'_n}\}$ converges weakly to $\nu_0$. By Theorem 1.3.3 of Silvestrov (47), we conclude that $\nu_\varepsilon$ converges weakly to $\nu_0$ as $\varepsilon \to 0$. $\square$

Due to Proposition 6.6.3, we can identify the case of complete information on $\mu$ with the case of inside information on $\mu$ with $\varepsilon = 0$.

## 6.7 A Discussion on Growth Rates

Let $\varepsilon \in (0, \infty]$. Extending the notation of Corollary 5.2.4, let $\gamma(V_\varepsilon^*)$ denote the largest possible long-run discounted growth rate of the wealth process in the case of inside information with parameter $\varepsilon$. Then

$$\begin{aligned}
\gamma(V_\varepsilon^*) &= \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{2\sigma^2 T}\int_0^T \left(\mathbb{E}\left[\mu_t | \mathcal{F}_t^{\mathbf{Y}}\right] - r\right)^2 dt\right] && \text{(by Cor 5.2.4)} \\
&= \frac{1}{2\sigma^2} \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left[b + (a-b)\phi_t^\varepsilon - r\right]^2 dt\right] \\
&= \frac{1}{2\sigma^2}\int_0^1 \left[(a-b)x + (b-r)\right]^2 \pi_\varepsilon(x)dx. && \text{(by Thm 2.4.3)}
\end{aligned}$$

Using the decomposition

$$\begin{aligned}
\left[(a-b)x + (b-r)\right]^2 &= -(a-b)^2(-x^2 + x) + (a-b)^2 x + 2(a-b)(b-r)x + (b-r)^2 \\
&= -(a-b)^2 x(1-x) + (a-b)(a+b-2r)x + (b-r)^2,
\end{aligned}$$

$$\begin{aligned}
2\sigma^2\gamma(V_\varepsilon^*) &= \int_0^1 \left[-(a-b)^2 x(1-x) + (a-b)(a+b-2r)x + (b-r)^2\right]\pi_\varepsilon(x)dx \\
&= -(a-b)^2\left[\int_0^1 x(1-x)\pi_\varepsilon(x)dx\right] + (a-b)(a+b-2r)\left[\int_0^1 x\pi_\varepsilon(x)dx\right] + (b-r)^2 \\
&= -(a-b)^2\left[\int_0^1 x(1-x)\pi_\varepsilon(x)dx\right] + (a-b)(a+b-2r)\lim_{t\to\infty}\mathbb{E}[\phi_t^\varepsilon] + (b-r)^2. \\
&\qquad\qquad \text{(as } \pi_\varepsilon \text{ is the invariant probability measure for } \phi^\varepsilon\text{)}
\end{aligned}$$

By working similar to that of the proof of Proposition 6.6.3,

$$\begin{aligned}
\Rightarrow \lim_{t\to\infty}\mathbb{E}[\phi_t^\varepsilon] &= \lim_{t\to\infty}\mathbb{P}\{\mu_t = a\} \\
&= \frac{\lambda_a}{\lambda_a + \lambda_b}.
\end{aligned}$$

Substituting this into the previous equation and combining the last two terms gives us the following proposition:

**Proposition 6.7.1.** *Let $\varepsilon$ be in $(0, \infty]$. In the case of inside information on mean return $\mu$ with parameter $\varepsilon$, the largest possible long-run discounted growth rate of the wealth process,* *denoted by $\gamma(V_\varepsilon^*)$, is given by*

$$\gamma(V_\varepsilon^*) = \frac{1}{2\sigma^2}\frac{\lambda_a(a-r)^2+\lambda_b(b-r)^2}{\lambda_a+\lambda_b} - \frac{(a-b)^2}{2\sigma^2}\int_0^1 x(1-x)\pi_\varepsilon(x)dx. \tag{6.5}$$

*If $\nu_\varepsilon$ is the invariant probability measure for $\phi^\varepsilon$, then we can rewrite the above equation as*

$$\gamma(V_\varepsilon^*) = \frac{1}{2\sigma^2}\frac{\lambda_a(a-r)^2+\lambda_b(b-r)^2}{\lambda_a+\lambda_b} - \frac{(a-b)^2}{2\sigma^2}\int_0^1 x(1-x)\nu_\varepsilon(dx). \tag{6.6}$$

Note that the first term in equation (6.5) is equal to the largest possible long-run discounted growth rate of the wealth process in the case of complete information on $\mu$. As the second term in equation (6.5) is always less than or equal to 0, the proposition confirms the intuition that in the long run, our trading strategy in the case of inside information will never outperform our trading strategy in the case of complete information, no matter how good our estimate of $\mu$ is. In addition, note that when $\nu_\varepsilon$ is replaced by $\nu_0$, the invariant probability measure of $\phi$ in the case of complete information (see Proposition 6.6.3), the second term on the RHS of equation (6.6) is equal to zero. As such, Proposition 6.7.1 holds for $\varepsilon = 0$ as well.

We can use Proposition 6.4.2 to obtain a more explicit expression for the growth rate:

**Proposition 6.7.2.** *Let $\varepsilon \in [0,\infty]$. Then*

$$\gamma(V_\varepsilon^*) = \frac{1}{2\sigma^2}\left[\frac{\lambda_a(a-r)^2+\lambda_b(b-r)^2}{\lambda_a+\lambda_b} - \frac{(a-b)^2K_q(\Lambda)}{2K_q(\Lambda)+\sqrt{\frac{\lambda_a}{\lambda_b}}K_{1+q}(\Lambda)+\sqrt{\frac{\lambda_b}{\lambda_a}}K_{1-q}(\Lambda)}\right],$$

*where $N = 4cd(\varepsilon)$, $\Lambda = N\sqrt{\lambda_a\lambda_b}$, $q = \frac{N}{2}(\lambda_a - \lambda_b)$, and $K$ is the modified Bessel function of the second kind. In the case where $\lambda_a = \lambda_b =: \lambda$,*

$$\gamma(V_\varepsilon^*) = \frac{1}{2\sigma^2}\left[\frac{(a-r)^2+(b-r)^2}{2} - \frac{(a-b)^2K_0(\Lambda)}{2K_0(\Lambda)+2K_1(\Lambda)}\right], \tag{6.7}$$

*where $\Lambda = N\lambda = 4cd(\varepsilon)\lambda$.*

The rest of this section is devoted to deriving an asymptotic for the long-run discounted growth rate $\gamma(V_\varepsilon^*)$ as $\varepsilon$ goes to zero in the case where $\lambda_a = \lambda_b =: \lambda$.

Recall that $d(\varepsilon) = \frac{\varepsilon^2}{\sigma^2+\varepsilon^2}$, which means that as $\varepsilon$ goes to zero, $\Lambda$ goes to zero. The following lemma is the key to deriving the asymptotic that we seek:

**Lemma 6.7.3.** *As $\Lambda$ goes to zero,*

$$\frac{K_0(\Lambda)}{K_0(\Lambda)+K_1(\Lambda)} = \Lambda\log\left(\frac{2}{\beta\Lambda}\right) + O\left((\Lambda\log\Lambda)^2\right), \tag{6.8}$$

*where $\beta = e^\gamma$, with $\gamma$ being the Euler-Mascheroni constant.*

*Proof.* Note that for small $x$,

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = 1 + O(x). \tag{6.9}$$

By Proposition 2.5.6, as $\Lambda$ goes to zero,

$$\begin{aligned}
\frac{K_0(\Lambda)}{K_0(\Lambda)+K_1(\Lambda)} &= \frac{-\left(\log\frac{\Lambda}{2}+\gamma\right)+O(\Lambda^2\log\Lambda)}{-\left(\log\frac{\Lambda}{2}+\gamma\right)+O(\Lambda^2\log\Lambda)+\frac{1}{\Lambda}+O(\Lambda\log\Lambda)} \\
&= \frac{-\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^2\log\Lambda)}{-\log\left(\frac{\beta\Lambda}{2}\right)+\frac{1}{\Lambda}+O(\Lambda\log\Lambda)} \\
&= \frac{-\Lambda\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^3\log\Lambda)}{1-\left[\Lambda\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^2\log\Lambda)\right]} \\
&= \left[-\Lambda\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^3\log\Lambda)\right]\left\{1+O\left[\Lambda\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^2\log\Lambda)\right]\right\} \\
&\qquad \text{(by Eqn (6.9))} \\
&= \left[-\Lambda\log\left(\frac{\beta\Lambda}{2}\right)+O(\Lambda^3\log\Lambda)\right]\left[1+O(\Lambda\log\Lambda)\right] \\
&= \Lambda\log\left(\frac{2}{\beta\Lambda}\right)+O\left((\Lambda\log\Lambda)^2\right).
\end{aligned}$$

□

**Lemma 6.7.4.** *As $\varepsilon$ goes to zero,*

$$\frac{K_0(\Lambda)}{K_0(\Lambda)+K_1(\Lambda)} = \frac{4c\lambda\varepsilon^2}{\sigma^2}\log\left(\frac{\sigma^2}{2c\lambda\beta\varepsilon^2}\right)+O\left(\varepsilon^4(\log\varepsilon^2)^2\right),$$

*where $\beta = e^\gamma$, with $\gamma$ being the Euler-Mascheroni constant.*

*Proof.* The proof basically involves substituting $\Lambda = \frac{4c\lambda\varepsilon^2}{\sigma^2+\varepsilon^2}$ into equation (6.8). See page 74 in the Appendix for the complete proof. □

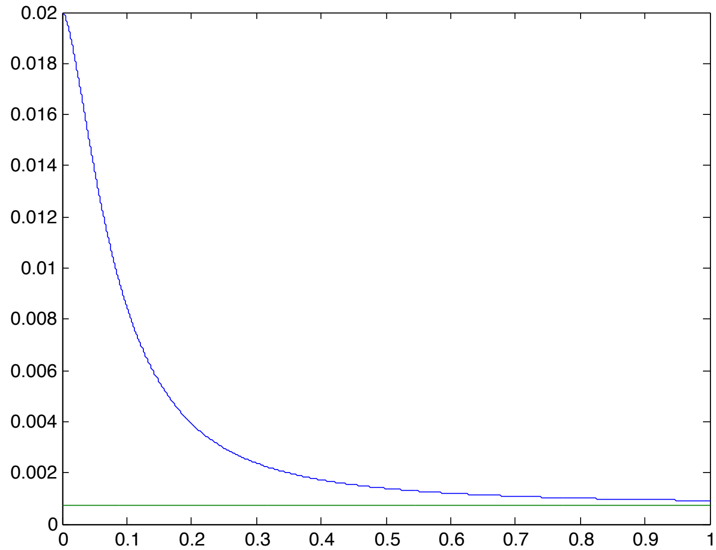
Applying Lemma 6.7.4 to equation (6.7) and using the fact that $c = \frac{\sigma^4}{(a-b)^2}$, we have the asymptotic of the largest-possible long-run discounted growth rate as $\varepsilon$ goes to zero:

**Proposition 6.7.5.** *In the case of $\lambda_a = \lambda_b =: \lambda$, as $\varepsilon$ goes to zero,*

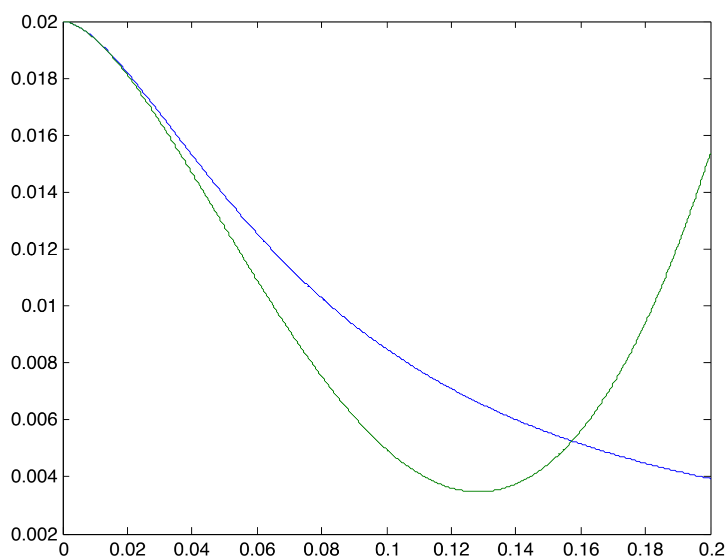
$$\gamma(V_\varepsilon^*) = \frac{(a-r)^2+(b-r)^2}{4\sigma^2} - \lambda\varepsilon^2\log\left(\frac{\sigma^2}{2c\lambda\beta\varepsilon^2}\right)+O\left(\varepsilon^4(\log\varepsilon^2)^2\right),$$

*where $\beta = e^\gamma$, with $\gamma$ being the Euler-Mascheroni constant.*

We conclude the discussion on the long-run discounted growth rate with two plots that gives an idea of how the parameter $\varepsilon$ affects the growth rate.




In the plot above, the blue line is the graph of $\gamma(V_\varepsilon^*)$ against $\varepsilon$, for parameters $a = 0.3$, $b = 0.1$, $r = 0.2$, and $\lambda_a = \lambda_b = 1$. The graph suggests that the mapping $\varepsilon \mapsto \gamma(V_\varepsilon^*)$ is continuous at $\varepsilon = 0$, which is what Proposition 6.6.3 predicts. It is also clear that $\gamma(V_\varepsilon^*)$ decreases as $\varepsilon$ increases. This is in line with our intuition that we must achieve greater portfolio gains on the average if our inside information on the stock's mean return is less noisy. The horizontal green line represents the maximum long-run discounted growth rate in the case of no information. It can be seen that as $\varepsilon$ grows large, the blue line tends toward the green line. This is in line with Proposition 6.6.1.




In the plot above, the blue line is again the mapping $\varepsilon \mapsto \gamma(V_\varepsilon^*)$ for the same parameters as before. The green line represents the asymptotic given in Proposition 6.7.5. It appears that the asymptotic holds for only very small values of $\varepsilon$.

# Chapter 7

# Conclusion & Future Research

In this chapter, we provide a summary of the results of this thesis, along with some ideas for future research.

This thesis has sought to model inside information within the context of a regime-switching model. Our financial market model consisted of a bond with a fixed interest rate r, and a stock whose price process $S$ is the solution of the SDE

$$dS_t = \mu_t S_t dt + \sigma S_t dW_t,$$

where $\mu$ is a continuous-time Markov chain with two states, and $\sigma$ is constant. In this context, we derived explicit expressions for the greatest possible expected utility at a given terminal time, the greatest possible long-run discounted growth rate of the wealth process, along with a system of SDEs from which the trading strategy that achieves these maximums can be obtained. These expressions were derived for the cases of complete, partial and inside information. We also showed that by modeling the inside information available to the investor as a noise-corrupted signal of the regime that $\mu$ is in, the cases of partial and complete information can be obtained (from the point of view of the invariant probability measure) by varying the amount of noise in the inside information. We also obtained an asymptotic for the largest possible long-run discounted growth rate as the noise parameter $\varepsilon$ goes to zero, in the case where the rates of entering each state of the Markov chain is equal.

There are a number of directions that future research can take place. The first direction would be to explore the case where the number of states $\mu$ can take is more than two. The reason why we were able to find explicit expressions in the case of two states for $\mu$ was because of the relationship

$$\mathbb{P}\{\mu_t = a | F_t^{\mathbf{Y}}\} = 1 - \mathbb{P}\{\mu_t = b | F_t^{\mathbf{Y}}\}.$$

As such, the system of SDEs obtained from the Shiryaev-Wonham filter became a single

SDE in one variable (equation (6.1)), which was straightforward to analyze. In the case where $\mu$ can take on values in the set $\{a_1, \ldots, a_d\}$ with $d \geq 3$, there is no obvious relationship between the quantities $\mathbb{P}\{\mu_t = a_1 | F_t^{\mathbf{Y}}\}, \ldots, \mathbb{P}\{\mu_t = a_d | F_t^{\mathbf{Y}}\}$ apart from the fact that they sum up to one. Future research could explore ways to solve the system of SDEs for the variables $\mathbb{P}\{\mu_t = a_1 | F_t^{\mathbf{Y}}\}, \ldots, \mathbb{P}\{\mu_t = a_d | F_t^{\mathbf{Y}}\}$ given by the Shiryaev-Wonham filter.

A second possible direction for future research would be to consider the financial market model with stochastic volatility. This means that instead of assuming that $\sigma$ is constant with time, $\sigma$ is modeled as a stochastic process. The stochastic filtering results presented in Chapter 3 are only applicable for the case where volatility is constant— as such, more general filtering results must be used. While these general stochastic filtering results are known (see Xiong (57)), to our knowledge these results have not been applied to this regime-switching model with stochastic volatility.

A third direction for future research would be to make the relationship between the drift rate of the stock price process and the additional signal received by the investor more complex. In the model discussed in this thesis (see Section 4.2), the investor received a noise-corrupted signal of $\mu$, which was exactly the drift rate of the stock price process. A more realistic model would be for the additional signal received by the investor to be a noise-corrupted version of some underlying factor (e.g. state of the economy), which would be called $\mu$. The drift rate of the stock price process would then be modeled as either a deterministic or random function of $\mu$.

Finally, one could consider a model that consists of not just one but $n$ stocks. The stock price processes, denoted by the vector $\mathbf{S}$, can be modeled as the solution to the system of SDEs

$$d\mathbf{S}_t = \mathbf{h}(\mu_t) \cdot \mathbf{S}_t dt + \boldsymbol{\sigma} \mathbf{S}_t d\mathbf{W}_t, \tag{7.1}$$

where $\mathbf{h}$ is some function, $\boldsymbol{\sigma}$ is an $n \times m$ matrix, $\mathbf{W}$ is an $m$-dimensional Wiener process and $\mu$ is the process of the underlying factor to which we alluded in the previous paragraph. In this case, a trading strategy is given by $n$ processes $\mathbf{p} = (p^1, \ldots, p^n)$, where $p_t^i$ is the fraction of the investor's wealth invested in stock $i$ at time $t$. It is worth noting that a generalization of Proposition 5.2.1 in the case of $n$ stocks is easy to derive: if $V$ is the wealth process associated with trading strategy $\mathbf{p}$, then one obtains

$$\int_0^T \frac{1}{\bar{V}_t} d\bar{V}_t = \sum_{i=1}^n \int_0^T \frac{p_t^i}{\bar{S}_t^i} d\bar{S}_t^i.$$

With a little bit more work, generalizations of the other propositions in Chapter 5 are not hard to derive either. The difficulty of the analysis in this case is the derivation of explicit expressions when the filtration used is that which is identified with the investor's knowledge. The complex relationships between the underlying factor and the stock prices, as well as that between the stock prices themselves, make analysis difficult.

# Appendix A

# Proofs of Technical Lemmas

This appendix consists of proofs of some lemmas and propositions which were presented in the thesis.

**Lemma 3.3.4.** *Let* $(\Omega, \mathcal{F}, \{F_t\}_{t\in[0,T]}, \mathbb{P})$ *be a complete filtered probability space. Let* $\mathbf{W}$ *be an* $n$*-dimensional* $\{\mathcal{F}_t\}$*-Wiener process on* $[0,T]$*, and let* $(\mathbf{g}_t)_{t\in[0,T]}$ *be a measurable process (in the sense of Definition 2.2.1) taking values in* $\mathbb{R}^n$*. let* $B$ *be some positive real constant. Assume that* $\mathbb{P}\left\{\omega : \int_0^T \mathbf{g}_t(\omega)^2 dt < \infty\right\} = 1$*, and that processes* $\mathbf{g}$ *and* $\mathbf{W}$ *are independent of each other. Then*

$$\mathbb{E}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^T \|\mathbf{g}_s\|^2 ds\right] = 1. \tag{A.1}$$

*If, in addition,* $\mathbf{g}$ *is adapted to the filtration* $\{\mathcal{F}_t\}$*, then by defining the probability measure* $\tilde{\mathbb{P}}$ *on* $(\Omega, \mathcal{F})$ *by*

$$d\tilde{\mathbb{P}} = \exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^T \|\mathbf{g}_s\|^2 ds\right] d\mathbb{P},$$

*The process* $\tilde{\mathbf{W}}$ *given by*

$$\tilde{\mathbf{W}}_t = \mathbf{W}_t - \frac{1}{B}\int_0^t \mathbf{g}_s ds$$

*is an* $\{\mathcal{F}_t\}$*-Wiener process under the probability measure* $\tilde{\mathbb{P}}$*.*

*Proof.* For each $i = 1, \ldots, n$, let $g^{(i)}$ and $W^{(i)}$ denote the $i^{\text{th}}$ component of $\mathbf{g}$ and $\mathbf{W}$ respectively. Due to the independence of $\mathbf{g}$ and $\mathbf{W}$, we can assume that the two processes are defined on a product space $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$, such that for each $\omega \in \Omega$, writing $\omega = (\omega_1, \omega_2)$, where $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$, we have

$$\mathbf{g}_s(\omega) = \mathbf{g}_s(\omega_1), \quad \mathbf{W}_s(\omega) = \mathbf{W}_s(\omega_2) \quad \text{for all } s.$$

As such,

$$
\begin{aligned}
&\mathbb{E}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s\cdot d\mathbf{W}_s-\frac{1}{2B^2}\int_0^T\|\mathbf{g}_s\|^2ds\right]\\
&=\int_{\Omega_1\times\Omega_2}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s(\omega_1)\cdot d\mathbf{W}_s(\omega_2)-\frac{1}{2B^2}\int_0^T\|\mathbf{g}_s(\omega_1)\|^2ds\right]d(\mathbb{P}_1\times\mathbb{P}_2)\\
&=\int_{\Omega_1}\exp\left[-\frac{1}{2B^2}\int_0^T\|\mathbf{g}_s(\omega_1)\|^2ds\right]\\
&\qquad\times\left\{\int_{\Omega_2}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s(\omega_1)\cdot d\mathbf{W}_s(\omega_2)\right]d\mathbb{P}_2\right\}d\mathbb{P}_1\\
&\hspace{16em}\text{(Fubini's theorem, see Thm I.6.4 in Cinlar (12))}
\end{aligned}
$$

For fixed $\omega_1\in\Omega_1$, for each $i=1,\dots,n$,

$$
\begin{aligned}
\int_0^T g_s^{(i)}(\omega_1)dW_s^{(i)}&=\lim_{n\to\infty}\sum_{k=1}^n g^{(i)}_{\frac{T(k-1)}{n}}(\omega_1)\left[W^{(i)}_{\frac{Tk}{n}}-W^{(i)}_{\frac{T(k-1)}{n}}\right]\\
&=\lim_{n\to\infty}\sum_{k=1}^n g^{(i)}_{\frac{T(k-1)}{n}}(\omega_1)Z_k,
\end{aligned}
$$

where each $Z_k$ is Gaussian distributed with mean 0 and variance $\frac{T}{n}$, and the $Z_k$'s are independent of each other. As such,

$$
\begin{aligned}
&\sum_{k=1}^n g^{(i)}_{\frac{T(k-1)}{n}}(\omega_1)Z_k\sim\mathcal{N}\left(0,\frac{T}{n}\sum_{k=1}^n g^{(i)}_{\frac{T(k-1)}{n}}(\omega_1)^2\right),\\
\Rightarrow&\int_0^T g_s^{(i)}(\omega_1)dW_s^{(i)}\sim\mathcal{N}\left(0,\lim_{n\to\infty}\frac{T}{n}\sum_{k=1}^n g^{(i)}_{\frac{T(k-1)}{n}}(\omega_1)^2\right),\\
\Rightarrow&\int_0^T g_s^{(i)}(\omega_1)dW_s^{(i)}\sim\mathcal{N}\left(0,\int_0^T g_s^{(i)}(\omega_1)^2ds\right).
\end{aligned}
$$

Noting further that the $W^{(i)}$'s are independent of each other,

$$
\begin{aligned}
\int_{\Omega_2}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s(\omega_1)\cdot d\mathbf{W}_s(\omega_2)\right]d\mathbb{P}_2&=\int_{\Omega_2}\left\{\prod_{i=1}^n\exp\left[\frac{1}{B}\int_0^T g_s^{(i)}(\omega_1)dW_s^{(i)}(\omega_2)\right]\right\}d\mathbb{P}_2\\
&=\prod_{i=1}^n\int_{\Omega_2}\exp\left[\frac{1}{B}\int_0^T g_s^{(i)}(\omega_1)dW_s^{(i)}(\omega_2)\right]d\mathbb{P}_2\\
&=\prod_{i=1}^n\mathbb{E}\exp\left[A_i\right],
\end{aligned}
$$

where $A_i$ is Gaussian distributed with mean 0 and variance $\frac{1}{B^2}\int_0^T g_s^{(i)}(\omega_1)^2ds$. For a random variable $A$ which is Gaussian with mean 0 and finite variance $\sigma^2$, the moment generating function of $A$ is given by $\mathbb{E}[\exp(tA)] = \exp\left(\frac{\sigma^2 t^2}{2}\right)$. Hence, by substituting $t = 1$,

$$\mathbb{E}\exp[A] = \exp\left(\frac{\sigma^2}{2}\right).$$

As such,

$$\begin{aligned}
&\mathbb{E}\exp\left[\frac{1}{B}\int_0^T \mathbf{g}_s \cdot d\mathbf{W}_s - \frac{1}{2B^2}\int_0^T \|\mathbf{g}_s\|^2 ds\right] \\
&= \int_{\Omega_1} \exp\left[-\frac{1}{2B^2}\int_0^T \|\mathbf{g}_s(\omega_1)\|^2 ds\right] \times \left\{\prod_{i=1}^n \mathbb{E}\exp\left[A_i\right]\right\} d\mathbb{P}_1 \\
&= \int_{\Omega_1} \exp\left[-\frac{1}{2B^2}\int_0^T \|\mathbf{g}_s(\omega_1)\|^2 ds\right] \times \left\{\prod_{i=1}^n \exp\left[\frac{1}{2B^2}\int_0^T g_s^{(i)}(\omega_1)^2 ds\right]\right\} d\mathbb{P}_1 \\
&= \int_{\Omega_1} \exp\left[-\frac{1}{2B^2}\int_0^T \sum_{i=1}^n g_s^{(i)}(\omega_1)^2 ds\right] \times \exp\left[\frac{1}{2B^2}\sum_{i=1}^n \int_0^T g_s^{(i)}(\omega_1)^2 ds\right] d\mathbb{P}_1 \\
&= 1,
\end{aligned}$$

as required. Now that $\mathbf{g}$ is adapted to the filtration $\{\mathcal{F}_t\}$ and equation (A.1) holds, the conditions for the multi-dimensional Girsanov theorem (Theorem 3.3.1) hold, and thus the rest of the lemma follows by a direct application of Girsanov's theorem for the function $\frac{\mathbf{g}}{B}$. ∎

**Proposition 6.4.2.** *For any $\varepsilon > 0$,*

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx = \frac{K_q(\Lambda)}{2K_q(\Lambda) + \sqrt{\frac{\lambda_a}{\lambda_b}}K_{1+q}(\Lambda) + \sqrt{\frac{\lambda_b}{\lambda_a}}K_{1-q}(\Lambda)},$$

*where $N = 4cd(\varepsilon)$, $\Lambda = N\sqrt{\lambda_a\lambda_b}$, $q = \frac{N}{2}(\lambda_a - \lambda_b)$, and $K$ is the modified Bessel function of the second kind.*

*Proof.* Writing out the normalization constant explicitly,

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx = \frac{\int_0^1 \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N[\lambda_a-(\lambda_a-\lambda_b)x]}{2x(1-x)}\right]dx}{\int_0^1 \frac{1}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N[\lambda_a-(\lambda_a-\lambda_b)x]}{2x(1-x)}\right]dx}. \tag{A.2}$$

Splitting up the numerator,

$$\begin{aligned}
&\int_0^1 \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right]dx \\
&\qquad = \int_0^{\frac{1}{2}} \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right]dx
\end{aligned}$$

$$
\begin{aligned}
&+\int_{\frac{1}{2}}^{1} \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right] dx \\
=&\int_{0}^{\frac{1}{2}} \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right] dx \\
&+\int_{0}^{\frac{1}{2}} \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_b-\lambda_a)} \exp\left[-\frac{N\left[\lambda_b-(\lambda_b-\lambda_a)x\right]}{2x(1-x)}\right] dx,
\end{aligned} \tag{A.3}
$$

where the last equality was obtained by replacing $x$ with $1-x$ in the second integral. Note that the first integral is exactly the same as the second, except that $\lambda_a$ and $\lambda_b$ are swtiched. As such, it is enough to evaluate the second integral because we would obtain an expression for the first integral as well. Consider the change of variables given by $\frac{1}{x(1-x)} = 4\cosh^2\frac{\varphi}{2}$. When $x=\frac{1}{2}, \varphi=0$, and when $x=0, \varphi=\infty$. Also,

$$
\begin{aligned}
& x^2 - x + \frac{1}{4\cosh^2\frac{\varphi}{2}} = 0, \\
\Rightarrow\ & x = \frac{1}{2}\left(1-\sqrt{1-\frac{1}{\cosh^2\frac{\varphi}{2}}}\right) \qquad \text{(we ignore the positive root as } x\in\left[0,\tfrac{1}{2}\right]) \\
& \phantom{x} = \frac{1}{2}\left(1-\tanh\frac{\varphi}{2}\right), \\
& \frac{dx}{d\varphi} = -\frac{1}{4\cosh^2\frac{\varphi}{2}}.
\end{aligned}
$$

Substituting this into the second integral of equation (A.3),

$$
\begin{aligned}
&\int_{0}^{\frac{1}{2}} \frac{1}{x(1-x)}\left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_b-\lambda_a)} \exp\left[-\frac{N\left[\lambda_b-(\lambda_b-\lambda_a)x\right]}{2x(1-x)}\right] dx \\
&= \int_0^\infty 4\cosh^2\frac{\varphi}{2}\left(\frac{1-\tanh\frac{\varphi}{2}}{1+\tanh\frac{\varphi}{2}}\right)^{\frac{N}{2}(\lambda_b-\lambda_a)} \\
&\qquad \times \exp\left\{-2N\cosh^2\frac{\varphi}{2}\left[\lambda_b-(\lambda_b-\lambda_a)\left(\frac{1-\tanh\frac{\varphi}{2}}{2}\right)\right]\right\}\frac{1}{4\cosh^2\frac{\varphi}{2}}d\varphi \\
&= \int_0^\infty (e^{-\varphi})^{\frac{N}{2}(\lambda_b-\lambda_a)} \exp\left\{-\frac{N}{2}(\cosh\varphi+1)\left[2\lambda_b-(\lambda_b-\lambda_a)\left(1-\frac{\sinh\varphi}{\cosh\varphi+1}\right)\right]\right\} d\varphi \\
&= \int_0^\infty (e^{\varphi})^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left\{-\frac{N}{2}\left[(\lambda_b+\lambda_a)(\cosh\varphi+1)+(\lambda_b-\lambda_a)\sinh\varphi\right]\right\} d\varphi \\
&= \exp\left[-\frac{N}{2}(\lambda_a+\lambda_b)\right]\int_0^\infty \exp\left\{\frac{N}{2}(\lambda_a-\lambda_b)(\varphi+\sinh\varphi)-\frac{N}{2}(\lambda_a+\lambda_b)\cosh\varphi\right\} d\varphi.
\end{aligned}
$$

As such, for the first integral on the RHS on equation (A.3),

$$\begin{aligned}
&\int_0^{\frac{1}{2}} \frac{1}{x(1-x)} \left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right] dx \\
&\quad = \exp\left[-\frac{N}{2}(\lambda_a+\lambda_b)\right] \int_0^{\infty} \exp\left\{\frac{N}{2}(\lambda_b-\lambda_a)(\varphi+\sinh\varphi) - \frac{N}{2}(\lambda_b+\lambda_a)\cosh\varphi\right\} d\varphi \\
&\quad = \exp\left[-\frac{N}{2}(\lambda_a+\lambda_b)\right] \int_{-\infty}^{0} \exp\left\{\frac{N}{2}(\lambda_a-\lambda_b)(\varphi+\sinh\varphi) - \frac{N}{2}(\lambda_a+\lambda_b)\cosh\varphi\right\} d\varphi,
\end{aligned}$$

where we replaced $\varphi$ with $-\varphi$ in the last step and used the fact that $\sinh(-\varphi) = -\sinh\varphi$, $\cosh(-\varphi) = \cosh\varphi$. This implies that

$$\begin{aligned}
&\int_0^{1} \frac{1}{x(1-x)} \left(\frac{x}{1-x}\right)^{\frac{N}{2}(\lambda_a-\lambda_b)} \exp\left[-\frac{N\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{2x(1-x)}\right] dx \\
&\quad = \exp\left[-\frac{N}{2}(\lambda_a+\lambda_b)\right] \int_{-\infty}^{\infty} \exp\left\{\frac{N}{2}(\lambda_a-\lambda_b)(\varphi+\sinh\varphi) - \frac{N}{2}(\lambda_a+\lambda_b)\cosh\varphi\right\} d\varphi.
\end{aligned}$$

We can perform the same method of splitting the integral into two and doing a change of variables on the denominator of the RHS of equation (A.2). As a result, equation (A.2) becomes

$$\begin{aligned}
&\int_0^1 x(1-x)\pi_\varepsilon(x)dx \\
&\quad = \frac{\int_{-\infty}^{\infty} \exp\left\{\frac{N}{2}(\lambda_a-\lambda_b)(\varphi+\sinh\varphi) - \frac{N}{2}(\lambda_a+\lambda_b)\cosh\varphi\right\} d\varphi}{\int_{-\infty}^{\infty} 2(1+\cosh\varphi)\exp\left\{\frac{N}{2}(\lambda_a-\lambda_b)(\varphi+\sinh\varphi) - \frac{N}{2}(\lambda_a+\lambda_b)\cosh\varphi\right\} d\varphi},
\end{aligned}$$

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx = \frac{\int_{-\infty}^{\infty} \exp\left\{q(\varphi+\sinh\varphi) - (q+N\lambda_b)\cosh\varphi\right\} d\varphi}{\int_{-\infty}^{\infty} 2(1+\cosh\varphi)\exp\left\{q(\varphi+\sinh\varphi) - (q+N\lambda_b)\cosh\varphi\right\} d\varphi} \tag{A.4}$$

Now, consider the change of variables given by $\phi = \exp\varphi$. Then $\varphi = \infty \Rightarrow \phi = \infty$, $\varphi = -\infty \Rightarrow \phi = 0$. Also,

$$\begin{aligned}
\frac{d\varphi}{d\phi} &= \frac{1}{\phi}, \\
\cosh\varphi &= \frac{1}{2}\left(\phi + \frac{1}{\phi}\right), \\
\sinh\varphi &= \frac{1}{2}\left(\phi - \frac{1}{\phi}\right).
\end{aligned}$$

Applying this change of variables to the numerator of the RHS of equation (A.4):

$$\begin{aligned}
&\int_{-\infty}^{\infty} \exp\left\{q(\varphi + \sinh\varphi) - (q + N\lambda_b)\cosh\varphi\right\} d\varphi \\
&\quad = \int_0^{\infty} \exp\left\{q\left[\log\phi + \frac{1}{2}\left(\phi - \frac{1}{\phi}\right)\right] - (q + N\lambda_b)\frac{1}{2}\left(\phi + \frac{1}{\phi}\right)\right\}\frac{1}{\phi}d\phi \\
&\quad = \int_0^{\infty} \phi^{q-1}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi.
\end{aligned}$$

Letting $\psi = \frac{N\lambda_b}{2}\phi$,

$$\begin{aligned}
\int_0^{\infty} \phi^{q-1}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi &= \int_0^{\infty}\left(\frac{2\psi}{N\lambda_b}\right)^{q-1}\exp\left[-\psi - \frac{N^2\lambda_a\lambda_b}{4\psi}\right]\frac{2}{N\lambda_b}d\psi \\
&= \left(\frac{2}{N\lambda_b}\right)^q \int_0^{\infty}\psi^{q-1}\exp\left[-\psi - \frac{(N\sqrt{\lambda_a\lambda_b})^2}{4\psi}\right] d\psi \\
&= \left(\frac{2}{N\lambda_b}\right)^q \times 2\left(\frac{N\sqrt{\lambda_a\lambda_b}}{2}\right)^q K_q(N\sqrt{\lambda_a\lambda_b}) \\
&= 2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^q K_q(N\sqrt{\lambda_a\lambda_b}),
\end{aligned}$$

where the second-last equality was obtained by applying formula 3.471.12 in Gradshteyn & Ryzhik (21). Applying this same change of variables to the denominator of the LHS of equation (A.4):

$$\begin{aligned}
&\int_{-\infty}^{\infty} 2(1 + \cosh\varphi)\exp\left\{q(\varphi + \sinh\varphi) - (q + N\lambda_b)\cosh\varphi\right\} d\varphi \\
&\quad = 2\int_0^{\infty}\phi^{q-1}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi + \int_0^{\infty}\phi^{q}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi \\
&\qquad + \int_0^{\infty}\phi^{q-2}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi.
\end{aligned}$$

The first integral on the RHS has already been computed. The other two integrals can computed in a similar manner:

$$\begin{aligned}
\int_0^{\infty}\phi^{q}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right] d\phi &= \left(\frac{2}{N\lambda_b}\right)^{q+1}\int_0^{\infty}\psi^{(q+1)-1}\exp\left[-\psi - \frac{(N\sqrt{\lambda_a\lambda_b})^2}{4\psi}\right] d\psi \\
&= \left(\frac{2}{N\lambda_b}\right)^{q+1} \times 2\left(\frac{N\sqrt{\lambda_a\lambda_b}}{2}\right)^{q+1} K_{1+q}(N\sqrt{\lambda_a\lambda_b}) \\
&= 2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^{q+1} K_{1+q}(N\sqrt{\lambda_a\lambda_b}),
\end{aligned}$$

$$\int_0^\infty \phi^{q-2}\exp\left[-\frac{N\lambda_b}{2}\phi - \frac{N\lambda_a}{2\phi}\right]d\phi = \left(\frac{2}{N\lambda_b}\right)^{q-1}\int_0^\infty \psi^{(q-1)-1}\exp\left[-\psi - \frac{(N\sqrt{\lambda_a\lambda_b})^2}{4\psi}\right]d\psi$$
$$= \left(\frac{2}{N\lambda_b}\right)^{q-1} \times 2\left(\frac{N\sqrt{\lambda_a\lambda_b}}{2}\right)^{q-1} K_{1-q}(N\sqrt{\lambda_a\lambda_b})$$
$$= 2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^{q-1} K_{1-q}(N\sqrt{\lambda_a\lambda_b}).$$

Substituting all these expressions into equation (A.4):

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx$$
$$= \frac{2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^q K_q(N\sqrt{\lambda_a\lambda_b})}{4\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^q K_q(N\sqrt{\lambda_a\lambda_b}) + 2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^{q+1} K_{1+q}(N\sqrt{\lambda_a\lambda_b}) + 2\left(\sqrt{\frac{\lambda_a}{\lambda_b}}\right)^{q-1} K_{1-q}(N\sqrt{\lambda_a\lambda_b})}$$
$$= \frac{K_q(N\sqrt{\lambda_a\lambda_b})}{2K_q(N\sqrt{\lambda_a\lambda_b}) + \sqrt{\frac{\lambda_a}{\lambda_b}}K_{1+q}(N\sqrt{\lambda_a\lambda_b}) + \sqrt{\frac{\lambda_b}{\lambda_a}}K_{1-q}(N\sqrt{\lambda_a\lambda_b})}.$$

This concludes the proof. □

**Lemma 6.6.2.** *For any $\varepsilon_0 > 0$, the family of functions $\left\{\frac{\pi_\varepsilon}{N_\varepsilon} : \varepsilon \geq \varepsilon_0\right\}$ is bounded above by an integrable function.*

*Proof.* For any $\varepsilon > 0$,

$$\frac{\pi_\varepsilon(x)}{N_\varepsilon} = \frac{cd(\varepsilon)}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2cd(\varepsilon)(\lambda_a-\lambda_b)}\exp\left[-\frac{2cd(\varepsilon)\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right]$$
$$= \frac{c}{x^2(1-x)^2}\left(\frac{x}{1-x}\right)^{2c(\lambda_a-\lambda_b)}\exp\left[-\frac{2c\left[\lambda_a-(\lambda_a-\lambda_b)x\right]}{x(1-x)}\right]$$
$$\times\left\{d(\varepsilon)\left(\frac{1-x}{x}\right)^{2c[1-d(\varepsilon)](\lambda_a-\lambda_b)}\exp\left[2c\left[1-d(\varepsilon)\right]\frac{\lambda_a-(\lambda_a-\lambda_b)x}{x(1-x)}\right]\right\}$$
$$= \frac{\pi(x)}{N}d(\varepsilon)\left\{\left(\frac{1-x}{x}\right)^{\lambda_a-\lambda_b}\exp\left[\frac{\lambda_a-(\lambda_a-\lambda_b)x}{x(1-x)}\right]\right\}^{2c[1-d(\varepsilon)]}$$
$$:= \frac{\pi(x)}{N}d(\varepsilon)\left[q(x)\right]^{2c[1-d(\varepsilon)]},$$

where $q$ is defined as

$$q(x) = \left(\frac{1-x}{x}\right)^{\lambda_a-\lambda_b}\exp\left[\frac{\lambda_a-(\lambda_a-\lambda_b)x}{x(1-x)}\right], \quad x\in(0,1).$$

Note that $d(\varepsilon)$ increases from 0 to 1 as $\varepsilon$ increases from 0 to $\infty$. Using the fact that the mapping $x \mapsto a^x$ is increasing when $a \geq 1$, for all $\varepsilon \geq \varepsilon_0$,

$$\begin{aligned} \frac{\pi_\varepsilon(x)}{N_\varepsilon} &\leq \frac{\pi(x)}{N} d(\varepsilon)\,[q(x) \vee 1]^{2c[1-d(\varepsilon)]} \\ &\leq \frac{\pi(x)}{N}\,[q(x) \vee 1]^{2c[1-d(\varepsilon_0)]}. \end{aligned}$$

But the function on the RHS is integrable:

$$\begin{aligned} \int_0^1 \frac{\pi(x)}{N}\,[q(x) \vee 1]^{2c[1-d(\varepsilon_0)]}\,dx &= \int_{\{q<1\}} \frac{\pi(x)}{N} dx + \int_{\{q\geq 1\}} \frac{\pi(x)}{N}\,[q(x)]^{2c[1-d(\varepsilon_0)]}\,dx \\ &\leq \int_0^1 \frac{\pi(x)}{N} dx + \int_0^1 \frac{\pi(x)}{N}\,[q(x)]^{2c[1-d(\varepsilon_0)]}\,dx \\ &= \frac{1}{N} + \int_0^1 \frac{\pi_\varepsilon(x)}{N_{\varepsilon_0} d(\varepsilon_0)} dx \\ &= \frac{1}{N} + \frac{1}{N_{\varepsilon_0} d(\varepsilon_0)} < \infty. \end{aligned}$$

□

**Lemma 6.6.4.**

$$\int_0^1 x(1-x)\pi_\varepsilon(x)dx \to 0 \qquad as \qquad \varepsilon \to 0.$$

*Proof.* As the integrand is non-negative, it is clear that the integral is non-negative for every $\varepsilon > 0$. Using Proposition 6.4.2 and the fact that $K_{1-q}(\Lambda) > 0$ (see Proposition 2.5.1),

$$\begin{aligned} \int_0^1 x(1-x)\pi_\varepsilon(x)dx &\leq \frac{K_q(\Lambda)}{2K_q(\Lambda) + \sqrt{\frac{\lambda_a}{\lambda_b}} K_{1+q}(\Lambda)}, \\ \int_0^1 x(1-x)\pi_\varepsilon(x)dx &\leq \frac{1}{2 + \sqrt{\frac{\lambda_a}{\lambda_b}} \frac{K_{1+q}(\Lambda)}{K_q(\Lambda)}}. \end{aligned} \tag{A.5}$$

Note that $q = \frac{N}{2}(\lambda_a - \lambda_b) = \frac{\lambda_a - \lambda_b}{2\sqrt{\lambda_a \lambda_b}} \Lambda$. Note also that $\Lambda = 4cd(\varepsilon)\sqrt{\lambda_a \lambda_b}$, hence $\Lambda \to 0$ as $\varepsilon \to 0$. As such, we can apply Proposition 2.5.4:

$$\begin{aligned} \lim_{\varepsilon \to 0} \frac{K_{1+q}(\Lambda)}{K_q(\Lambda)} &= \lim_{\Lambda \to 0} \frac{K_{1+\alpha\Lambda}(\Lambda)}{K_{\alpha\Lambda}(\Lambda)} \qquad \left(\text{where } \alpha = \frac{\lambda_a - \lambda_b}{2\sqrt{\lambda_a \lambda_b}}\right) \\ &= \infty. \end{aligned}$$

Substituting this into equation (A.5),

$$\begin{aligned} 0 &\leq \lim_{\varepsilon \to 0} \int_0^1 x(1-x)\pi_\varepsilon(x)dx \\ &\leq \frac{1}{2 + \sqrt{\frac{\lambda_a}{\lambda_b}} \lim_{\varepsilon \to \infty} \frac{K_{1+q}(\Lambda)}{K_q(\Lambda)}} = 0. \end{aligned}$$

□

**Lemma 6.7.4.** *As $\varepsilon$ goes to zero,*

$$\frac{K_0(\Lambda)}{K_0(\Lambda)+K_1(\Lambda)} = \frac{4c\lambda\varepsilon^2}{\sigma^2}\log\left(\frac{\sigma^2}{2c\lambda\beta\varepsilon^2}\right) + O\left(\varepsilon^4(\log\varepsilon^2)^2\right),$$

*where $\beta = e^\gamma$, with $\gamma$ being the Euler-Mascheroni constant.*

Recall that $\Lambda = 4c\lambda d(\varepsilon) = \frac{4c\lambda\varepsilon^2}{\sigma^2+\varepsilon^2}$.

*Proof.* Recall that for small $x$,

$$\frac{1}{1+x} = 1 - x + x^2 - \cdots = 1 + O(x), \tag{A.6}$$

and

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots = x + O(x^2). \tag{A.7}$$

As such, as $\varepsilon$ goes to zero,

$$\begin{aligned} \Lambda &= \frac{4c\lambda\varepsilon^2}{\sigma^2+\varepsilon^2} \\ &= \frac{4c\lambda\frac{\varepsilon^2}{\sigma^2}}{1+\frac{\varepsilon^2}{\sigma^2}} \\ &= \frac{4c\lambda\varepsilon^2}{\sigma^2}\left[1 + O\left(\frac{\varepsilon^2}{\sigma^2}\right)\right] && \text{(by Eqn (A.6))} \\ &= \frac{4c\lambda\varepsilon^2}{\sigma^2} + O(\varepsilon^4), \end{aligned}$$

and

$$\begin{aligned} \log\Lambda &= \log\left[\frac{4c\lambda\varepsilon^2}{\sigma^2} + O(\varepsilon^4)\right] \\ &= \log\left[\frac{4c\lambda\varepsilon^2}{\sigma^2}\left(1 + O(\varepsilon^2)\right)\right] \\ &= \log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + \log\left[1 + O(\varepsilon^2)\right] \\ &= \log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + O(\varepsilon^2). && \text{(by Eqn (A.7))} \end{aligned}$$

Thus, as $\varepsilon$ goes to zero,

$$\begin{aligned} \Lambda\log\Lambda &= \left[\frac{4c\lambda\varepsilon^2}{\sigma^2} + O(\varepsilon^4)\right]\left[\log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + O(\varepsilon^2)\right] \\ &= \frac{4c\lambda\varepsilon^2}{\sigma^2}\log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + O(\varepsilon^4\log\varepsilon^2), \end{aligned}$$

$$
\begin{aligned}
\Rightarrow \frac{K_0(\Lambda)}{K_0(\Lambda)+K_1(\Lambda)} &= \Lambda \log\left(\frac{2}{\beta\Lambda}\right) + O\left((\Lambda\log\Lambda)^2\right) \qquad\qquad \text{(by Lemma 6.7.3)} \\
&= -\Lambda\left[\log\Lambda + \log\frac{\beta}{2}\right] + O\left(\varepsilon^4(\log\varepsilon^2)^2\right) \\
&= -\frac{4c\lambda\varepsilon^2}{\sigma^2}\log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + O(\varepsilon^4\log\varepsilon^2) \\
&\quad - \left[\frac{4c\lambda\varepsilon^2}{\sigma^2} + O(\varepsilon^4)\right]\log\frac{\beta}{2} + O\left(\varepsilon^4(\log\varepsilon^2)^2\right) \\
&= -\frac{4c\lambda\varepsilon^2}{\sigma^2}\left[\log\left(\frac{4c\lambda\varepsilon^2}{\sigma^2}\right) + \log\frac{\beta}{2}\right] + O\left(\varepsilon^4(\log\varepsilon^2)^2\right) \\
&= \frac{4c\lambda\varepsilon^2}{\sigma^2}\log\left(\frac{\sigma^2}{2c\lambda\beta\varepsilon^2}\right) + O\left(\varepsilon^4(\log\varepsilon^2)^2\right). \qquad \square
\end{aligned}
$$

# Bibliography


[1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions*. U.S. Government Printing Office, 1972.

[2] Jürgen Amendinger, Peter Imkeller, and Martin Schweizer. Additional logarithmic utility of an insider. *Stochastic Processes and their Applications*, 75(2):263–286, 1998.

[3] Andrew Ang and Geert Bekaert. Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182, 2002.

[4] Stefan Ankirchner, Steffen Dereich, and Peter Imkeller. The Shannon information of filtrations and the additional logarithmic utility of insiders. *The Annals of Probability*, 34(2):743–778, 2006.

[5] Robert Bell and Thomas M. Cover. Game-theoretic optimal portfolios. *Management Science*, 34(6):724–733, 1988.

[6] Carl M. Bender and Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. Springer-Verlag, New York, NY, 1999.

[7] Daniel Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36, 1954.

[8] Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, Princeton, NJ, 2007.

[9] Tomas Björk, Mark H. A. Davis, and Camilla Landén. Optimal investment under partial information. *Mathematical Methods of Operations Research*, 71(2):371–399, 2010.

[10] Xinfu Chen. Optimal portfolio growth. `http://www.math.pitt.edu/~xfc/math3226/chapter4.ps`.

[11] Pavel Chigansky. Introduction to nonlinear filtering. `http://pluto.huji.ac.il/~pchiga/nlf/filtering-v0.1.pdf`.

[12] E. Cinlar. *Lecture Notes on Probability and Stochastics.* 2009. Lecture notes for ORF 551, Fall 2009–10.

[13] José M. Corcuera, Peter Imkeller, Artura Kohatsu-Higa, and David Nualart. Additional utility of insiders with imperfect dynamical information. *Finance and Stochastics*, 8(3):437–450, 2004.

[14] John C. Cox and Chi-Fu Huang. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83, 1989.

[15] E. B. Dynkin. *Markov Processes. Vols. I, II.* Academic Press Inc., Publishers, New York, NY, 1965.

[16] Robert J. Elliott, Lakhdar Aggoun, and John B. Moore. *Hidden Markov Models, Estimation and Control.* Springer-Verlag, New York, NY, 1995.

[17] Robert J. Elliott and Raymond W. Rishel. Estimating the implicit interest rate of a risky asset. *Stochastic Processes and their Applications*, 49(2):199–206, 1994.

[18] Masatoshi Fujisaki, Gopinath Kallianpur, and Hiroshi Kunita. Stochastic differential equations for the non linear filtering problem. *Osaka Journal of Mathematics*, 9(1):19–40, 1972.

[19] Thomas Goll and Jan Kallsen. Optimal portfolios for logarithmic utility. *Stochastic Processes and their Applications*, 89(1):31–48, 2000.

[20] Thomas Goll and Jan Kallsen. A complete explicit solution to the log-optimal portfolio problem. *The Annals of Applied Probability*, 13(2):774–799, 2003.

[21] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products.* Academic Press, New York, NY, 1980.

[22] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

[23] Mary R. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53, 2001.

[24] J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1981.

[25] R. Z. Has'minskiĭ. *Stochastic Stability of Differential Equations.* Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

[26] Ulrich G. Haussmann and Jörn Sass. Optimal terminal wealth under partial information for HMM stock returns. In George Yin and Qing Zhang, editors, *Mathematics of Finance*, Contemporary Mathematics, pages 171–185. American Mathematical Society, 2004.

[27] Hua He and Neil D. Pearson. Consumption and portfolio policies with incomplete markets and short-sale constraints: The infinite dimensional case. *Journal of Economic Theory*, 54(2):259–304, 1991.

[28] Gopinath Kallianpur. *Stochastic Filtering Theory*. Springer-Verlag, New York, NY, 1980.

[29] Ioannis Karatzas, John P. Lehoczky, and Steven E. Shreve. Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM Journal on Control and Optimization*, 25(6):1557–1586, 1987.

[30] Ioannis Karatzas, John P. Lehoczky, Steven E. Shreve, and Gan-Lin Xu. Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal on Control and Optimization*, 29(3):702–730, 1991.

[31] Ioannis Karatzas and Steven E. Shreve. *Methods of Mathematical Finance*. Springer-Verlag, New York, NY, 1998.

[32] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

[33] D. Kramkov and W. Schachermayer. The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *The Annals of Applied Probability*, 9(3):904–950, 1999.

[34] Harold J. Kushner. *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*. Springer-Verlag, New York, NY, 2001.

[35] Peter Lakner. Utility maximization with partial information. *Stochastic Processes and their Applications*, 56(2):247–273, 1995.

[36] Peter Lakner. Optimal trading strategy for an investor: The case of partial information. *Stochastic Processes and their Applications*, 76(1):77–97, 1998.

[37] Rahul R. Marathe and Sarah M. Ryan. On the validity of the geometric Brownian motion assumption. *The Engineering Economist*, 50(2):159–192, 2005.

[38] Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969.

[39] Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

[40] Michel Métivier. *Semimartingales: A Course on Stochastic Processes.* W. de Gruyter, Berlin; New York, 1982.

[41] J. R. Norris. *Markov Chains.* Cambridge University Press, Cambridge, UK, 1997.

[42] Frank W. J. Olver. *Asymptotics and Special Functions.* A K Peters, Natick, MA, 1997.

[43] Igor Pikovsky and Ioannis Karatzas. Anticipative portfolio optimization. *Advances in Applied Probability*, 28(4):1095–1122, 1996.

[44] L. C. G. Rogers. *Mathematics of Derivative Securities*, pages 15–40. Cambridge University Press, Cambridge, UK, 1997.

[45] Jörn Sass and Ulrich G. Haussmann. Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain. *Finance and Stochastics*, 8(4):553–577, 2004.

[46] Huntley Schaller and Simon van Norden. Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177–191, 1997.

[47] Dmitrii S. Silvestrov. *Probability and Its Applications.* Springer-Verlag, London, UK, 2004.

[48] A. V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations.* American Mathematical Society, 1989.

[49] Kazimierz Sobczyk. *Stochastic Differential Equations with Applications to Physics and Engineering.* Kluwer Academic Publishers, 1991.

[50] Christian Soize. *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions.* World Scientific Publishing, Singapore, 1994.

[51] J. Michael Steele. *Stochastic Calculus and Financial Applications.* Springer, New York, NY, 2001.

[52] R. L. Stratonovich. *Conditional Markov Processes and Their Application to the Theory of Optimal Control.* American Elsevier Publishing Company, New York, NY, 1968.

[53] Edward O. Thorp. Portfolio choice and the Kelly criterion. In *Proceedings of the 1971 Business and Economics Section of the American Statistical Association*, pages 215–224, 1971.

[54] Ramon van Handel. Lecture notes for ORF 474, Spring 2008–09: Special topics in operations research and financial engineering– stochastic methods for quantitative finance.

[55] Ramon van Handel. Stochastic calculus, filtering, and stochastic control. `http://www.princeton.edu/~rvan/acm217/ACM217.pdf`.

[56] S. R. S. Varadhan. *Stochastic Processes.* American Mathematical Society, 2007.

[57] Jie Xiong. *An Introduction to Stochastic Filtering Theory.* Oxford University Press, Oxford, UK, 2008.